INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Inhalations in speech: acoustic and physiological characteristics

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

[1]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

This paper examines the acoustic properties of breath noises in speech pauses in relation to similar speech segments and with regard to their inhalation speed. We measured intensity, center of gravity, and formants, as well as kinematic data (via Respiratory Inductance Plethysmography) for inhalations, aspirations of stops, glottal fricatives, and schwa vowels. We find that inhalations within speech are louder than those initiating speech, share spectral properties (center of gravity) with the aspiration phase of /k/-realizations, and generally involve a more open vocal tract (higher F1) than schwa-realizations. Intensity, center of gravity, and F1 are found to be positively correlated to inhalation speed. Overall, we conclude that jaw openness and inhalation speed are major contributors to inhalation noises in speech pauses.

Index Terms: speech breathing, respiration, breathing kinematics, breath acoustics.

1. Introduction

Breath noises became particularly interesting in the current COVID-19 pandemic [1, 2]. In general, breathing noise is mostly examined from a medical point of view, e.g., as during auscultation (listening through a stethoscope) [3], where the sound is recorded at the chest (anterior or posterior), the trachea, or the nasal cavity – rather than at the mouth – to study lung functions. It has also been used in investigating closure or constrictions in the vocal tract for sleep apnea (and often snoring, see e.g., [4]).

Despite their high frequency during articulatory activity [5], a comprehensive acoustic description of breath noises produced by healthy individuals' breathing during speech (see Fig. 1 for an example) has not yet been provided. Hence, the aim of this study is to examine the acoustic characteristics of breath noises in speech compared to similar speech segments.

While breath noise is so abundantly related to speech, it is surprising that its spectral properties are understudied. Knowing these properties, and in particular, the similarities and differences between breath noises and selected speech sounds would improve word aligners, which may mix up breath noises with speech events and then reorganize the whole speech stream accordingly. If breath noise and inhalation in physiological signals are linearly related, it might allow us to draw certain conclusions about the underlying respiratory signals even without directly measuring them, as attempted in [6].

Several mechanisms may underlie the production of breath noise: First, the speed of inhalation may influence breath noise, because speech breathing is more audible than breathing at rest [7]. Breathing cycles during speech are characterized by short, rapid inhalations and long, slow exhalations used for speech production, resulting in sawtooth-shaped breathing profiles. Audible breath noises are typically produced by these

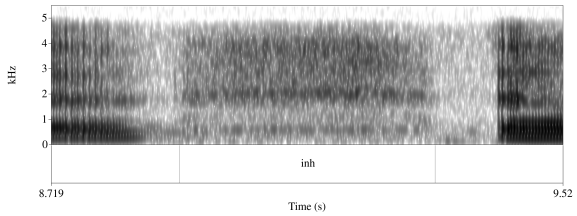

Figure 1: *Example of a breath noise ('inh') with visible formants.*

short and deep inhalations during speech. At rest, listening or performing inner speech, inhalations are only slightly shorter than exhalations within breathing cycles, giving their profiles a more symmetric and rounded shape than during speech production [8].

Second, and probably more importantly, breath noises can be generated by any constriction in the vocal tract that may result from different coordinations of the respiratory system along with the glottis, the velum, the lips, and the tongue.

At rest, the vocal tract is usually closed and the velar port is open, allowing an ingressive airstream from the atmosphere through the nose into the lungs for gas exchange. The closed mouth at rest should be the default in healthy individuals. Frequent inhalation through the mouth, not the nose, during childhood can even lead to craniofacial pathologies [9]. During speech production, purely nasal inhalation can occur, but it is probably limited because the nostrils can cover only a smaller volume of air in comparison to mouth aperture. Therefore, mouth breathing may be typical for speech, but nose breathing is preferred at rest.

Inhalation through the mouth is accompanied by the lowering of the jaw and the opening of the glottis (a closed glottis would not allow air to pass to the lungs and a closed mouth would only allow nasal breathing). While the jaw needs to be lowered, it is unclear whether the tongue simply rests on the jaw, as in the configurations for /a/ or /ə/ or whether it is affected by coarticulation with the preceding and following segments. Mouth opening and inhalation onset are generally coupled, too, but there may be some variability [10].

For speech, there is also a third option (neglecting the options of sequential nasal-oral or oral-nasal inhalation) of simultaneous oral and nasal inhalations to deal with the larger air volume requirement in speech breathing while preserving some of the benefits of nasal breathing, such as filtering, humidifying, and warming the air [11].

Another crucial factor for audible noise is the coordination between inhalation and glottal aperture. In principle, noise can be generated due to a glottal constriction, similar to the production of /h/, but with ingressive instead of egressive airflow. At rest, the vocal tract is closed and the glottis is open, but in run-

http://dx.doi.org/10.21437/Interspeech.2021-1262

ning speech the glottis can also be closed before inhalation, because it is surrounded by segments involving phonation. With a preceding voiced segment, the glottis needs to move from a closed to an open configuration for inhalation. When the mouth is open as well, the opening of the glottis and the ingressive airstream may result in audible inhalation. When the mouth is closed, the inhalation may not be audible. In some cases, opening the mouth leads to weak clicks, sometimes referred to as percussives [12]. The inhalation is often marked by a sudden and strong vertical downwards movement of the larynx. In combination with an increased glottal opening, sufficient negative pressure can be reached to generate tongue clicks [13, 14].

Breath noises may be used for speaker identification [15, 16]. When looking at breath noises produced by singers, [17] found high variability for duration (from 50 ms to 1225 ms) and spectral peaks at 1.7 kHz in female participants. They have also received some attention in the context of automatic breath detectors [18, 19, 20, 21] for a variety of applications.

The current work aims to investigate the spectral properties of breath noises in speech breathing using acoustic and respiratory data. We will focus on three main questions:

1. Do breath noises have similar spectral properties as [h] and aspirations of stops?

If noise is generated at the glottal level, we may expect similarities with the voiceless glottal fricative. If noise is generated in the upper vocal tract, we may find similarities with aspiration noise in stops. We do not consider sibilants, because for the production of these sounds, the jaw needs to be in a high position so that the lower incisors can function as an obstacle source.

2. Does breath noise reveal similar formant structures as /ə/-realizations?

Initial inspections of the breath noise revealed a formant structure, even in the absence of phonation (cf. Fig. 1). This would speak for vowel-like vocal tract configurations during the lowering of the jaw. We chose /ə/ because we do not assume any specific articulatory target for either /ə/ or breath noises.

3. Is there a relation between the speed of inhalation and certain acoustic parameters? Specifically, does speed of inhalation reveal a correlation with the first formant frequency (corresponding to the lowering of the jaw), center of gravity, and acoustic intensity?

We aim to investigate the relation between spectral properties and the physiological breathing signal since speech breathing is often related to the opening of the mouth.

2. Methodology

2.1. Material

We used a subset of the material described in [22] where five fables were retold in German by each subject. All 31 participants analyzed in the present study were female native speakers of German with a mean age of 25 years (age range: 21–32 years, normal body mass index).

The files generally consisted of three phases: Speech phases with a mean duration of 41.2 s $\pm$ 12.3 s (standard deviation) were preceded by pre-speech inactivity (7.6 s $\pm$ 2.6 s) and followed by post-speech inactivity (8.0 s $\pm$ 2.8 s).

The data include audio as well as kinematic data collected via Respiratory Inductance Plethysmography (RIP). Movement of the rib cage (RC) and the abdomen (AB) were measured by placing one elastic band at the level of the axilla and another one at the umbilicus. Compression and expansion of the AB and RC are monitored to infer in- and exhalations, respectively. We used the sum of AB and 2$\times$RC movements to get a more realistic representation of the lung volume (see [22] for a detailed justification).

Due to recording conditions, the audio files were sampled with 11,030 Hz resulting in a frequency range from 0 to 5515 Hz. This, however, does not pose a major problem for the acoustic analysis as this range is sufficient for the segments inspected here, i.e., breath noises, glottal fricatives, [ə], and aspiration phases of plosives [23].

2.2. Annotation

In this study, we only focus on audible inhalation noises; therefore, we hand-annotated them to separate them from the surrounding edges of silence [24, 25]. The breath noises were categorized according to their position: *inh* for inhalation within speech; *inh-ini* when immediately preceding speech initiation; and *n-inh* when outside of speech phases, i.e., in articulatory inactivity. We decided to separate *inh-ini* from *inh* as they might differ based on the involvement of inhalation noises in turn-taking [26, 27], which are typically louder than those in tidal breathing [28]. Thus, there might be an intensity difference based on 'turn' taking (even though there is no real dialogue situation in the data here) or speech initiation [29], as speech planning is connected to both inhalation duration and depth [30].

As for speech segments, we annotated aspiration phases of fortis plosives (as *p-asp, t-asp, k-asp*, depending on place of articulation), voiceless glottal fricatives (*[h]*; voiced variants not included), and mid central vowels (*[ə]*; the more open [ɐ] was not included). These were chosen because of their potential similarity to breath noises either due to the glottal opening in production (aspirations and glottal fricatives) or the neutral configuration of the vocal tract (*[ə]*). *p-asp* was removed due to a small number of data points. *[ə]* and *n-inh* are only used in the section on formants. Overall, we found 690 instances of *inh*, 138 *inh-ini*, 101 *n-inh*, 259 *[h]*, 185 *k-asp*, 537 *t-asp*, and 675 *[ə]*.

The assessment of the airway used (nasal or oral or combined) was not part of the experiment when the material was collected. An assessment based on audio alone does not seem to be reliable; for this reason a distinction between nasal and oral was not included here. For the time being, we adopt the findings of [11] for a prevalence of around 90% or more depending on the task for simultaneous usage of nasal and oral airways (over nasal only, oral only, and alternating nasal and oral) as a working hypothesis for breath noises.

Even though intensity is highly sensitive to several factors (such as distance to microphone, acoustic conditions, ambient noise), it is included here. Preserving the same distance to the microphone was controlled for in the experimental setup (cf. [22]). All participants in this study are female, so biological sex as a potential factor (via higher respiratory flow leading to louder breathing, as is the case in auscultation [3]) is eliminated. To account for local speaker-dependent differences we normalized the intensity of the examined segments by subtracting the mean intensity of the segment from the mean intensity of the entire speech activity in the respective file. As a result of this normalization, the 'normalized intensity' will be lower for more intense and higher for less intense segments.

Table 1: *Mean duration and standard deviation of breathing events and speech segments (h, k-asp, t-asp) without [ə] in ms.*

Segment type	mean	sd
inh	408	150
inh-ini	535	241
speech segments	68	36

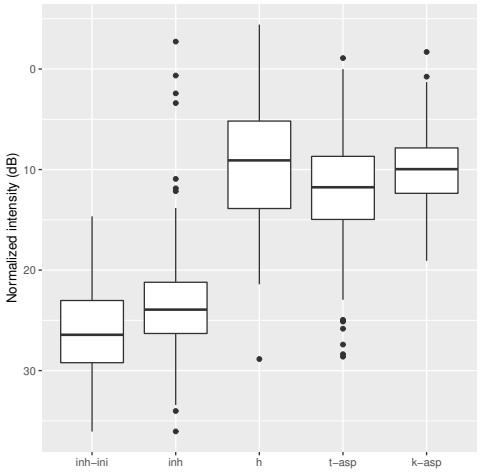

Figure 2: *Normalized intensity (dB) of the respective segments. y-axis is reversed, since lower values represent higher intensity.*

2.3. Procedure

The acoustic data were extracted from the audio signal using a Praat [31] script and the kinematic breathing signals using MATLAB (2017b). Acoustic parameters were taken as averages over the duration of the segment, except for the formants, where the parameters were extracted and averaged for the central third of the segment to control for potential coarticulation effects. Formant objects were created with maximum frequency set to 5,500 Hz and 5 formants. From the temporal segmentation, we obtained the corresponding lung volume (sum signal) at the onset and offset of the segment. From these temporal (x) and displacement (y) events, we calculated the respiratory slope for each segment, using the formula $slope = (y_2 - y_1)/(x_2 - x_1)$. Inhalation slope thus corresponds to the speed of inhalation.

All statistical results reported here come from linear mixed effects models calculated with the lme4 [32] package in R [33]. For the formant data, the Pillai score was calculated to measure vowel overlap [34, 35], with lower values indicating higher degrees of overlap.

3. Results

3.1. Duration and intensity

Both types of inhalations are longer than the speech segments (Table 1). Inhalations right before speech (*inh-ini*) tend to be longer than those sandwiched between speech (*inh*).

For intensity (Fig. 2), there is a separation between breath segments and speech segments, with the latter being more intense. Within the breath noises, *inh* tends to be slightly louder than *inh-ini*. This is reflected in the model $lmer(norm\_int \sim segment + (1 + segment \mid speaker))$ using *inh* as the intercept (23.62, t=47.3, p<0.001), suggesting that *inh-ini* is less loud (2.46, t=5.39, p<0.001), whereas the speech segments are

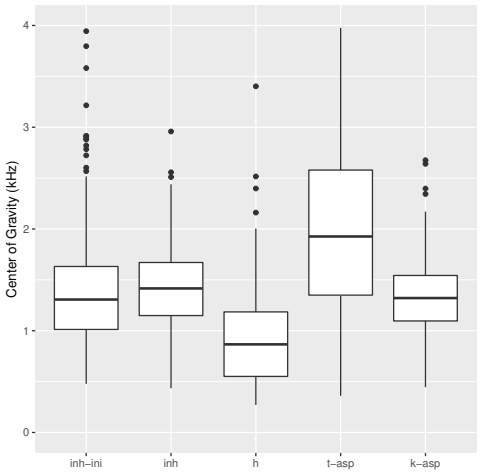

Figure 3: *Center of Gravity of the respective segments in kHz.*

Figure 4: *F1-F2 vowel chart for inhalations within speech (inh, orange) and outside of speech (n-inh, blue) compared to [ə] (green). Ellipses refer to density.*

all louder: *h* (-14.18, t=-21.71, p<0.001), *t-asp* (-11.21, t=-18.63, p<0.001), *k-asp* (-13.28, t=-22.96, p<0.001). The difference between the two types of inhalation may be related to shorter durations, assuming that a similar amount of air is being inhaled.

3.2. Center of Gravity

The differences in center of gravity (CoG) between the inhalation and speech segments can be seen in Fig. 3. We used the log-transformed CoG to account for linearity of the residuals in the model $lmer(logCoG \sim segment + (1 \mid speaker))$ with *inh* as intercept (7.23, t=451.08, p<0.001), showing an effect for *h* (-0.51, t=-20.94, p<0.001) and *t-asp* (0.28, t=14.27, p<0.001) but no significant effect for *inh-ini* (-0.04, t=-1.1, p=0.253) and *k-asp* (-0.05, t=-1.82, p=0.069).

3.3. Formants

Fig. 4 shows a vowel chart plotting F1 and F2 values for inhalations surrounded by speech (*inh*) and in phases of articulatory inactivity (*n-inh*) compared to /ə/-realizations.

While F2 tends to be less variable and generally higher in

inh, i.e., slightly more front than *[ə]*, F1 separates them with *inh* involving a more open vocal tract than the neutral vowels. While there is virtually no density overlap for these two types of inhalation, inhalations outside of speech occupy a position that overlaps with both the *inh* and the *[ə]* regions with a separate peak in each. To measure the overlap between these three segments, we calculated Pillai scores using F1, F2, and F3: There are similar degrees of overlap between both *inh* and *n-inh* (V=.28, F(1, 789)=102.44***) and *n-inh* and *[ə]* (V=.21, F(1, 774)=70.03***). The degree of overlap is lowest between *inh* and *[ə]* (V=.74, F(1, 1363)=1319.1***).

3.4. Relation between inhalation speed and acoustic properties

We assume that deeper inhalations are related to faster inhalations since in articulatory kinematics, movement velocity and displacement are positively correlated [36]. Taking a deep breath involves lowering the jaw to allow air to pass through the vocal tract without large obstructions. In the current data set, jaw motion was not obtained, but the first formant might approximate the degree of jaw opening (larger jaw opening for higher F1 values). To test the effect of inhalation speed on intensity, CoG, and F1, we subset the data to include only *inh* with the centralized inhalation slope as a continuous predictor. For each of those three parameters, a separate model was run with speaker-specific random slopes for inhalation slope as random effects on 30 speakers (n=667) without further normalization.

For intensity, the model $lmer(norm\_int \sim slope\_cen + (1 + slope\_cen \mid speaker))$ returns an intercept of 27.03 (t=26.53, p<0.001) and shows an effect on intensity for slope (-2.27, t=-4.58, p<0.001). The effect for faster inhalation or higher volume intake, which leads to more intense inhalation noise, is visualized in Fig. 5 (left).

For CoG, the model $lmer(logCoG \sim slope\_cen + (1 + slope\_cen \mid speaker))$ gives an intercept of 6.80 (t=198.9, p<0.001) and shows an effect for slope (0.30, t=10.6, p<0.001). Shorter inhalations or situations where larger amounts of air are inhaled in the same time thus lead to a higher center of gravity (Fig. 5, middle).

The model for inhalation slope and F1 ($lmer(F1 \sim slope\_cen + (1 + slope\_cen \mid speaker))$) finds an intercept of 742.34 (t=60.64, p<0.001) and reveals a significant effect (34.73, t=6.16, p<0.001) for slope. Their relationship is visualized in Fig. 5 (right). Z-transformation was only carried out for visualizing the data, i.e., pooling all speakers together while making sure the correlation is not a by-product of individual differences in breathing slope or F1. The output is encouraging and suggests that a large part of the breathing noise is related to the motion of the jaw.

4. Discussion

The present study found intensity to be higher for inhalations within speech (*inh*) than for speech initiation (*inh-ini*). We assume this is caused by a steeper inhalation slope that was shown to increase intensity in *inh*. The CoG of *inh* is not significantly different from those of *inh-ini* or *k-asp*, which may have implications for modeling the 'place of articulation' for breath noises. Higher inhalation speed also increases CoG. When looking at F1, F2, and F3, we find F1 to separate *inh* from a neutral configuration of the vocal tract as seen in *[ə]*. This difference may be linked to the wider mouth opening in inhalations. This idea is also supported by the positive correlation between inhalation slope (as a measure of inhalation speed) and F1.

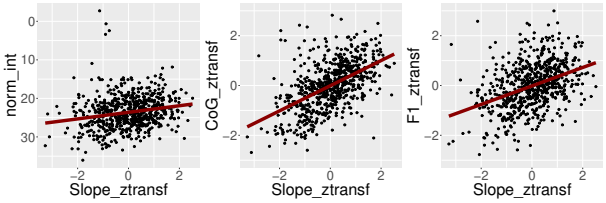

Figure 5: *Correlations between inhalation slope and normalized intensity (norm_int, left), inhalation slope and CoG (middle), and inhalation slope and F1 (right) for all 'inh' data. All speakers are pooled together. To account for different anatomical properties, data were z-transformed by speaker. norm_int was not transformed as it already underwent a normalization procedure; its y-axis is reversed to visualize the underlying positive correlation (cf. Fig. 2 and section 2.2).*

We found a relationship between inhalation slope and the acoustic parameters of the inhalation noise, suggesting that degree of jaw lowering and inhalation speed are major contributors to the creation of these noises. For confirmation, these findings should be tested by including articulatory data (e.g., electromagnetic articulography). Our findings held in speaker-wise examination and should thus also be applicable to male speakers whose breathing is expected to be similar ([37], p. 57).

A limitation of this study is the absence of the source component (i.e., vocal fold vibration) from the source-filter model since breath noises are typically voiceless. However, even being unable to derive a clear position of the tongue in the vocal tract through breathing noise, there is still a positive relationship between inhalation slope and F1 within a given speaker. This indicates that each speaker has a specific lung volume and vocal tract properties that are anatomically determined. Extracting formants seems to work in breath noises even though F2 is generally more prominent than F1 (Fig. 1). Finally, the results for CoG suggest that averages of F1 and F2 could be merged here.

5. Conclusions

The present study found that inhalations within speech share spectral properties (CoG) with the aspiration phase of /k/-realizations and generally involve a more open vocal tract (higher F1) than /ə/-realizations. Intensity, CoG, and F1 were found to be positively correlated with inhalation speed.

These findings have implications for models of speech production in general, the automatic classification of breath noises, and the non-invasive detection of abnormal behavior for the diagnosis of diseases or disorders.

Breath noises should also be evaluated in tasks other than the pseudo-spontaneous setting used here. Another area of interest is coarticulation between breath noises and surrounding speech. Finally, categorizing breath noises based on airway usage may lead to different outcomes concerning the acoustic parameters.

6. 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 thank Beeke Muhlack, Mikey Elmers, and our student helpers Christin Weiß and Sven Kirchner for their assistance.

7. References

[1] E. G. Furman, A. Charushin, E. Eirikh, S. Malinin, V. Sheludko, V. Sokolovsky, and G. Furman, "The remote analysis of breath sound in COVID-19 patients: A series of clinical cases," *medRxiv*, 2020.

[2] B. W. Schuller, D. M. Schuller, K. Qian, J. Liu, H. Zheng, and X. Li, "COVID-19 and computer audition: An overview on what speech sound analysis could contribute in the SARS-COV-2 corona crisis," *arXiv*, 2020.

[3] A. Oliveira and A. Marques, "Respiratory sounds in healthy people: A systematic review," *Respir. Med.*, vol. 108, no. 4, pp. 550–570, 2014.

[4] J. Solà-Soler, R. Jané, J. A. Fiz, and J. Morera, "Formant frequencies of normal breath sounds of snorers may indicate the risk of obstructive sleep apnea syndrome," *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS'08 - "Personalized Healthc. through Technol.*, pp. 3500–3503, 2008.

[5] J. Trouvain, B. Möbius, and R. Werner, "An Acoustic Analysis of Inbreath Noises in Read and Spontaneous Speech," in *Proc. 10th Int. Conf. Speech Prosody*, Tokyo, 2020, pp. 789–793.

[6] Z. Mostaani, V. S. Nallanthighal, A. Härmä, H. Strik, and M. Magimai-Doss, "On the relationship between speech-based breathing signal prediction evaluation measures and breathing parameters estimation," in *ICASSP 2021 - IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021.

[7] R. Werner, J. Trouvain, S. Fuchs, and B. Möbius, "Exploring the presence and absence of inhalation noises when speaking and when listening," in *Proc. 12th Int. Semin. Speech Prod. ISSP 2020*, In Press.

[8] B. Conrad and P. Schönle, "Speech and respiration," *Archiv für Psychiatrie und Nervenkrankheiten*, vol. 226, no. 4, pp. 251–268, 1979.

[9] D. Harari, M. Redlich, S. Miri, T. Hamud, and M. Gross, "The effect of mouth breathing versus nasal breathing on dentofacial and craniofacial development in orthodontic patients," *The Laryngoscope*, vol. 120, no. 10, pp. 2089–2093, 2010.

[10] O. Rasskazova, C. Mooshammer, and S. Fuchs, "Temporal coordination of articulatory and respiratory events prior to speech initiation," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, pp. 884–888, 2019.

[11] R. A. Lester and J. D. Hoit, "Nasal and oral inspiration during natural speech breathing," *J. Speech, Lang. Hear. Res.*, vol. 57, no. 3, pp. 734–742, 2014.

[12] R. Ogden, "Clicks and percussives in English conversation," *J. Int. Phon. Assoc.*, vol. 43, no. 3, pp. 299–320, 2013.

[13] S. Fuchs and B. Rodgers, "Negative intraoral pressure in German: Evidence from an exploratory study," *J. Int. Phon. Assoc.*, vol. 43, no. 3, pp. 321–337, 2013.

[14] J. Trouvain, "Laughing, Breathing, Clicking - The Prosody of Nonverbal Vocalisations," *Proc. Int. Conf. Speech Prosody*, pp. 598–602, 2014.

[15] M. Kienast and F. Glitza, "Respiratory sounds as an idiosyncratic feature in speaker recognition," in *Proc. 15th ICPhS*, Barcelona, 2003, pp. 1607–1610.

[16] J. Chauhan, Y. Hu, S. Seneviratne, A. Misra, A. Seneviratne, and Y. Lee, "BreathPrint: Breathing acoustics-based user authentication," in *Proc. 15th Annu. Int. Conf. Mob. Syst. Appl. Serv.* ACM, 2017, pp. 278–291.

[17] T. Nakano, J. Ogata, M. Goto, and Y. Hiraga, "Analysis and Automatic Detection of Breath Sounds in Unaccompanied Singing Voice," *ICMPC10*, no. 10, pp. 387–390, 2008.

[18] N. Braunschweiler and L. Chen, "Automatic detection of inhalation breath pauses for improved pause modelling in HMM-TTS," in *8th ISCA Work. Speech Synth.*, 2013, pp. 1–6.

[19] É. Székely, G. E. Henter, and J. Gustafson, "Casting to Corpus: Segmenting and Selecting Spontaneous Dialogue for TTS with a CNN-LSTM Speaker-dependent Breath Detector," *ICASSP 2019 - 2019 IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 6925–6929, 2019.

[20] E. E. Hamke, R. Jordan, and M. Ramon-Martinez, "Breath activity detection algorithm," 2016.

[21] S. H. Dumpala and K. N. Alluri, "An algorithm for detection of breath sounds in spontaneous speech with application to speaker recognition," in *Speech Comput. 19th Int. Conf. SPECOM*, A. Karpov, R. Potapova, and I. Mporas, Eds., 2017, pp. 98–108.

[22] A. Rochet-Capellan and S. Fuchs, "Changes in breathing while listening to read speech: the effect of reader and speech mode," *Frontiers in psychology*, vol. 4, p. 906, 2013.

[23] M. Zellers and B. Schuppler, "Microprosodic variability in plosives in German and Austrian German," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, pp. 656–660, 2020.

[24] D. Ruinskiy and Y. Lavner, "An effective algorithm for automatic detection and exact demarcation of breath sounds in speech and song signals," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 3, pp. 838–850, 2007.

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

[26] D. H. McFarland, "Respiratory Markers of Conversational Interaction," *J. Speech, Lang. Hear. Res.*, vol. 44, no. 1, pp. 128–143, 2001.

[27] A. Rochet-Capellan and S. Fuchs, "Take a breath and take the turn: How breathing meets turns in spontaneous dialogue," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 369, no. 1658, 2014.

[28] M. Włodarczak and M. Heldner, "Respiratory belts and whistles: A preliminary study of breathing acoustics for turn-taking," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, pp. 510–514, 2016.

[29] J. M. Scobbie, S. Schaeffler, and I. Mennen, "Audible aspects of speech preparation," *ICPhS XVII*, pp. 1782–1785, 2011.

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

[31] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2020. [Online]. Available: http://www.praat.org/

[32] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[33] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: https://www.R-project.org/

[34] J. Hay, P. Warren, and K. Drager, "Factors influencing speech perception in the context of a merger-in-progress," *J. Phon.*, vol. 34, no. 4, pp. 458–484, 2006.

[35] J. Nycz and L. Hall-Lew, "Best practices in measuring vowel merger," *Proc. Meet. Acoust.*, vol. 20, no. 1, 2014.

[36] D. J. Ostry and K. G. Munhall, "Control of rate and duration of speech movements," *The Journal of the Acoustical Society of America*, vol. 77, no. 2, pp. 640–648, 1985.

[37] T. J. Hixon, G. Weismer, and J. D. Hoit, "Preclinical Speech Science: Anatomy, Physiology, Acoustics, and Perception," *Plur. Publ.*, p. 728, 2020.