On Acoustic Features of Inhalation Noises in Read and Spontaneous Speech

Jürgen Trouvain, Bernd Möbius and Raphael Werner

Language Science and Technology, Saarland University, Saarbrücken, Germany
trouvain@lst.uni-saarland.de

Breath noises are probably the most common non-verbal vocalisations in spoken communication. They can occur in a multitude of contexts and can serve as functional markers in various ways. Our goal in this paper is to offer some acoustic-phonetic descriptions of inhalation noises by considering read speech and spontaneous dialogues.

INTRODUCTION

Breath noises as acoustic and audible reflections of inhalation and exhalation are probably the most common non-verbal vocalisations in spoken communication. Breath noises can occur in a multitude of contexts and they can serve as functional markers in various ways. In contrast to acoustic and audible correlates of phonemes there are hardly any acoustic descriptions of inhalation noises and other respiratory signals from a phonetic perspective. Thus, the aim of this paper is to suggest some acoustic descriptors of inhalation noises to fill this particular research gap. This will be done for two different speech modes: on the one hand for read and highly controlled speech and on the other hand for dialogical and spontaneous speech. We start with a review of different functions of respiratory noises in spoken communication.

Inhalation noises frequently occur in speech pauses. Here, breath noises function as markers of prosodic-syntactic boundaries, which has motivated the use of the term breath-groups for intonation (or prosodic) phrases [13]. Phonetic studies have shown that duration and intensity of inhalation noises can be indicators of utterance planning in speech production and inform listeners about the length of the upcoming phrase [6,7]. A recent study also suggests that in read speech duration and intensity of inhalation noises are due to a 'recovery' from the effort of the prior utterance [10]. Interestingly, when speakers are under physical stress they show different forms of breath noises in speech pauses, e.g. with many exhalation noises [18].

A typical non-verbal vocalisation in spontaneous speech is laughter of which various forms can be described with characteristic noises of ex- and inhalation [1,20]. A strong inhalation noise can mark the offset of a long and complex laugh [4,20]. Also in (other) affect bursts, breath noises can play a crucial role, such as startle or in crying [16].

On the level of pragmatics, breath noises can be used as discourse markers, signalling an intent to take the turn, and in some cultures respiratory noises are markers of politeness, e.g. in Korean [22]. Breath noises also have a high potential of signalling individuality, either by idiosyncratic acoustics, e.g. by inhalation noises with [s↓], an ingressive alveolar fricative [15], or by different patterns of inhalation and exhalation [11,12]. The incomplete list above shows that breath noises are a rather rich source of information on the linguistic but also on the non-linguistic level.

Surprisingly, breath noises are often and maybe systematically ignored in speech analysis, speech synthesis and speech recognition. This is reflected for instance by the fact that in speech fluency research, pauses that contain breath noises are regarded as 'silent', although they are not silent from an acoustic point of view [3]. In some conversational corpora the annotation schemes do not have a category for breath noises [17]. Likewise, speech prosodists regularly ignore breath noises as important acoustic cues of prosodic phrase boundaries.

Pauses in synthesised speech are often not modelled naturalistically [19] and they virtually never contain breath noises. However, breath noises are likely to be beneficial for speech synthesis that is pleasant and memorisable [21], and they are necessary for expressive speech synthesis. Breath noise in automatic speech recognition is still an under-researched topic, although there are various approaches for explicit breath detection, e.g. [8].

While there are research groups working on the physiological, particularly the kinematic, bases of respiration in speech, e.g. [2,7,21], the link between kinematic and acoustic signals of inhalation and exhalation in speech is not yet fully understood. The disctintion between in- and exhalation in this paper is based on auditory assessment of acoustic data which were recorded under laboratory conditions. Adverse acoustic conditions might be challenging for this task.

GENERAL OBSERVATIONS

In this paper, which can be considered as a preliminary study, we offer some acoustic descriptions of inhalation noises. For read speech we selected some news produced by professional news casters [5]. Here, all breath noises investigated were inhalation noises that used a combination of oral and nasal airstream. Nearly all pauses were marked with these inbreath noises.

A typical acoustic feature of an inbreath noise is that it is sandwiched between short intervals of silence. The edges to the left and right of the breath noises show an average duration of 50 ms whereas the breath noises themselves have a duration between 200 and 500 ms (see Fig. 1).

Inhalation noises in the read speech samples reveal a relatively low intensity and the values for centre of gravity (COG) are below 2 kHz. The formant values seem to have rather stable values.

It might be of interest to compare inhalation noises with other 'breath' sounds, i.e. unvoiced fricative segments with inhalation or exhalation as their primary sound source. For German, two types of segments can play a role here: aspiration phases of the closure release of unvoiced stops, and unvoiced variants of the glottal fricative /h/. Regarding realisations of /h/, a preceding voiced context, for instance a vowel or a sonorant, usually leads to a voiced instantiation of /h/, which is more similar to a glide. Unvoiced productions obviously require a voiceless left context, for instance an unvoiced obstruent or a silence. This has been shown to be a regular pattern in German [14, 24] that probably functions in a similar way in other Germanic languages.

Figure 1 depicts an example where these three kinds of respiratory noises occur in close vicinity. In contrast to inhalation noises, aspiration phases of unvoiced stops are much shorter and rarely exceed 60 ms. Their intensity is much higher than those of breath noises. The COG values are above 2 kHz. The formant values do show more variable values than those for breath noises. Cases of unvoiced variants of /h/ were rather infrequent and therefore not considered for this preliminary investigation.

The spontaneous speech comes from the Lindenstraße dialogue corpus [9]. It contains dyads of friends (same sex) with the task of talking about video clips of an episode of a soap opera. The interlocutors could not see each other and were recorded by separate channels.

For these spontaneous dialogues the pattern regarding the inhalation noises is by far more variable. Breath noises are only observed in phases of vocal activity, for instance when having the turn, giving a comment, or providing a feedback (or backchannel) utterance. This means that for a given speaker a significant portion of the recorded dialogue is marked by the absence of vocalisations, which should not be confused with regular pauses in speech.

In contrast to the read speech of the professional news readers, the breath noises in the dialogues are sometimes only nasal inhalation. In some cases the dialogues also show some exhalations. In addition, some inbreath noises are apparently produced with an [s]-like tongue position giving this fricative an additional sound source.

A substantial difference to read speech is that laughter may occur in spontaneous speech. Often, the laughter episodes are marked by strong inhalation noises. In Figure 2 we show an example of a typical offset of a 'voiced' or 'song-like' laugh with a long duration.

It is no surprise that inhalation noises often occur at the turn-initial position. However, a form with a higher intensity can be assumed to function as a turn-claiming cue, as exemplified in Figure 3.

Finally, it should be mentioned that some inhalation noises are 'enriched' with tongue clicks [3], a discourse-related pause-internal particle that also occurs in languages that do not have clicks as phonemes.

DISCUSSION AND CONCLUSION

Although this study has only an exploratory character, it suggests that inhalation noises differ acoustically from other segments in spoken communication. Appropriate acoustic parameters that establish this difference include intensity, COG, duration, and formants of which often the first four are visible in the spectrogram. A special feature of inbreath noises in pauses is that inbreath noises are accompanied at the edges of short silent sections that separate the frication section from the prior and the upcoming speech sequences.

There are manifold functions in which inhalation noises are involved. A typical inbreath noise that occurs in a pause that marks a syntactic-prosodic break usually has a rather different acoustic form than an inbreath noise that marks the offset of a longer laugh. It is of general interest in phonetics to learn more about how a given phonetic form reflects certain functions, and vice versa. However, for the time being it is unclear how complex or simple the relationship between the acoustic shapes of inhalation noises and their (presumed) functions really are.

Thus, the next step is to perform a detailed and systematic study of the proposed acoustic parameters of inhalation noises. This should entail various speech styles as the above sketched differences between read and spontaneous speech samples have shown. Ideally, such a study would also compare speech data across languages.

ACKNOWLEDGMENTS

This research was funded in part by the German Research Foundation (DFG) under grants TR 468/3-1 and MO 597/10-1.

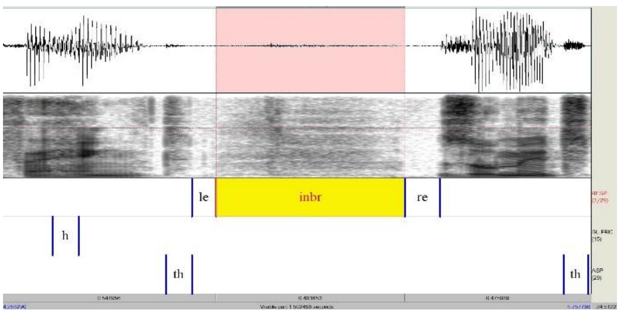

Fig. 1. Typical inbreath noise (spectrogram 0-8 kHz) with short silent edges on each side (top tier) in a 1.5-second section of read speech (... verhängt. Somit ...); realisation of /h/ as a glide (middle tier); aspiration of unvoiced alveolar stop (bottom tier).

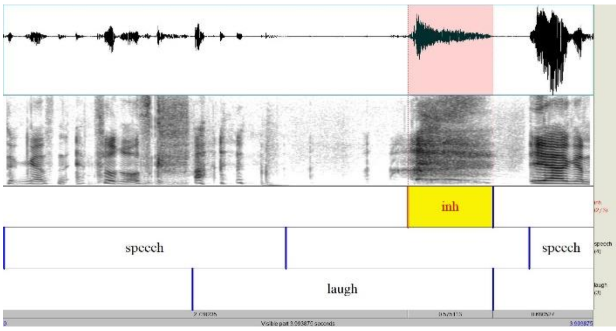

Fig. 2. Laugh with an inbreath noise (inh) as an offset (spectrogram 0-8 kHz).

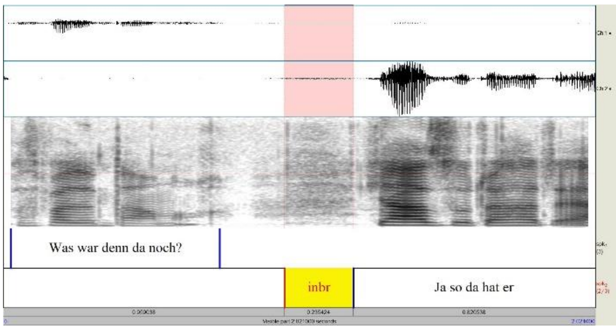

Fig. 3. 2-sec extract from a dialogue: speaker at top ends turn, speaker at bottom takes turn with turn-initial inhalation noise (spectrogram 0-8 kHz).

REFERENCES

[1] Bachorowski, J.A. & Owren, M.J. (2001). Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science* 12, pp. 252-257.

[2] Bailly, G., Rochet-Capellan, A. & Vilain, C. (2013). Adaptation of respiratory patterns in collaborative reading. Proc. *Interspeech*, Lyon. pp. 1653-1657.

[3] Belz, M. & Trouvain, J. (2019). Are 'silent' pauses always silent? Proc. *19th International Congress of Phonetic Sciences* (ICPhS), Melbourne, pp. 2744-2748.

[4] Chafe, W. (2007). *The Importance of Not Being Earnest*. Amsterdam: Benjamins.

[5] Eckart, K., Riester, A., Schweitzer, K. 2012. A Discourse Information Radio News Database for Linguistic Analysis. In: Chiarcos, C., Nordhoff, S., Hellmann, S., (eds), *Linked Data in Linguistics*. Springer, pp. 65–75.

[6] Fuchs, S., Petrone, C., Krivokapic, J. & Hoole, Ph. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics* 41, pp. 9-47.

[7] Fuchs, S., Petrone, C., Rochet-Capellan, A., Reichel, U. & Koenig. L. (2015). Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. *Journal of Phonetics* 52, pp. 35-45.

[8] Fukuda, T., Ichikawa, O. & Nishimura, M. (2018). Detecting breathing sounds in realistic Japanese telephone conversations and its application to automatic speech recognition. *Speech Communication* 98, pp. 95-103.

[9] IPDS (2006). *Video Task Scenario: Lindenstraße – The Kiel Corpus of Spontaneous Speech*, Volume 4, DVD, Institut für Phonetik und Digitale Sprachsignalverarbeitung Universität Kiel.

[10] Kallay, J.E., Mayr, U. & Redford, M.A. (2019). Characterizing the coordination of speech production and breathing. Proc. *19th International Congress of Phonetic Sciences*, Melbourne, pp. 1412-1416.

[11] Kienast, M. & Glitza, F. (2003). Respiratory sounds as an idiosyncratic feature in speaker recognition. Proc. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 1607-1610.

[12] Lauf, R. (2001). Aspekte der Sprechatmung: Zur Verteilung, Dauer und Struktur von Atemgeräuschen in abgelesenen Texten. In: Braun, A. (ed) *Beiträge zu Linguistik und Phonetik*. Stuttgart: Franz Steiner Verlag, pp. 406-420.

[13] Lieberman, Ph. (1967). *Intonation, Perception and Language*. Cambridge, Mass.: MIT Press.

[14] Möbius, B. (2004). Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica* 38, pp. 5-26.

[15] Trouvain, J. 2010. Affektäußerungen in Sprachkorpora. Proc. *21st Konferenz Elektronische Sprachsignalverarbeitung*, Berlin, pp. 64-70.

[16] Trouvain, J. (2011). Zur Wahrnehmung von manipuliertem Weinen als Lachen. Proc. *22nd Konferenz Elektronische Sprachsignalverarbeitung*, Aachen, pp. 253-260.

[17] Trouvain, J. & Truong, K. (2012). Comparing non-verbal vocalisations in conversational speech corpora. Proc. *4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals*, Istanbul, pp. 36-39.

[18] Trouvain, J. & Truong, K. (2015). Prosodic characteristics of read speech before and after treadmill running. Proc. *Interspeech*, Dresden, pp. 3700-3704.

[19] Trouvain, J. & Möbius, B. (2018). Zu Mustern der Pausengestaltung in natürlicher und synthetischer Lesesprache. Proc. *29th Konferenz Elektronische Sprachsignalverarbeitung*, Ulm, pp. 334-341.

[20] Truong, K.P., Trouvain, J. & Jansen, M.-P. (2019). Towards an annotation scheme for complex laughter in speech corpora. Proc. *Interspeech*, Graz, pp. 529-533.

[21] Whalen, D.H., Hoequist, Ch.E. & Sheffert, S. (1995). The effects of breath sounds on the perception of synthetic speech. *Journal of the Acoustical Society of America* 97, pp. 3147-315.

[22] Winter, B. & Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics* 40, pp. 808-815.

[23] Włodarczak, M. & Heldner, M. (2017). Respiratory constraints in verbal and non-verbal communication. *Frontiers in Psychology* 8, article id 708.

[24] Zimmerer, F. & Trouvain, J. 2015. Productions of /h/ in German: French vs. German speakers. Proc. *Interspeech*, Dresden, pp. 1922-1926.