10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

An Acoustic Analysis of Inbreath Noises in Read and Spontaneous Speech

Jürgen Trouvain, Raphael Werner and Bernd Möbius

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

Abstract

Inbreath noises are very common non-verbal vocalisations in spoken communication. They can occur in a multitude of contexts and can serve as functional markers in various ways, with indicating syntactic-prosodic breaks as its salient but not exclusive function. In this paper we first describe some acoustic-phonetic features of inbreath noises such as intensity and duration. Second we analyse read speech and spontaneous dialogues from four corpora with data of twenty speakers of German. It is found that the majority of pauses contain inbreath noises, which are typically soft in intensity and extremely variable in duration. The link between the duration of the entire pause and the breath noise is stronger in read speech than in spontaneous dialogues. It is suggested that ingressive frication can be rather informative for various types of prosodic analysis.

Index Terms: speech respiration, speech pauses, inhalation noises, prosodic breaks

1. 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, and to analyse samples of speech. 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 or inbreath 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 [1]. 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 [2, 3]. 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 [4]. Interestingly, when speakers are under physical stress they show different forms of breath noises in speech pauses, e.g. with many exhalation noises [5].

A typical non-verbal vocalisation in spontaneous speech is laughter of which various forms can be described with characteristic noises of ex- and inhalation [6, 7]. A strong inhalation noise can mark the offset of a long and complex laugh [8, 7]. Also in (other) affect bursts, breath noises can play a crucial role, such as startle or in crying [9].

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 [10]. 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 [11], or by different patterns of inhalation and exhalation [12, 13]. 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 [14]. In some conversational corpora the annotation schemes do not have a category for breath noises [15]. Likewise, speech prosodists regularly ignore breath noises as important acoustic cues of prosodic phrase boundaries.

Pauses in synthesised speech are often not modelled naturalistically [16] and they virtually never contain breath noises. However, breath noises are likely to be beneficial for speech synthesis that is pleasant and memorisable [17], 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. [18].

While there are research groups working on the physiological, particularly the kinematic, bases of respiration in speech, e.g. [2, 3, 19], 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 the auditory assessment of acoustic data which were recorded under laboratory conditions. Adverse acoustic conditions might be challenging for this task.

2. Preliminary observations

For our preliminary observations, we explored several acoustic parameters of inhalation noises. For read speech we selected news items produced by professional news casters [20]. 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 have an average duration of 50 ms, whereas the breath noises themselves have a duration between 200 and 500 ms (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

10.21437/SpeechProsody.2020-161

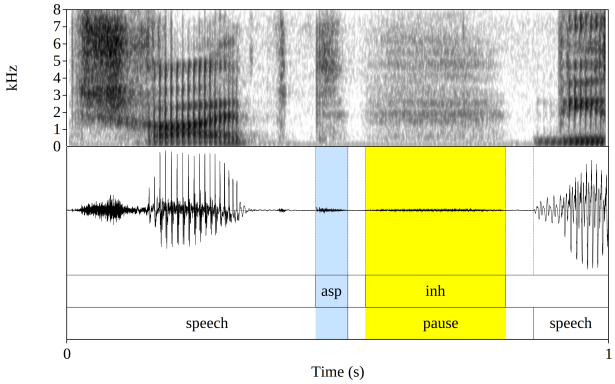

Figure 1: *1-sec section of (read) speech containing a pause with a typical inbreath noise between short edges of silence.*

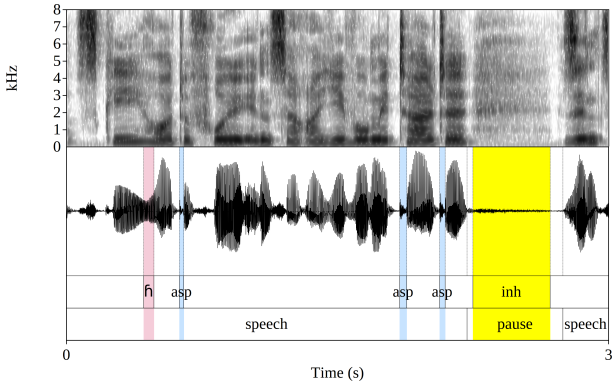

Figure 2: *3-sec section of (read) speech with inbreath noise (inh) in comparison with a voiced variant of the glottal fricative /h/ and three aspiration phases of unvoiced stops (asp).*

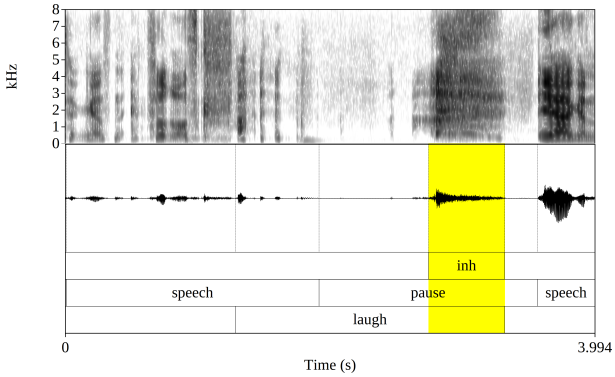

Figure 3: *4-sec section of (spontaneous) speech containing a laugh with an inbreath noise (inh) as an offset.*

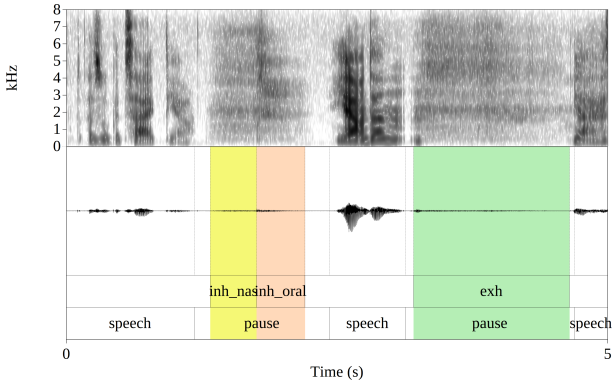

Figure 4: *Change from nasal to oral inhalation followed by speech and exhalation.*

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 [21, 22], which probably functions similarly in other Germanic languages.

Figure 2 depicts an example in which 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 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 [23]. It contains dyads of friends (same sex) given 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 (e.g. 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, some of the breath noises in the dialogues include only nasal inhalation. Occasionally 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. Figure 3 shows 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 turn-initial position. However, a form with a higher intensity can be assumed to function as a turn-claiming cue. The inhalation noise may also consist of two parts, for instance starting with a nasal inhalation noise immediately followed by an oral inhalation noise, as in Figure 4.

Finally, it should be mentioned that some inhalation noises are 'enriched' with tongue clicks [14], a discourse-related pause-internal particle that also occurs in languages that do not have clicks as phonemes.

3. Study

The aim of this analysis is to verify the rather general preliminary observations described in the previous section in different corpora and across different speakers. In addition, acoustic features of the rather frequent aspiration noises will be compared with inbreath noises.

3.1. Material and annotation

Four corpora of German speech were selected as the basis for our analysis. The first two corpora contain read speech, the other two contain spontaneous dialogues:

1. Read 1: the DIRNDL corpus [20] with recordings of professional newscasters
2. Read 2: read narrations (13 sentences) from the IFCASL corpus [24] recorded by native speakers of German
3. Spont 1: the GECO corpus [25] containing dyads of female students who do not know each other and who are talking about topics of their choice
4. Spont 2: the Lindenstraße corpus [23]

From each corpus we randomly selected five samples of exactly 60 seconds duration. Each sample contained speech of a different speaker. To avoid longer stretches of vocal inactivity or just short feedback utterances, we decided for the samples of spontaneous dialogues to randomly take one-minute sections with running speech where the speaker clearly has the turn.

Annotation was performed by hand using PRAAT [26]. In a first step correlates of perceived pauses were annotated whereby there was no specific durational threshold for pauses, i.e. pauses can also be shorter than, say, 100 ms, to mention a typical cut-off point for pauses. Each correlate of a perceived *pause* can have sub-parts, such as *oral inhalation noise*, *nasal inhalation noise*, *exhalation noise*, or *clicks*. Empty sub-parts of pauses were considered as *silence*.

For each oral inhalation noise, if present, its surrounding silences were labelled as *left edge* and *right edge*, respectively. *Aspiration* was annotated when an unvoiced stop consonant shows an aspiration noise longer than 20 ms.

In Praat the relevant values for *duration* were extracted. For *intensity* the difference of the following two values in dB was calculated: the mean of the values of 2 secs before and 2 secs after the corresponding pause minus the value for the entire breath noise. Similarly, the dB value of the aspiration noise was calculated relative to the mean of the values of 2 secs before and 2 secs after the aspiration noise.

3.2. Results

Frequency of occurrence. Table 1 shows that the number of pauses can differ between the different corpora. On an individual level, the number of pauses widely differs between speakers, with extreme values of 10 and 32 pauses per minute, respectively.

In general, there are many more pauses with breath noises than without. Only three out of the 20 speakers show more pauses *without* breath noises than pauses *with* breath noises.

The typical breath noise is an oral inhalation noise. Among all 20 speakers there are only three who have more breath noises other than oral inhalation. A more detailed view of the pause types per speaker and genre is offered in Figure 5.

Left and right edges are sometimes omitted. There seems to be a tendency for this to happen more frequently at the left edge of an oral inhalation noise than at its right edge. For read speech these values are higher than for spontaneous speech.

Table 1: *The four corpora with mean frequencies of occurrence per minute (standard deviation in parentheses) for pauses in general (pau), containing breath noise (br. n.), exclusively oral inhalation (oral), left edge (l. e.), and right edge (r. e.).*

Corpus	Read 1	Read 2	Spont 1	Spont 2
pau	16.4 (2.6)	22.4 (3.6)	19.0 (7.9)	16.6 (6.9)
br. n.	13.6 (3.8)	16.8 (2.4)	9.0 (3.4)	10.8 (4.1)
oral	10.6 (4.9)	10.6 (4.6)	6.6 (2.2)	6.2 (2.3)
l. e.	10.2 (1.9)	13.4 (3.0)	7.2 (3.6)	9.0 (3.3)
r. e.	12.0 (4.2)	13.2 (3.5)	8.4 (3.8)	10.0 (3.8)

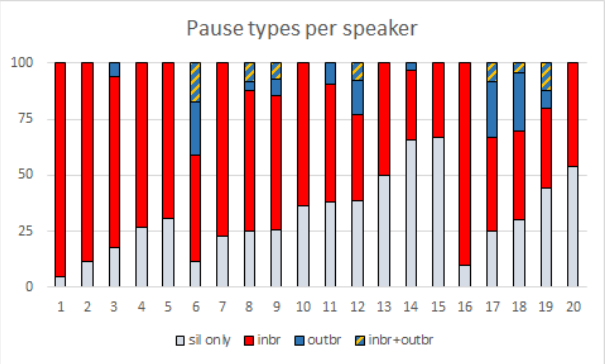

Figure 5: *Individual distribution of pause types in percent: without any breath noise (sil only), with inbreath noise, with outbreath noise, or with both. Speakers 1-5 are from Read 1, 6-10 from Read 2, 11-15 from Spont 1, and 16-20 from Spont 2.*

Duration. Figure 6 shows the durations of inbreath noises in comparison with the durations of the corresponding pauses. Table 2 shows that the professional speakers in Read 1 have substantially shorter breath noises than the non-professional speakers in the other three corpora.

The duration values for the left and right edges strongly vary between 23 and 96 ms with three corpora which show longer values for the left edge.

The mean duration for aspiration noises varies as well with again the professional speakers having the shortest values. The numbers clearly reveal that aspiration noises are by far shorter in duration than inbreath noises.

Intensity. The mean intensity difference between frication noises and their surroundings is clearly larger for oral inhalation noises than for aspiration noises. This can be seen for all four corpora in Table 3. The 'loudest' inhalation noises, in Spont 2,

Table 2: *The four corpora with mean values of duration in ms for inbreath noises (inbr. n.), left edge (l. e.), right edge (r. e.), and aspiration noise (asp. n.).*

Corpus	Read 1	Read 2	Spont 1	Spont 2
inbr. n.	291 (112)	418 (89)	408 (114)	441 (229)
l. e.	67 (70)	75 (79)	61 (17)	92 (63)
r. e.	42 (11)	23 (35)	96 (25)	63 (25)
asp. n.	44 (12)	62 (23)	67 (31)	81 (32)

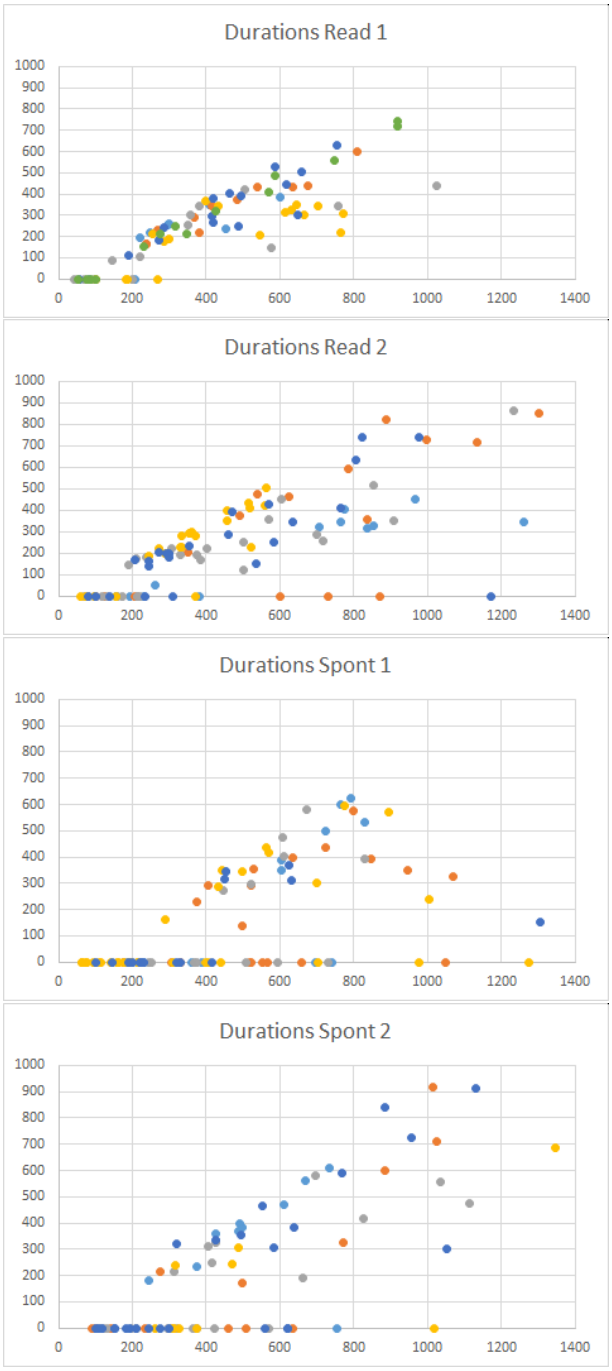

Figure 6: *Duration in ms for entire pauses (x-axis) and inbreath noises (y-axis) for the speech data from top to bottom: Read 1, Read 2, Spont 1, Spont 2.*

Table 3: *The four corpora with mean values of intensity differences in dB for inbreath noises (inbr. n.) and aspiration noise (asp. n.).*

Corpus	Read 1	Read 2	Spont 1	Spont 2
inbr. n.	29 (5)	28 (2)	28 (4)	16 (8)
asp. n.	13 (6)	9 (4)	10 (5)	7 (4)

are still softer than the 'softest' aspiration noises (in Read 1).

For the vast majority of data points there is a clear segregation between both categories with inbreath noises showing a lower intensity compared to aspiration noises. Only 5 out of 20 speakers showed single values that were located in the range of the respective other category.

3.3. Discussion

The investigated sample of 20 speakers clearly shows that the majority of *all* pauses that are often called 'silent' pauses are marked by breath noises. Thus, there is a minority of *phonetically* silent pauses. Read speech shows the fewest phonetically silent pauses. Among the observed breath noises the oral inhalation type markedly dominates, though a clear acoustic and auditory separation between purely oral and combinations of oral and nasal inhalation seems to be challenging.

The inbreath noises show a high degree of flexibility in their temporal extension, mostly within the range between 200 and 600 ms. There seems to be a weak correlation between breath noise duration and the duration of the entire pause. In spontaneous speech the duration values for breath noises, for phonetically silent pauses, and for pauses in general are longer compared to read speech.

Inbreath noises typically show a much lower intensity than the surrounding articulated speech. This is also valid for aspiration noises of stop consonants which also differ in duration.

4. Summary and conclusion

Although this study covers only a limited data set and thus has an exploratory character, it suggests that inhalation noises differ acoustically from other segments in spoken communication. Appropriate acoustic parameters that establish this difference are intensity and duration. They also include COG 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 by 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 from 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.

5. Acknowledgements

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

6. References

[1] P. Lieberman, *Intonation, Perception and Language*. Cambridge, Massachussetts: MIT Press, 1967.

[2] S. Fuchs, C. Petrone, J. Krivokapić, and P. Hoole, "Acoustic and respiratory evidence for utterance planning in German," *Journal of Phonetics*, vol. 41, no. 1, pp. 29–47, 2013.

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

[4] J. E. Kallay, U. Mayr, and M. A. Redford, "Characterizing the coordination of speech production and breathing," in *Proc. 19th International Congress of Phonetic Sciences (ICPhS)*, vol. 2019. Melbourne: NIH Public Access, 2019, pp. 1412–1416.

[5] J. Trouvain and K. P. Truong, "Prosodic characteristics of read speech before and after treadmill running," in *Interspeech*, Dresden, 2015, pp. 3700–3704.

[6] K. P. Truong, J. Trouvain, and M.-P. Jansen, "Towards an annotation scheme for complex laughter in speech corpora," in *Interspeech*, Graz, 2019, pp. 529–533.

[7] J.-A. Bachorowski and M. J. Owren, "Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect," *Psychological Science*, vol. 12, no. 3, pp. 252–257, 2001.

[8] W. L. Chafe, *The importance of not being earnest: The feeling behind laughter and humor*. John Benjamins Publishing, 2007, vol. 3.

[9] J. Trouvain, "Zur Wahrnehmung von manipuliertem Weinen als Lachen," in *22nd Konferenz Elektronische Sprachsignalverarbeitung*, Aachen, 2011, pp. 253–260.

[10] B. Winter and S. Grawunder, "The phonetic profile of Korean formal and informal speech registers," *Journal of Phonetics*, vol. 40, no. 6, pp. 808–815, 2012.

[11] J. Trouvain, "Affektäußerungen in Sprachkorpora," in *21st Konferenz Elektronische Sprachsignalverarbeitung*, Berlin, 2010, pp. 64–70.

[12] M. Kienast and F. Glitza, "Respiratory sounds as an idiosyncratic feature in speaker recognition," in *15th International Congress of Phonetic Sciences*, Barcelona, 2003, pp. 1607–1610.

[13] R. Lauf, "Aspekte der Sprechatmung: Zur Verteilung, Dauer und Struktur von Atemgeräuschen in abgelesenen Texten," in *Beiträge zu Linguistik und Phonetik*, A. Braun, Ed. Franz Steiner Verlag, 2001, pp. 406–420.

[14] M. Belz and J. Trouvain, "Are 'silent' pauses always silent?" in *19th International Congress of Phonetic Sciences (ICPhS)*, Melbourne, 2019, pp. 2744–2748.

[15] J. Trouvain and K. P. Truong, "Comparing non-verbal vocalisations in conversational speech corpora," in *Proceedings of the LREC Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, Istanbul, 2012, pp. 36–39.

[16] J. Trouvain and B. Möbius, "Zu Mustern der Pausengestaltung in natürlicher und synthetischer Lesesprache," in *29th Konferenz Elektronische Sprachsignalverarbeitung*, Ulm, 2018, pp. 334–341.

[17] D. H. Whalen, C. E. Hoequist, and S. M. Sheffert, "The effects of breath sounds on the perception of synthetic speech," *The Journal of the Acoustical Society of America*, vol. 97, no. 5, pp. 3147–3153, 1995.

[18] T. Fukuda, O. Ichikawa, and M. Nishimura, "Detecting breathing sounds in realistic japanese telephone conversations and its application to automatic speech recognition," *Speech Communication*, vol. 98, pp. 95–103, 2018.

[19] M. Włodarczak and M. Heldner, "Respiratory constraints in verbal and non-verbal communication," *Frontiers in psychology*, vol. 8, 2017, article id 708.

[20] K. Eckart, A. Riester, and K. Schweitzer, "A discourse information radio news database for linguistic analysis," in *Linked Data in Linguistics*. Springer, 2012, pp. 65–75.

[21] B. Möbius, "Corpus-based investigations on the phonetics of consonant voicing," *Folia Linguistica*, vol. 38, no. 1-2, pp. 5–26, 2004.

[22] F. Zimmerer and J. Trouvain, "Productions of */h/* in German: French vs. German speakers," in *Interspeech*, Dresden, 2015, pp. 1922–1926.

[23] IPDS, "Video Task Scenario: Lindenstraße – The Kiel Corpus of Spontaneous Speech," Institut für Phonetik und Digitale Sprachsignalverarbeitung Universität Kiel, DVD 4, 2006.

[24] J. Trouvain, A. Bonneau, V. Colotte, C. Fauth, D. Fohr, D. Jouvet, J. Jügler, Y. Laprie, O. Mella, B. Möbius, and F. Zimmerer, "The IFCASL corpus of French and German non-native and native read speech," in *Proc. 9th Language Resources and Evaluation Conference (LREC)*, Portorož, 2016, pp. 1333–1338.

[25] A. Schweitzer and N. Lewandowski, "Convergence of Articulation Rate in Spontaneous Speech," in *Interspeech*, Lyon, 2013, pp. 525–529.

[26] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, pp. 341–345, 2001.