



Creak in the respiratory cycle

Kätlin Aare^{1,2}, Pärtel Lippus¹, Marcin Włodarczak², Mattias Heldner²

¹University of Tartu, Estonia

²Stockholm University, Sweden

{katlin.aare, partel.lippus}@ut.ee, {wlodarczak, heldner}@ling.su.se

Abstract

Creakiness is a well-known turn-taking cue and has been observed to systematically accompany phrase and turn ends in several languages. In Estonian, creaky voice is frequently used by all speakers without any obvious evidence for its systematic use as a turn-taking cue. Rather, it signals a lack of prominence and is favored by lengthening and later timing in phrases. In this paper, we analyze the occurrence of creak with respect to properties of the respiratory cycle. We show that creak is more likely to accompany longer exhalations. Furthermore, the results suggest there is little difference in lung volume values regardless of the presence of creak, indicating that creaky voice might be employed to preserve air over the course of longer utterances. We discuss the results in connection to processes of speech planning in spontaneous speech.

Index Terms: respiration, creaky voice, spontaneous Estonian, multiparty conversation, speech planning

1. Introduction

Creaky voice is a voice quality associated with tightly adducted vocal folds open along a portion of their length to allow for voicing. Acoustically, it is characterized by a series of irregularly spaced vocal pulses typically with a decreased acoustic intensity; a lower fundamental frequency; and a decreased spectral tilt compared to modal voice [1]. In this paper, we will use the term creak to denote both creaky voice and creak (i.e. without voice).

Apart from its typological and sociolinguistic variation, creaky voice has been extensively studied as a turn-taking, and in particular as a turn-yielding cue in conversations (see e.g. [2] and references mentioned therein). Creak is regularly located at phrase boundaries in several languages (e.g. in Finnish [2], Swedish [3], English RP [4], etc.). In Estonian, creak has not been found to be systematically used for signalling turn-yielding [5]. Furthermore, the presence of creak in Estonian appears to be governed by a number of time-related properties, such as lengthening and timing in words and phrases. In particular, creak co-occurs with (non-final) lengthening and later timing among the syllables in words as well as within phrases. It is possible that listeners interpret a creaky phrase end as a signal of turn end, triggering a turn change between interlocutors, and speakers are capable of manipulating the use of creak depending on their conversational goals [5].

However, creakiness can easily appear throughout an entire utterance. As such, it seems to be a relatively flexible phenomenon and not necessarily connected to f_0 declination. It has been hypothesized that generally f_0 declination is a passive result of gradually decreasing subglottal pressure during the course of ongoing speech activity [6]. A number of later studies claim that mechanisms involved in f_0 variation are more

complex and the variation cannot be explained by subglottal pressure changes alone. Importantly, it has been suggested that the results of these studies imply that f_0 changes are primarily controlled by laryngeal muscle activity rather than by respiratory activity (see [7] for a more thorough discussion).

Phonation types can additionally be characterized by glottal resistance (transglottal pressure differential divided by glottal airflow). Compared to modal voice, creak has been found to have higher average glottal resistance values due to increased vocal fold thickness, along with a significantly reduced airflow for both men and women [8]. The average airflow rates for creaky and modal voice have been reported to be approximately 40 ml/sec and 110 ml/sec, respectively, with no overlap in the individual subjects' average flow rates during the two types of phonation [9]. A study modelling airflow conservation and reduction of respiratory effort in phonation revealed that increasing glottal resistance allows finishing exhalations containing speech at a higher lung volume value, reducing the overall expiratory muscular pressure [10]. The authors concluded that in order to maintain a desirable value of subglottal pressure with a target of 800 Pa for the average time of a breath group in conversational speech (4 s), a deeper inhalation is necessary or the glottal resistance may need to be increased. However, the author pointed out that the conservation of airflow and respiratory effort is not normally a concern in conversational speech.

Communicative behavior is governed by the principle of spending least effort [11], or in other words, achieving minimally sufficient clarity [12]. Evidence has also been found in support of a flexibly incremental language planning system in which speakers tend to look for balance between planning and initiating speech quickly, indicating that speakers can plan larger portions of the utterance beyond the immediate phonological word [13]. Accordingly, speakers can continue the planning process of an utterance in a dynamic way during the ongoing utterance, choosing the optimal way of delivering their thought with minimum (physical) effort. A study on spontaneous conversations showed evidence for this in the form of pause locations in speech – according to [14], 72% of inhalations followed clause structure. In another study, inhalation locations indicated that speech is structured into breath groups (exhalations containing speech [6]), taking into account not only respiratory demands, but also grammatical structure. Hence, a breath group, i.e. a single exhalation containing speech tends to consist of relatively complete clauses, phrases or sentences, especially in read speech.

Given the characteristics of creaky voice, which include reduced air flow, it is plausible that this phonation type can function as a device for using respiratory resources more economically. When incorporating creak, the speaker can finish their utterance at a grammatically and semantically suitable place without needing to increase muscular effort. Muscular

effort increases rapidly after the speaker reaches the relaxation pressure lung volume level (REL), i.e. the balance point between internal and external air pressures when the breathing apparatus is in mechanically neutral position, after which a continued exhalation requires the increasing contraction of abdominal muscles. Conversational speech usually ends near REL, but may extend beyond that [15]. Importantly, the conversational situation contains several types of feedback mechanisms between the speaker and listener(s). Some of these can indicate that the listener is not following or understanding the speaker, in which case speakers can make their message more explicit. In circumstances where the speaker needs to adjust details, for example, by adding a few words or a phrase to make their meaning clearer, creaky voice could serve as a way to be able to finish longer utterances on the residual amount of air, without needing to increase muscular effort.

In this paper, we were interested in the interplay of creaky voice and respiratory activity in spontaneous Estonian three-party conversations to investigate the potential use of creak as an economy strategy. Firstly, we looked into what happens before an exhalation containing creaky speech and whether creakiness is somehow indicated in the preceding inhalation. We predicted that creak is not planned in the initial stages and, as such, would not be reflected in the lung volume values. Secondly, we explored how an exhalation containing creaky speech was different from an exhalation that did not contain creakiness. We examined this by assessing durations of the exhalations. Finally, we investigated where creak onset was located within exhalations in terms of both timing and lung volume level. If creak functions as a device for saving air, we expect it to be used whenever the speakers realize they need to minimize their air flow. Therefore, we anticipated that creak occurs before reaching REL and below REL, and that the presence of creak would lead to longer sentences.

2. Material & methods

The data acquisition was carried out by recording respiratory activity synchronized with audio and video in spontaneous three-party conversations, each lasting approximately 15-20 minutes. The conversations were recorded in a sound-treated room in the Phonetics Laboratory at Stockholm University.

The data set used here was gathered from 25 unique speakers (16F and 9M; $M = 24.5$ years, $SD = 3.6$) participating in 10 different conversations. The participants were all healthy native speakers of Estonian with an average Body Mass Index of 22.1. The speakers did not report any speech, language, hearing or respiratory disorders, and had never been smokers or professional singers. The speakers were asked to wear tight-fitting shirts. During the experiment, the speakers were standing around a 1-meter-high round table for the calibration and entire conversation duration, and were instructed to avoid large movements.

Respiratory activity was measured with Respiratory Inductance Plethysmography [16], which quantifies changes in the rib cage and abdomen cross-sectional areas by means of two elastic transducer belts (Ambu RIP-mate) placed at the level of the armpits and the navel. The overall lung volume change was estimated by isovolume manoeuvre [17]. A more detailed setup description is provided in [18].

The respiratory signal was recorded using PowerLab [19]. Audio was captured using head-worn microphones with a cardioid polar pattern (Sennheiser HSP 4). Each speaker was

facing a GoPRO Hero 3+ camera recording the upper part of the torso.

Annotation of the data was carried out semi-automatically using Praat [20] and Python scripts [21]. The sum of the rib cage and abdomen signals was used to segment the breathing signal into periods of inhalations and exhalations (for details see [22]). A total of 12.6 % of the automatically assigned borders were either moved or added manually due to some inaccuracy of the automatic annotation. Lung volume levels were normalized between 5 % – 95 % for each speaker with respect to the speaker’s speaking lung volumes measured during the entire recording phase. Similar to [22], REL was estimated dynamically as the median lung volume of valleys in the previous 20 cycles.

Creak has been identified automatically in the data set. Automatic creak detection is done in Matlab with Voice Analysis Toolkit, which is specially designed to carry out voice quality analysis [23]. Finally, all labels marked as creaky by automatic detection were checked by a trained phonetician.

We focused on exhalations containing speech, each of which could either contain creak or not, and measured their durations (s) as well as starting and ending lung volume values (% VC). Lung volume values were also recorded for inhalations preceding the exhalations that contained speech. For each inhalation, the amplitude was calculated by subtracting the starting lung volume value from ending value (% VC). Creak was treated as a binary factor based on its presence in exhalations containing speech. For each creak, we logged the duration (s), as well as starting and ending lung volume values (% VC). The distance of creak onset to REL was measured for each creak in terms of timing (s) and lung volume (% VC). The results were included in a binomial logistic regression model to predict the presence of creak in an exhalation containing speech. The statistical analysis was performed with R [24].

3. Results

Out of all analyzed exhalations containing speech ($N = 4085$), 46.8 % ($N = 1912$) contained creak. There was a considerable amount of individual variation. All exhalations containing speech are shown in Figure 1 for each speaker. The number of exhalations containing non-creaky speech and creaky speech are side by side for comparison.

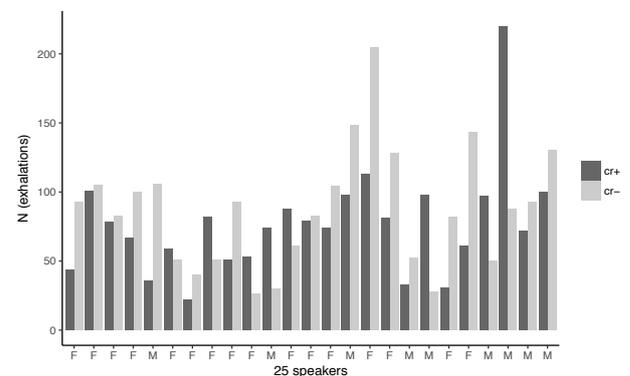


Figure 1: The number of exhalation phases containing speech produced with creak (cr+, dark grey) and without creak (cr-, light grey) per speaker. Speakers’ gender is marked with M (male) or F (female).

As can be seen in Figure 1, there was much variation in how many of the exhalations produced in the conversations contain creak. This was true for men and women alike. In addition, there was much variation in terms of how much time per exhalation was spent by speaking with a creaky voice, which was also individual and dependent on how long utterances and exhalations are. The statistical model in Table 1 describes the occurrence of creak in more detail.

Table 1: *The fixed effects of the logistic regression model estimating the likelihood of producing an exhalation without creak.*

	Est.	Std. Error	z value	Pr(> z) value
(Intercept)	1.646	0.091	18.090	< 0.001
Gender <i>M</i>	-0.577	0.070	-8.201	< 0.001
Exh_duration	-0.467	0.022	-21.236	< 0.001
Inh_amplitude	0.008	0.002	4.351	< 0.001

The base value of the intercept was taken to be a non-creaky exhalation produced by a woman. The results indicated that speech with creak was more likely to occur in a male speaker's speech, was more likely to be produced during a longer exhalation and was less likely if accompanied by a slightly lower inhalation amplitude before speech initiation. In the following sections, the two latter results are discussed in more detail, while the gender effect of male speakers exhibiting significantly more creak has been reported and confirmed by several earlier studies on Estonian, e.g. in [5], [24].

3.1. Exhalation duration

The model showed that exhalation duration plays an important role in determining the likelihood of creak occurrence in speech. In particular, creak was more likely if the exhalation containing speech was longer. Figure 2 illustrates the difference between the durations of exhalations that contained creak and those that did not.

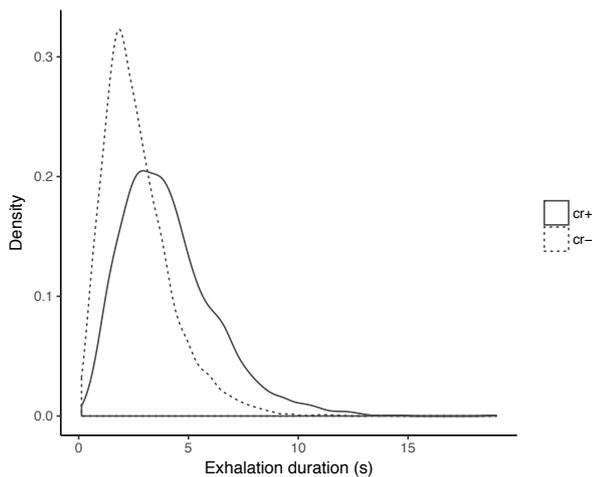


Figure 2: *Exhalation duration with creak (cr+, solid line) and without creak (cr-, dotted line).*

The results showed that while most exhalations could last up to 5 s, the exhalations that contained at least one interval of creak could last significantly longer than exhalations that did not contain creak.

3.2. Inhalation amplitude

As pointed out before, inhalation depth has been shown to correlate strongly to the upcoming exhalation duration in several languages, especially in read speech. To see whether the durational change in exhalations is indicated in the preceding inhalation, we compared the inhalation amplitudes ahead of exhalations consisting of speech with and without creak.

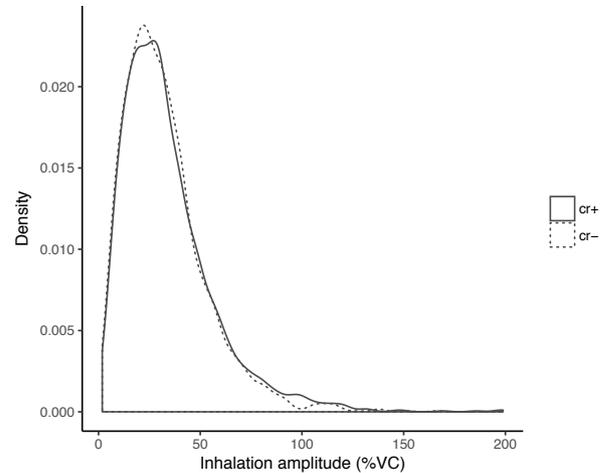


Figure 3: *Inhalation amplitude before exhalations with creak (cr+, solid line), and without creak (cr-, dotted line).*

The logistic regression model indicated a significant difference between the inhalation amplitudes before creaky and non-creaky speech. A larger inhalation amplitude was less likely to lead to the occurrence of creak during the speech spurt(s) in the following exhalation. Figure 3 shows that the difference between inhalation amplitudes was minute, although statistically significant.

3.3. Creak onset in relation to REL

The comparison of creak onset and REL lung volume values is shown on Figure 4.

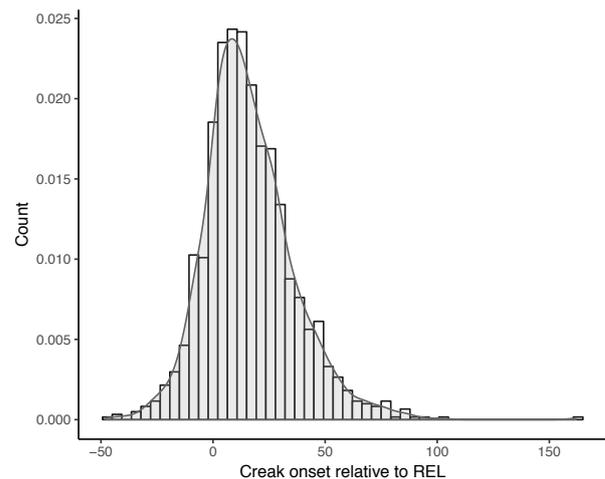


Figure 4: *Creak onset lung volume value (% VC) relative to the relaxation lung volume level (% VC), with 0 shifted to the value of REL (normalized by the speakers' individual lung volume ranges, individual dynamic REL values and the timing of reaching REL in an exhalation).*

In the exhalations that contained speech with at least one instance of creak, we wanted to see where the value of lung volume at the onset of creak was in relation to the respective relaxation lung volume values. If there were several creaks in the exhalation, we compared the data from only the first occurrence to REL.

The data showed that most creak onsets were located after reaching lung volume levels below REL. More specifically, during an exhalation, the tendency of creak onset was rising gradually while lung volume levels were approaching REL, with a sharp rise directly after reaching REL. After that, the probability of creak onset gradually declined. It can be seen that it was likely for creak to start relatively deep after passing REL. On the whole, the results showed that it is more likely for creak to start below REL.

3.4. Duration of exhalation below REL

In order to have a clearer understanding of how long exhalations lasted after passing REL, we investigated the duration of exhalations below REL. We analyzed the durations of all exhalations that contain speech and divided them into two groups based on the presence or lack of creak. The results are illustrated on Figure 5.

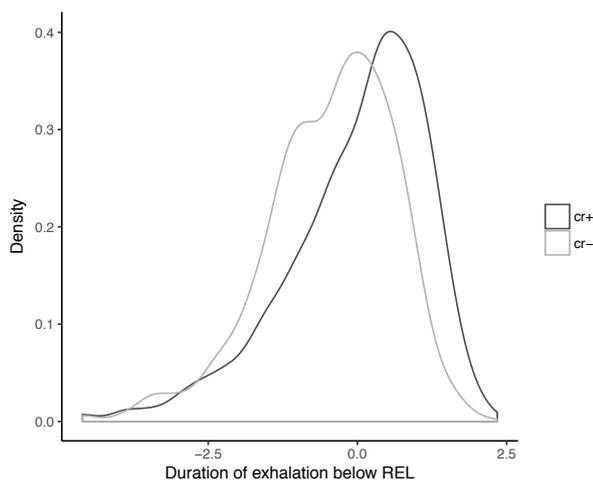


Figure 5: The durations of exhalation (log(s)) reaching below the relaxation lung volume level (normalized by the speakers' individual lung volume range). Exhalations with creak are shown in black, exhalations without creak in grey.

The density curves demonstrate that if the exhalation contained creak, it was more frequently longer after REL. In particular, the mode of the distribution was located after REL, while the mode for non-creaky exhalations was located slightly before reaching REL.

To test the difference of the two distributions, a t-test was performed. The variances of the distributions were equal ($F(1175) = .84$, 95% CI [.75, .94]). The t-test indicated a significant difference in the distributions of exhalation durations below REL for exhalations with non-creaky speech ($M = -.46$, $SD = 1.08$) and creaky speech ($M = -.01$, $SD = 1.18$), 95% CI [-.54, -.36].

4. Discussion

The results presented in this study suggest that the use of creaky voice in spontaneous Estonian can among sociolinguistic [25]

and prosodic [5] reasons be the result of avoiding physical discomfort and excess effort while trying to finish longer utterances. Understandably, these results do not cover the speaking behavior of speakers who use creak constantly for pathological reasons, nor does the corpus include speakers who use creak constantly for habitual reasons.

Firstly, in relation to exhalation duration, the results indicated that the exhalations that contained creak lasted longer than the exhalations containing non-creaky speech. If creak operates as an economy device, this would allow speakers to finish longer turns, regardless of whether the utterance is long due to speaking slowly or due to attempting to produce an utterance consisting of a large amount of smaller speech units. This result indicated there is basis to look deeper into the variation of creak and its effects on other respiratory properties with regard to speech planning.

Secondly, by investigating read speech, it has been well established that speech planning is reflected in respiratory patterns through the strong correlation of inhalation amplitudes and following exhalation durations. This is true for spontaneous speech as well, although to a smaller extent. Therefore, accordingly to our hypothesis that creak is not planned in the initial stages of speech planning, we did not expect to see differences in inhalation amplitudes. While the changes in inhalation amplitudes were statistically significant in the regression model, they were also very small, allowing us to interpret it as support for the hypothesis. As such, creak can accommodate for increased articulatory demands during ongoing speech planning in running speech.

Finally, the lung volumes and timing of creak onset in relation to REL show that creak allows speaking below REL longer than without creak. We only analyzed the onset of the first interval of creak – if there were more per exhalation, they were located even later in the exhalation. This result serves as additional support for the hypothesis that speakers can speak longer than normally by using creak. Furthermore, as creak can occur before reaching REL, speakers seem to use it relatively flexibly, indicating that creak is not necessarily bound to REL.

5. Conclusions

The purpose of this study was to observe the occurrence of creak in spontaneous Estonian in relation to the respiratory cycle. Our results give support for the hypothesis that creak can be employed as an air preservation strategy over the course of longer sentences that extend below REL. Exhalations containing speech with creak were found to be longer, but have similar inhalation amplitude in comparison to exhalations that contained non-creaky speech. Creak onset was in large part located around REL in terms of timing and lung volumes. We acknowledge that the topic needs more research beyond what this small corpus on Estonian spontaneous speech allows, including studies on read speech and in relation to the content of speech in order to verify current findings.

6. Acknowledgements

This work was funded by the Swedish Research Council project 2014-1072 *Andning i samtal (Breathing in conversation)* and supported by the Estonian Research Council grant IUT2-37, the National Program for the Estonian Language Technology project EKT71.

7. References

- [1] M. Gordon and P. Ladefoged, "Phonation types: a cross-linguistic overview," *Journal of Phonetics*, 29, no. 4, pp. 383–406, 2001.
- [2] R. Ogden, "Turn transition, creak and glottal stop in Finnish talk-in-interaction," *Journal of the International Phonetic Association*, 31, no. 1, pp. 139–152, 2001.
- [3] R. Carlson, J. Hirschberg, and M. Swerts, "Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates," *Speech Communication*, 46, no. 3–4, pp. 326–333, 2005.
- [4] J. Laver, *Principles of phonetics*. Cambridge; New York, NY: Cambridge University Press, 1994.
- [5] K. Aare, P. Lippus, and J. Šimko, "Creak as a feature of lexical stress in Estonian." *Proc. Interspeech 2017*, pp. 1029-1033, 2017.
- [6] P. Lieberman, *Intonation, perception, and language*; MIT Research Monograph, 1967.
- [7] C. Petrone, S. Fuchs, and L. L. Koenig, "Relations among subglottal pressure, breathing, and acoustic parameters of sentence-level prominence in German", *The Journal of the Acoustical Society of America* 141, no. 3, pp. 1715-1725, 2017.
- [8] M. Blomgren, Y. Chen, M. L. Ng, and H. R. Gilbert, "Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers", *The Journal of the Acoustical Society of America* 103, no. 5, pp. 2649-2658, 1998.
- [9] T. Murry, "Subglottal pressure and airflow measures during vocal fry phonation", *Journal of Speech, Language, and Hearing Research* 14, no. 3, pp. 544-551, 1971.
- [10] Z. Zhang, "Respiratory laryngeal coordination in airflow conservation and reduction of respiratory effort of phonation", *Journal of Voice* 30, no. 6, pp. 760-e7, 2016.
- [11] G. K. Zipf, *Human behaviour and the principle of least-effort*, Cambridge MA edn, Reading: Addison-Wesley, 1949.
- [12] B. Lindblom, "Explaining phonetic variation: A sketch of the H&H theory", *Speech production and speech modelling*, pp. 403-439. Springer, Dordrecht, 1990.
- [13] F. Ferreira and B. Swets, "How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums", *Journal of Memory and Language* 46, no. 1, pp. 57-84, 2002.
- [14] A. L. Winkworth, P. J. Davis, R. D. Adams, and E. Ellis, "Breathing patterns during spontaneous speech", *Journal of Speech, Language, and Hearing Research* 38, no. 1, pp. 124-144, 1995.
- [15] T. J. Hixon, M.D. Goldman, and J. Mean, "Kinematics of the chest wall during speech production: Volume displacements of the rib cage, abdomen, and lung", *Journal of Speech, Language, and Hearing Research*, 16(1), pp.78-115, 1973.
- [16] H. Watson, "The technology of respiratory inductive plethysmography", *Proceeding of the Second International Symposium on Ambulatory Monitoring (ISAM 1979)*, pp. 537-563, 1980.
- [17] K. Konno and J. Mead, "Measurement of the separate volume changes of rib cage and abdomen during breathing", *Journal of applied physiology* 22, no. 3, pp. 407-422, 1967.
- [18] J. Edlund, M. Heldner, and M. Włodarczak, "Catching wind of multiparty conversation", pp. 35-36, 2014.
- [19] ADInstruments, LabChart software and PowerLab hardware (Version 8). New South Wales, Australia: ADInstruments, 2014.
- [20] P. Boersma and D. Weenink, Praat: doing phonetics by computer [Computer program], version 6.0.37, retrieved from <http://www.praat.org/>, 2018.
- [21] H. Buschmeier and M. Włodarczak, "TextGridTools: A TextGrid processing and analysis toolkit for Python", *Tagungsband der 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV 2013)*, 2013.
- [22] M. Włodarczak and M. Heldner, "Respiratory Constraints in Verbal and Non-verbal Communication", *Frontiers in psychology* 8, pp. 708, 2017.
- [23] T. Drugman, J. Kane, and C. Gobl, "Data-driven detection and analysis of the patterns of creaky voice," *Computer Speech & Language*, 28, no. 5, pp. 1233–1253, 2014.
- [24] R Core Team, R: A language and environment for statistical computing. [computer program] R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>, 2015.
- [25] K. Aare and P. Lippus. "Some gender patterns in Estonian dyadic conversations." In *Nordic prosody. Proceedings of the XIIIth conference, Trondheim*, pp. 29–38, 2017