Advancing Speech Breathing Analysis: Benefits of Using EMA

Tabea Thies, Philipp Buech, Anne Hermes

ISSP 2024 - 13th International Seminar on Speech Production
13-17 May 2024, Autrans, France

Advancing Speech Breathing Analysis: Benefits of Using EMA

Tabea Thies[1], Philipp Buech[2], Anne Hermes[2]

[1]*IfL Phonetics & Department of Neurology, University Hospital Cologne, Germany*
[2]*Laboratoire de Phonétique et Phonologie, CNRS & Sorbonne Nouvelle, Paris, France*
tabea.thies@uni-koeln.de, {philipp.buech; anne.hermes}@sorbonne-nouvelle.fr

Abstract

This study presents an innovative approach to speech breathing analysis, emphasizing the potential of Electromagnetic Articulography (EMA) as a viable tool. We compared the widely used Respiratory Inductive Plethysmography (RIP) with EMA by collecting speech breathing data from 18 speakers during sustained vowel productions of /a/ under habitual and loud speech conditions. Our findings indicate that EMA signals can effectively track temporal patterns of speech breathing movements, which do not differ from the RIP system. With this study, we would like to emphasize the potential of using (existing) EMA systems in laboratories to analyze speech breathing patterns. This paper explores the advantages and opportunities that arise from integrating EMA systems into speech breathing research. The findings suggest that such integration can enhance our understanding of speech production and contribute to advancements in related fields.

Keywords*: speech production, speech breathing, inductive plethysmography, electromagnetic articulography*

1. Introduction

The respiratory inductive plethysmography (RIP) is a popular technique and a validated, common tool for studying speech breathing patterns (Winkworth et al. 1995, Fuchs & Rochet-Capellan 2021, Charuau et al. 2022). Two elastic bands (with insulated wires) are positioned around the chest and the abdomen to track breathing patterns. Although different sizes of bands exist, wearing the bands may affect participants' comfort and awareness of the equipment which could further lead to alterations in breathing behavior. Another limitation is that body movements can generate artifacts in the signal that can affect the accuracy of the data (Fuchs & Rochet-Capellan 2021). Additionally, Fuchs and Rochet-Capellan (2021) pointed out that the development of smaller and/or wireless sensors could improve comfort during breathing recordings, which has been recently developed by Columbi Computers AB (Sweden) for the RespTrack system. To simultaneously capture kinematic speech data, one is currently dependent on using two systems, such as RIP and e.g., an Electromagnetic Articulograph (EMA) as it has been done by e.g., Rasskazova et al. (2019).

Here, we present the use of EMA as a new applied technique for tracking speech breathing patterns, entailing high-resolution contours with better comfort and fewer artifacts. We conducted a study comparing the RIP system (Inductotrace®) and the EMA system (Carstens AG501) to track and analyze speech breathing patterns. The goal was to assess the similarity of the kinematic trajectories for capturing speech breathing patterns recorded by both systems. The data used for comparisons are sustained vowel productions in two different conditions, i.e., in habitual and loud speech.

In a first step, we analyze data from all applied EMA sensors to identify the most suitable sensors for accurately tracking speech breathing. This initial assessment ensures that the selected sensors provide reliable and precise measurements. The second step involves comparing the signals obtained from both the RIP system and the EMA system. By examining the signals from these two systems, we evaluated the consistency and accuracy of the EMA system in capturing speech breathing patterns. Finally, in the third step, we identify similarities in the signals to analyze the robustness of the tracking methods.

By conducting this comprehensive analysis, we aim to highlight the reliability and effectiveness of the EMA system for tracking speech breathing. The findings from this study will contribute to advancing research in speech production and to enhance our understanding of the intricate mechanisms involved in speech breathing.

2. Methods

2.1. Participants

We collected acoustic and kinematic data from 18 native German speaking participants (9 males, 9 females). The age ranged from 23 to 54 years with a mean age of 33 years.

2.2. Experimental Set-up

The kinematic breathing data were collected using the (a) EMA (AG 501) and (b) RIP (Inductotrace®) at the same time with a sampling rate of 1250 Hz. To track breathing data with EMA, sensors were placed at different positions and fixed with tape (**Figure 1**). One sensor on the lowest vertebra of the cervical spine functioned as the reference sensor. Sensors on the sternum and three on the chest were used to track (speech) breathing kinematics. Sensors tracking thorax movements were positioned at the axilla level on the chest (on clothes); one in the middle and two at the height of each papilla. After placing the EMA sensors (**Figure 1** left), the RIP band (only upper band for thorax movement) was put around the participants' chest (**Figure 1** right). Three different band sizes were used (7 x small, 5 x medium, and 6 x regular), thus representing different body sizes.

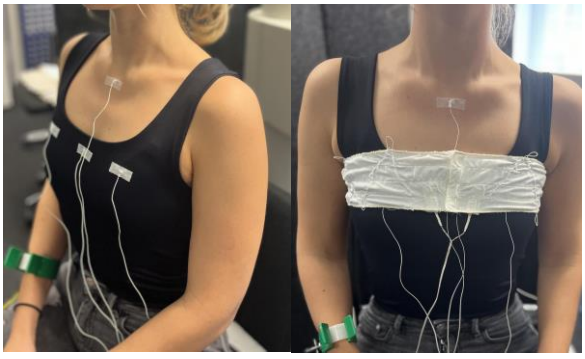

Figure 1: *EMA sensors on subject – (left) before the RIP belt is put on and (right) with the RIP belt put on.*

10.21437/issp.2024-32

2.3. Speech Material

In this paper, only data of sustained productions of the vowel /a/ in habitual and loud speech are presented. The data analyzed here is part of a larger data set. Participants were asked to take a deep breath and to produce maximum phonation of the vowel /a/ in habitual speech and loud speech. Tracking of speech loudness was done via a sound level meter that was set up 1.25m away from the participants. For loud speech, participants were asked to keep a constant level of 80dB. The sustained vowel /a/ phonation was repeated three times per condition.

2.4. Data Processing and Analysis

Since the RIP and EMA recordings started asynchronous, we aligned the audio tracks of the EMA and RIP by an acoustic impulse at the beginning of the recording. The acoustic boundaries of both habitual and loud /a/ were manually segmented using Praat (Boersma & Weenink, 2024). For the EMA system, different distances between sensors were calculated and analyzed in the vertical (low-high, y) and horizontal (front-back, x) dimension (**Figure 2**):

- D1: Distance of the chest's middle sensor to the reference sensor (chest mid to R) → EMA_{D1}
- D2: Distance of the calculated midpoint between left sensor and right sensor on the chest to the reference sensor (midpoint to R) → EMA_{D2}
- D3: Distance of sternum to the reference sensor (sternum to R) → EMA_{D3}

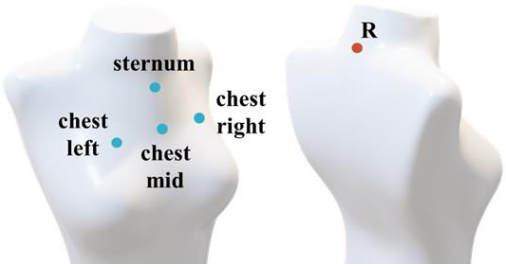

Figure 2: *Schematized EMA sensors on the front and on the back (R = reference sensor).*

For the calculated distances, three landmarks were automatically determined in the RIP and the EMA signal: (i) inhalation onset, (ii) inhalation peak, and (iii) exhalation offset (**Figure 3**). The landmark detection was as follows: The signals were prepared first by resampling them to 100 Hz and applying a Savitzky-Golay filter using a window of 101 samples and polynomial order 3 afterwards. The basis for the landmark detection was then the processed signal within a window of the acoustic boundaries of the target vowels ± 7s.

The signals' velocity was used for the detection of the inhalation onset and the exhalation offset. For the inhalation onset, the maximum velocity left to the inhalation peak was determined first and then the first zero crossing in the velocity was used for the landmark detection of the onset. The detection of the offset was based on the velocity multiplied by a window function consisting of two half Gaussians and a stable region during the acoustic segment. The last zero crossing left to the velocity maximum in the second half of the window was used as the offsets' landmark. The inhalation peak was defined as the maximum in the signal.

Figure 3 displays examples of synchronized RIP and EMA data during the production of sustained /a/ in habitual speech, namely the raw filtered signal, the resampled and filtered signal, the signals' velocity and the windowed velocity, along with the detected landmarks in vertical dashed lines.

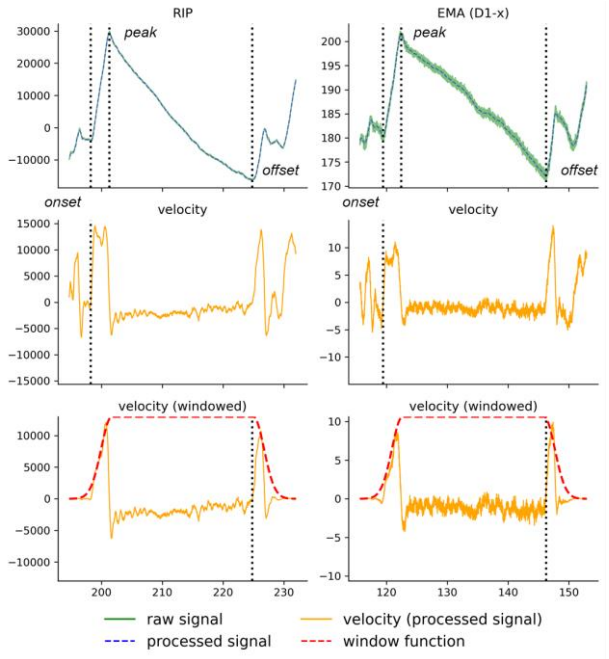

Figure 3: *Example of landmark detection in RIP (left) and EMA signal* $EMA_{D1}x$ *(right). Vertical dotted lines refer to landmarks (onset, peak and offset). Rows show the raw filtered and the processed signal (top), the velocity (mid), and the window function (bottom).*

To compare the RIP and EMA signal and to determine which EMA distance trajectories are most comparable to the RIP system, the procedure was as follows:

First, the following two parameters were calculated to analyze temporal breathing patterns:

1) Inhalation phase (s): Interval between inhalation onset and inhalation peak.
2) Exhalation phase (s): Interval between inhalation peak and inhalation offset.

To compare each of the two parameters, we run hierarchical Bayesian regression models for the two temporal parameters and speaking styles (loud, habitual) with the SIGNAL TYPE (RIP vs. $EMA_{D1}x$, $EMA_{D1}y$, $EMA_{D2}x$, $EMA_{D2}y$, $EMA_{D3}x$, $EMA_{D3}y$) as independent variables with by-speaker intercepts and slopes. We used default priors in all models. Results are reported under section 3.1.

Second, we compared the RIP and EMA trajectories based on 100 equally distanced time points from the inhalation onset to the exhalation offset and standardized the trajectories by token and signal type. For visual inspection, we calculated Euclidean-distance matrices showing the (dis-)similarity between RIP and the EMA dimensions across speakers and repetitions, and speaking styles (section 3.2.).

Third, we run Gaussian Process regression models for each speaking style on a subset of the standardized signal trajectories (steps of 5% from inhalation onset to exhalation offset). We used separate covariances for each SIGNAL TYPE with exponential priors for amplitude (lambda=1) and length scale (lambda=3) and a by-SIGNAL TYPE intercept with a default prior. The models were run with 2000 samples for tuning and

2000 samples in four chains, thus leading to 8000 iterations for the analysis. We computed the difference between the posterior of the RIP and the posterior of each EMA distance afterwards. Results are reported under section 3.3. We report the mean and the 95% highest density interval (HDI) of the posterior estimates for all regression analyses.

3. Results

3.1. Parameter comparisons

Table 1 contains the averaged results for the parameters of interest for the different signals (RIP vs. $EMA_{D1\text{-}D3}$) in both the x- and the y-dimension.

Table 1: *Mean durations of inhalation and exhalation phases in seconds (standard deviations in brackets) for the RIP and EMA distance signals.*

Condition	Signal	Inhalation phase	Exhalation phase
habitual	RIP	2.78 (1.18)	22.62 (8.50)
	$EMA_{D1}x$	2.36 (1.00)	22.49 (8.68)
	$EMA_{D1}y$	2.58 (1.02)	22.63 (8.63)
	$EMA_{D2}x$	2.81 (1.05)	22.49 (8.69)
	$EMA_{D2}y$	2.53 (1.07)	22.68 (8.59)
	$EMA_{D3}x$	2.29 (1.22)	22.10 (8.66)
	$EMA_{D3}y$	2.58 (1.02)	22.53 (8.59)
loud	RIP	2.41 (0.93)	23.46 (10.15)
	$EMA_{D1}x$	2.15 (0.99)	22.61 (10.37)
	$EMA_{D1}y$	2.14 (1.00)	23.23 (10.42)
	$EMA_{D2}x$	2.18 (1.02)	23.07 (10.59)
	$EMA_{D2}y$	2.15 (0.98)	23.36 (10.59)
	$EMA_{D3}x$	2.25 (1.10)	23.21 (10.17)
	$EMA_{D3}y$	2.22 (0.97)	23.16 (10.68)

No durational differences in the exhalation phases of the EMA signal (and its related differences) compared to RIP's in the production of sustained vowel /a/ in habitual and loud speech were found. However, regarding the inhalation phases, the models reveal slightly shorter inhalation phases in $EMA_{D1}x$ (β=-0.96 [-1.6, -0.35]) and $EMA_{D2}x$ (β=-0.45 [-0.82, -0.09]) in habitual and $EMA_{D1}x$ (β=-0.5 [-0.79, -0.18]), $EMA_{D2}x$ (β=-0.32 [-0.59, -0.05]) and $EMA_{D3}y$ (β=-0.35 [-0.67, -0.3]) in loud speech compared to the RIP signal.

3.2. Distance plots for visual inspection

Figure 4 and **Figure 5** display distance plots comparing RIP and EMA signals averaged across all speakers during sustained vowel productions in habitual speech (**Figure 4**) and loud speech (**Figure 5**). For the signal comparison in habitual and loud speech, the EMA_{D2y} signal was chosen as an example, as this EMA distance signal is most similar to the phases of the RIP signal - particularly in habitual speech (**Table 1**). The color coding indicates the continuum from similar (black; 0 of the normalized Euclidean distance) to dissimilar (white, 1 of the normalized Euclidean distance). The diagonal of each matrix represents the comparison of the trajectories at the corresponding time points. In both conditions (habitual and loud), a black diagonal beam can be observed indicating a clear similarity between the trajectories of RIP and EMA.

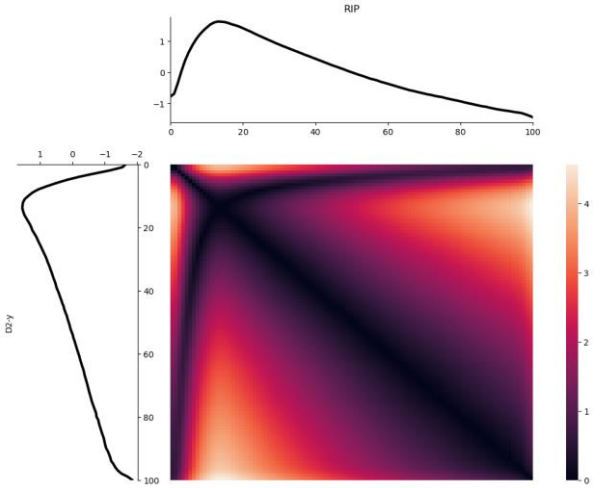

Figure 4: *Distance plot ($EMA_{D2}y$) comparing RIP and EMA signals in habitual speech.*

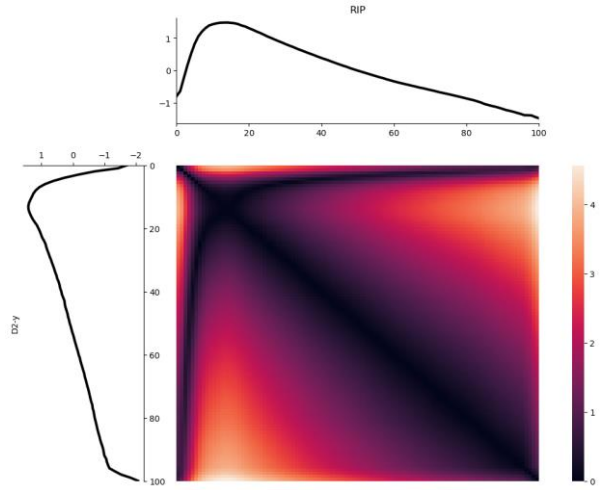

Figure 5: *Distance plot ($EMA_{D2}y$) comparing RIP and EMA signals in loud speech.*

3.3. Trajectory comparisons: Regression analysis

To investigate which distance signal is most suitable to track speech breathing patterns with EMA, we compare the contours of the RIP signal with all EMA distance signals by means of Gaussian Process regression models. **Figure 6** shows the output of the models for habitual (left column) and loud (right column) speech. Each panel shows the comparison of the RIP signal with the respective EMA signal. The top of each panel depicts the 95% posterior estimate for the RIP (blue, hatched) and the EMA signal (red), and the plot below shows the difference (orange) between the RIP signal and the EMA signal.

Our regression analyses revealed that none of the EMA distance signals significantly differs from the RIP signal in shape across the speech breathing movements. As can be seen in **Figure 6**, the 95% HDI of the posterior differences between the RIP and EMA contours is centered around zero, thus indicating no difference at each of the evaluated time points. If a significant deviation between the signals was detected, this would be marked by a red area (which is not the case here).

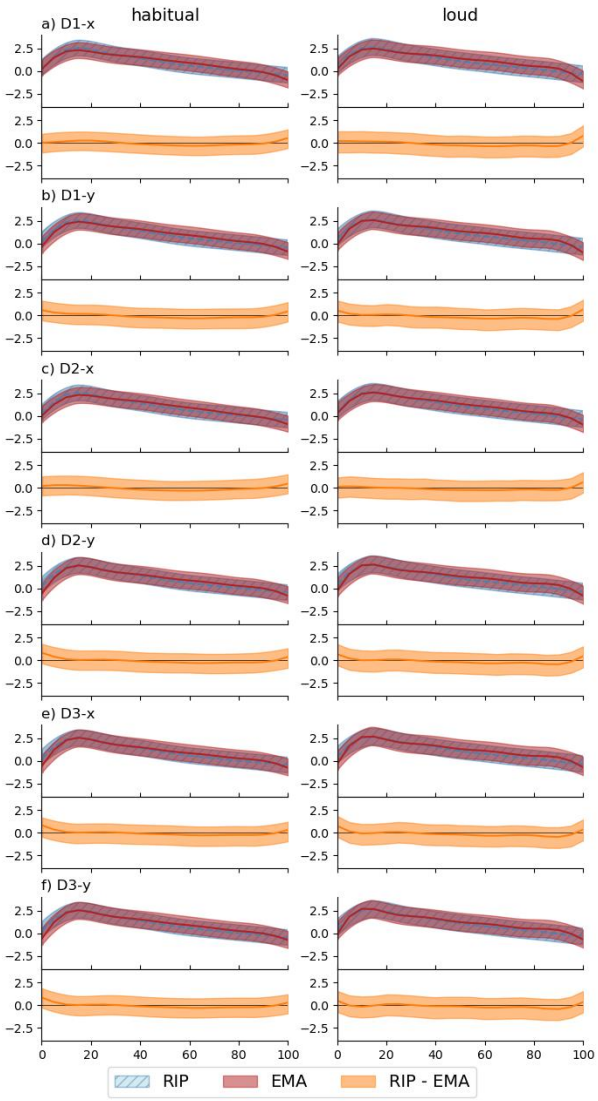

Figure 6: *Regression results for RIP compared to various EMA signals (rows) for habitual (left) and loud speech (right). The top of each panel shows 95% of the posterior estimates for RIP and the EMA signal, and the lower plot shows the difference between RIP and the EMA signal.*

4. Discussion

This study reveals that EMA sensors are capable of tracking speech breathing patterns that are comparable with the commonly used RIP signal. We were able to show that temporal parameters, such as inhalation and exhalation phases do not differ between the EMA and RIP signal. However, slightly longer durations were detected for some parameters. This could be explained by the fact that the expansion of the RIP band is measured in a three-dimensional space, whereas the EMA signal only measured one-dimensional distances. As EMA also allows for the analysis of 3D movement patterns, possible parameters need to be developed to capture 3D patterns in the future. Nonetheless, since the movement trajectories did not differ between RIP and EMA, we postulate that EMA is a potential method to collect speech breathing data.

As we attached EMA sensors to various positions on the chest, we were able to show that in principle, the signal from all sensors can be used. A subsequent analysis will determine which sensors are most suitable to give a recommendation on the minimum number of EMA sensors that should be used in future studies. In general, when doing EMA recordings, sensors for tracking speech breathing are easily addable to the sensor set-up when tracking articulation, making EMA a promising tool for research in speech breathing production studies. As breathing is the basic requirement for speech production and as it has a linguistic and communicative role (Fuchs & Rochet-Capellan 2021), the relevance of examining speech breathing patterns, breath cycle coordination and the interaction between breathing with other speech systems is given (Werner 2023).

Due to the significant cost difference between an EMA system and an RIP, laboratories that already possess an EMA device can derive practical advantages from utilizing EMA instead of the traditionally employed RIP. The experimental process becomes simplified since there is no longer a requirement for diverse belts (as for Inductotrace®), resulting in enhanced convenience and reduced intrusiveness.

We will pursue the analyses of speaker-specific behaviors and look more into natural speech production, such as sentence productions and text reading.

5. Conclusion

Previous research has demonstrated that the respiratory inductive plethysmography (RIP) is a widely accepted and validated tool for studying speech breathing patterns. However, it also has its limitations, such as potential discomfort for participants and the possibility of body movements generating artifacts in the signal. This study is the first comparing speech breathing patterns assessed with Electromagnetic Articulography (EMA) to RIP signals. Results underscore the benefits and ease of using EMA for analyzing speech breathing pattern and paves the way for further studies which are using EMA systems to also easily collect data on speech breathing simultaneously to speech production kinematics.

6. Acknowledgements

This work has been carried out within the framework of the ANR-23-CE28-0017 INSPECT supported by the French ANR. Further, this work is partially supported by a public ANR grant as part of the program “Investissements d’Avenir” (ANR-10-LABX-0083). It contributes to the IdEx Université de Paris - ANR-18-IDEX-0001.

7. References

Charuau, D., Vaxelaire, B., & Sock, R. (2022). L’organisation spatio-temporelle de la respiration chez l’enfant. In SHS Web of Conferences (Vol. 138, p. 08005). EDP Sciences.

Fuchs, S. & Rochet-Capellan, A. (2021). The Respiratory Foundations of Spoken Language. *Annual Review of Linguistics*, 7(1), 13-30.

Rasskazova, O., Mooshammer, C. & Fuchs, S. (2019). Temporal coordination of articulatory and respiratory events prior to speech initiation. *Proceedings of 20th Interspeech 2019*, 7(1), 884-888.

Werner, Raphael Johannes (2023). *The phonetics of speech breathing: pauses, physiology, acoustics, and perception*. Doctoral dissertation. doi: 10.22028/D291-41147

Winkworth, Alison L.; Davis, Pamela J.; Adams, Roger D.; Ellis, Elizabeth (1995). Breathing Patterns During Spontaneous Speech. *Journal of Speech Language and Hearing Research*, 38(1), 124-144. doi:10.1044/jshr.3801.12