BREATH HOLDS IN SPONTANEOUS SPEECH

Kätlin Aare[1, 2], Marcin Włodarczak[1], and Mattias Heldner[1]
[1]Stockholm University and [2]University of Tartu

Abstract. This article provides a first quantitative overview of the timing and volume-related properties of breath holds in spontaneous conversations. Firstly, we investigate breath holds based on their position within the coinciding respiratory interval amplitude. Secondly, we investigate breath holds based on their timing within the respiratory intervals and in relation to communicative activity following breath holds. We hypothesise that breath holds occur in different regions of the lung capacity range and at different times during the respiratory phase, depending on the conversational and physiological activity following breath holds. The results suggest there is not only considerable variation in both the time and lung capacity scales, but detectable differences are also present in breath holding characteristics involving laughter and speech preparation, while breath holds coinciding with swallowing are difficult to separate from the rest of the data based on temporal and volume information alone.

1. Introduction

Speech breathing commonly refers to the special manner of using the respiratory mechanisms to produce airflow for speech sounds. During speech, the rate and volume of inhalation and rate of exhalations are mostly governed by the speech controlling system in the cerebral cortex (von Euler 1982); this takes into account requirements for phrasing, loudness, and articulation. Breathing in general and speech breathing consist of cycles containing an inhalation and exhalation phase. In quiet breathing, the cycles repeat 12 times per minute on average, with relatively equal inhalation and exhalation phases. By contrast, speech breathing cycles are commonly characterised by relatively short inhalations and relatively long exhalations mostly due to higher resistances in the upper airway during speech production, which

prevents air from simply flowing out (see Figure 1 for examples) (Hixon 1987: 45–46). Speech breathing operates in the lower midrange of the vital capacity (henceforth VC, defined as the maximum volume of air that can be exhaled after a maximum inhalation). According to Hixon (1982) and Hixon et al. (1973: 93), conversational speech range is normally around approximately 40–60% of VC, with speech mostly initiated at 50–60% VC and terminated approximately at 30–50% VC. These changes can be traced by using Respiratory Inductance Plethysmography (Watson 1980), which uses two elastic belts to quantify changes in the circumferences of the chest and abdomen (illustrated on the bottom signals in Figure 1).
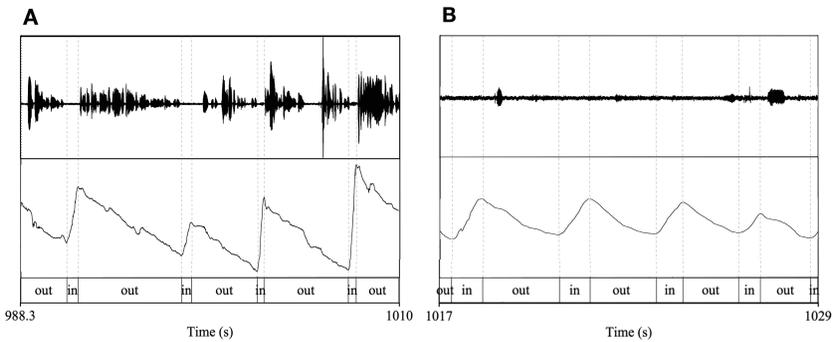

Figure 1. Excerpts from a spontaneous conversation with the speech (top) and breathing (bottom) signals of one participant. The text level on the panels should be interpreted as follows: "in" stands for inhalation and "out" stands for exhalation. The absolute values of the signals on the two panels are not calibrated in absolute units and should not be compared to each other.

In addition to the inhalation and exhalation phases, the respiratory signal can include portions where air is neither inhaled or exhaled – breath holds (see Figure 2). These are often overlooked or incorporated into the inhalation or the exhalation phase. Breath holds are considered a special type of breathing, as they are not directly controlled by the brainstem in one of its primary functions of maintaining the optimal gas exchange system for life purposes (Shea 1996 and von Euler 1982). Special acts of breathing can be voluntary, like breath holds or breathing exercises, they can be practiced, such as for wind instrument playing, or they can be driven by emotions, for instance, during crying

or laughter. Hixon et al. (2018) explain that such special breathing acts are controlled by higher brain centers that override or bypass the activity of the brainstem in maintaining control over breathing. This is only possible for short periods at a time: according to Schneider (1930: 467), "it is practically impossible for a man at sea level to voluntarily hold his breath until he becomes unconscious", and this has been supported by later research (e.g., Cooper et al. 2003). Breath holds are categorised as voluntary, conscious, and are known to be controlled by the cerebral cortex, similar to speech breathing in general.
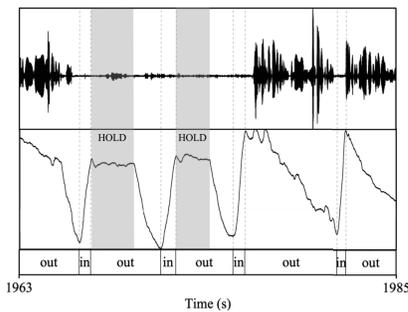

Figure 2. Excerpt from a spontaneous conversation with the speech (top) and breathing (bottom) signals of one participant. The text level on the panels should be interpreted as follows: "in" stands for inhalation and "out" stands for exhalation. Breath holds are marked with a grey box and the label "hold".

According to Parkes (2006), it is not exactly clear what happens to respiratory muscles when people hold their breath. The author points out that, in breath holding at high lung volumes, there may be some voluntary muscular contribution to keep the chest open and air from flowing out. Further, breath holds cannot be explained by closure of the glottis and airway, as breath holds can be continued with these structures open (Godfrey et al. 1969). Parkes (2006) notes that, while the precise activity of the respiratory muscles has not been established, the diaphragm has been reported to contribute as a potential "holding" muscle after numerous experiments (see Parkes 2006 for an overview). Importantly, Hårdemark-Cedborg et al. (2009) found that diaphragmatic and abdominal muscles display demonstrable electrical activity, indicating that breath holds are active in nature and not simply stops in respiratory activity. In addition to the uncertainty around the mechanism

behind the realisation of breath holds, there is also great variation involved in the maximum duration of breath holds from subject to subject and even within subjects in different experiment conditions. Godfrey and Campbell (1968) report that this variation is partly due to psychological factors but can also be traced to chemical and mechanical factors. For example, the authors note that breath may be held longer at maximum lung capacity than at resting level, while Muxworthy (1951) has shown that the relationship between lung volume and maximum breath holding time is linear.

In conversation, some parallel physiological processes are relevant for understanding the occurrence of breath holds. One of them is swallowing, which is always accompanied by a breath hold. A laryngeal closure is an essential mechanism to protect the airway during swallowing to prevent aspiration of foreign materials into the lungs, especially during eating and drinking (Cherniack 2009). It has been reported that like most speech, most swallowing takes place during exhalations (Smith et al. 1989). Similarly, the larynx is closed during other defensive airway reflexes common in naturally occurring speech, such as before the onset of the exhalation phase in a cough or sneeze (Poliacek et al. 2016).

Although little evidence exists, breath holds have also been hypothesised to be important for turn organisation in conversations. It has been suggested that breath holding may function as a marker of turn incompleteness (French and Local 1983). Local and Kelly (1986) discuss "holding silences", which refer to intra-turn pauses following conjunctionals, ending in glottal closures maintained through silence and released at the beginning of the following word produced by the same speaker. Jefferson (1983) notes that if the current speaker produces this "holding silence" between their consecutive speech spurts, it would still be considered a maintained turn even if overlapping talk occurs. At the moment, breath holds are commonly not included in work on communicative functions of breathing or conventions as they are not easily identifiable from audio alone and require special equipment to detect.

The purpose of this study is to describe the occurrence of breath holds in spontaneous multiparty conversations held in Estonian. Since little is known about breath holds in communicative situations, in this preliminary study we provide a basic statistical description of the phenomenon, leaving statistical modelling for the future. Breath holds

are measured in terms of their durational properties and timing within the respiratory phase with which they coincide, as well as by determining their location within the vital capacity range. We want to determine if different categories of breath holds, which are potentially associated with different functions (such as breath holds coinciding with swallowing, breath holds related to speech preparation, etc.), display different patterns in terms of placement in the breathing signal and/or timing in the speech flow. In particular, we hypothesise that speech- and laughter-preparatory breath holds occur high in the lung capacity range and early in the respiratory interval. Secondly, we assume the pre-production holds (e.g., before any speech from the speaker in the same respiratory cycle) will occur higher and earlier than between-production (e.g., between two speech spurts from the same speaker in the same respiratory cycle) preparatory breath holds. Additionally, we hope to see that breath holds coinciding with swallowing can be separated from others based on their temporal and volume-related properties. The results of this study could be used to develop speech technology applications, such as conversational agents, and to extend established conversation analytical methods by using the breathing signal.

2. Material and method

This section describes the data under observation and the procedures applied during data collection and processing.

2.1. Material

The data consisted of three-party conversations held in Estonian, deposited to the Phonetic Corpus of Estonian Spontaneous Speech at the University of Tartu. The data were collected from 30 unique speakers (17F and 13M; M_{age} = 25.1 years, SD_{age} = 3.4) participating in 10 different conversations. Participants were all healthy native speakers of Estonian with an average Body Mass Index of 22.5 (SD = 2.5). The participants in each conversation knew each other well. The conversations lasted about 15–20 minutes. No speakers reported speech, language, hearing, or respiratory disorders, and none of them had been smokers or professional singers.

2.2. Method

The conversations were recorded in a sound-treated room in the Phonetics Laboratory at Stockholm University (see Figure 3 for recording set-up). Respiratory signals, audio and video were recorded synchronously for each participant. The participants were asked to wear tight-fitting clothes on their upper body for the recording procedure in order to maximise the accuracy of the signals from the transducer belts. The participants stood around a 1-meter-high round table, facing each other for the entire calibration and experiment duration. The participants were instructed to avoid large movements to minimise disrupting the respiratory signals.

Figure 3. Recording set-up.

Respiratory activity was captured with Respiratory Inductance Plethysmography, using two elastic transducer belts (Ambu RIP-mate). The belts were placed at the level of the armpits and the navel. The overall lung volume change was estimated by asking the participants to perform the isovolume maneuver (Konno and Mead 1967). The technical setup is described in more detail in Edlund et al. (2014).

The respiratory signal was recorded with PowerLab (ADInstruments). Audio was recorded with head-worn microphones with a cardioid polar pattern (Sennheiser HSP 4). Each speaker was facing a GoPRO Hero 3+ wide angle camera to record movements of the upper part of the torso and the head.

The audio signal was annotated semi-automatically by first automatically identifying sections of speech and silence using an intensity-based threshold in ELAN and then manually adjusting the resulting interval borders as well as changing interval labels according to their function, i.e. separating short feedback vocalisations (backchannels) from the rest of speech intervals, marking laughter and coughing, etc. in Praat (Boersma and Weenink 2019). The reason backchannels were labelled separately is that they are normally not considered as separate turns in conversation organisation (Yngve 1970). Annotation of the respiratory data was carried out automatically with Praat and Python scripts (Włodarczak 2019). The weighted sum of the rib cage and abdomen signals was used to automatically segment the breathing signal into periods of inhalations and exhalations. A total of 12.6% of the automatically assigned borders were either moved or added manually due to some inaccuracy in the automatic annotation.

Breath holds were identified manually from the respiratory signal during silence and marked if their duration was at least 250 ms. Breath holds which occurred in the same respiratory phase as speech or laughter were associated with speech- or laughter-preparatory activity and were assigned a dedicated tag to mark it. Additionally, video footage from the conversations was checked to mark breath holds coinciding with swallowing with a dedicated tag. This was possible for 27 participants, as for 3 participants there were no video recordings available due to technical reasons.

3. Results

In total, there were 1008 breath holds in the data set on which we present results. Prior to reaching this amount, a number of breath holds were filtered out due to annotation mistakes, i.e. if their boundaries exceeded the boundaries of the coinciding respiratory phase and thus spun over two phases. One breath hold coincided with a yawn, but was excluded due to infrequency of this category. The only breath hold produced in an inhalation phase was also excluded. Consequently, all breath holds in the resulting data set were produced during the exhalation phase.

Of the final sample, 109 (11%) were associated with swallowing ("sw") on the basis of video footage. Preparatory breath holds were divided into two groups: the first ("prep.init") included 663 (66%) breath holds that occurred before the first speech/laughter interval during the exhalation, the second ("prep.pause") included 98 (10%) breath holds which occurred between two speech/laughter segments in the same exhalation. The leftover 138 (14%) breath holds were uncategorised ("other") and occurred during a longer stretch of silence, after speech production, together with smacking or coughing, etc. All participants produced breath holds (between 10–82 per person, $M = 34.7$). There was some inter-speaker variation involved in the occurrence of breath-hold categories: median number of occurrences per speaker for breath holds coinciding with swallowing was 3 ($M = 4.4$, $SD = 4.7$); for uncategorised breath holds the median was 4 ($M = 4.9$, $SD = 3.5$), for preparatory initial breath holds, the median value was 21 ($M = 22.9$, $SD = 13.1$), and for preparatory-pause breath holds, the median was at 3 ($M = 4.1$, $SD = 2.6$).

3.1. Breath holds in lung capacity

Firstly, we look at the volume-related properties of breath holds. Breath holds were measured in terms of the relative amount of air in the lungs at the onset of their occurrence. Breath hold position was normalised within the minimum and maximum values of the exhalation it coincided with. Figure 4 summarises the findings in the four identified breath hold categories.
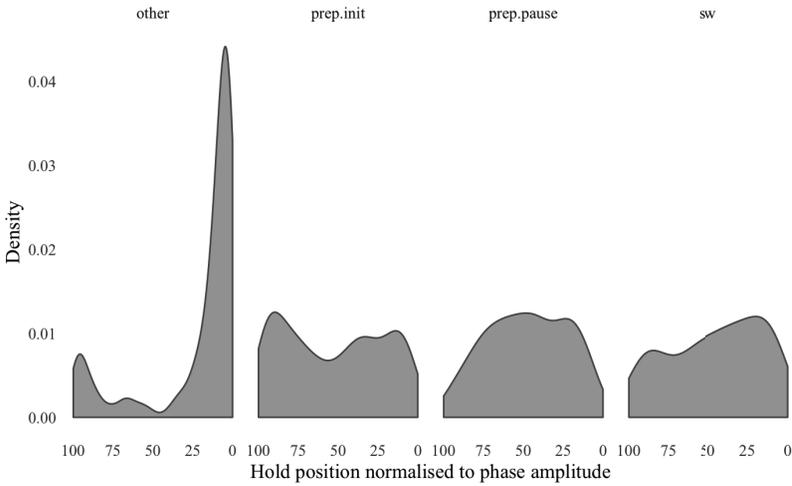
Figure 4. Densities of breath holds within the minimum and maximum values of the respiratory interval with which they coincided. Breath holds divided into four categories: "other" for uncategorised holds, "prep.init" for preparatory initial holds, "prep.pause" for preparatory pause holds, and "sw" for holds coinciding with swallowing. 0 and 100 values on the x-axis correspond to the top and bottom values of the respiratory phase amplitude, respectively.

The distributions for preparatory initial and preparatory pause breath holds are markedly different, as most preparatory initial breath holds occur either at the bottom half of the respiratory interval amplitude or at the very top. Preparatory pause breath holds, on the other hand, mostly tend to occur in the midsection of the respiratory range. Breath holds coinciding with swallowing are most frequent between the bottom 20–30% of the respiratory phase amplitude. Uncategorised breath holds are very frequent up to the bottom 20%, but exhibit a second, smaller mode at the top 90%.

Additionally, because the preparatory pause category contained much fewer data points (93 vs 615 in preparatory initial category), we pooled the two preparatory categories together to see if there were differences between the hold positions before speech-, backchannel-, and laughter-initial breath holds (Figure 5 illustrates the results). The pooled categories amount to 708 breath holds followed by 115 backchannels, 164 laughters, and 429 speech spurts.
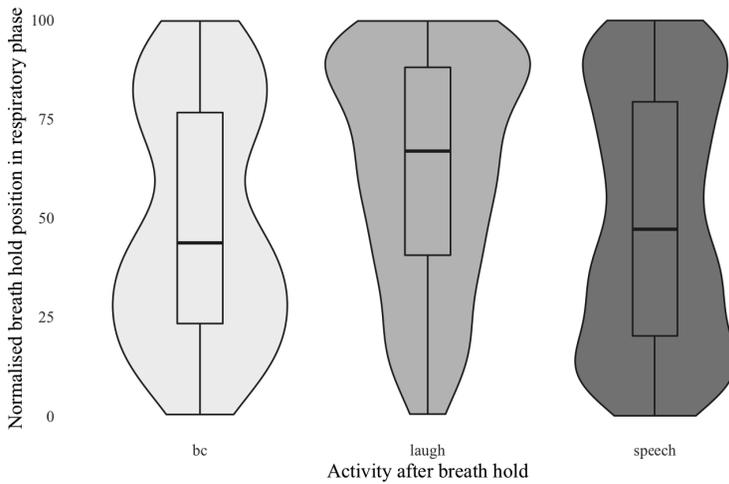
Figure 5. Breath hold positions in the respiratory phase before speech, laughter, or backchannels ("bc"). Breath hold position is normalised to the start and end values of the coinciding respiratory phase. The plot combines so-called violin shapes to describe the distribution of the data and boxplots to represent information about the main statistics of the sample.

Breath holds before laughter, on average, were produced at the highest level of the three categories, with a median at 67% ($M = 62.2$, $SD = 28.3$), while breath holds before speech and backchannels had bimodal shapes. Breath holds before backchannels have their modes around 25–30% and 80% of the respiratory phase amplitude and breath holds before speech have modes around 15–20% and 80–90% of the respiratory phase amplitude.

3.2. Temporal properties of breath holds

Secondly, we looked at timing-related characteristics of breath holds. Breath holds were measured in terms of where they started and ended within the exhalation they coincided with, as well as in terms of duration. Additionally, for preparatory breath holds distance from following speech or laughter was measured. Figures 6 and 7 summarise findings for all breath holds.

Measurements of start and end time of breath holds were normalised with respect to the start and end times of the coinciding respiratory interval.
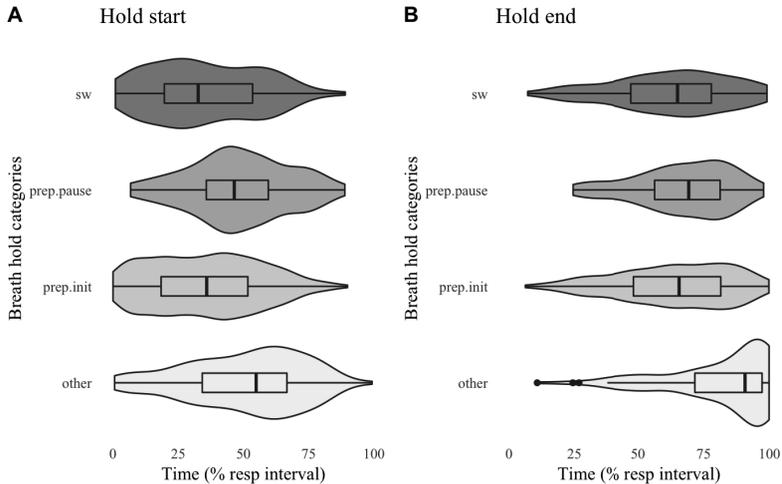

Figure 6. Breath hold start (panel A) and end (panel B) times relative to the start and end times of the coinciding respiratory interval. Breath hold categories are as follows: breath holds coinciding with swallowing ("sw"), preparatory initial breath holds ("prep.init"), preparatory pause breath holds ("prep.pause"), uncategorised breath holds ("other").

Normalised breath hold start median values (shown in Figure 6, panel A) were as follows: 33% ($M = 35.1$, $SD = 22.3$) for breath holds coinciding with swallowing, 36% ($M = 35.5$, $SD = 20.9$) for preparatory initial breath holds, 47% ($M = 48.4$, $SD = 19.1$) for preparatory pause breath holds, and 55% ($M = 50.4$, $SD = 22.3$) for uncategorised breath holds. The distributions of the breath hold start values are unimodal with modes around 25% for breath holds coinciding with swallowing, 65% for uncategorised breath holds, and 45% for both preparatory initial and pause breath holds. However, the preparatory initial breath holds show a relatively uniform distribution from the beginning of the respiratory interval up to its mode, whereas the preparatory pause breath holds exhibit a more distinct mode from the rest of the data.

Normalised breath hold end median values (shown in Figure 6, panel B) were 67% ($M = 62.8$, $SD = 22.5$), 66% ($M = 64.1$, $SD = 22.1$), 69% ($M = 66.8$, $SD = 18$), and 92% ($M = 83.4$, $SD = 19.8$), respectively. The distributions show differences from the relative breath hold start time values. Breath hold end time values show modes around 85% for both preparatory initial and preparatory pause breath holds, 75% for breath holds coinciding with swallowing, and 95% for uncategorised breath holds.

Duration of breath holds was measured in seconds, but also relative to the entire exhalation duration (Figure 7).
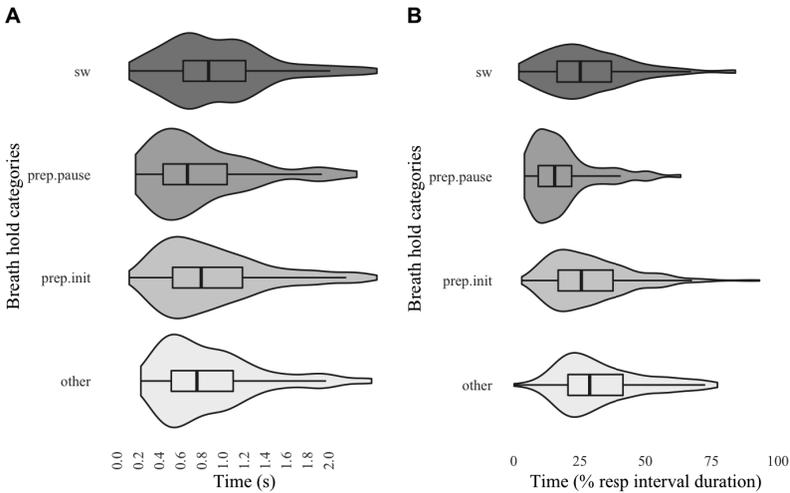

Figure 7. Absolute (panel A) and normalised (panel B) breath hold durations (outliers of more than 1.5* interquartile range were excluded from both panels for presentation purposes). Breath hold categories were as follows: breath holds coinciding with swallowing ("sw"), preparatory initial breath holds ("prep. init"), preparatory pause breath holds ("prep.pause"), uncategorised breath holds ("other").

Measuring absolute breath hold durations (shown in Figure 7, panel A) provided median values of 0.82 s ($M = 1.1$, $SD = 0.96$) for breath holds coinciding with swallowing, 0.87 s ($M = 1.1$, $SD = 0.87$) for preparatory initial breath holds, 0.74 s ($M = 1.1$, $SD = 1.1$) for preparatory pause breath holds, and 0.88 s ($M = 1.3$, $SD = 1.2$) for uncategorised breath holds. The data is distributed as follows: the

modes are around 0.5–0.6 s for preparatory initial, preparatory pause, and uncategorised breath holds, and around 0.8 s for breath holds coinciding with swallowing.

The median values of normalised breath hold durations (shown in Figure 7, panel B) were 25% ($M = 27.8$, $SD = 16.5$) for breath holds coinciding with swallowing, 26% ($M = 28.6$, $SD = 15.3$) for preparatory initial breath holds, 15% ($M = 18.3$, $SD = 12.7$) for preparatory pause breath holds, and 29% ($M = 33$, $SD = 16.3$) for uncategorised breath holds. Relative breath hold times had modes around 25% of the respiratory interval duration for breath holds coinciding with swallowing and uncategorised breath holds, 20% for preparatory initial breath holds, and 10–15% for preparatory pause breath holds.

Additionally, the distance of preparatory initial breath hold start and end times from the beginning of following speech or laughter was measured. The median values for the distance for breath hold start and end values were 2.4 s ($M = 4.4$, $SD = 5.6$) and 1.1 s ($M = 3.3$, $SD = 5.6$), respectively.

4. Discussion

The aim of this study was to analyse temporal and volume-related data in order to identify patterns involving breath holds in spontaneous speech. This was achieved by measuring breath holds on two scales: time and lung capacity. The analysis demonstrated considerable variation on both scales. In addition, the occurrence of breath holds was assessed on the basis of their communicative context – namely, if they were preceded or followed by speech, laughter or silence, or if they coincided with parallel physiological processes common in naturally occurring speech, such as swallowing. Here, too, the results demonstrated much variation between the categories.

The analysis focusing on the volume-related properties of breath holds compared where breath holds occurred within the lung capacity range. More specifically, the location of each breath hold was normalised between the minimum and maximum value of the exhalation.

Comparison between the preparatory initial and preparatory pause categories attempted to determine whether the pre-speech breath holds occurred higher in the range than between-speech breath holds. The results demonstrated a difference that partly corresponds to our

hypothesis: a large proportion of pre-production holds occurs towards the top the exhalatory range. However, in addition to the mode around 95%, preparatory initial breath holds have a second, lower mode spanning around a third of the bottom range from 10 to 30%. One possible explanation for this second mode could be that people simply started speaking very late in the respiratory range, producing short speech spurts, such as backchannels (which accounted for around a sixth of the breath holds in the prep.int category). The data offer another explanation pointing to laughter instead, as there was actually a decrease in the amount of backchannels produced around the second, 10–30% mode (from 18% around the first mode to 14% around the second); and a similar pattern can be observed for following speech intervals, but a considerable increase was found for laughter (from 18% around the first mode to 30% around the second mode). By contrast, preparatory pause breath holds have one mode around 25–60% of the respiratory phase range, fittingly allowing a second speech spurt after the pause.

The respiratory range analysis results also indicated that breath holds coinciding with swallowing displayed a different distribution: breath holds during swallowing are more frequent towards the bottom range of the exhalation with a mode around 20–35%. On the basis of these results, swallowing would be difficult to distinguish from other breath holds by considering respiratory range only. Some clarification could be provided by zooming in on the breath holds coinciding with swallowing and investigating their communicative context further to establish potential factors affecting their occurrence. In this data set, it appears that breath holds coinciding with swallowing are immediately preceded by speech in the same respiratory interval for 60% of the occurrences, with an additional 19% of breath holds preceded by backchannels and 11% by laughter. This information gives a potential explanation to why the breath holds coinciding with swallowing overlap with the second mode of the preparatory initial breath holds. Two-thirds of breath holds are immediately followed by speech in the same respiratory interval, which can account for the partial overlap of distributions in the bottom range of the exhalation with preparatory pause breath holds.

The uncategorised breath holds are connected to a mixture of different factors, but one of the most common features they have is that they are often produced after speech production or laughter. In fact, many of them are produced towards the very end of the exhalation

which, in most cases, coincides with the resting level of the respiratory apparatus (an equilibrium where there is balance between the lungs pulling inward and the rib cage pulling outward). It is easy to reach this position by producing a sigh and letting air flow out on its own – the resulting position is the equilibrium. Considering the physical ease and comfort in this position, it is perhaps not surprising to find breath holds so low in the respiratory range of the coinciding phase. In this case, the breath hold occurs after the exhalation and potentially functions as a resting position before a new inhalation. An example of this from the data set is provided on Figure 8. A factor which might be partly responsible for the second mode in the upper 90–100% level could have to do with speech planning and conversational turn organisation. Namely, these breath holds could represent situations where a person in a conversation inhales as they intend to start speaking or plan to take the turn but fail at it and produce a silent exhalation instead. An example is provided on Figure 9. Both Figures 8 and 9 include speech activity from all three participants to provide conversational context.

Additionally, we looked at how different communicative activities affect the position of the breath hold in the respiratory range. Specifically, we focused on laughter, speech, and backchannels – very short vocal feedback units. The breath holds were again measured in relation to the exhalation amplitude. The results showed that, while all three categories exhibited much variation, on average, laughter took place much higher in the relative respiratory range than speech or backchannels. Furthermore, speech and backchannels were relatively similar in their bimodal distribution and median value. Backchannels have been found to be frequently produced on residual breath late in the respiratory cycle (Włodarczak and Heldner 2017), and as such, the bimodal shape of the distribution of the breath holds preceding them is not a surprise but rather a confirmation of earlier results. Arguably, breath holds preceding speech might have the same feature provided the speech spurt is short enough to be produced on residual air, as it has also been shown that backchannels cannot be distinguished from other very short utterances on the basis of respiratory properties alone (Aare et al. 2014). The data show a tendency for this, but 10% of speech intervals lasting for over 50% of the respiratory interval still begin as low as 70% in the exhalatory range. It is possible that speakers use methods like laryngeal constriction to diminish their air use.
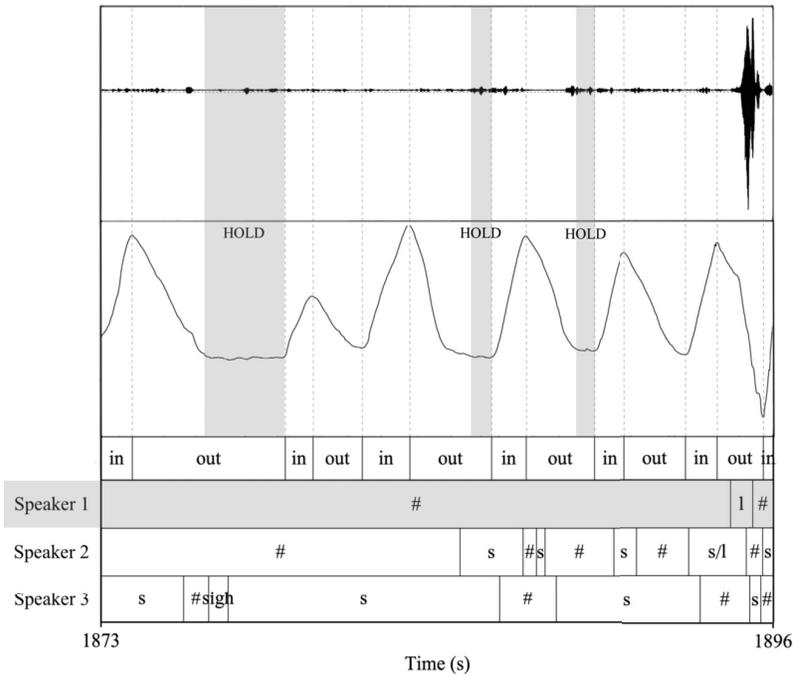
Figure 8. An excerpt from a conversation with the speech and breathing signals of Speaker 1, showing an example of uncategorised breath holds occurring low in the respiratory range. On the topmost tier, "in" and "out" correspond to inhalation and exhalation phases of Speaker 1, respectively. Breath holds are indicated with the label "hold" and a grey box over the speech and respiratory signal. The second tier shows speech activity from Speaker 1, the third tier shows speech activity from Speaker 2, and the bottom tier shows speech activity from Speaker 3. On all speaker tiers, "s" marks speech, "#" marks pauses, "l" marks laughter, "s/l" marks speech/laughter, "sigh" stands for sighing, and "bc" for backchannels.
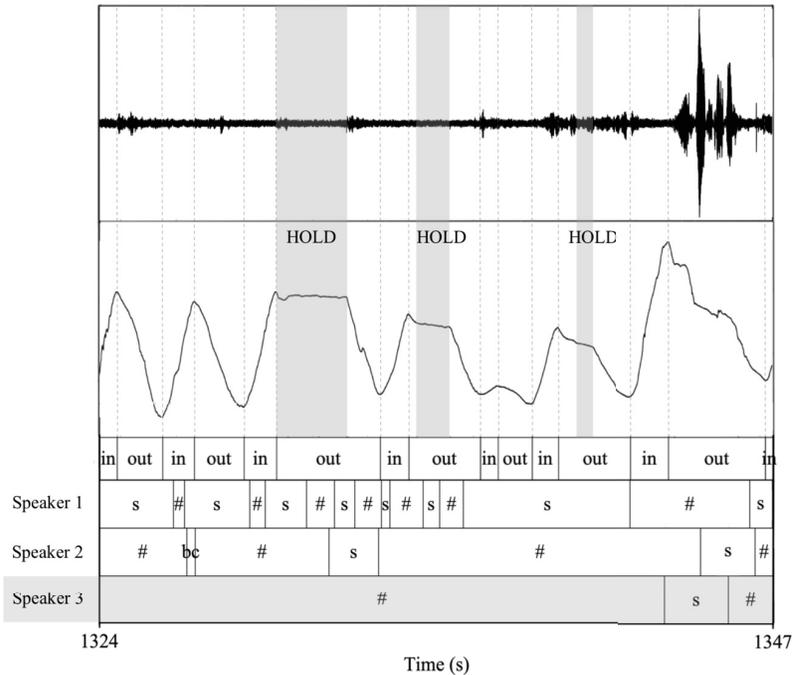
Figure 9. An excerpt from a conversation with the speech and breathing signals of Speaker 3, showing an example of uncategorised breath holds occurring high in the respiratory range. On the topmost tier, "in" and "out" correspond to inhalation and exhalation phases of Speaker 3, respectively. Breath holds are indicated with the label "hold" and a grey box over the speech and respiratory signal. The second tier shows speech activity from Speaker 1, the third tier shows speech activity from Speaker 2, and the bottom tier shows speech activity from Speaker 3. On all speaker tiers, "s" marks speech, "#" marks pauses, and "bc" marks backchannels.

In terms of temporal characteristics, we were interested to see if we could separate categories on the basis of how early or late they occurred in the respiratory phase. First, we looked at breath hold start and end times in relation to the start and end times of the coinciding respiratory phase and compared the four categories of breath holds. Expectedly, preparatory initial breath holds started earlier and ended earlier in the respiratory phase than preparatory pause breath holds. Importantly, these two preparatory groups likely reflect two processes relevant in conversation organisation: holding the turn (Local and Kelly 1986) and queueing up for it, a difference which is also responsible for the resulting statistical differences. Similar to results from the respiratory range analysis, breath holds coinciding with swallowing exhibit great variation and are similar in median values to preparatory initial breath holds. As explained for swallowing before, many of the occurrences included speech directly before or after the breath hold coinciding with swallowing, making the breath hold characteristics similar to pre-speech breath holds. Therefore, swallowing-induced breath holds cannot be reliably distinguished from other breath holds using temporal information either. In fact, as this data has provided evidence that swallowing and speech often occur together in combination, we can expect breath holds involved with both swallowing and speech preparation to have similar characteristics. Lastly, uncategorised breath holds on average start slightly later than other categories and end much later than other categories. In fact, often the respiratory cycles end with such breath holds (see example in Figure 7).

Durational measurements were carried out in seconds and as a percentage of coinciding respiratory phase duration. Analysis on the absolute time scale demonstrated a relatively uniform distribution across all categories, with the average median duration of the breath hold categories placed at around 0.8 s. The results on relative duration show that, while other categories usually make up around 25–28% of the entire phase duration, preparatory pause breath holds take up less of the total exhalation time at around 15% of the phase duration. However, this outcome can be explained by the indication that this category occurs during a respiratory phase with at least two separate speech spurts and that the phase can therefore be expected to be longer, making the relative duration of the breath hold appear shorter.

Theoretically, breath holds could occur anywhere in the respiratory cycle in terms of both range and timing, although some ranges are more likely. As such, they can be found from inhalation phases as well, although the data set analysed here only included breath holds produced during exhalation phases. A separate technical question is whether breath holds should be considered as part of the inhalation or exhalation phase in the respiratory cycle or as separate holding phases. In principle, breath holds could be categorised as another commonly-occurring phase between the other two respiratory phases. On the basis of this manually labeled data, the automatic breath hold detection method proposed in the BreathMetrics toolbox (Noto et al. 2018) has been adapted to the RIP signal (Włodarczak 2019). The method allows including breath holds into the automatic annotation of respiratory signals.

Timing and position in lung capacity are two scales representing the breath holds in this study. Mostly, they tend to reflect similar changes, as events involving higher lung capacity values usually also happen earlier in the respiratory phase. However, as these data have shown, these do not always match perfectly. For example, the volume-related findings we report are bimodal in some cases, such as for the preparatory initial and uncategorised breath hold positions in the exhalation phase. In addition, we saw that, for the positions of preparatory breath holds in the respiratory range, there is bimodality present in the distributions of breath holds followed by backchannels and speech. Such patterns are not present in the timing-related properties of breath holds. While this discrepancy is slightly surprising, volume-related information seems to account for more intricate characteristics of the breath holds than what temporal information can provide.

5. Conclusion

This study provides a first look into breath holds by analysing temporal and volume-related respiratory data in spontaneous conversations held in Estonian. The study investigated breath hold positions within the coinciding respiratory interval amplitudes, as well as based on their timing within the respiratory intervals and in relation to communicative activity following breath holds. The results indicate variation between breath-hold categories on both the time and lung capacity scale, with more complicated patterns present in the lung capacity section. In

addition, we found detectable differences in breath-holding characteristics involving laughter and speech preparation, while breath holds coinciding with swallowing share characteristics with preparatory breath holds and are therefore difficult to separate from preparatory breath holds.

Acknowledgements

Addresses:
Kätlin Aare
Department of Linguistics
Stockholm University
Universitetsvägen 10c
11418 Stockholm, Sweden
E-mail: katlin.aare@ling.su.se

Marcin Włodarczak
E-mail: wlodarczak@ling.su.se

Mattias Heldner
E-mail: heldner@ling.su.se

References

Aare, Kätlin, Marcin Włodarczak, and Mattias Heldner (2014) "Backchannels and breathing". In Mattias Heldner, ed. *Proceedings from FONETIK 2014 Stockholm, June 9-11, 2014, PERILUS XXIV*, 47–52. Stockholm, Sweden: Department of Linguistics, Stockholm University.

Boersma, Paul and David Weenink (2019) *Praat: doing phonetics by computer* [Computer program]. Version 6.1. Available online at <http://www.praat.org/>. Accessed on 13.07.19.

Cherniack, Neil S. (2009) "Inspirational lessons from watching swallows". *Experimental Physiology* 94, 4, 398–399.

Cooper, H. E., M. J. Parkes, and T. H. Clutton-Brock (2003) "CO2-dependent components of sinus arrhythmia from the start of breath holding in humans". *American Journal of Physiology-Heart and Circulatory Physiology* 285, 2, H841–H848.

Edlund, Jens, Mattias Heldner, and Marcin Włodarczak (2014) "Catching wind of multiparty conversation". In Jens Edlund, Dirk Heylen, and Patrizia Paggio, eds. *Proceedings of Multimodal Corpora: Combining applied and basic research targets (MMC 2014)*, 35–36. Reykjavik, Iceland.

French, Peter and John Local (1983) "Turn-competitive incomings". *Journal of Pragmatics* 7, 1, 17–38.

Godfrey, S. and E. J. M. Campbell (1968) "The control of breath holding". *Respiration physiology* 5, 3, 385–400.

Godfrey, S., R. H. T. Edwards, and D. A. Warrell (1969) "The influence of lung shrinkage on breath holding time". *Quarterly Journal of Experimental Physiology and Cognate Medical Sciences: Translation and Integration* 54, 2, 129–140.

Hårdemark-Cedborg, Anna I., Eva Sundman, Katarina Bodén, Hanne Witt Hedström, Richard Kuylenstierna, Olle Ekberg, and Lars I. Eriksson (2009) "Co-ordination of spontaneous swallowing with respiratory airflow and diaphragmatic and abdominal muscle activity in healthy adult humans". *Experimental Physiology* 94, 4, 459–468.

Hixon, Thomas J. (1982) "Speech breathing kinematics and mechanism inferences therefrom". *Speech motor control*, 75–93. New York: Pergamon Press.

Hixon, Thomas J. (1987) "Respiratory function in speech". In Thomas J. Hixon and Collaborators, eds. *Respiratory function in speech and song*, 1–54. Boston, MA: College-Hill Press.

Hixon, Thomas J., Michael D. Goldman, and Jere Mead (1973) "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, 78–115.

Hixon, Thomas J., Gary Weismer, and Jeannette D. Hoit (2018) *Preclinical speech science: Anatomy, physiology, acoustics, and perception*. San Diego, CA: Plural Publishing.

Jefferson, Gail (1983) "On a failed hypothesis:'Conjunctionals as overlap-vulnerable'". *Tilburg Papers in Language and Literature* 28, 1–33. Tilburg: Tilburg University.

Konno, Kimio and Jere Mead (1967) "Measurement of the separate volume changes of rib cage and abdomen during breathing". *Journal of Applied Physiology* 22, 3, 407–422.

Local, John and John Kelly (1986) "Projection and 'silences': notes on phonetic and conversational structure". *Human Studies* 9, 2–3, 185–204. https://doi.org/10.1007/BF00148126

Muxworthy, J. F. (1951) "1. Breath holding studies: relationship to lung volume". *USAF Technical Report* 6528, 452–456. Dayton, Ohio: Wright-Patterson Air Force Base.

Noto, Torben, Guangyu Zhou, Stephan Schuele, Jessica Templer, and Christina Zelano (2018) "Automated analysis of breathing waveforms using BreathMetrics: A respiratory signal processing toolbox". *Chemical Senses* 43, 8, 583–597.

Parkes, M. J. (2006) "Breath-holding and its breakpoint". *Experimental Physiology* 91, 1, 1–15.

Poliacek, I., M. Simera, V. Calkovsky, and J. Jakus (2016) "Upper airway control in airway defense". *Acta Medica Martiniana* 16, 1, 5–16.

Schneider, Edward C. (1930) "Observations on holding the breath". *American Journal of Physiology-Legacy Content* 94, 2, 464–470.

Shea, Steven A. (1996) "Behavioural and arousal-related influences on breathing in humans". *Experimental Physiology: Translation and Integration* 81, 1, 1–26.

Smith, John, Norman Wolkove, Antoinette Colacone, and Harvey Kreisman (1989) "Coordination of eating, drinking and breathing in adults". *Chest* 96, 3, 578–582.

Von Euler, Curt (1982) "Some aspects of speech breathing physiology". In S. Grillner, B. Lindblom, J. Lubker, and A. Persson, eds. *Speech motor control*, 95–103. New York: Pergamon Press.

Watson, H. (1980) "The technology of respiratory inductive plethysmography". *Proceedings of the Second International Symposium on Ambulatory Monitoring (ISAM 1979)*, 537–563. London: Academic Press.

Włodarczak, Marcin (2019) "RespInPeace: toolkit for processing respiratory belt data". *Proceedings from FONETIK 2019 Stockholm, June 10–12*, 115–118. https://doi.org/10.5281/zenodo.3246019

Włodarczak, Marcin and Mattias Heldner (2017) "Respiratory constraints in verbal and non-verbal communication". *Frontiers in Psychology* 8, 708. https://doi.org/10.3389/fpsyg.2017.00708

Yngve, Victor H. (1970) "On getting a word in edgewise". *Papers from the sixth regional meeting of the Chicago Linguistic Society*, 567–578. Chicago, Illinois: Chicago Linguistic Society.

Kokkuvõte. Kätlin Aare, Marcin Włodarczak ja Mattias Heldner: Hinge kinni hoidmine spontaanses kõnes. Artikkel kirjeldab hinge kinni hoidmist spontaansetes vestlustes ajaliste ja kopsumahuga seotud omaduste kaudu. Hinge kinni hoidmist analüüsitakse esmalt selle põhjal, kus see käimasoleva hingamisfaasi kopsumahu ulatuse suhtes asub. Teine fookus on ajalisel faktoril: kus hoitakse hinge kinni käimasoleva hingamisfaasi alguse ja lõpu ning vestlustes sisalduva kõne vm suhtes. Hüpoteeside kohaselt peaks hinge kinni hoidmine ajalisel ja kopsumahuga seotud skaalal toimuma erinevas kohas sõltuvalt sellest, milline kontekst hinge kinni hoidmist vestluses ümbritseb. Tulemused näitavad, et kuigi hinge kinni hoidmine esineb suure varieeruvusega mõlemal skaalal, sisaldab andmestik mustreid, mis eristavad kõneplaneerimisega seotud hinge kinni hoidmist teistest alternatiividest. Tulemustest selgub ka, et neelatustega seotud hinge kinni hoidmist on keeruline muudest eristada vaid aja ja kopsumahuga seotud informatsiooni abil.

Märksõnad: hinge kinni hoidmine, kõnehingamine, spontaanne kõne, vestlused kolme osalejaga