Exploring the presence and absence of inhalation noises when speaking and when listening

Raphael Werner[1], Jürgen Trouvain[1], Susanne Fuchs[2], Bernd Möbius[1]

[1]*Dept. of Language Science and Technology, Saarland University, Saarbrücken (Germany)*
[2]*Laborphonologie, Leibniz Centre General Linguistics (ZAS), Berlin (Germany)*
[rwerner|trouvain|moebius]@lst.uni-saarland.de, fuchs@leibniz-zas.de

Abstract

In this paper we look at the temporal coordination of acoustic and respiratory events when listening and speaking. We first look at 4 German female speakers individually and then test our observations on 14 speakers. Apart from general observations on different timing between listening and speaking (fewer breath cycles and longer inhalations when listening) it is found that when speaking, rib cage breathing (in relation to abdomen) tends to happen a bit earlier as compared to listening but there is high individuality. Further, there seems to be a link between the onset of audible inhalation and both the rib cage and the abdomen expanding. Audible inhalations are sandwiched between two short edges on each side. When either of them is lengthened, the other one remains relatively short.

Keywords: speech respiration, speech pauses, breath noises, respiratory inductance plethysmography

1. Introduction

Breath noises in speech communication can usually be observed during articulatory activity but not when speakers let their articulation rest, for instance when they are listening (Trouvain, Werner, and Möbius 2020).

The presence of audible breath noises is so prevalent in speech breathing that it can even make synthetic speech be perceived as more natural (Whalen, Hoequist, and Sheffert 1995). The question is why inhalation is made audible only in active (or planned) articulation in comparison to listening even though respiration is permanently at work. This question can only be addressed by an analysis of acoustic and physiological (respiratory) data. Thus, the aim of this exploratory study is to look more closely at breathing and the interplay between synchronously recorded acoustic and kinematic respiratory signals.

Respiration during speech and rest differs in various ways. Breathing cycles during speech are characterized by a short, rapid inhalation and a long, slow exhalation, giving the breathing profile a sawtooth shape. Breathing cycles during rest, listening or inner speech follow a more symmetric shape with only slightly shorter inhalations than exhalations (cf. Conrad and Schönle 1979). The short and deep inhalations during speech frequently coincide with audible noise. The audible noise can be generated by any constriction in the vocal tract resulting from a specific coordination of the respiratory system with the glottis, the velum, the tongue, and the lips .

In speech breathing, audible inhalations are often surrounded by *edges*, i.e., short silent stretches between the inspiration and the preceding or following articulation (Fukuda, Ichikawa, and Nishimura 2018).

The current work aims to investigate the temporal properties of breathing, especially inhalation noise, using acoustic and respiratory data. In particular, we compare general aspects of breathing when listening and when speaking. For speech breathing, we further examine the interplay between inhalation noises and articulation and how they are temporally aligned with expansions of the abdomen and the rib cage.

In a first exploratory step, we had a closer look at 4 speakers and their respiratory behavior, observed a pattern that can be seen in Fig. 1 and modelled the findings in Fig. 2.

In the final part, we try to check this general observation with speech data featuring synchronously recorded acoustic and kinematic (respiratory) signals from 14 speakers (including the other 4).

2. Methods

This exploratory study builds on German material originally elicited for a different study (Rochet-Capellan and Fuchs 2013). All participants analyzed here were female and engaged in two tasks: listening to a fable (LN) and re-telling this fable (SN). Along with the audio, kinematic data were also collected via Respiratory Inductance Plethysmography (RIP): one elastic band was placed at the level of the axilla to measure movement of the rib cage (RC) and the other one at the level of the umbilicus to measure movement of the abdomen (AB). By that, compression and expansion of AB and RC during in- and exhalation can be monitored. Participants were told to stand still during the experiments to avoid the RIP signal picking up arm movement.

In post-processing, the RIP signal was transformed into on- and offsets of in- and exhalations resulting in time-aligned annotation for these events split into phases of inhalation and exhalation. Inhalation onsets were detected automatically at 10% of the velocity peak, while offsets were detected at 90%. The remaining stretch from the end of inhalation to the start of the next inhalation is regarded as exhalation and thus potentially includes phases of breath holding.

We further annotated audible inhalation noises from the speech signal to relate them to the on- and offsets of articulation, AB, and RC. It is important to note that by articulation here we refer to the acoustic result of speech production, which is therefore based solely on the audio signal. Since it is not possible to reliably annotate breath noises in LN (due to less loud breathing and masking from the fable being played), analysis of audible breath noises is only done in SN, whereas the analysis of LN is fully based on kinematic data.

The succession of first and second author was changed in the time between the conference and submission of this paper.

The dataset used here includes 108 corresponding files (54 LN, 54 SN) from 14 participants. Analyses are based on the duration and coordination of these intervals as annotated in Praat TextGrids (Boersma and Weenink 2019).

3. Preliminary analysis

3.1. Inhalation in listening vs. in speaking

A first comparison between the kinematic respiration patterns of both tasks (for an illustration of a sample see Fig. 3) reveals that subjects have shorter and more variable breath cycles (inhalation phase plus following exhalation phase) in LN than in SN. Consequently, there are more breath cycles per minute in LN than in SN (see Table 1).

Table 1: *Overview of the respiratory time quotient (RTQ) for abdomen (AB) and rib cage (RC), mean durations of breath cycles, and number of mean breath cycles per minute while speaking and listening*

	speaking	**listening**
RTQ AB: mean (sd)	0.18 (0.41)	0.59 (0.90)
duration of AB breath cycle	5.45 (2.79)	3.68 (1.37)
AB breath cycles per minute	11.0	16.3
RTQ RC: mean (sd)	0.16 (0.13)	0.61 (0.72)
duration of RC breath cycle	5.26 (2.47)	3.67 (1.37)
RC breath cycles per minute	11.4	16.3

In addition, the duration of the inhalation phase is substantially longer in LN (as opposed to SN) while the exhalation phase in SN is much longer and more variable in its duration. This can be seen in Table 1 where the respiratory time quotient (RTQ; duration of inhalation divided by duration of exhalation), as proposed by (Conrad and Schönle 1979), is much lower in SN than in LN, reflecting the less symmetric respiratory behavior there. The high standard deviation of AB in speaking is partly caused by two RTQs that are around 1.2, created by a very short speaking phase between 2 exhalations. Thus, in both conditions exhalations are generally longer than inhalations but in speech breathing, the imbalance is much stronger. Additionally, the duration of a breath cycle is shorter in listening.

As expected, only few breath noises in LN were observable (in the phases before and after the playback of the tale to be listened to; during the playback an observation was not possible). The few instances of breath noises in LN were very soft compared to those in SN. In addition, in SN all inhalation phases were acoustically reflected by a salient breath noise.

3.2. Edges

A typical acoustic feature of an inbreath noise is that it is sandwiched between short intervals of silence. These "edges" (Fukuda, Ichikawa, and Nishimura 2018; Trouvain, Werner, and Möbius 2020) to the left and right of the breath noise typically have an average duration of 50 ms, whereas the breath noises themselves have a duration between 200 and 500 ms (Trouvain, Werner, and Möbius 2020). Thus, we aimed to find a link between the timing of the "edges" in the acoustic signal and the respiratory activities of RC and/or AB in the inhalation phase in breath pauses in speech.

3.3. Timing of AB and RC activities

Interestingly, several speakers shifted their AB phase in relation to RC to an earlier time point when speaking (as compared to listening) as illustrated schematically in Fig. 2.

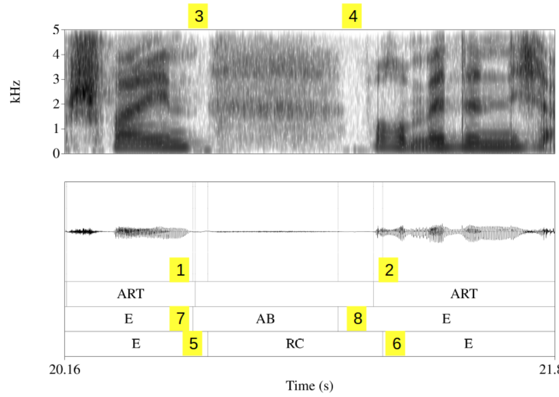

Figure 1: *The temporal alignment of articulation phases, inhalation noises and AB and RC in an inhalation phase between two articulation phases in a speech signal (left, for numbers see text, speaker S04).*

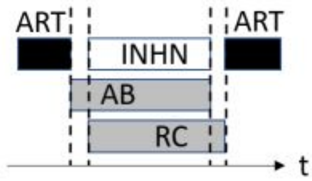

Figure 2: *A schematic view of the temporal alignment of articulation phases (ART) and the abdominal activity (AB) and the rib cage activity (RC) in an inhalation phase (INHN) in speech.*

When inspecting the temporal structure of the acoustic signal (articulation phases and inhalation noises) together with the kinematic signal (RC and AB) the following pattern was observed for two of the four speakers (as illustrated in Fig. 1): The end of articulation in the acoustic signal [1] seems to be aligned with the start of AB [7], whereas the start of an articulation phase [2] seems to be aligned with the end of RC [6]. In contrast, the start of the inhalation noise [3] seems to be synchronous with the start of RC [5] whereas the end of the inhalation noise [4] and the end of AB [8] seem to be synchronized.

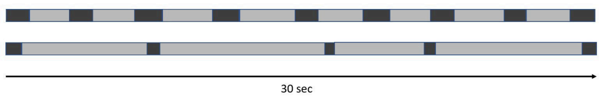

Figure 3: *Exhalation (light grey) and inhalation phases (black) in two 30-sec excerpts from the inspected kinematic data of one speaker in both conditions (top: LN, bottom: SN).*

4. Advanced analysis

To test the observed patterns described in the previous section, the following questions will be addressed:

(1) Are AB & RC of similar length?

(2) Does AB inhalation start earlier than RC inhalation?

(3) Does AB inhalation generally begin when the preceding articulation ends? Does RC inhalation generally end when the following articulation starts?

(4) Is inhalation only audible when both AB & RC are synchronously inhaling, i.e., from when the later contributor

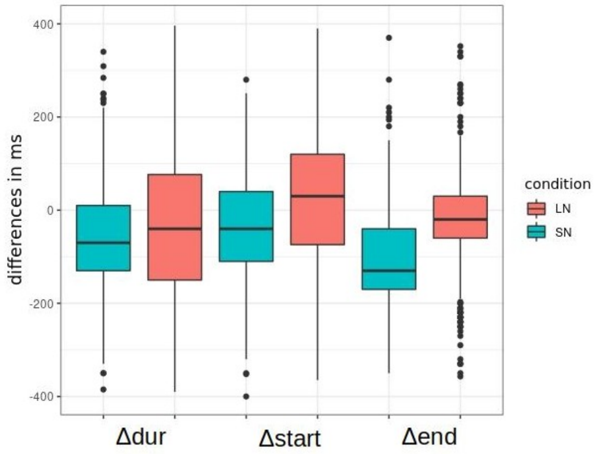

Figure 4: *Differences between durations and start/end times of inhalation in the two conditions. Boxes 1 & 2 show differences between duration of AB minus duration of RC (Δdur), boxes 3 & 4 show differences between starting times of AB minus RC (Δstart), and boxes 5 & 6 show differences between end times of AB minus RC (Δend). Positive values thus show a longer/later AB, while negative values indicate a longer/later RC.*

(presumably RC) starts to when the earlier contributor (presumably AB) ends?

(5) How do the durations of the edges left and right of inhalation noises relate to each other?

To test observations (1) to (4), for every inhalation noise in speaking we calculated the difference between the respective events. The values for LN are based on kinematics while listening. The results can be seen in Fig. 4 and 5 showing the difference of the respective subtraction.

4.1. Durations of AB and RC

For observation (1) we used the duration values of AB and RC. In Fig. 4, Δdur shows that RC generally tends to be slightly longer than AB in both SN and LN. For LN, it is less clear with more variance (probably coming from longer total durations) and values closer to 0.

4.2. Start and end of AB inhalation relative to RC inhalation

Δstart and Δend from Fig. 4 show results for (2): to see if AB was typically earlier than RC we compared the start and end times of both. While the start times show a difference that is close to 0 with a tendency towards RC being later for SN and the opposite for LN, the difference is clearer when looking at the end times: here RC is later than AB for SN but for LN the results are close to 0 with little variance. For SN, the difference between Δstart and Δend can be explained by Δdur that showed RC being longer.

Further inspection of the SN data revealed that the pattern illustrated in Fig. 2 with start of AB first, RC second is valid for 6 out of the 14 speakers, whereas 3 show the exact opposite pattern. The remaining 5 speakers had no clear tendency.

Fig. 2 also includes the pattern that AB ends before RC ends. 10 speakers follow this pattern. Among the 4 with the opposite behaviour here were also the 3 'outliers' with the opposite starting pattern.

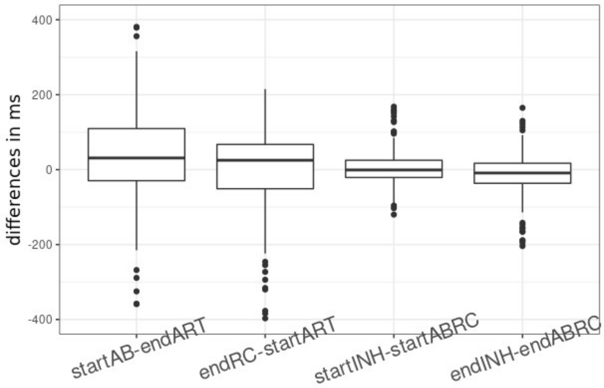

Figure 5: *Differences between starting times (start) and end times (end) of inhalation noises, acoustic articulation, AB breathing, and RC breathing. In 'startABRC' the value is taken from the one starting later and in 'endABRC' from the one ending earlier to have both synchronously inhaling.*

4.3. AB and RC inhalations relative to the articulation

Observation (3) is addressed in the boxes 1 and 2 of Fig. 5: Concerning the start of AB and the end of the preceding articulation, the inhalation in AB starts a little later than the end of the articulation, leading to a short gap here. For the end of RC and the beginning of the preceding articulation, the difference is less clear with the box being at 0 but with a positive median, suggesting a slightly smaller gap than for end of articulation and start of AB.

4.4. AB and RC inhalations relative to the inhalation noise

Observation (4) is about inhalations only being audible when both AB & RC are expanding. To test this we looked at two time points: First, the difference between the onset of audible inhalation and the begin of inhalation in AB or RC (whichever started later to ensure both were expanding; Fig. 5, box 3); second, the difference between the offset of audible inhalation and the end of AB or RC (whichever ended earlier; Fig. 5, box 4). The results of both subtractions are very close to 0 with little variance, suggesting that there is only a very small gap between those events happening.

This suggests that there is a link between both AB and RC being synchronously active and an audible breath noise being produced even though AB and RC are displaced slightly, with RC being later than AB.

4.5. Timing of edges

As concerns the edges surrounding a breath noise (5), it can be seen in Fig. 6 that they have a similar duration on both sides, with a slight tendency for longer edges following an inhalation. The mean duration for a left edge is 116 ms (sd: 107 ms) and 160 ms (sd: 164 ms) for a right edge.

Most inhalation noises (79%) are accompanied by edges that are shorter than 250 ms both left and right. Only 7% have one or two edges that are longer than 500 ms. There are hardly any combinations of both edges being long, meaning that an inhalation noise is typically not surrounded by two longer silent phases. For all except 4 cases, at least one edge always remains shorter than 250 ms. As a consequence, the inhalation noises here are only central when both edges are short – otherwise, one edge is longer.

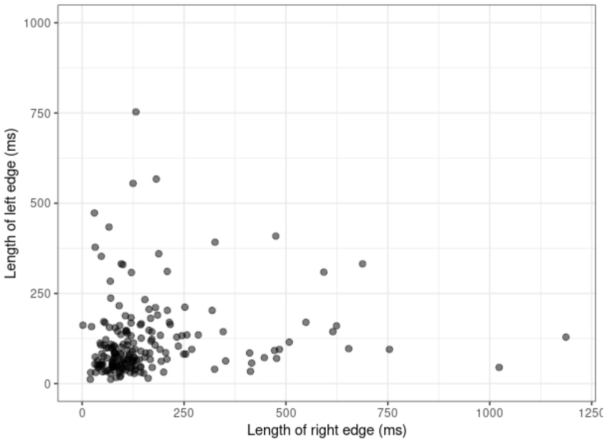

Figure 6: *Corresponding sections of silence (edges) left and right of an audible inhalation noise.*

5. Discussion

We worked with a coupled approach of observing patterns in the data and then trying to test them by looking at the respective times in the data. While this shows general tendencies, there are different strategies at work here, especially for the cooardination of AB and RC. This high degree of individuality was also observed for prephonatory chest wall posturing by (Hixon et al. 1988).

When comparing the relation of breathing and articulation, we compared kinematic data for breathing and the speech signal for articulation. The gaps we found there might thus be due to a delay between articulatory and acoustic onset (Rasskazova, Mooshammer, and Fuchs 2019).

As concerns acoustic and kinematic inhalation, it appears to be the case that the acoustic inhalation is closely coupled to breathing, happening synchronously at both abdomen and rib cage. This study cannot answer why that is and it should further be studied if that also applies to speakers with a clear preference for either AB or RC.

The edges we found here are partly longer than the ones found in previous studies (Fukuda, Ichikawa, and Nishimura 2018, e.g.), with some of them even exceeding a duration of 1 s. The reason for this is that we defined edges to be the time between preceding articulation and inhalation noise (left edge) and inhalation noise and following articulation (right edge). This thus includes potential hesitations that are not as clearly attributable to motor control reasons as edges of 20 ms length. However, since it is not clear where the boundary between an edge in a narrow sense and a hesitation following inhalation lies, we decided to include them.

6. Conclusion and Outlook

In summary, it has been shown that when retelling a fable as compared to listening to it, participants have fewer breath cycles which in turn are longer but also more variable in their duration. When listening, the ratio of duration of inhalation to duration of exhalation is about 6:10, whereas in speaking it is less than 2:10.

As expected, both articulatory phases and inhalation noises seem to be closely coupled to the activity of RC and AB, which often leads to short near-silent gaps around the inhalation noise. It appears to be the case that an audible inhalation noise is only generated when both AB & RC are expanding at the same time.

Finally, it has been shown that the edges left and right of the breath noise are generally short and have a similar duration (both edges are <250 ms in about 80% of the cases). When one of them is longer, the other typically remains relatively short, meaning that the inhalation noise in the speaking condition is only central when neither of the edges is long. This aspect should be investigated in a different experimental setting where the cognitive load is higher and/or the elicited speech is more spontaneous as opposed to pseudo-spontaneous data in our study. The question as to why edges of silence can be found on both sides of an audible inhalation noise remains open. It may be related to motor control when switching from exhaling (i.e., speech) to inhaling and vice versa. This should also be looked at in speakers who show a clear preference for either abdominal or thoracical breathing. The findings reported here are all based on younger, female participants who were standing during the experiment. It would be worthwhile to verify them by using a more diverse group of participants and a different experimental setup, as breathing movements can vary by age, sex, and posture (Kaneko and Horie 2012).

Furthermore, breath noises were only regarded as either present or absent in our study, but a closer look at their spectral properties may yield important insights.

7. Acknowledgements

This research was funded in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID MO 597/10-1.

We would like to give special thanks to Amélie Rochet-Capellan for giving us access to the data. We also thank the anonymous reviewers as well as the poster visitors at the conference for their valuable feedback.

8. References

Boersma, Paul and David Weenink (2019). *Praat: doing phonetics by computer*. URL: `http://www.praat.org/`.

Conrad, B and P Schönle (1979). "Speech and respiration". In: *Archiv für Psychiatrie und Nervenkrankheiten* 226.4, pp. 251–268.

Fukuda, Takashi, Osamu Ichikawa, and Masafumi Nishimura (2018). "Detecting breathing sounds in realistic Japanese telephone conversations and its application to automatic speech recognition". In: *Speech Comm.* 98, pp. 95–103.

Hixon, Thomas J., Peter J. Watson, Frances P. Harris, and Nancy B. Pearl (1988). "Relative volume changes of the rib cage and abdomen during prephonatory chest wall posturing". In: *J. Voice* 2.1, pp. 13–19.

Kaneko, Hideo and Jun Horie (2012). "Breathing movements of the chest and abdominal wall in healthy subjects". In: *Respir. Care* 57.9, pp. 1442–1451.

Rasskazova, Oksana, Christine Mooshammer, and Susanne Fuchs (2019). "Temporal coordination of articulatory and respiratory events prior to speech initiation". In: *INTERSPEECH* 2019, pp. 884–888.

Rochet-Capellan, Amélie and Susanne Fuchs (2013). "Changes in breathing while listening to read speech: the effect of reader and speech mode". In: *Frontiers in Psychology* 4, p. 906.

Trouvain, Jürgen, Raphael Werner, and Bernd Möbius (2020). "An acoustic analysis of inbreath noises in read and spontaneous speech". In: *Proceedings of Speech Prosody, Tokyo*, pp. 789–793.

Whalen, Doug H., Charles E. Hoequist, and Sonya M. Sheffert (1995). "The effects of breath sounds on the perception of synthetic speech". In: *J. Acoust. Soc. Am.* 97.5, pp. 3147–3153.