INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Intonation of contrastive topic in Estonian

Heete Sahkai, Meelis Mihkla

Institute of the Estonian Language, Estonia
heete.sahkai@eki.ee, meelis.mihkla@eki.ee

Abstract

Contrastive topic is an information structural category that is usually associated with a specific intonation, which tends to be similar across languages (a rising pitch accent). The aim of the present study is to examine whether this also true of Estonian. Three potential prosodic correlates of contrastive topics are examined: marking with a particular pitch accent type, an emphatic realization of the pitch accent, and a following prosodic boundary. With respect to pitch accent types, it is found that only two subjects out of eight distinguish sentences with a contrastive topic from other types of information structure; the contour bears resemblance to contrastive topic intonation in other languages (consisting of an H* accent on the contrastive topic and an HL* accent on the focus), but is not restricted to sentences with contrastive topics. A more consistent correlate turns out to be an emphatic realization of the pitch accent carried by the contrastive topic constituent. No evidence is found of a tendency to produce contrastive topics as separate prosodic phrases.

Index Terms: Estonian, intonation, contrastive topic

1. Introduction

The aim of this study is to determine the prosodic properties of contrastive topics in Estonian. Contrastive topic (henceforth CT) can be described as an information structural category that evokes identifiable alternative questions relevant to, but not answered by, the sentence containing the CT [1]. A typical context where CT occurs are answers to multiple constituent questions: the CT signals that the answer is partial, and that there are pertinent alternative questions that are left unanswered by it (1).

(1) Q: Who ate what? A: Fred_{CT} ate the $\text{beans}_{\text{FOCUS}}$.

Although there is a debate as to whether the CT should be treated as a separate information structural category (e.g. [1]), or as a kind of focus (e.g. [2], [3], [4]), it is generally recognized to be intonationally different from the focus. In many languages, CT has been described as having a distinctive intonation, which tends to be similar across languages, usually consisting of a rising pitch accent and separate prosodic phrasing. For example, English has been described as using an LH* pitch accent followed by an L phrase accent and an H boundary tone on the CT, and an H* pitch accent followed by an L phrase accent and boundary tone on the focus [1]; German has been reported to have an L*H accent on the CT, and an H*L accent on the focus [5], the rising accent on the CT being additionally followed by a boundary tone [6]; Hungarian CT intonation has been characterized as consisting of a brief fall followed by a long rise on the CT, and a fall on the following predicate [7]; Italian displays an H* pitch accent followed by an L boundary tone on the CT [8]; Russian is characterized by a rise in pitch on the stressed vowel of the CT, and a falling intonation on the focus [3]. However, studies that aim to verify the correspondence between CTs and the proposed intonation by examining either spontaneous or elicited data tend to find that the prosodic realization of CTs is considerably less distinctive and uniform than has been suggested; this has been observed in English [9], German [10], and Hungarian [11].

In many languages, CTs can also be marked syntactically, usually by fronting; this is the case for instance in German (e.g. [6]), Hungarian [7], and Russian [3]. Topicalization also occurs in Estonian, a language with discourse-configurational features [12]. The syntactic and prosodic marking of CTs have not been claimed to be complementary: syntactically marked CTs are assumed to be prosodically marked as well.

The present study tests for three possible prosodic markers of CTs in Estonian: pitch accent type, emphatic realization of the pitch accent, and a following prosodic boundary.

Estonian has two main nuclear accent types, H*L and HL*, and three main pre-nuclear accents, H*L, HL* and H*, which have been identified and described by [13] and [14]. H*L is considered to be the default accent, a series of H*L accents constituting the default contour; a nuclear H*L can also be preceded by an H* accent, but not by an HL* accent: a pre-nuclear HL* is possible only if the subsequent accents are of the same type. The nuclear HL* is considered to be marked and associated with declarative force and finality: it does not occur in questions [13], and it has been hypothesized to be excluded from non-utterance-final intonation phrases [15]; it also cannot appear as a non-final nuclear accent marking early narrow focus [16]. A nuclear HL* accent can be preceded by a series of pre-nuclear HL* accents, which is optionally introduced by a single H* accent; it can also be preceded by H*L accents, in which case the transition to HL* accents is mediated by a single H* accent. In summary, Estonian possesses a series of pre-nuclear pitch accents, which could in principle consistently mark the CT: H*L, H*, and HL*.

Previous studies have examined whether the two nuclear accent types correlate with narrow vs. broad information focus [16], or with contrastive vs. information focus [15]. No clear correlation between an accent type and one of these categories was found: both accent types occurred in all these categories. Instead, it was found that sentence-final narrow focus is distinguished from broad focus by an emphatic realization of the pitch accent, while contrastive and information focus are prosodically and perceptually, as well as syntactically, undistinguishable ([15], [16], [17], [18]). Since information structural categories have previously not been found to correlate with pitch accent types in Estonian, we will additionally examine whether the CT is expressed with an emphatic realization of the pitch accent. Emphasis has also been found to characterize CTs in German [10].

http://dx.doi.org/10.21437/Interspeech.2017-840

Finally, as a third possibility, it will be examined whether CTs are produced in a separate prosodic phrase, as has been found for many other languages (see above).

2. Data and procedure

The data was elicited using the design and materials of the "Who does what" production task of the "Questionnaire on Information Structure" [19]. The task is designed to elicit identical sentences with a large number of different types of information structure. The subjects were shown pictures and asked to answer recorded questions about the pictures.

The present study uses data elicited with four types of question. For example, the sentence *Mees joob kokat* 'The/A man is drinking coke' was produced in answer to the following questions: (i) the question 'Who is drinking what?', asked about a picture representing a parallel scene with a man drinking coke and a woman drinking wine, eliciting paired sentences with parallel CTs and foci (respectively, 'the man' and 'the woman', and 'coke' and wine'); (ii) the question 'What is the MAN drinking?', asked again about a parallel scene and with an additional focus on 'man' (see [2, pp. 388–397] on how foci in questions evoke a set of alternative questions), eliciting a single sentence with a CT ('the man') and a narrow information focus ('coke'); (iii) the question 'What is happening?', asked about a picture representing a single scene with a man drinking coke, eliciting a single broad focus sentence; and (iv) the question 'What is the man drinking?', asked again about a single scene and eliciting a single sentence with an aboutness topic ('the man') and a narrow information focus ('coke').

The data includes thus five conditions; three conditions include CTs: (i) the first conjuncts of the parallel sentences, (ii) the second conjuncts of the parallel sentences, and (iii) the single CT sentences; two conditions are used for comparison: (iv) the broad focus sentences, and (v) the sentences with an aboutness topic and a narrow information focus.

All sentences consist of three words and have the structure subject - verb - object, which is superficially unmarked. The CT is always the subject, and it is compared with the corresponding subjects in the baseline conditions. Each condition contains 8 sentences produced by 8 speakers (4 women and 4 men), altogether 319 sentences since one of the broad focus sentences of one speaker (M4) was missing for technical reasons.

The data were automatically segmented into words and segments, using the automatic aligner created at the Institute of Cybernetics at the Tallinn University of Technology (https://phon.ioc.ee/dokuwiki/doku.php?id=projects:tuvastus:est-align.et). The segmentation was manually corrected and the data were analyzed using Praat [20]. Using a Praat script, the following values were extracted for each test sentence: the beginning and end time of each word, which were used to calculate the duration of the word, and the F0 maximum and minimum of each target word, which were used to calculate their F0 range.

To determine whether the CTs are marked with a particular pitch accent type, we manually annotated the sentences for pitch accents, assuming the inventory and description of Estonian pitch accents in [13] and [14]. Figure 1 illustrates the two main contours identified in the data: a series of H*L accents, considered by [13] to represent the Estonian default intonation, and the sequence of a pre-nuclear H* followed by a nuclear HL*.

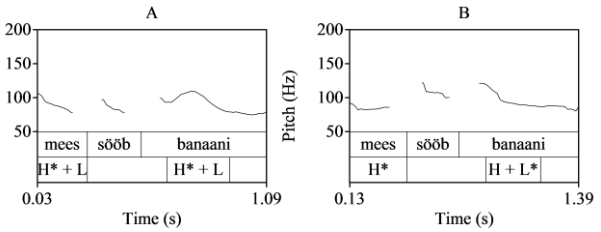

Figure 1: *The sentence* Mees sööb banaani *'The man is eating a banana' produced by speaker M4 in the aboutness topic condition (A) and the single CT condition (B). The second tier indicates the limits of the stressed syllables and the pitch accent labels.*

In order to determine whether the CTs are marked with an emphatic realization of the pitch accent, we compared the relative durations of both the subjects and the objects in the different conditions, as lengthening has previously been found to be the major correlate of emphasis in Estonian [21]. Prior to the comparisons, we equalized the durations of the four versions of a sentence for each informant, and then calculated the relative difference between the normalized duration of a target word in a particular condition and its mean duration in the four conditions. For the purpose of the measurements, some sentences containing minor disfluencies had to be discarded. From the parallel sentences, only the clauses identical to the single-clause sentences were used, provided they were produced as a separate intonation phrase. As additional potential correlates of emphasis, we compared the F0 range and the F0 maximum of the subject constituent in the different conditions (pitch accent alignment was not included in the study as it signals primarily word quantity in Estonian [22]). For the purpose of the comparison of the F0 maxima of the target words in the different conditions, we first calculated the average F0 maximum of each target word for each speaker across the four conditions, and then the difference between this average and the F0 maximum in each condition. From the parallel sentences, only the first conjuncts were included, as the second conjuncts were sometimes produced with a reduced pitch range. In total, duration was measured in 240 sentences and the F0 parameters in 229 sentences.

To determine whether the CTs are distinguished by separate prosodic phrasing, we checked the annotations for pauses between the subject and the verb in the different conditions. Pauses were chosen as the most reliable boundary signal, given that the Estonian tonal inventory posited by [13] does not include boundary tones, and that the prosodic structure below the intonation phrase and above the prosodic word has not been studied.

3. Results and discussion

3.1. Distribution of pitch accent types

The type of pitch accent found on the CTs and the corresponding constituents in the broad focus and aboutness topic sentences depends primarily on the type of the nuclear accent, rather than the information structural status of the constituent. This can be seen from Table 1, which presents the

combinations of the first accent and the nuclear accent occurring in the data.

Table 1: *Contours (first accent + nuclear accent) in the sentences with parallel contrastive topics (PCT), a single contrastive topic (SCT), broad focus (BF), and an aboutness topic (AT).*

Contour	PCT 1	PCT 2	SCT	BF	AT	Total
H*L + H*L[1]	54	20	19	29	32	154
H* + H*L	7	4	5	3	4	23
H* + HL*	3	38	36	21	23	121
HL* + HL*	0	1	2	5	3	11
H*L + HL*	0	1	2	5	2	10
Total	64	64	64	63	64	319

It can be seen that an initial H*L accent occurs primarily with a nuclear H*L and only marginally with a nuclear HL*, whereas an initial H* occurs primarily with a nuclear HL* and only rarely with an H*L. Grouping together the two nuclear contours, it can be seen from Table 2 that the CT sentences prefer the marked contour, with the exception of the first clause of the parallel sentences. The strong preference for the default contour in the first clause of the parallel sentences suggests that the marked contour is not compatible with non-utterance-final intonation phrases, meaning that non-finality is a much stronger determinant of accent distribution than the presence of a CT. The broad focus sentences show an equal distribution of the two contours, and the aboutness topic sentences a slight preference for the default contour.

Table 2: *Distribution of the nuclear contours in the sentences with parallel contrastive topics (PCT), a single contrastive topic (SCT), broad focus (BF), and an aboutness topic (AT).*

Nucleus	PCT 1	PCT 2	SCT	BF	AT	Total
H*L/H*	61	24	24	32	36	177
HL*	3	40	40	31	28	142
Total	64	64	64	63	64	319

A closer look at the data reveals that in addition to non-finality, speaker preference is also a stronger determinant of accent type than the presence of a CT. Table 3 shows the distribution of nuclear accent types in the data of individual speakers.

Table 3: *Distribution of nuclear accents by speakers.*

Acc.	F1	F2	F3	F4	M1	M2	M3	M4	Tot.
H*L	5	40	26	11	40	23	14	18	177
HL*	35	0	14	29	0	17	26	21	142
Total	40	40	40	40	40	40	40	39	319

Two speakers (F2 and M1) have an absolute preference for the default contour. A third speaker (F3) also prefers the default pattern, but uses the marked contour as well; however, the marked contour occurs in all the conditions except the non-utterance-final one and hence does not distinguish CTs from the other types of information structure. Three speakers (F1, F4, and M3) in turn prefer the marked contour, abandoning it primarily in the non-final intonation phrases. Two speakers (M2 and M4) show a more or less equal distribution of the two contours. Factoring out the non-final sentences (where they exclusively use the default pattern) and taking together the remaining two CT conditions, these speakers show a preference for the marked contour in the CT sentences: in the data of M2, the marked contour occurs in 11 out of the 16 CT sentences; in M4, the marked contour occurs in all the CT sentences. The remaining instances of the marked pattern occur mostly in the broad focus sentences, in which both speakers use the two contours in an equal proportion. In aboutness topic sentences, the marked contour occurs twice for M2 and once for M4. The strongest distinction is thus made between CTs and aboutness topics, the broad focus sentences showing no preference in terms of the contour type.

In summary, the choice of the initial pitch accent is primarily determined by the global contour. The choice of the contour in turn is determined first and foremost by (non-)utterance-finality and speaker preference, the former overriding the latter. However, in two speakers, the choice is also affected by the presence of a CT. Both speakers show a preference for the marked pattern in the (utterance-final) CT sentences. The marked contour bears some similarity to the CT intonation described in other languages: it usually contrasts a small rise to the H* pitch accent on the topic with a fall to an L tone (HL*) on the focus. However, a longer CT sentence with intervening accents between the CT and the focus is predicted to not have a similar hat or bridge shape, as HL* accents have been described to be preceded by a single H* accent. Also, the marked pattern is far from being restricted to CT sentences, contrasting the latter more strongly with aboutness topic than with broad focus sentences.

3.2. Emphasis

Figure 2 presents the relative lengthening/shortening of the subject and object in aboutness topic, broad focus, and CT sentences, taking together the three CT conditions, which showed similar behavior. In CT sentences, the subject (the CT) is in average 31 ms longer than the aboutness topic, and 15 ms longer than the subject of the broad focus sentences. The object of the CT sentences is in average 21 ms shorter than in the aboutness topic sentences, and 11 ms longer than in the broad focus sentences.

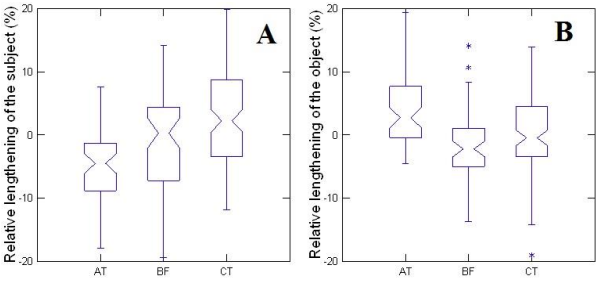

Figure 2: *Boxplots of the relative lengthening/shortening of the subject (A) and the object (B) in the aboutness topic (AT), broad focus (BF) and contrastive topic (CT) sentences.*

All the duration differences between the categories are statistically significant (see Table 4). However, neither of the

[1] This category includes nuclear H* accents, which occurred only in the first conjuncts of the parallel sentences, alternating with the H*L accents and marking continuation [13].

F0 parameters (the F0 range and the changes in the F0 maximum of the subject) differed significantly.

Table 4: *The significance (p-values) of the differences between the conditions in terms of the relative lengthening/shortening of the subject and the object (AT – aboutness topic, BF – broad focus).*

	Subject		Object	
	BF	**CT**	**BF**	**CT**
AT	<.0005	<.0005	<.0005	<.0005
BF		0.0183		0.0007

In order to further evaluate the discriminatory power of the emphatic realization of the accent, we performed a linear discriminant analysis (LDA), using the temporal and F0 parameters of the subjects and objects as predictors for the classification of the three information structure categories. Using the LDA automatic classification algorithm, the number of correct classifications (above chance) serves as a good approximation of the relative strength of an acoustic correlate of a category [23]. The results are presented in Figure 3. The F0 range and the changes in the F0 maximum yielded only 39% of correct classifications (chance = 33%). The relative duration of the subject and the object gave a much better classification (63%). They were the best predictors, although the precision was not very high. Also, Figure 3 shows that there are considerable overlaps between the three categories.

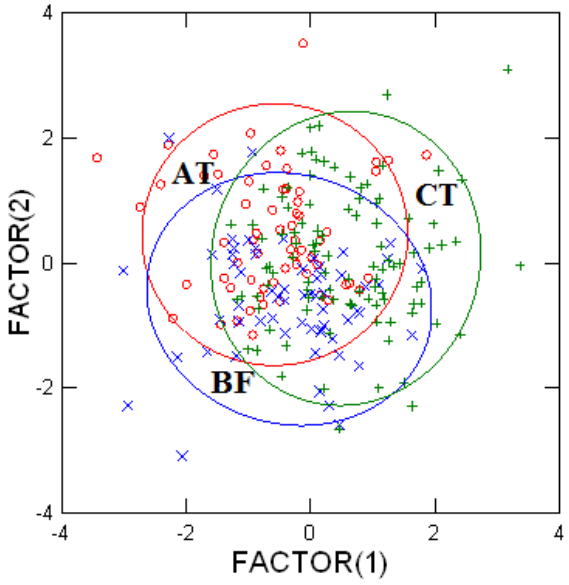

Figure 3: *Canonical scores plot of the discriminant analysis in a two factor space with confidence ellipses (AT – aboutness topic, BF – broad focus).*

In conclusion, emphasis is indeed a correlate of the CT in Estonian. Emphasis itself is expressed by lengthening in Estonian, as has been found previously [21], while F0 does not play a significant role, unlike in many other languages (cf. e.g. [10]). CT and aboutness topic sentences are distinguished by the lengthening of the topic in the former and its shortening in the latter. Although both of these sentence types contain a narrow focus, which is also signaled by emphasis in Estonian, it lengthens significantly more in the aboutness topic sentences, suggesting that the CT is also signaled by a reduced emphasis on the focus. CT and broad focus sentences in turn are distinguished by the lengthening of both the subject (CT) and the object (narrow focus) in the former. However, emphasis/lengthening does not make an absolute distinction between the information structure categories, as is shown by the relatively low percentage of the correct predictions and the considerable overlap between the categories in the LDA analysis. This can be explained by the fact that emphasis is a gradual phonetic feature and not a phonological category.

3.3. Prosodic phrasing

None of the conditions revealed a tendency for a pause occurring after the subject constituent. Only two sentences in the data display a pause after the subject, and it can be related to hesitation rather than boundary marking. It can thus be concluded that there is no evidence of a tendency to produce the CT as a separate prosodic phrase.

4. Conclusions

Estonian CT sentences are indeed prosodically distinguished from broad focus sentences and sentences with an aboutness topic and a narrow focus. This makes CT a more firmly established information structural category in Estonian than contrastive focus, which was found to be neither syntactically nor prosodically distinct from information focus.

The most consistent prosodic correlate of the CT was found to be the emphatic realization of the pitch accent carried by the CT constituent, accompanied by a reduced emphasis on the narrow focus (which is also signaled by emphasis in Estonian). As a result, a CT constituent is relatively longer than the corresponding constituent in broad focus and, especially, aboutness topic sentences. The narrow focus constituent in turn is relatively shorter than in aboutness topic sentences, but longer than the corresponding constituent in broad focus sentences. As predicted by the earlier studies on the acoustic correlates of emphasis in Estonian ([21]), emphasis was found to be expressed by lengthening rather than an increased F0 range or maximum.

As for pitch accent type, some speakers prefer to produce CT sentences with the marked contour usually involving an H* pitch accent on the topic and an HL* accent on the focus. However, the marked pattern is far from being restricted to CT sentences. Furthermore, CT is a much weaker determinant of accent choice than speech act type (question vs. declarative, cf. [13]), (non-)utterance-final position of the intonation phrase, and speaker preference. This is in line with the previous findings to the effect that pitch accent types are not strongly associated with information structural categories in Estonian.

CT was not found to be distinguished by separate prosodic phrasing.

5. Acknowledgements

We thank Eva Liina Asu-Garcia for having discussed with us the principles of pitch accent annotation. This research has been supported by the Centre of Excellence in Estonian Studies (CEES, European Regional Development Fund) and is related to the research project IUT35-1 (Estonian Research Council).

6. References

[1] D. Büring, "(Contrastive) Topic," in C. Féry and S. Ishihara (eds) *The Oxford Handbook of Information Structure*. New York: Oxford University Press, 2016, pp. 64–85.

[2] N. Kadmon, *Formal pragmatics: semantics, pragmatics, presupposition, and focus*. Malden: Blackwell, 2001.

[3] E. Titov, "Do contrastive topics exist?" *Journal of Linguistics*, 2013, 49, pp. 413–454.

[4] M. Wagner, "Contrastive topics decomposed," *Semantics and Pragmatics*, 2012, 5, pp. 1–54.

[5] D. Büring, *The meaning of topic and focus: The 59th Street Bridge accent*. London, New York: Routledge, 1997.

[6] C. Féry, "The prosody of topicalization," in K. Schwabe and S. Winkler (eds) *On Information Structure, Meaning and Form: Generalizations across languages*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2007, pp. 69–86.

[7] K. É. Kiss, *The Syntax of Hungarian*. Cambridge, New York: Cambridge University Press, 2002.

[8] M. Frascarelli and R. Hinterhölzl, "Types of topics in German and Italian," in K. Schwabe and S. Winkler (eds) *On Information Structure, Meaning and Form: Generalizations across languages*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2007, pp. 87–116.

[9] N. Hedberg and J. M. Sosa, "The Prosody of Topic and Focus in Spontaneous English Dialogue," in C. Lee, M. Gordon, and D. Büring (eds), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*. Dordrecht: Springer, 2007, pp. 101–120.

[10] B. Braun, "Phonetics and Phonology of Thematic Contrast in German," *Language and Speech*, 2006, 49 (4), pp. 451–493.

[11] B. Gyuris and K. Mády, "Contrastive topics between syntax and pragmatics in Hungarian: an experimental analysis," in R. Baglini, T. Grinsell, J. Keane, A. R. Singerman, and J. Thomas (eds) *CLS 46-1: Proceedings of the forty-sixth annual meeting of the Chicago Linguistic Society. The main session*. Chicago: Chicago Linguistic Society, 2014, pp. 147–162.

[12] K. Tael, *An approach to word order problems in Estonian*. Tallinn: Eesti Teaduste Akadeemia, 1990.

[13] E. L. Asu, *The phonetics and phonology of Estonian intonation*. PhD dissertation, University of Cambridge, 2004.

[14] E. L. Asu and F. Nolan, "The Analysis of Low Accentuation in Estonian," *Language and speech*, 2007, 50 (4), pp. 567–588.

[15] H. Sahkai, M.-L. Kalvik and M. Mihkla, "Prosody of contrastive focus in Estonian," *Proceedings of INTERSPEECH 2013*, Lyon, 2013, pp. 315–319.

[16] H. Sahkai, M.-L. Kalvik, and M. Mihkla, "Prosodic effects of Information Structure in Estonian," in E. L. Asu and P. Lippus (eds) *Nordic Prosody. Proceedings of the XIth Conference, Tartu 2012*. Frankfurt am Main: Peter Lang Verlag, 2013, pp. 323–332.

[17] H. Sahkai, M.-L. Kalvik, and M. Mihkla, "Perception of focus size and focus type in Estonian," in K. Jähi and L. Taimi (eds) *Proceedings of the XXVIII Fonetiikan päivät Turku*, 2014, pp. 72–79. <http://urn.fi/URN:ISBN:978-951-29-5980-8>

[18] N. Salveste, *On the pragmatic functions of Estonian sentence intonation*. PhD dissertation, Ludwig-Maximilians-Universität München, 2015.

[19] S. Skopeteas, I. Fiedler, S. Hellmuth, A. Schwarz, R. Stoel, G. Fanselow, C. Féry, and M. Krifka, "Questionnaire on Information Structure: Reference Manual," in *Interdisciplinary studies on information structure*, Vol. 4. Potsdam: Universitätsverlag Potsdam, 2006.

[20] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer. Version 5.3.14* [Computer program]. Retrieved on 13.04.2012 from http://www.praat.org/.

[21] M. Mihkla, H. Sahkai, and Mari-Liis Kalvik, "Acoustic correlates of emphasis in Estonian," *Proceedings of the 18th International Congress of Phonetic Sciences*, 2015, 1–5.

[22] I. Lehiste, "Segmental and syllabic quantity in Estonian," *American Studies in Uralic Linguistics*, Bloomington: Indiana University, 1960, pp. 21–82.

[23] V. J. van Heuven, "Acoustic correlates and perceptual cues of word and sentence stress: mainly English and Dutch," *Proceedings of 4th International Symposium on Tonal Aspects of Languages (TAL-2014)*, Nijmegen, 2014, pp. 211–217.