

## THE PERCEPTION OF PROMINENCE: A RAPID PROSODY TRANSCRIPTION STUDY WITH ESTONIAN LISTENERS

Nele Ots<sup>1</sup> and Eva Liina Asu<sup>2</sup>

*<sup>1</sup>Goethe-University of Frankfurt am Main and <sup>2</sup>University of  
Tartu*

**Abstract.** This study investigates the effect of different prosodic variables (e.g., pitch, syllable duration, presence of pitch accent and intonation boundary) and non-prosodic variables (e.g., type of verbal argument, presence of clause boundary, part-of-speech and number of syllables) on the perception of prominence in spontaneous Estonian. Following the methodology of Rapid Prosody Transcription, 396 randomly selected speech fragments from ten speakers were presented over the internet to 51 prosodically untrained listeners, whose task was to highlight the words they heard as prominent. The same dataset was annotated for intonational pitch accents and boundary tones by two experts. The results demonstrate that the strongest predictors of prominence perception are the pitch and duration of stressed syllables together with the presence of a pitch accent while the non-prosodic variables are somewhat weaker. The study corroborates earlier findings in that the perceptual salience in spoken language processing depends largely on the acoustic signal-based cues.

**Keywords:** prominence, perception, intonation, pitch accents, part-of-speech, argument structure, Rapid Prosody Transcription, Estonian

**DOI:** <https://doi.org/10.12697/jeful.2019.10.1.08>

### 1. Introduction

Perceived prosodic prominence is a relational property of a word or phrase that can have a number of linguistic sources (Baumann and Winter 2018, Cole et al., 2010, and Turnbull et al. 2017). Speakers employ prominence in order to facilitate the production and processing of verbally transmitted information. All languages exhibit prominence but prominence patterning varies strongly among languages in relation to differences in stress and contextual factors. For example, in some languages, prominence relations can be expressed by word order while

in others (the so called intonation languages) pitch accents can be used. Previous research has shown that what is perceived as highlighted in speech depends on various prosodic and non-prosodic cues (e.g., Cole et al. 2010, and Turnbull et al. 2017).

For better understanding of linguistic prominence, it is crucial to acquire more data from typologically different languages (Cole et al. 2017) by using comparable methodologies. This paper investigates prominence perception with the *Rapid Prosody Transcription* (RPT) methodology (see e.g., Cole et al. 2010), which entails that a number of ‘naïve’ listeners transcribe a chunk of text by indicating words they hear as prominent while simultaneously listening to the audio recording of the same text. In the RPT methodology, listeners do not receive any instructions on prominence but instead are expected to follow or develop their individual internal criteria for resolving multiple cues to prominence during the transcription task. The degree of prominence of a word is measured based on the number of listeners and how frequently it was labelled as prominent. The RPT methodology has been used for the study of numerous typologically different languages (for the Germanic languages see Baumann and Winter 2018, Cole et al. 2010 and 2017; for the Romance languages Hualde et al. 2016 and Roux et al. 2016, and for Russian Luchkina and Cole 2016), enabling meaningful cross-linguistic comparisons.

This study is the first to investigate prominence perception in Estonian using the RPT methodology. The goal is to investigate the effect of prosodic and non-prosodic sources of prominence in Estonian.

### **1.1. Prosodic sources of prominence**

Prosodic sources of prominence are relatively well studied. A number of studies have shown that in many languages semantically prominent words (i.e., words in sentence focus) bear a pitch accent (see subsection 2.1.1. for a definition) in read-aloud speech (for English see e.g., Breen et al. 2010, for German Féry and Kügler 2008, and for Dutch Swerts et al. 2002). The presence of pitch accents very often implies that the F0 of pitch-accented syllables differs from the speakers’ or utterances’ mean F0 (the so-called reference F0). F0 can be either lower or higher than the reference F0, depending on the type of pitch accent (e.g., rising or falling pitch). Also, vowels of accented syllables tend to be longer

than in unaccented positions. This combination of acoustic cues (F0 and duration) makes pitch-accented syllables stand out perceptually relative to adjacent non-accented syllables.

A number of perception studies have demonstrated that listeners are highly sensitive to the presence of a pitch accent while interpreting an utterance within context (Birch and Clifton 1995, and Bock and Mazzella 1983). For example, the study by Birch and Clifton (1995) that asked native English speakers to rate the appropriateness of pitch accents in utterances in relation to context utterances showed that the participants were sensitive to the semantically expected locations of pitch accents. In a follow-up experiment, the listeners were much faster at assessing the correct meaning when the pitch accents were appropriately located than when they were not. These results can be taken to demonstrate that listeners of intonation languages are highly sensitive to prosodic prominence and use acoustic information for fast and effective processing of spoken utterances.

The evidence that (a) speakers employ prosodic prominence, and (b) listeners are sensitive to it in processing spoken utterances, supports the view that linguistic prominence can be observed directly in the acoustic signal (in F0, duration and intensity measures). However, this is not as straightforward as it may seem. Previous research suggests that the acoustic cues (e.g., F0 maximum) interact with lexical, syntactic and discourse context in triggering prominence perception (see Cole et al. 2010). Baumann and Winter (2018) tested a set of 17 different linguistic variables and showed that prominence is signalled simultaneously by multiple factors but the most predictive of perceived prominence were the ones relating to intonational phonology such as the presence of a pitch accent, its position and type (e.g., rising or falling pitch). Their results imply that phonological categories play a role in determining listeners' prominence judgements in that listeners employ them to approximate the gradient phonetic parameters such as F0, duration and intensity.

## 1.2. Non-prosodic sources of prominence

Contextual or non-prosodic influences on prominence and prominence perception have received less attention. Calhoun (2006) argues

that some types of sentence constituents (verbal arguments) are more likely to form a separate syntactic and prosodic phrase and be perceived as prominent. In this regard, some theoretical accounts have argued that there is a higher probability for sentence subjects to form separate phrases whereas verbs and objects are mostly phrased together into one prosodic phrase (e.g., Arnhold 2014, and Truckenbrodt 1999), implying that sentence subjects are more likely to carry pitch accents than objects. Therefore, due to the differing accentuation patterns, types of verb arguments might also affect the perception of prominence.

Baumann and Winter (2018) show that prominence perception is also influenced by part-of-speech. In particular, function words are highly predictable in utterances, and pronouns frequently refer to given information in discourse while content words such as nouns, proper names, adjectives and lexical verbs are much less predictable and more informative than function words. Nominal constituents (e.g., nouns, adjectives) have been demonstrated to sound much more prominent than verbs, pronouns and function words (e.g., Cole et al. 2017). Consequently, it can be concluded that prominence perception is also influenced by lexical information.

A number of theoretical accounts of languages with flexible word order claim that the peripheral positions within syntactic phrases (e.g., clause-initial and clause-final) are for information-structural reasons more prominent than phrase-internal positions (e.g., Kiss 1995, and Vilkuna 1989). For example, Vainio and Järvikivi (2006) showed that in Finnish the adverbials at the clause boundaries were perceived as more prominent than the objects in the same position (32.2% of prominence responses for object vs. 42.3% of prominence responses for adverbial). This result lends some support to the expectation that clausal position interacts with the type of verb argument in the prominence perception.

The above reviewed heterogeneous evidence on the production and perception of prominence suggests that there are a number of prosodic and non-prosodic factors to consider as predictors of perceived prominence. We have seen that not only continuous acoustic variables such as duration and F0 but also discrete prosodic variables such as the presence of a pitch accent and its position and type have the potential to modulate the perceptual salience of a word within a spoken utterance. The fact that phonological information which is more abstract than signal-based information can have an impact on prominence perception strongly

suggests that contextual properties (e.g., syntactic function, syntactic phrase boundary or part-of-speech) might also cue prominence in spoken language.

### 1.3. Aims of the current study

The current study aims to investigate the linguistic sources of perceived prominence in Estonian. We consider the following to potentially influence the perception of prominence: F0 change across the word and duration of the first syllable, presence of pitch accents and intonation phrase boundary (prosodic variables), argument type, part-of-speech, presence of clause boundary and number of syllables (non-prosodic variables).

Typologically, Estonian is an SVO language that exhibits flexible word order. In appropriate contexts, virtually all possible orderings of major sentence constituents (subject (S), object (O) and verb (V)) are accepted (Lindström 2006). Corpus studies of spoken Estonian have demonstrated that, for the one-argument-sentences, sentence-final subjects are as common as sentence-initial subjects (e.g., Lindström 2002, and 2004). A number of theoretical accounts hypothesize that when a syntactic constituent occurs in clause-peripheral position, it is more prominent than when it is in clause-internal position (Erelt et al. 1993, Lindström 2006, and 2017, and Tael 1988). As far as we know, there is no study seeking empirical evidence for this theoretical claim. The current study tests whether the verbal arguments differ in the degree of prominence and to what degree the presence of a clause boundary modulates the perceived prominence of these arguments.

Estonian makes it possible to investigate the abstract semantico-syntactic cues weighted against the signal-based and phonological cues to prominence because the few earlier studies have also indicated that Estonian employs pitch accents in a functional way (see Sakhai et al. 2013ab, Salveste 2015, and Ots 2017). For example, Sakhai et al. (2013b) found in a production study that different types of focus had an effect on word duration. Additionally, Ots (2017) showed that F0 change across the word was greater in focused than in unfocused words. Thus, based on speech production, it is reasonable to expect that signal-based acoustic cues such as duration and F0 are also available for prominence perception in Estonian.

Sahkai et al. (2013b) also show that words in focus carry pitch accents. A perception experiment in Salveste (2013) that investigated the effect of word order and the position of pitch accents on congruency judgements in Estonian found that it was the position of the nuclear pitch accent that had the strongest effect.

In sum, this study examines the weight of prosodic and non-prosodic variables in the perceived prominence based on spontaneous Estonian. As to the prosodic variables, it will be tested how well the continuous signal-based acoustic cues predict prominence perception and how the perception of prominence is affected by the presence of a phonological pitch accent. We will investigate whether the verbal arguments vary in the degree of perceived prominence, and whether the prominence in arguments is affected by the presence of a clause boundary. We will also test the replicability of the finding from earlier studies that perceived prominence correlates with the part-of-speech and the number of syllables in a word (see Baumann and Winter 2018, and Cole et al. 2017). Finally, Estonian offers a good testbed for the reliability of the RPT methodology because its grammatical properties differ somewhat from the properties of the languages tested so far.

## **2. Materials and method**

A perception experiment with naturally spoken stimuli was devised along the lines of the RPT task (Cole et al. 2010 and 2017).

### **2.1. Materials**

We selected 396 speech fragments from 10 speakers (5 females, 5 males, average age 25.3 years), comprising in total 4,727 words, from the Phonetic Corpus of Estonian Spontaneous Speech (Lippus et al. 2016). The corpus is annotated on multiple levels including words (in orthographic spelling), lemmas (derived automatically from the orthographic spelling with the FiloSoft morphological analyser, FiloSoft 2011), speech sounds (phonetic transcription in Estonian SAMPA), sound structures (CV-structures), syllables (short – long, open – closed), feet and utterances. Words and sounds are manually segmented and also contain some rhythmic information such as the lengthening of word-internal sounds.

The selection of fragments was based on the number of syllables and duration of interpausal units (a pause was defined as a period of silence longer than 400 ms in order to allow for the inclusion of shorter pauses in the fragments). Each speech fragment was between 1.8 and 5.3 s long and consisted of 18 to 24 syllables. The number of syllables and the duration of fragments were controlled for because due to the higher processing load listeners could mark fewer prominences in longer than in shorter speech fragments.

### *2.1.1. Transcription of intonation*

Relying on the consensus approach, two prosodically trained listeners (the authors of the paper) transcribed the materials for the intonational pitch accents and boundary tones. In the Autosegmental-Metrical theory of intonation, pitch accents are abstract phonological categories associated with stressed syllables that separate the continuous F<sub>0</sub> contour into discrete units of tunes (Pierrehumbert and Hirschberg 1990). The basic components of pitch accents are low and high tones (L and H respectively; see Ladd 2008) which can occur as simple tones (e.g., H\* where the asterisk indicates the alignment of the high tone with the stressed syllable) or combine into complex tones (e.g., H\*+L marking a fall from a high to a low tone). These tones need to be perceptually assessed by means of careful listening and observation of the F<sub>0</sub> contour. Intonational phonology of the present materials was transcribed using the inventory proposed for the phonological transcription of Estonian intonation in Asu (2004, and 2005).

While pitch accents were analyzed manually, intonation phrase (IP) boundaries were identified semi-automatically with the help of a Praat script and subsequently manually checked. Intonational boundaries in Estonian can be signalled by pauses, final lengthening, hesitations or interrupted words, creaky voice and pitch reset (Asu et al. 2016). Each transcriber independently produced her own annotation that was later compared to the other transcriber's. Only pitch accents and boundary tones where the two expert annotators reached a consensus were included in the analysis.

### 2.1.2. Scoring of syntactic elements

For the observation of syntactic effects, all utterances were scored for the type of sentence constituents and clause boundaries independently by two linguists (one of whom was the first author of the paper). A clause was defined to be a finite verb (i.e., a verb that agrees with the sentence subject and contains suffixes for person and number; V) together with its arguments. Following a recent overview of Estonian syntax (Erelt and Metslang 2017), each word was analyzed for the type of verb argument. The analysis did not take into account the fact that some words were part of multiple-word syntactic phrases, for example, *mingi peavalu tableti* ('some headache pill') in the utterance *Ma võtsin mingi peavalu tableti* ('I took some headache pill'). All words in such a phrase were labelled as objects but only the last word was labelled for the presence of a boundary.

## 2.2. Procedure and participants

LMEDS software (Mahrt 2016) was used for the audio presentation of spontaneous utterances together with the transcription on the screen over the internet. The 396 fragments were randomly subdivided into four lists containing 99 utterances. Listening to a list took about 40–60 minutes. Each participant was assigned to only one list creating four listener groups. Fifty-one native speakers of Estonian (40 females, 11 males between 20 and 72 years, average age 34.8) participated voluntarily in the online listening and transcription task. There were 12 transcribers in the first, 11 transcribers in the second, 13 transcribers in the third, and 15 transcribers in the fourth listener group. The task was to click on the words that sounded stressed (*'Klikkige sõnadel, mis kõlavad rõhulisena'*). No additional instructions for prominence assessment were given.

## 2.3. Analysis

The prosodically untrained transcribers marked each word they heard as either prominent or not, a binary distinction encoded as 1 and 0 respectively. For each word a prominence score (p-score) was calculated by dividing the sum of scores by the number of annotators. As such, the



p-score varied between 0 (no listeners marked the word as prominent) and 1 (all listeners marked the word as prominent). The p-score, thus, assesses the probability of a word to be perceived as prominent.

Statistical inferences were drawn from general additive models (as implemented in the *mgcv* package in the software R, Wood et al. 2016, and R Core Team 2019) as they are suitable for detecting functional relationships between the dependent and independent variables without assuming the normal distribution of data or of the residuals of the fitted model (see Cole et al. 2017 for a similar analysis procedure). The dependent variable was the p-score. The explanatory variables included the categorical factors *content word*, *presence of pitch accent*, *presence of IP boundary*, *argument type*, *presence of clause boundary*, *speaker* and *list*, and continuous variables *syllable duration*, *F0 excursion* and *number of syllables*. F0 excursion was the difference between the F0 maximum of the first syllable and the F0 minimum of the last syllable of a word. The continuous variables were scaled prior to the analysis. The model also contained interactions between the presence of pitch accent and intonation boundary, and between argument type and presence of clause boundary. The continuous variables (F0, duration) were fitted by using the smoothing function while the categorical variables have the standard OLS regression analysis interpretation.

### 3. Results

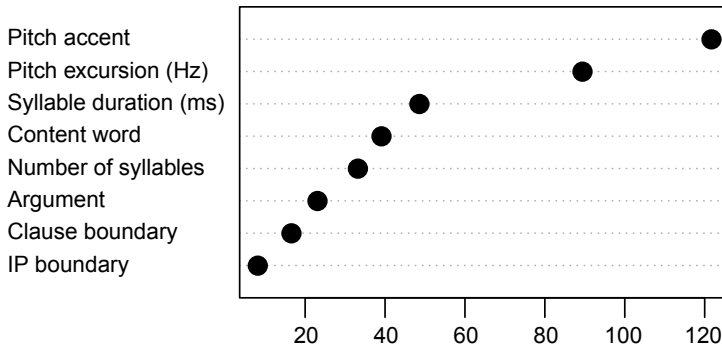
For estimating the inter-rater agreement between the annotators, we calculated the Fleiss' Kappa for each subset of annotators. The Fleiss' statistic measures the degree of agreement among annotators in relation to the expected agreement based on the number of annotators and the number of items in a list. Fleiss' Kappa coefficients and their normalized z-scores are shown in Table 1.

**Table 1.** Multi-transcriber reliability by four lists of short excerpts (and subsets of listeners), with Fleiss' Kappa coefficients and their normalized z-scores. At  $\alpha = .01$ , significance is reached at  $z = 2.32$ . All z-scores are highly significant.

Subset	Fleiss' Kappa	z-score	p-value
1	.39	110.7	< .0001
2	.27	68.96	< .0001
3	.39	122.31	< .0001
4	.36	125.91	< .0001

The Fleiss' Kappa coefficients demonstrate a rather fair agreement among the four subsets of listeners. The poor general agreement might indicate that the cues to prominence in Estonian are rather subtle and listeners vary a lot in how they resolve these potentially conflicting cues.

We first examine the influences from the various linguistic features by means of the random forest analysis (Breiman's random forest algorithm implemented in the R package *RandomForest*; Liaw and Wiener 2002). Each word represented one data point in the analysis. The model parameters were tested and selected by using the F-fold cross-validation feature. The best parameters for fitting the random forest were 800 trees and 3 features per tree. The random forest was trained on a random subset of 70% of the data (training set) and its predictions were tested on the remaining 30% (test set). The correlation between the actual p-scores and the predicted p-scores in the test set was comparatively high ( $r = .77$ ,  $R^2 = .59$ ). See Figure 1 for the importance ranking of the different variables.



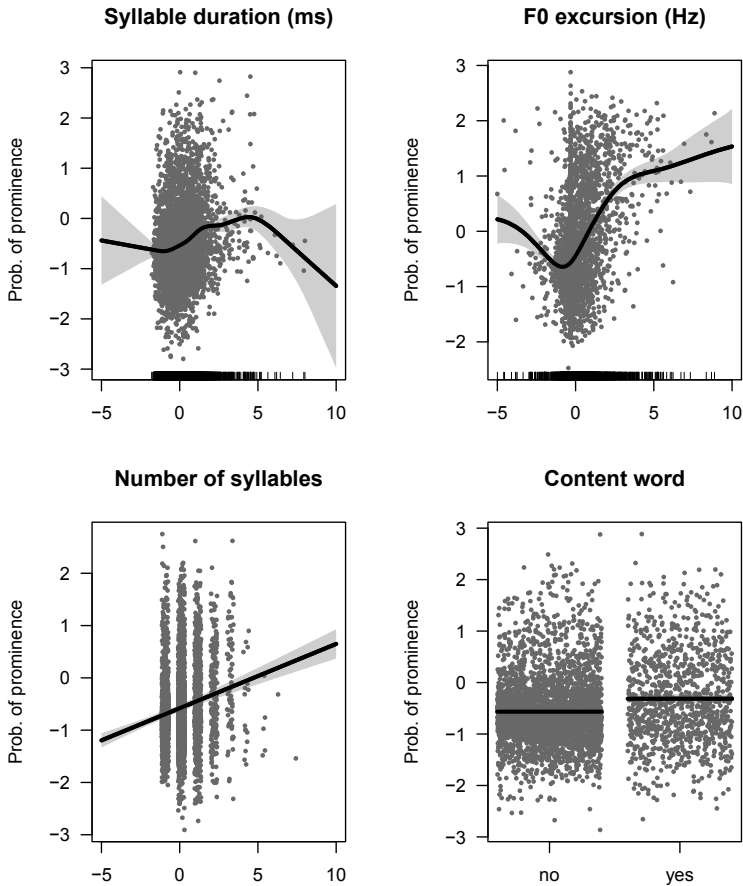
**Figure 1.** Relative variable importance based on a random forest analysis with 800 trees and 3 features per tree.

Figure 1 shows that the presence of the pitch accent is ranked the highest and is followed by the continuous prosodic variables pitch excursion and syllable duration. Other sources of prominence are ranked lower.

See Table 2 for the estimates of significant effects and Figures 2 and 3 for the visualization of the nonparametric model parameters.

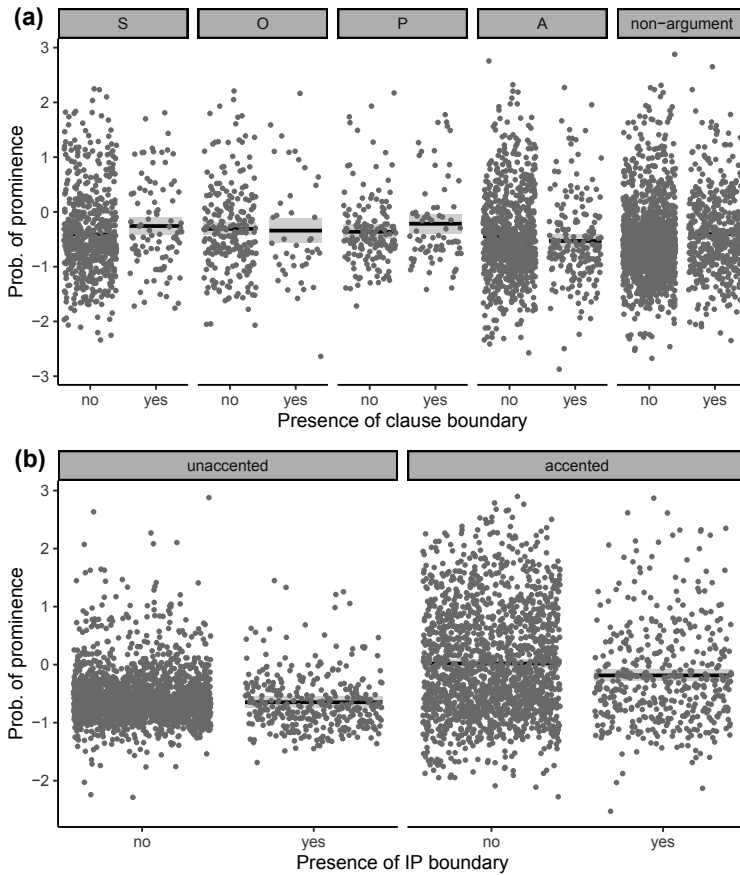
**Table 2.** Nonparametric estimates of categorical and continuous predictors. Only significant predictors are reported. The model explained 51.5% of deviance.

Model term	Estimate	Std. Error	<i>t</i> value	<i>P</i> -value
Pitch accent yes	-0.29	.08	-3.63	<.0001
Argument non-argument	-0.15	.03	-4.84	<.0001
Number of syllables	0.12	.01	9.21	<.0001
Content word yes	0.25	.03	8.63	<.0001
Pitch accent yes: IP boundary yes	-0.12	.06	-2.21	= .0300
Argument A: Clause boundary yes	-0.24	.1	-2.50	= .0126
	<b>Effective</b>	<b>Residual</b>		
	<b>df</b>	<b>effects DF</b>	<b><i>F</i></b>	<b><i>p</i>-value</b>
s(First syllable duration, ms)	6.44	7.59	25.51	<.0001
s(F0 excursion, Hz)	7.12	8.22	110.68	<.0001



**Figure 2.** Estimated probability of prominence perception. The black line shows the model estimated probability and the grey band shows the confidence interval around the estimate. The distributions of duration and F0 values in the data are shown by the thickness of the short black bars along the x-axis, with white intervals at values where there are no data.

Figure 2 indicates for the syllable duration that within the range of most data (between -0.5 and 2), as the syllable duration increases, also the probability of prominence increases. For F0 excursions, the range of most data (between -0.5 and 2) shows that as the F0 excursions increase, the p-scores increase. The p-scores are significantly higher for words with more syllables. Finally, the model estimates indicate that the prominence perception in content words is higher than in function words.



**Figure 3.** Estimated probability of prominence perception. The black line shows the model estimated probability and the grey band shows the confidence interval around the estimate. Plot (a) demonstrates the effect of the interaction between the argument type (sentence subject (S), object (O), predicate complement (P), adverbial (A)) and presence of clause boundary (x-axis) on the p-scores; plot (b) shows the effect of the interaction between the presence of pitch accent (unaccented vs. accented) and IP boundary (x-axis) on the p-scores.

Plot (a) in Figure 3 indicates that the presence of clause boundary affects the p-scores of the non-arguments, subjects, objects and predicate complements. The probability of prominence increases as the subjects, objects and predicate complements occur at the clause boundary.

In adverbials, the presence of clause boundary influences the probability of prominence perception in an opposite way: adverbial at the clause boundary is less prominent than the adverbial in the phrase-internal position. Estimated probabilities in the left panel of plot (b) demonstrate the significant interaction between the presence of IP boundary and the accentuation: the p-scores for the unaccented words at the IP boundary are higher than in the non-boundary position, while the prominence of accented words is higher in the absence of the IP boundary.

#### 4. Discussion

The present study investigated which prosodic and non-prosodic features make words or phrases perceptually salient in spontaneously spoken Estonian. The results show that prosodic variables such as syllable duration and F0 considerably contributed to the variation in p-scores. Perceived prominence was more likely as the duration of stressed syllable and F0 range increased. Similarly, the presence of a pitch accent increased the likelihood of prominence perception. This result corroborates earlier evidence of pitch accents being important cues to prominence perception in Estonian (Salveste 2013). The presence of a pitch accent was by far the most important predictor of the prominence perception in the random forest analysis of variable importance. This result indicates that it is both the concrete (i.e., acoustical signal-based) and abstract (phonological) prosodic variables that contribute to prominence perception. This also implies that the inventory of Estonian pitch accents (see e.g., Asu 2004) has a perceptual basis.

The non-prosodic variables were somewhat weaker predictors of prominence perception. The results, though, reveal important empirical support for the hypotheses present in the theoretical literature about the Estonian language. We found that the presence of the clause boundary has an effect on the prominence of different types of verbal arguments. First, the prominence of subjects, objects and predicate complements increased together with the presence of clause boundary. In accordance with earlier analyses (e.g., Tael 1988 and Lindström 2017), the perceived prominence of the clause-final subjects turned out to be greater than that of non-clause-final subjects. Second, the adverbials were

perceived as more prominent in the absence of a clause boundary. This result might be affected by the method that was used for the analysis of verbal arguments and clause boundaries. Namely, all words in multiple-word syntactic phrases were scored for the argument type but only the last word was scored for the presence of a boundary. It might be that the adverbials were more frequently composed of several words and this obscured the effect of phrase boundary. Alternatively, the clause-internal and -final positions of adverbials might differ in information-structural characteristics but this needs to be examined in future studies.

Finally, corroborating the evidence from other languages (see Baumann and Winter 2018 for German, and Cole et al. 2017 for English) we found that in prosodically untrained prominence perception, content words are more prominent than function words. Baumann and Winter (2018) predict a positive effect for content words based on information-structural criteria. The current study did not investigate the effect of information structure on the p-scores. However, we think that content words induce stronger perceived prominence because they are on average longer and lexically more specified.

The results gained for prominence perception in Estonian using the RPT methodology comply well with those from the studies of other languages. The low inter-rater scores (fairly good) are somewhat surprising but not only characteristic to prominence perception in Estonian. For instance, in two studies with American English speakers, Cole et al. (2010, and 2017) report that the inter-rater agreement oscillates around the mean Fleiss' Kappa score of .3. Baumann and Winter (2018) report a slightly higher Fleiss' Kappa score (.53) for the German annotators. We think that this difference arises from the characteristics of the materials used. Namely, prosodically untrained annotators in the study by Baumann and Winter (2018) listened to the read-aloud speech elicited under experimental conditions. The listeners in the current study and in the studies by Cole et al. (2010, and 2017) labelled prominent words in randomly selected fragments of natural conversations. Spontaneous speech production is essentially different from read-aloud speech production because it is accompanied by simultaneous processes of sentence planning (Konopka and Brown-Schmidt 2014). Therefore, under high cognitive load the speakers might not be able to produce reliable cues to prominence or the presence of cues to prominence depends on

the activation of concepts in speakers rather than in the ongoing discourse (for a discussion of speaker- and listener-oriented characteristics in speech production see Arnold 2008). Thus, as also suggested in Cole et al. (2017), listeners of naturally occurring speech are dealing with conflicting cues, and due to this, they need to vary their strategies to resolve the conflicting cues to prominence.

Another general characteristic of our data is the quite low probability estimates. This is, again, something that we observe not only in the Estonian data but also in the data reported for English and German (see in Baumann and Winter 2018, Cole et al. 2010, and 2017). The probability functions tend to oscillate below .5% (or slightly higher for some variables in Baumann and Winter 2018) independent of the language. Thus, probabilities for Estonian match quite well with those found for other languages indicating that the RPT methodology worked well for investigating the prominence perception in Estonian. The results gained using this methodology enable us to better understand the phenomenon of prominence as such, and to draw conclusions applicable for languages differing in their grammatical properties.

The results of the current study indicate that prosodic variables are better predictors of perceived prominence in Estonian than non-prosodic variables. This observation lends support to the idea that perceptual salience in speech processing is more strongly related to (a) signal-based cues and via them to (b) phonological categories than to semantico-syntactic cues. Phonological pitch accents, although linguistic abstractions, are directly related to the features of the signal (e.g., the shape of the F0 curve within a given time period) and are therefore different from semantico-syntactic cues, the retrieval of which relies to a large degree on the resolution of the acoustic cues in the incremental speech processing. Therefore, semantico-syntactic cues such as argument type and presence of clause boundary are significant for prominence perception, but they do worse as predictors and can offer only a supportive role for the resolution of relational acoustic cues such as duration and F0.

The present results could have been influenced by the nature of the materials. The listeners annotated very short fragments taken out of context, which might have boosted the strong effect of prosodic cues in the perception of prominence. Lexical, syntactic and discourse context would probably influence the perceived prominence more if longer



speech fragments with more easily traceable semantic and pragmatic relationships between the words were provided for annotation.

## 5. Conclusions

The present study investigated prominence perception in spontaneous Estonian by prosodically untrained listeners using the Rapid Prosody Transcription (RPT), which has not been used for Estonian before. In line with earlier research on other languages, the results show that prominence perception in Estonian is above all guided by prosodic variables such as the presence of a pitch accent and the size of the F0 excursions. The non-prosodic characteristics of words tested, such as the type of verbal argument, presence of clause boundary and part-of-speech, influence the perception of prominence to a lesser degree. Future studies on Estonian need to tackle among other aspects the influence of larger discourse context on the perception of prominence.

## Acknowledgements

The authors are extremely thankful to all the participants who took part in this study. We would also like to thank Jennifer Cole and Francis Nolan for their valuable feedback on an earlier draft of this paper, and Piia Taremaa for her help with the syntactic scoring of the excerpts. This research was partly supported by the grant awarded to the first author by the Fritz-Thyssen-Stiftung in Germany (“Planning sentences and sentence intonation cross-linguistically”, 10.18.2.040SL).

### Addresses:

Nele Ots

Linguistic Institute Goethe-University

Norbert-Wollheim-Platz 1–4.352

D-60629 Frankfurt, Germany

E-mail: Ots@em.uni-frankfurt.de

Eva Liina Asu  
Institute of Estonian and General Linguistics  
University of Tartu  
Ülikooli 18  
50090 Tartu, Estonia  
E-mail: eva-liina.asu@ut.ee

## References

- Arnhold, Anja (2014) *Finnish prosody: studies in intonation and phrasing* (PhD thesis). Frankfurt am Main: Johann Wolfgang Goethe-Universität.
- Arnold, Jennifer E. (2008) “Reference production: production-internal and addressee-oriented processes”. *Language and Cognitive Processes* 4, 23, 495–527.
- Asu, Eva Liina (2004) *The phonetics and phonology of Estonian intonation* (PhD thesis). Cambridge: University of Cambridge.
- Asu, Eva Liina (2005) “Towards a phonological model of Estonian intonation”. In Margit Langemets and Priit Penjam, eds. *Proceedings of the second Baltic conference on human language technologies*, Tallinn 4–5 May 2005, 95–100. Tallinn.
- Asu, Eva Liina, Pärtel Lippus, Nele Salveste, and Heete Sahkai (2016) “F0 declination in spontaneous Estonian: implications for pitch-related preplanning in speech production”. *Proceedings of the 8th International Conference on Speech Prosody 2016*, 1139–1142.
- Baumann, Stefan and Bodo Winter (2018) “What makes a word prominent? Predicting untrained German listeners’ perceptual judgments”. *Journal of Phonetics* 70, 20–38.
- Birch, Stacy L. and JR. Charles Clifton (1995) “Focus, accent, and argument structure: effects on language comprehension”. *Language and Speech* 38, 4, 365–391.
- Bock, J. Kathryn and Joanne R. Mazzella (1983) “Intonational marking of given and new information: some consequences for comprehension”. *Memory and Cognition* 11, 64–76.
- Breen, Mara, Evelina G. Fedorenko, Michael Wagner, and Edward A. Gibson (2010) “Acoustic correlates of information structure”. *Language and Cognitive Processes* 25, 7, 1044–1098.
- Calhoun, Sasha (2006) *Information structure and the prosodic structure of English: a probabilistic relationship* (PhD Thesis). Edinburgh: The University of Edinburgh.
- Cole, Jennifer, Yoonsook Mo, and Mark Hasegawa-Johnson (2010) “Signal-based and expectation-based factors in the perception of prosodic prominence”. *Laboratory Phonology* 1, 425–452. <https://doi.org/10.1515/labphon.2010.022>
- Cole, Jennifer, Timothy Mahrt and Joseph Roy (2017) “Crowd-sourcing prosodic annotation”. *Computer, Speech and Language* 45, 300–325.
- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, and Silvi Vare (1993) *Eesti keele grammatika. Süntaks*. Tallinn: Eesti Teaduste Akadeemia, Keele ja Kirjanduse Instituut.

- Erelt, Mati and Helle Metslang (2017) *Eesti keele süntaks*. Tartu: Tartu Ülikooli Kirjastus.
- Fery, Caroline and Frank Kügler (2008) “Pitch accent scaling on given, new and focused constituents in German”. *Journal of Phonetics* 36, 680–703.
- Filosoft (2011) *Morfoloogilise analüsaatori ESTMORF kasutamine*. Available online at <[http://www.filosoft.ee/html\\_morf\\_et/morfoutinfo.html#2](http://www.filosoft.ee/html_morf_et/morfoutinfo.html#2)>. Accessed on 19.08.2019.
- Hualde, José I., Jennifer Cole, Caroline L. Smith, Christopher D. Eager, Timothy Mahrt, and Ricardo Napoleão de Souza (2016) “The perception of phrasal prominence in English, Spanish and French conversational speech”. *Proceedings of the 8th International Conference on Speech Prosody 2016*, 459–463.  
<https://doi.org/10.21437/SpeechProsody.2016-94>
- Kiss, É. Katalin (1995) “Introduction”. In Katalin É. Kiss, ed. *Discourse configurational languages*, 3–27. (Oxford Studies in Comparative Syntax.) Oxford: Oxford University Press.
- Konopka, Agnieszka E. and Sarah Brown-Schmidt (2014) “Message encoding”. In Matthew Goldrick, Victor S. Ferreira, and Michele Miozzo, eds. *The Oxford handbook of language production*, 1–22. New York: Oxford University Press.
- Ladd, D. Robert (2008) *Intonational phonology*. (Cambridge Studies in Linguistics.) Cambridge: Cambridge University Press.
- Liaw, Andy and Matthew Wiener (2002) “Classification and regression by random-Forest”. *R News* 2, 3, 18–22.
- Lindström, Liina (2002) “Veel kord subjekti ja predikaadi vastastikusest asendist laiendi järel”. *Emakeele Seltsi aastaraamat* (Tallinn) 47, 87–105.
- Lindström, Liina (2004) “Sõnajärg lause tuumaargumentide eristajana eesti keeles!”. In Liina Lindström, eds. *Lauseliikmeist eesti keeles*, 40–49. (Tartu Ülikooli eesti keele õppetooli preprintid.) Tartu.
- Lindström, Liina (2006) “Infostruktuuri osast eesti keele sõnajarje muutumisel”. *Keel ja Kirjandus* 11, 875–888.
- Lindström, Liina (2017) “Lause infostruktuur ja sõnajärg”. In Mati Erelt and Helle Metslang, eds. *Eesti keele süntaks*. Tartu: Tartu Ülikooli Kirjastus.
- Lippus, Pärtel, Tuuli Tuisk, Nele Salveste, and Pire Teras (2016) *Phonetic Corpus of Estonian Spontaneous Speech v.1.0.0*. Available online at <<https://www.keel.ut.ee/et/foneetikakorpus>>. Accessed on 19.08.2019.
- Luchkina, Tatiana and Jennifer Cole (2016) “Structural and referent-based effects on prosodic expression in Russian”. *Phonetica* 73, 279–313.
- Mahrt, Timothy (2016) *LMEDS: Language Markup and Experimental Design Software*. Available online at <<https://github.com/timmahrt/LMEDS>>. Accessed on 19.08.2019.
- Ots, Nele (2017) “On the phrase-level function of f0 in Estonian”. *Journal of Phonetics* 65, 77–93.
- Pierrehumbert, Janet and Julia Hirschberg (1990) “The meaning of intonational contours in the interpretation of discourse”. In Philip R. Cohen, Jerry Morgan, and

- Martha E. Pollack, eds. *Intensions in communication*, 271–311. Cambridge, Mass. and London: The MIT Press.
- R Core Team (2019) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available online at <<https://www.R-project.org/>>. Accessed on 19.08.2019.
- Roux, Guillaume, Roxane Bertrand, Alain Ghio, and Corine Astésano (2016) “Naïve listeners’ perception of prominence and boundary in French spontaneous speech”. *Proceedings of the 8th International Conference on Speech Prosody 2016*, 912–916. <https://doi.org/10.21437/SpeechProsody.2016-187>
- Sahkai, Heete, Mari-Liis Kalvik, and Meelis Mihkla (2013a) “Prosodic effects of the information structure in Estonian”. In Eva-Liina Asu and Pärtel Lippus, eds. *Nordic Prosody – Proceedings of the XIth Conference, Tartu 2012*, 323–332, Frankfurt am Main: Peter Lang Verlag.
- Sahkai, Heete, Mari-Liis Kalvik, and Meelis Mihkla (2013b) “Prosody of contrastive focus in Estonian”. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association, 25-29 August*, 315–319. Lyon, France.
- Salveste, Nele (2013) “Focus perception in Estonian: is it governed by syntax or by prosody?” In Eva Liina Asu and Pärtel Lippus, eds. *Nordic Prosody – Proceedings of the XIth Conference, Tartu 2012*, 333–342. Frankfurt am Main: Peter Lang.
- Salveste, Nele (2015) *On the pragmatic and semantic functions of Estonian sentence prosody*. PhD Thesis. Munich: Ludwig-Maximilians-University of Munich.
- Swerts Marc, Emiel Krahmer, and Cinzia Avesani (2002) “Prosodic marking of information status in Dutch and Italian: a comparative analysis”. *Journal of Phonetics* 30, 629–654.
- Tael, Kaja (1988) *Sõnajärjemallid eesti keeles (võrrelduna soome keelega)*. (Preprint, 56.) Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Truckenbrodt, Hubert (1999) “On the relation between syntactic phrases and phonological phrases”. *Linguistic Inquiry* 2, 219–255.
- Turnbull, Rory, Adam J. Royer, Kiwako Ito, and Shari R. Speer (2017) “Prominence perception is dependent on phonology, semantics, and awareness of discourse”. *Language, Cognition and Neuroscience* 32, 8, 1017–1033.
- Vainio, Martti and Juhani Järvikivi (2006) “Tonal features, intensity, and word order in the perception of prominence”. *Journal of Phonetics* 34, 319–342.
- Vilkuna, Maria (1989) *Free word order in Finnish: its syntax and discourse functions*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Wood, Simon N., Natalya Pya, and Benjamin Säfken (2016) “Smoothing parameter and model selection for general smooth models (with discussion)”. *Journal of the American Statistical Association* 111, 1548–1575.

**Kokkuvõte. Nele Ots ja Eva Liina Asu: Prominentsuse tajude eesti keeles: prosoodia kiirtranskribeerimise uurimus eesti keele kõnelejadega.** Artikkel uurib eri prosoodiliste tunnuste (põhitoon, rõhulise silbi kestus, tooniaktsendi ja intonatsioonipiiri olemasolu) ning mitteprosoodiliste tunnuste (verbi argument, klausipiiri olemasolu, sõnaliik ja silpide arv) mõju prominentsuse tajule spontaanses eesti keeles. Lähtudes prosoodia kiirtranskriptsiooni (*Rapid Prosody Transcription*, RPT) metodoloogiast, esitati 396 juhuslikult valitud spontaanse kõne lõiku interneti teel 51-le prosoodia valdkonnas koolitamata kuulajale, kelle ülesandeks oli tekstis esile tõsta sõnad, mis kuuldusid neile rõhulisena. Kaks kõneprosoodia uurijat transkribeerisid samades kõnelõikudes tooniaktsendid ja intonatsioonipiirid. Tulemused näitavad, et prominentsuse tajude tähtsaimad ennustajad on tooniaktsent, põhitoon ning silbi kestus. Mitteprosoodilised tunnused mõjutavad prominentsuse tajude mõnevõrra vähem. Katsetulemused kinnitavad varasemate uurimuste järeldust, et tajutud prominentsus sõltub eelkõige akustilistest signaalipõhistest tunnustest.

**Märksõnad:** prominentsus, tajude, intonatsioon, tooniaktsent, sõnaliik, argumentidstruktuur, prosoodia kiirtranskriptsioon (RPT), eesti keel