# RESPIRATORY AND SUPRALARYNGEAL EFFECTS ON SPEECH BREATHING NOISE ACROSS LOUDNESS CONDITIONS AND SPEAKING TASKS

Susanne Fuchs[1], Laura L. Koenig[1,2,3], Raphael Werner[4]

[1]Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS) Berlin (Germany), [2]Adelphi University, New York (USA), [3]Haskins Labs, New Haven (USA), [4]Universität des Saarlandes, Saarbrücken (Germany)

fuchs@leibniz-zas.de, koenig@haskins.yale.edu, rwerner@lst.uni-saarland.de

## ABSTRACT

Respiratory and supralaryngeal actions differ between speech and vegetative breathing, and speech-breathing sounds convey information to listeners. To date, little work has explored how breath noise is shaped by respiratory and supralaryngeal actions. Here, we assess respiratory movement, obtained using inductive plethysmography, and lip apertures, collected using electrostomatography, for eight typical German-speaking women. We varied loudness and speaking task since these can be expected to affect breathing behavior. Participants were recorded while reading, producing a spontaneous narrative, and engaging in an interactive memory game with an interlocutor. The data show widespread effects of speaking task on breath acoustics and respiratory kinematics, with more naturalistic tasks generally yielding louder breath noise, and steeper inhalation slopes. Louder speech is associated with greater oral apertures. Breath intensity correlates with inhalation slope. Lip apertures do not correlate with breath acoustics during reading, but may do so for other tasks.

**Keywords:** speech breathing, speech kinematics, loudness, task effect

## 1. INTRODUCTION

Breathing is a natural rhythm that structures the flow of speech temporally [1]. Breath sounds are ubiquitous in spoken language. They mostly occur during inhalation, but can also be produced under physical effort, i.e. forced exhalation. Speech usually involves egressive airflow, but ingressive phonation can also be observed [2] and may convey information about intense emotional states [3]. In general, breath noise may provide listeners with considerable information about the speaker [4] (e.g., Is it a human or a machine? What is the level of physical activity or emotion?) and may also be crucial for management of turn-taking.

In rest breathing, nasal inhalation is the default. However, speech breathing requires faster gas exchange. Previous work has shown that, besides oral-nasal sequences, breath noise can be a consequence of oral, nasal or simultaneous oral-nasal inhalation. In [5], speech was elicited by asking the participant to count, read a paragraph, speak spontaneously about their favorite activity, and do a phone call with a familiar person. The authors also assessed coarticulatory effects of surrounding sounds. The findings showed that speakers typically produce simultaneous oral-nasal breathing in speech. The data also showed a non-significant trend suggesting possible differences in speech breathing patterns as a function of speech task.

Although there have been some studies on articulation during pauses [6, 7, 8, 9] it is neither clear whether inhalations have been produced during these pauses nor what acoustic consequences a breath-related articulation has. Only a few studies have analyzed the relationship between acoustics and articulation of breath noise.

Our work follows up on Werner and colleagues [10]. They compared acoustic characteristics of breath noises to speech sounds with similar acoustic or articulatory characteristics, such as aspiration noise in stops and formants in vowels. The authors also investigated the relationship between acoustic measures of inhalation noise and respiratory kinematics using Inductance Plethysmography. They showed that the Centre of Gravity of breathing noise was similar to the aspiration noise of /k/, but not to that of /t/, nor to the glottal frication noise of /h/. The authors [10] found that breath noises had a higher F1 and slightly higher F2 than /ə/ and thus seemed to involve more open, slightly fronted vocal tract configurations. Results revealed that inhalation slope, i.e. how fast participants inhale, was positively correlated with the Center of Gravity, F1, and intensity of the breath noise. However, articulatory data on mouth openings were not recorded.

The current study extends previous work in three directions to investigate 1) breath noises in different speech tasks (reading, an interactive game, a narrative); 2) the effect of loudness (speech produced with normal and loud intensity) on breath noise; and 3) the relation between breath noise, inhalation slope, and lip opening.

Exploration of different speech tasks is motivated by [5] and the fact that in a conversation an interlocutor may breathe silently through the nose when listening, but may change her breathing pattern when taking a turn. Breathing while telling a story may add more variability to the acoustic properties of the breath signal because inhalation depth roughly corresponds to the length of the upcoming sentence which may vary considerably in spontaneous speech [11].

Varying loudness as a condition is motivated by the fact that loudness is known to modify breathing behavior [12] as well as acoustics [13] and supralaryngeal articulation [14, 15]. It is, however, unclear whether loudness affects only speech itself, or also enhances voiceless inhalation noise. Finally, using a multimodal instrumental setup permits investigating acoustic-supralaryngeal-respiratory relations directly.

We predict that speaking with higher intensity (in the loud condition) leads to more rapid inhalation and also affects the intensity of the breath noise. However, since there is no phonation in breath noise, the effect might be much smaller than in speech itself. Furthermore, we expect that F1 of the breath noise is related to lip aperture and inhalation slope.

For speech task, our predictions are rather tentative. However, following [5], we suspect that more spontaneous speech tasks may differ from read speech, and in particular may show greater variability in acoustic and kinematic measures.

## 2. METHODOLOGY

### 2.1. Experimental design

Two different conditions were experimentally varied: *Speech Task* and *Loudness*. Speech Task comprised the following:

1. Reading sentences (Read) of comparable length varying a target word (sentences were taken from [13]). All utterances began with "Ich mag" (I like) followed by a target word beginning with a bilabial oral or nasal stop (e.g., Mate, Paten, Buesum).

2. Speaking about a summer holiday (Holiday). Participants were instructed: "Do you remember a summer holiday? Please tell us about this holiday. You can also talk about other things. It is important that you speak for at least 2-3 minutes. It may also be longer."

3. Playing the game "I pack my suitcase and take with me ..." (Game). In this game, a participant has to propose one thing that she will take in her suitcase and the next participant needs to repeat the sentence and adds another item to the suitcase. In the end, the suitcase will be full of items and the aim of the game is to memorize everything in the right order. The game ends when one of the participants does not remember everything or puts items in the wrong order. The experimenter served as an interlocutor for the game task.

The reading sentences and the game were repeated twice for each loudness condition. Speech tasks were presented randomly to different participants. All participants produced these tasks first using their normal comfortable loudness and in the second part of the experiment, they were instructed to speak louder, because their interlocutor (the experimenter) was wearing fully ear-covering headphones. In case participants became softer after some time, the interlocutor gave signs that she could not understand and the participant should try to maintain the loudness level.

### 2.2. Participants

Eight female speakers, all native speakers of German, were recorded. They spoke Standard German (Northern variety). Speakers were selected on the basis of an available custom-made palate and tolerance to wearing this palate. The palates are part of an Electro-Optical Stomatographic System (EOS) [16]. Since the costs for these palates were relatively high, we could not afford to record more participants. Participant ages ranged from 21-34 years and Body Mass Indices were between 19 and 27.

### 2.3. Equipment

Acoustic data were recorded via two devices: a high-quality Sennheiser condenser microphone that was connected to an amplifier and a PC that simultaneously acquired respiratory and acoustic data. The second acoustic signal was of low quality and was recorded with the internal microphone of the laptop controlling the Electro-Optical

Stomatographic system (see next paragraph). The acoustic data from the two microphones were taken to post-synchronize the data between the PC recording breathing and the PC for the EOS device.

Respiratory kinematics were recorded via Inductance Plethysmography. Two flexible respiratory belts were wrapped around the participant's torso, one at the abdomen and one at the thorax. The EOS device was used to record lip opening during breathing. The EOS system includes a custom-made artificial palate that can record tongue-palatal contact patterns during speech, but additionally senses the mid-sagittal distance from the tongue to the palate, in case there is no contact. Two lip sensors were connected to the EOS system that record lip motions (protrusion and aperture). We will here concentrate on the lip aperture only.

### 2.4. Data pre-processing and annotation

The on- and offsets of breathing noise were labelled in Praat [17]. We then extracted acoustic parameters from the audio signal using a Praat script. F1, F2, F3 were averaged over the central third of the breath noise, to avoid the influence of potential coarticulation as done in [10], using the Burg algorithm with max. number of formants 5, max. formant 5.5 kHz, window length 0.025 s, and dynamic range 50 Hz. Intensity was averaged over the whole interval. Total lung volume was computed as $2 \times thorax + abdomen$ [18]. We then took the acoustically defined on- and offset of breathing noise ($x$) and obtained the corresponding respiratory values ($y$) for these time points. Inhalation slope was calculated as $slope = \Delta y / \Delta x$. It is thus a measure of how quickly speakers inhale.

As noted above, the EOS data processing focused on the lip aperture signal only. Given the poor quality of the associated acoustic signal, we also restricted ourselves to the reading task, where every utterance began with a pair of bilabial consonants (the /m/ in 'Mag' followed by the bilabial-initial target word). These two bilabial closures allowed us to orient to the EOS data and delimit mouth openings that preceded utterances. For these mouth openings, we obtained the aperture as the distance between the lip closure of /m/ and the maximum breath opening. Since lip opening movements during breathing showed considerable spatial and temporal variability, we used overall aperture instead of an opening slope or velocity.

Linear mixed-effect models were run in R [19] using the packages *lme4* [20] and *lmerTest* [21]; graphical exploration was done using *ggplot2* [22]. We describe the statistical models in the respective results section. We used $\alpha = 0.05$ for significance testing.

Our sample includes 1,518 inhalations overall. We excluded cases where inhalations were preceded or followed by laughter because this can coincide with some extreme respiratory dynamics [23] and changes in intensity. We also excluded cases where some obvious spectral changes occurred and perceptually these sounded like oral-to-nasal or nasal-to-oral inhalations. Our final sample consisted of 1,297 breath noises.

## 3. RESULTS

### 3.1. Do loudness and speech tasks affect acoustic parameters of breath noise and speech?

To answer this question, we ran the model $lmer(intensity \sim Loudness * Task + (1 + Loudness * Task|Speaker))$. Participants did change their speech intensity between the normal and the loud conditions. On average speech was 7.3 dB louder ($t$=7.205, $p$<0.001***) in the loud than in the normal condition. Speech task (Read speech, Game, Holiday) had no effect on this intensity level.

The results look different for speech breathing (Fig. 1): Here we ran the same model as described above, but used breath noise intensity as the dependent measure instead. In the loud condition, breath noise was not significantly louder; only a trend towards higher intensity could be found ($t$=2.17, $p$=0.067). However, contrary to speech intensity, an effect of the task was found. Overall, breath noise in the interactive Game task was louder ($t$=2.51, $p$=0.039*) than in Read speech and the Holiday task was also louder ($t$=8.252, $p$<0.001***) than Read speech.

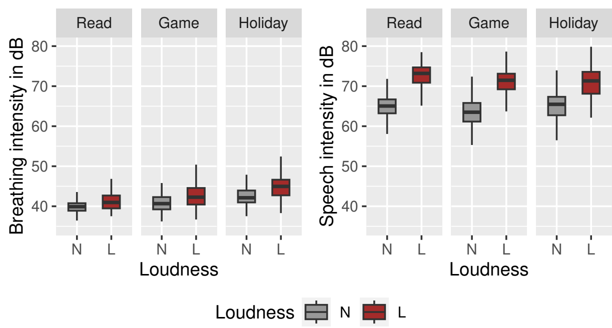


**Figure 1:** Boxplot for breathing noise intensity (left) and speech intensity (right) in different conditions.

F1 values of the breath noise only showed a trend towards slightly higher F1s (average 28 Hz) in loud speech ($t$=2.16, $p$=0.057) than in normal speech. A trend, but no effect was also found for speech task with F1 being slightly higher in the Game ($t$=2.07, $p$=0.075) and Holiday ($t$=2.001, $p$= 0.082) tasks than in Read speech. These outputs were obtained using the formula $lmer(F1 \sim Loudness * Task + (1 + Loudness * Task|Speaker))$.

### 3.2. Do loudness and speech tasks affect breathing slope?

Results for breathing slope are similar to those for intensity, as they only show a trend for loudness ($t$=2.14, $p$=0.07), but a main effect for speech task: breath noise in the Holiday task had a steeper slope than in the reading task ($t$=2.96, $p$=0.0211*).

### 3.3. Does loudness affect lip opening?

Finally, the effect of loudness on lip aperture was calculated as follows: $lmer(LipAperture \sim Loudness + (1 + Task|Speaker))$. Since only data for Read speech were analyzed, speech task did not serve as a fixed factor. Loudness affected lip opening, i.e. louder speech was produced with a more open lip configuration during inhalation ($t$=3.77, $p$=0.009**).

### 3.4. Relation between acoustic and articulatory parameters

To what extent are acoustic and articulatory parameters of breath noise correlated? To better understand the relations, we ran the following model: $lmer(intensity \sim slopeCentralized + Loudness * Task + (1 + Loudness * Task|Speaker) + (1 + slopeCentralized|Speaker))$. The predictor Breathing slope was centralized. Results revealed a strong effect (Fig. 2) of breathing slope on intensity of the breath noise ($t$=6.22, $p$=0.0005***).

A similar model with F1 as the dependent variable returned no significant effect of breathing slope.

Finally looking at the relation between lip aperture and F1, which was possible only for the Read speech data ($lmer(aperture \sim F1centralized * Loudness + (0 + F1centralized * Loudness|Speaker))$), there was only an effect of Loudness ($t$=2.55, $p$=0.048*).

## 4. DISCUSSION AND CONCLUSION

This study found the following effects: Breath noise intensity differed across speech tasks. It did not vary significantly across loudness conditions, which is probably due to the voiceless nature of

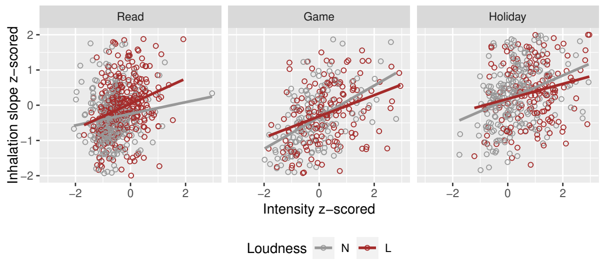


**Figure 2:** Scatterplot with linear regression lines showing a relation between inhalation slopes (z-scored by speaker) and breathing intensity (z-scored by speaker) in the three different speech tasks. Note, z-scoring is only done for graphical display. Grey colour corresponds to normal speech and dark red to loud speech.

the breath noise and speaker-specific realizations of the loudness condition. Similarly, F1 was higher in the Game and Holiday task than in Reading, i.e. it varied according to the task, but only showed an insignificant trend for normal versus loud speech. This was also true for inhalation slope, which showed no effect of Loudness but differences among tasks. A different pattern emerges from the lip aperture data, which were only analyzed for the reading task. In the loud speech condition, speakers opened their mouth wider when producing breath noise than in the normal condition. Thus, speech task had greater effects on acoustic parameters of breath noise than changing between normal and loud speech. In particular, reading had lower breath intensity and F1 values of breath noise and a shallower inhalation slope than one or both other tasks. One reason for this may be that the sentences in the Read task were relatively short and we kept them fairly constant in length. We assume that no deep inhalation was required for this reason and overall, data may be less variable than in the other tasks.

Looking at how acoustic and articulatory measures relate, we could confirm the relation between intensity and inhalation slope as proposed by [10] but could not find a relation between F1 and inhalation slope. We think that no correlation between F1 and lip aperture might be due to nasal coupling in some speakers (cf. [5]). However, data for the other speech tasks need to be investigated before any reliable conclusions can be drawn.

Speech breathing involves the coordinated actions among different articulators and all drivers of respiration. Using multimodal data to study different aspects of this coordination may be a fruitful topic for further investigations.

## 5. ACKNOWLEDGMENTS

This work was supported by the Leibniz Association. We like to thank our participants and Jörg Dreyer for technical support, Peter Birkholz and Alexander Wilbrandt for advice on the EOS system.

## 6. REFERENCES


[1] S. Fuchs and A. Rochet-Capellan, "The respiratory foundations of spoken language," *Annual Review of Linguistics*, vol. 7, no. 1, pp. 13–30, 2021.

[2] R. F. Orlikoff, R. J. Baken, and D. H. Kraus, "Acoustic and physiologic characteristics of inspiratory phonation," *The Journal of the Acoustical Society of America*, vol. 102, no. 3, pp. 1838–1845, 1997.

[3] A. Anikin and D. Reby, "Ingressive phonation conveys arousal in human nonverbal vocalizations," *Bioacoustics*, pp. 1–16, 2022.

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

[5] R. A. Lester and J. D. Hoit, "Nasal and oral inspiration during natural speech breathing," *Journal of Speech, Language, and Hearing Research*, vol. 57, pp. 734–742, 2014.

[6] B. Gick, I. Wilson, K. Koch, and C. Cook, "Language-specific articulatory settings: Evidence from inter-utterance rest position," *Phonetica*, vol. 61, no. 4, pp. 220–233, 2004.

[7] V. Ramanarayanan, E. Bresch, D. Byrd, L. Goldstein, and S. S. Narayanan, "Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. EL160–EL165, 2009.

[8] O. Rasskazova, C. Mooshammer, and S. Fuchs, "Temporal coordination of articulatory and respiratory events prior to speech initiation." in *Interspeech*, 2019, pp. 884–888.

[9] J. Krivokapić, W. Styler, and D. Byrd, "The role of speech planning in the articulation of pauses," *The Journal of the Acoustical Society of America*, vol. 151, no. 1, pp. 402–413, 2022.

[10] R. Werner, S. Fuchs, J. Trouvain, and B. Möbius, "Inhalations in speech: Acoustic and physiological characteristics," in *Interspeech 2021*. ISCA: ISCA, 2021, pp. 3186–3190. [Online]. Available: https://www.isca-speech.org/archive/interspeech_2021/werner21_interspeech.html

[11] 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.

[12] J. E. Huber, "Effects of utterance length and vocal loudness on speech breathing in older adults," *Respiratory Physiology & Neurobiology*, vol. 164, no. 3, pp. 323–330, 2008.

[13] L. L. Koenig and S. Fuchs, "Vowel formants in normal and loud speech," *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 5, pp. 1278–1295, 2019.

[14] R. Schulman, "Articulatory dynamics of loud and normal speech," *The Journal of the Acoustical Society of America*, vol. 85, no. 1, pp. 295–312, 1989.

[15] J. E. Huber and B. Chandrasekaran, "Effects of increasing sound pressure level on lip and jaw movement parameters and consistency in young adults," *Journal of Speech, Language, and Hearing Research*, vol. 49, no. 6, pp. 1368–1379, 2006.

[16] S. Stone and P. Birkholz, "Silent-speech command word recognition using electro-optical stomatography." in *Interspeech*, 2016, pp. 2350–2351.

[17] P. Boersma, "Praat: doing phonetics by computer [computer program]," *http://www. praat. org/*, 2022.

[18] R. B. Banzett, S. T. Mahan, D. M. Garner, A. Brughera, and S. H. Loring, "A simple and reliable method to calibrate respiratory magnetometers and respitrace," *Journal of Applied Physiology*, vol. 79, no. 6, pp. 2169–2176, 1995.

[19] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[20] 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.

[21] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[22] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [Online]. Available: https://ggplot2.tidyverse.org

[23] M. Filippelli, R. Pellegrino, I. Iandelli, G. Misuri, J. R. Rodarte, R. Duranti, V. Brusasco, and G. Scano, "Respiratory dynamics during laughter," *Journal of Applied Physiology*, vol. 90, no. 4, pp. 1441–1446, 2001.