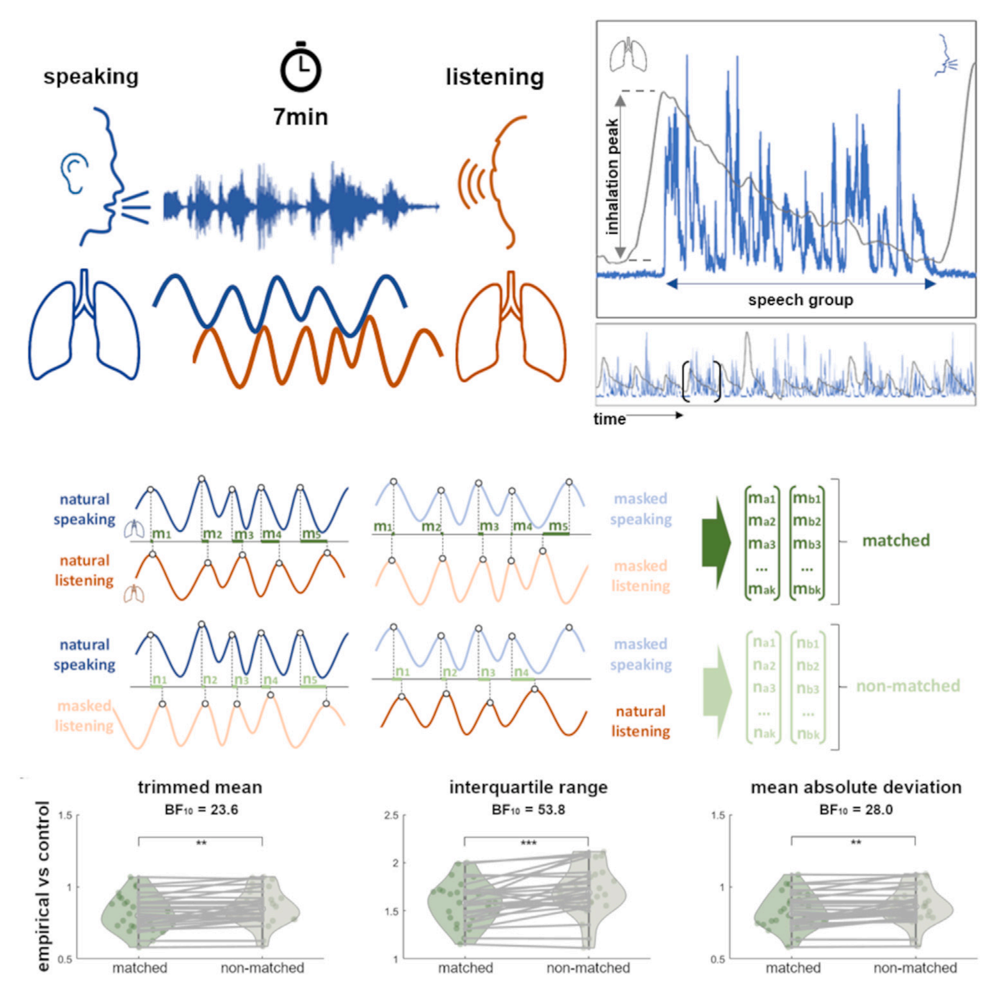
iScience

CellPress
OPEN ACCESS

Article

Predictive coordination of breathing during intra-personal speaking and listening



Omid Abbasi, Daniel S. Kluger, Nikos Chalas, Nadine Steingräber, Lars Meyer, Joachim Gross

daniel.kluger@wwu.de

Highlights

Human breathing is altered during listening and speaking compared to rest

In speaking, inhalation correlates with speech envelope in the following utterance

We find similar timing of inhalations during speaking and listening to one's own speech

Findings support hypothesized alignment of internal forward models of interlocutors

Abbasi et al., iScience *26*, 107281
August 18, 2023 © 2023 The Authors.
https://doi.org/10.1016/j.isci.2023.107281

iScience

CellPress
OPEN ACCESS

Article

Predictive coordination of breathing during intra-personal speaking and listening

Omid Abbasi,[1,4] Daniel S. Kluger,[1,2,4,5,*] Nikos Chalas,[1,2] Nadine Steingräber,[1] Lars Meyer,[3] and Joachim Gross[1,2]

SUMMARY

It has long been known that human breathing is altered during listening and speaking compared to rest: during speaking, inhalation depth is adjusted to the air volume required for the upcoming utterance. During listening, inhalation is temporally aligned to inhalation of the speaker. While evidence for the former is relatively strong, it is virtually absent for the latter. We address both phenomena using recordings of speech envelope and respiration in 30 participants during 14 min of speaking and listening to one's own speech. First, we show that inhalation depth is positively correlated with the total power of the speech envelope in the following utterance. Second, we provide evidence that inhalation during listening to one's own speech is significantly more likely at time points of inhalation during speaking. These findings are compatible with models that postulate alignment of internal forward models of interlocutors with the aim to facilitate communication.

INTRODUCTION

Human speech production fundamentally relies on respiration which—in turn—is critical for preserving homeostasis. It is therefore noteworthy that during speaking the stereotypical rhythmic breathing process changes: Breathing during speaking is more variable in peak inhalation amplitude and breathing rate; it is also characterized by a more asymmetric pattern of short inhalations and long exhalations.[1] However, there is a practical upper limit to the duration of a respiration cycle during speaking.[2,3] Supplying the body with sufficient oxygen as well as the capacity of the lung to supply sufficient air pressure for the articulators to resonate and produce speech sounds determine this limit individually.

Within this limit, breathing in conversation can be affected by cognitive factors of the speech production process. Here, we focus on two putative mechanisms for controlling breathing that both rely on predictions—one in speaking and another in listening. Both these mechanisms were put forward in a recent active inference model[4] which describes human communication as two dynamic systems that are coupled via sensory information and aim to minimize prediction errors: a generative (forward) model that is likely supported by cerebello-thalamo-cortical connections computes predictions in the speaker and listener. In the speaker, the forward model represents predicted sensory consequences of their own speech and allows the speaker to adjust parameters like speech volume, speed, or articulation based on the proprioceptive and auditory feedback. Pertaining to speech breathing, there is evidence that the forward model also informs respiration based on upcoming utterances. This idea is supported by the fact that the higher variability of speech breathing compared to restful breathing is due to the fact that inhalation during speaking does not occur at regular intervals (as in restful breathing) but rather adapts to linguistic components in speech and is strongest at the beginning of a new utterance.[5] This suggests a high level of fine control of speech breathing that requires speech planning to be tightly coordinated with breathing. Specifically, it has been proposed that during speaking, efficient speech breathing would adapt depth of inhalation at the beginning of a breath group (i.e., the words produced after a single inhalation) to provide sufficient air for the specific subsequent vocalization.[6]

In the listener's brain, the forward model generates predictions about the timing and content of upcoming speech. These predictions are constantly compared to incoming sensory information and updated accordingly.[7] Two interrelated models (detailed below) suggest that interpersonal alignment between speaker and listener may facilitate predictions and, in turn, comprehension. If this interpersonal alignment extends

[1] Institute for Biomagnetism and Biosignal Analysis, University of Münster, Münster, Germany

[2] Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany

[3] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[4] These authors contributed equally

[5] Lead contact

*Correspondence: daniel.kluger@wwu.de

https://doi.org/10.1016/j.isci.2023.107281

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

to breathing then we would expect that a listener could partly adapt their respiration to the respiratory dynamics of the speaker. The two models highlighted below provide further insights about the possible underlying mechanisms and their functional consequences.

First, the ''interactive alignment account'' posits that in a conversation, speech production and comprehension are facilitated by alignment between interlocutors at various levels. While the original account has demonstrated alignment at every linguistic level,[8] this concept also extends to temporal features in conversation such as speech rate, inter-speaker pause duration, and turn duration.[9] Furthermore, during listening, brain areas associated with speech production are activated and likely improve comprehension.[10–12] Indeed, activity in listeners' motor areas is at least partly temporally aligned with activity in the speaker's motor system.[13–15] This involvement of the cortical motor system in the alignment between speaker and listener could very well extend to some aspects of respiration (as a motor act) as well. As for other types of alignment, respiratory alignment could facilitate comprehension.

Second, ''active sensing'' refers to the idea that sensory signals are not just received passively but rather actively sampled in a way that is modulated by the statistics of the received signals and internally generated predictions about the to-be-perceived signals and their current relevance.[16–18] Pertaining to this study it is interesting to note that this concept has been recently extended to encompass respiration: animal studies have shown that respiration modulates spike rates in a variety of brain regions[19,20] suggesting that dynamic brain states of cortical excitability fluctuate with the breathing rhythm. Recent evidence from non-invasive magnetoencephalography (MEG) work indicates that this coupling of respiratory and neural rhythms may apply to human brain function as well.[21]

Such a respiratory alignment might be achieved by combining different sources of information. First, speech breathing can be perceived by the listener, and the duration and intensity of speech breathing might indicate the duration of the subsequent exhalation (see earlier text). Also, listeners can predict the length of a spoken sentence based on prosodic (and possibly breathing) cues.[22,23] Furthermore, it was recently shown that listeners utilize the perception of speech breathing to form temporal predictions about upcoming speech.[24] Second, the listener's internal forward models afford predictions about the end of a turn of the speaker.[25] Since inhalation happens frequently at the beginning of a sentence, these predictions might be used for respiratory alignment.

In what follows, we present primary evidence for intra-individual, predictive coordination of breathing during speaking and listening to one's own speech. This way, we aimed to address a central gap in the existing literature: while previous work has demonstrated that the amplitude of peak inhalation is related to the duration of the subsequent utterance during reading[26] and spontaneous speech,[27,28] these studies were solely focused on the amount of air required for a particular breath group. Here, we jointly investigated breathing and speech envelope as a more direct measure for investigating potentially predictive respiration-speech coupling.

RESULTS

Our analysis is based on latencies of peak inhalation which we extracted from 7-min respiration time series acquired for N = 27 participants in four conditions: 1) masked speech production (participants produced speech in the presence of white noise presented through ear tubes such that they could not hear their own voice); 2) normal speech production (without noise); 3) listening to their own masked speech; 4) listening to their own normal speech.

First, we assessed the breathing cycle duration for each condition (Figure 1A). We computed a linear mixed-effect model (LMEM) for our 2 x 2 design with the factors *condition* (speaking, listening) and *masking* (yes, no) for the mean cycle duration because these data were normally distributed (Lillie test). Mean duration of respiration cycle duration was significantly shorter during listening than during speaking ($t(104) = 5.3$, $p \ll 0.001$). In other words, breathing rate was significantly faster in listening than that in speaking. Neither the main effect of masking nor the interaction of masking X condition was significant ($t(104) = 1.8$ and $t(104) = 1.2$; both $p > 0.05$). In addition, changes of the variability (measured as standard deviation) of cycle duration were tested across conditions. Since these data were not normally distributed, we used the Kruskall-Wallis test. We observed a significant effect of condition ($\chi^2(3) = 58$, $p \ll 0.0001$). Pairwise testing of the four conditions revealed that standard deviation of cycle duration was significantly higher during speaking compared to listening (all $p \ll 0.001$, Tukey-Kramer correction for multiple comparisons).

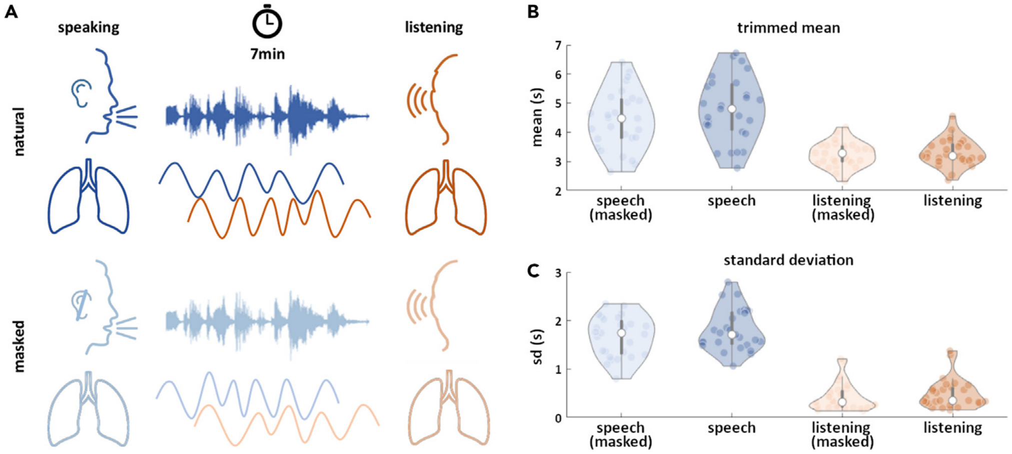


Figure 1. Data acquisition and respiratory cycle durations during speaking and listening

(A) We recorded 7 min of respiratory data while participants were either speaking (top left) or listening to their own speech (top right). This procedure was repeated with white noise masking applied via earphones so that participants were speaking without hearing themselves speak (*masked speech*, bottom left) and later listened to their masked speech (*masked listening*, bottom right).

(B) As expected, the LMEM model revealed significantly shorter durations of respiratory cycles for listening vs. speaking (all $p < 0.001$, trimmed mean with 10% exclusion).

(C) Complementing the shorter cycle durations during listening, the variability of cycle durations during listening was significantly reduced compared to speaking ($p = 0.026$).

As expected, masked speaking was not different from normal speaking ($p = 0.25$). This pattern of results was to be expected; during speaking, respiration is constrained by the linguistic structure of the produced speech leading to longer and more variable intervals between peak inhalation (Figures 1B and 1C).

We expected that the constraining effect of to-be-produced speech would partly determine the peak inhalation amplitude: Producing a longer or louder speech segment during one exhalation requires more air and should therefore lead to a higher peak inhalation amplitude. We tested this by employing a second LMEM to investigate the relationship between inhalation peak amplitude and the speech envelope summed over the subsequent respiration cycle (see STAR Methods section for details). The LMEM confirmed that higher peak inhalation amplitude was significantly associated with higher total speech envelope amplitude across the subsequent respiration cycle ($t(26) = 10.23$, $p \ll 0.001$; see Figure 2).

Next, we addressed the main question of the study and tested if respiration during listening to one's own speech partly follows the respiration timings during speaking. Due to the significantly different breathing cycle durations between the speaking and listening conditions, we cannot expect a strong synchronization of breathing time courses where each inhalation in listening is temporally aligned to a corresponding inhalation in speaking. However, relevant events such as inhalation during listening could still have a higher probability to occur close to peak inhalation during speaking. Therefore, we identified for each inhalation peak in the speaking condition the temporally closest peak (before or after) in the corresponding listening condition and extracted the temporal peak-to-peak distance (see Figure 3A).

Importantly, the delay between inhalation peaks during speaking and listening (which can be positive or negative) was not significantly different from 0 ms ($t(26) = 0.07$, $p = 0.94$), indicating that inhalation during listening is centered around inhalation latencies during speaking.

Next, we tested our main hypothesis that the temporal distance (i.e., the absolute delay) between inhalation peaks in speaking and listening is smaller than what can be expected by chance. This would indicate that, when listening to their own speech, participants are more likely to inhale at time points when they also inhaled during speaking.

This was tested in two ways. First, we constructed a new distribution of temporal peak-to-peak distances using *non-matched* stimuli (Figure 3A). Specifically, we computed the temporal distance between

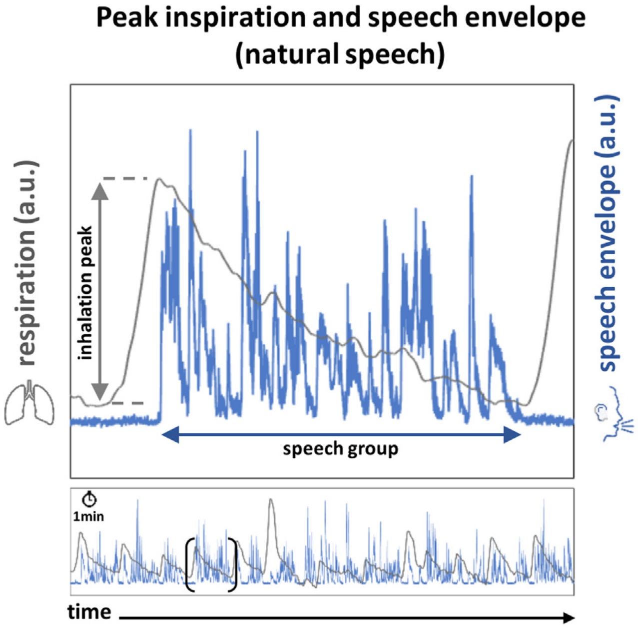


Figure 2. Relationship between peak inhalation and subsequent speech envelope

Exemplary respiration time course (top panel, gray line) shows the typical fast inhalation followed by slow exhalation during speaking. The blue line represents the corresponding speech envelope of this breath group. Data taken from a single 1-min trial (see bottom panel) of a single participant.

inhalation peaks during speaking and listening of the opposite conditions (i.e., natural speaking-masked listening and masked speaking-natural listening). These distances were then compared to those within the *matched* stimuli (i.e., natural speaking-natural listening and masked speaking-masked listening).

Second, we constructed an artificial sequence of inhalation time points with the same distribution of respiration cycle durations as that of the individual listening condition (see STAR Methods section and Figure 4 for design of these surrogate data). It is important to recall that the mean duration of the breathing cycle (i.e., inverse breathing rates) had a strong effect on the distribution of temporal distances of inhalation peaks. Therefore, both control distributions were specifically designed to preserve the mean cycle durations (see STAR Methods section for more details).

The results indicate that inhalation peaks during listening to one's own speech were significantly closer to inhalation peaks during speaking that can be expected by chance (control 2: surrogate data; Figure 4) or when inhalation peaks were taken from a different (non-matched) speech condition (control 1; Figure 3).

This is true for robust estimates of the mean absolute temporal distance but also for the interquartile range as a robust measure of the spread. The effect size of "inhalation alignment" is small but highly significant, indicating that there was a significant bias (i.e., higher probability) for inhalation during listening to occur at the time of inhalation during speaking.

Finally, we investigated if, beyond the temporal alignment of respiration, the depth of inhalation was also related in speaking and listening. We tested this with an LMEM relating peak inhalation amplitude in listening to peak inhalation amplitude in speaking. The model yielded a significant relationship between depth of inhalation in speaking and listening ($t(4373) = 2.69$, $p = 0.007$). Taken together, our results indicate that listeners mimic not only the timing but also the depth of inhalation of their previously produced speech.

DISCUSSION

Our results indicate predictive coordination of respiration during speaking and listening to one's own speech. During speaking the peak inhalation amplitude was related to the total speech envelope summed

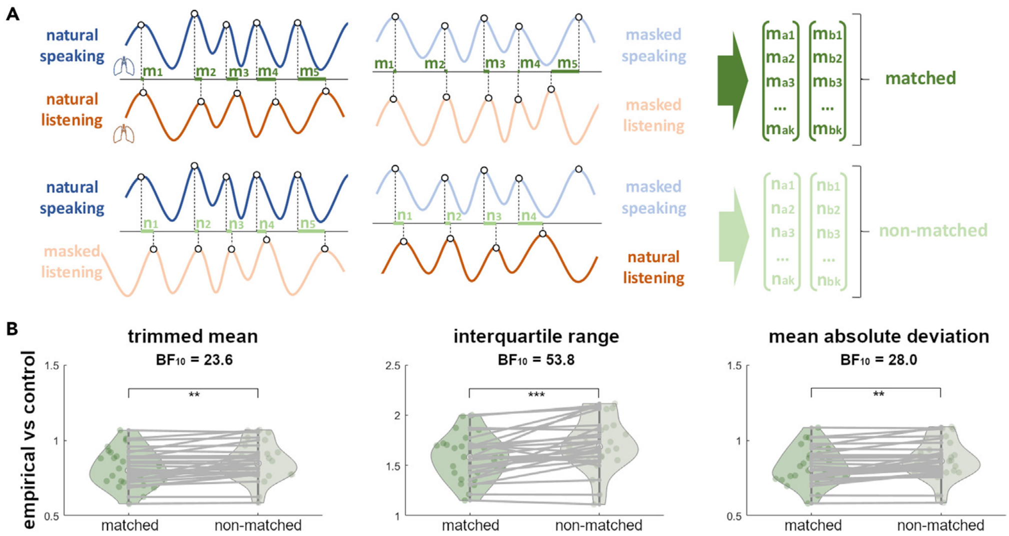


Figure 3. Contingencies between respiratory time courses during speaking and listening

(A) To quantify the contingencies between breathing patterns during speaking and listening to the same speech, we computed the temporal distances between inhalation peaks in both domains: For each inhalation peak during speaking (identified by the peak detection algorithm), we computed the temporal distance to the nearest inhalation peak (before or after) in the corresponding listening condition. For the *matched* condition, we pooled first-level differences computed for [natural speaking, natural listening] and [masked speaking, masked listening] (top). For the *non-matched* condition, we computed first-level differences to the counterpart of each domain, i.e., [natural speaking, masked listening] and [masked speaking, natural listening] (bottom).

(B) Paired t tests demonstrated a significantly closer correspondence (i.e., shorter distances) between speaking and listening for the matched (vs. non-matched) condition (left panel, trimmed mean with 10% exclusion). The effect size was Cohen's *d* = 0.44. The consistency of this decrease was corroborated by significantly lowered, variance measures like the interquartile range (middle) and mean absolute deviation (right) for matched vs. non-matched speaking and listening. BF = Bayes factor. ** represents p < 0.005; *** represents p < 0.001.

across the breath group. The positive coefficient of the linear mixed-effects model indicates that larger peak inhalation amplitude is associated with higher summed speech envelope. This means that in preparation of vocalization of a breath group, speakers adapt inhalation to the air volume required for the respective breath group. Previous studies have demonstrated that peak inhalation amplitude is correlated with the duration of the subsequent utterance during reading,[26] single sentence utterance,[29] and spontaneous speech.[27,28] However, the evidence is not unambiguous since in a recent study inhalation amplitude did not differ significantly between very short utterances and longer speech.[6] We improve on previous studies by using the speech envelope instead of breath group duration only. By relating inhalation amplitude to the summed speech envelope over the subsequent breath group (instead of its duration), we are getting closer to the hypothesized mechanism underlying predictive coordination of respiration during speaking because the summed speech envelope is a better measure of speech output compared to the duration of the breath group. The required air volume for a breath group correlates with its duration but also depends on speech loudness[30] and more generally the speech-specific sound pressure that is adequately quantified with the speech envelope. Therefore, our results based on the speech envelope and peak inhalation amplitude further support the notion that inhalation is finely controlled based on the upcoming breath group.

While this mechanism is strikingly intuitive for energy-efficient speech production, it requires very sophisticated computations. Specifically, while a speaker initiates inhalation, planning of the content of the upcoming breath group needs to be largely completed. In addition, the speaker needs a model that provides a mapping of this content to the estimated required air volume which in turn depends on loudness, the current physiological state (which is different, e.g., for resting, walking, running), and other factors. And, as alluded to above, this joint speech and breath planning needs to be conducted within the constraints of the individual lungs' vital capacity (i.e., volume of air available for vocalization).

The detected positive correlation between peak inhalation and subsequent speech envelope can alternatively be interpreted as speakers adjusting their speech in real time depending on the amount of air

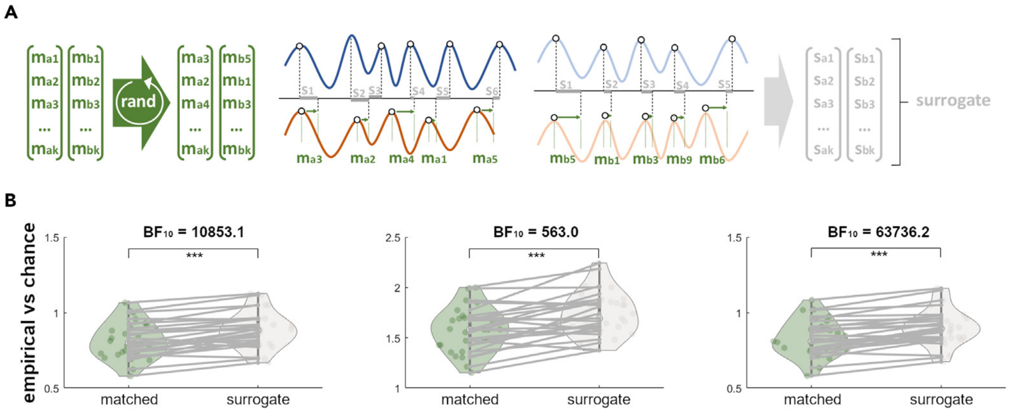


Figure 4. Contingency statistics against surrogate data

(A) Individual surrogate distributions of temporal distances were constructed from the empirical distributions described in Figure 2: For each participant, the vectors of peak-to-peak distances computed between [natural speaking, natural listening] and [masked speaking, masked listening] were shuffled separately. The elements of the resulting shuffled vectors were consecutively used to shift the empirical inhalation peaks during listening by a random (yet physiologically plausible) distance. In line with the procedure described above, we then identified the closest inhalation peak during speaking for each of the shifted time points, thus constructing vectors of surrogate peak-to-peak distances.

(B) Compared to these surrogate distributions, the empirical peak-to-peak distances were found to be significantly shorter on average (left). The effect size was Cohen's $d = 0.33$. The consistency of this effect was indicated by a significant lowering of interquartile range (middle) and mean absolute deviation (right) for the empirical (vs. surrogate) distances. BF = Bayes factor. *** = $p < 0.001$.

available to them. As an example, a speaker may shorten their sentence or use simpler words when they are not able to produce a longer sentence due to lack of air. In the same way, a speaker with more air may choose to use more complex language or longer sentences. Compared to the first interpretation which explained the role of internal forward models in speech-respiration relationships, this type of real-time adjustment could occur on a much shorter look-ahead time. In contrast to predicting the volume of air that will be needed, speakers may continuously monitor their air volume and adjust their speech accordingly.

The second aspect of predictive coordination of respiration studied here pertains to listening—in our case, to one's own speech. Our results indicate that the initiation of inhalation during listening is more likely at time points that correspond to inhalations during speaking. This is very different from a pure 1:1 phase synchronization of respiration between speaker and listener. Such strict phase synchronization is not possible in the case of speaker-listener respiratory alignment given the very different cycle durations between speaking and listening (see Figure 1). As a consequence there is no one-to-one mapping of inhalations between speaking and listening. Instead, our results are consistent with the idea that listeners have a preference to inhale at time points close to the inhalation of speakers. However, we like to note several caveats. First, while this partial temporal alignment is highly significant (i.e., very consistent across the group of participants), the actual effect (difference of temporal distances between real data and surrogate data [see Figures 3B and 4B]) in each individual is rather small. Second, in our study participants were listening to their own speech and they might have anticipated some parts during the listening condition. To prevent recollection of content during listening, we designed our experiment such that anticipation of upcoming speech was minimized by several means: participants were measured in separate sessions for speech production and perception tasks with several days' intervals between performing these two conditions. Further, our questions were mainly about a common/general topic. Consequently, participants were highly unlikely to remember their answers in detail. Still, it remains to be seen if our results generalize to unknown speech from a different speaker.

As outlined in the introduction, respiratory coordination between listener and speaker would be consistent with several models. All these models are to some degree based on the notion of a coupling of internal forward models of speaker and listener via sensory signals produced by the speaker's motor system (such as sound of speech or respiration or visual cues of respiration). Therefore, in the listener parts of the motor system are aligned to the speaker possibly leading to enhanced comprehension through coordination of internal excitability states and simulation of the speaker's internal model (see also Barsalou[31]).

There is convincing evidence that simultaneously perceived sensory signals lead to interpersonal synchrony. Recently, Madsen and Parra performed a comprehensive study showing that watching the same movie induces intersubject correlation in participants of electroencephalogram (EEG) signals, gaze position, pupil size, and heart rate, but not respiration and head movements.[32] In other studies interpersonal physiological synchrony has been observed for electrodermal activity and heart rate,[33] eye movement,[34] and movement and respiration.[35,36] Pertaining to respiration there is also plenty of evidence that it is adjusted to motor activity within an individual.[37–39] Evidence for auditory-motor alignment within individuals in the context of continuous speech is however sparse and has received relatively little attention. Garssen reported that the number of respiration cycles where inhalation is aligned between speaker and listener is higher than that expected by chance but only when watching a video of an actor where respiration is clearly visible and audible.[40] However, the authors employed a rather lenient criterion of respiratory coordination by comparing the data against the mean of three instantiations of surrogate data instead of a comparison to the 95th percentile of a large number of surrogates. More recently this question was revisited by Rochet-Capellan and Fuchs.[32] They asked participants to listen to read speech and studied to what extent a listener aligns inhalation onset to that of the reader. Alignment was observed when listening to the female reader but not the male reader, and authors concluded that findings ‘‘did not support stable or continuous temporal alignment of listener breathing to reader breathing.’’

Limitations of the study

The absence of a continuous alignment is consistent with our results and—given the different cycle durations (or breathing rates) between speaking and listening—can be expected. However, using a different methodology and two control conditions, we find significant respiratory alignment in the sense described at the beginning of this section. Our results therefore indicate that inhalation in listeners is modulated by attended speech not only in general aspects such as breathing cycle duration and amplitude but also in the timing, leading to a preferred inhalation of the listener near time points of inhalation in speakers. Clearly, not every inhalation in the speaker is matched with an inhalation in the listener. It therefore remains an intriguing question for further studies if the probability of speaker-listener alignment for each inhalation can be predicted—e.g., from factors such as momentary attention, emotional engagement, predictability of speakers' breathing pattern, or acoustic or linguistic aspects of listened speech.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Data acquisition
  - Paradigm
  - Preprocessing and data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

ACKNOWLEDGMENTS

We acknowledge support from the Interdisciplinary Center for Clinical Research (IZKF) of the Medical Faculty Münster (grant number Gro3/001/19). OA (EFRE-0400394) was supported by the EFRE. DSK (KL 3580/1-1) and JG (GR 2024/5-1; GR 2024/11-1; GR 2024/12 -1) were further supported by the DFG.

AUTHOR CONTRIBUTIONS

Conceptualization, OA, JG; Methodology, OA, LM, JG; Investigation, OA, NC; NS; Writing – Original Draft, OA, DSK, JG; Writing – Review & Editing, OA, DSK, NC, NS, LM, JG; Visualization, DSK; Funding Acquisition, OA, DSK, JG.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 9, 2023
Revised: May 4, 2023
Accepted: June 30, 2023
Published: July 4, 2023

REFERENCES

1. Fuchs, S., and Rochet-Capellan, A. (2021). The respiratory foundations of spoken language. Annu. Rev. Linguist. *7*, 13–30. https://doi.org/10.1146/annurev-linguistics-031720-103907.
2. Grosjean, F., Grosjean, L., and Lane, H. (1979). The patterns of silence: Performance structures in sentence production. Cognit. Psychol. *11*, 58–81. https://doi.org/10.1016/0010-0285(79)90004-5.
3. Pierrehumbert, J. (1979). The perception of fundamental frequency declination. J. Acoust. Soc. Am. *66*, 363–369.
4. Friston, K.J., and Frith, C.D. (2015). Active inference, communication and hermeneutics. Cortex *68*, 129–143.
5. Wang, Y.-T., Green, J.R., Nip, I.S.B., Kent, R.D., and Kent, J.F. (2010). Breath group analysis for reading and spontaneous speech in healthy adults. Folia Phoniatr. Logop. *62*, 297–302. https://doi.org/10.1159/000316976.
6. Włodarczak, M., and Heldner, M. (2017). Respiratory Constraints in Verbal and Non-verbal Communication. Front. Psychol. *8*, 708. https://doi.org/10.3389/fpsyg.2017.00708.
7. Arnal, L.H., and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. Trends Cogn. Sci. *16*, 390–398. https://doi.org/10.1016/j.tics.2012.05.003.
8. Pickering, M.J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. Behav. Brain Sci. *27*, 169–190. https://doi.org/10.1017/S0140525X04000056.
9. Ostrand, R., and Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. J. Phonetics *88*, 101074. https://doi.org/10.1016/j.wocn.2021.101074.
10. Pickering, M.J., and Garrod, S. (2013). An integrated theory of language production and comprehension. Behav. Brain Sci. *36*, 329–347. https://doi.org/10.1017/S0140525X12001495.
11. Möttönen, R., Dutton, R., and Watkins, K.E. (2013). Auditory-motor processing of speech sounds. Cerebr. Cortex *23*, 1190–1197. https://doi.org/10.1093/cercor/bhs110.
12. Watkins, K.E., Strafella, A.P., and Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia *41*, 989–994. https://doi.org/10.1016/s0028-3932(02)00316-0.
13. Park, H., Thut, G., and Gross, J. (2018). Predictive entrainment of natural speech through two fronto- motor top-down channels. Lang. Cogn. Neurosci. *35*, 739–751. https://doi.org/10.1080/23273798.2018.1506589.
14. Park, H., Ince, R.A.A., Schyns, P.G., Thut, G., and Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr. Biol. *25*, 1649–1653. https://doi.org/10.1016/j.cub.2015.04.049.
15. Keitel, A., Gross, J., and Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol. *16*, e2004473. https://doi.org/10.1371/journal.pbio.2004473.
16. Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of Active Sensing and perceptual selection. Curr. Opin. Neurobiol. *20*, 172–176. https://doi.org/10.1016/j.conb.2010.02.010.
17. De Kock, R., Gladhill, K.A., Ali, M.N., Joiner, W.M., and Wiener, M. (2021). How movements shape the perception of time. Trends Cogn. Sci. *25*, 950–963. https://doi.org/10.1016/j.tics.2021.08.002.
18. Yang, S.C.-H., Wolpert, D.M., and Lengyel, M. (2018). Theoretical perspectives on active sensing. Curr. Opin. Behav. Sci. *11*, 100–108. https://doi.org/10.1016/j.cobeha.2016.06.009.
19. Ito, J., Roy, S., Liu, Y., Cao, Y., Fletcher, M., Lu, L., Boughter, J.D., Grün, S., and Heck, D.H. (2014). Whisker barrel cortex delta oscillations and gamma power in the awake mouse are linked to respiration. Nat. Commun. *5*, 3572. https://doi.org/10.1038/ncomms4572.
20. Yanovsky, Y., Ciatipis, M., Draguhn, A., Tort, A.B.L., and Brankačk, J. (2014). Slow oscillations in the mouse hippocampus entrained by nasal respiration. J. Neurosci. *34*, 5949–5964. https://doi.org/10.1523/JNEUROSCI.5287-13.2014.
21. Kluger, D.S., Balestrieri, E., Busch, N.A., and Gross, J. (2021). Respiration aligns perception with neural excitability. Elife *10*, e70907. https://doi.org/10.7554/eLife.70907.
22. Grosjean, F. (1983). How long is the sentence? prediction and prosody in the on-line processing of language. Linguistics *21*, 501–529.
23. Lamekina, Y., and Meyer, L. (2022). Entrainment to speech prosody influences subsequent sentence comprehension. Lang. Cogn. Neurosci. *38*, 263–276. https://doi.org/10.1080/23273798.2022.2107689.
24. MacIntyre, A.D., and Scott, S.K. (2022). Listeners are sensitive to the speech breathing time series: Evidence from a gap detection task. Cognition *225*, 105171. https://doi.org/10.1016/j.cognition.2022.105171.
25. Levinson, S.C., and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. Front. Psychol. *6*, 731. https://doi.org/10.3389/fpsyg.2015.00731.
26. Fuchs, S., Petrone, C., Krivokapić, J., and Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. J. Phonetics *41*, 29–47. https://doi.org/10.1016/j.wocn.2012.08.007.
27. Rochet-Capellan, A., and Fuchs, S. (2013). The interplay of linguistic structure and breathing in German spontaneous speech. In Interspeech 2013 (ISCA), pp. 2014–2018. https://doi.org/10.21437/Interspeech.2013-478.
28. Winkworth, A.L., Davis, P.J., Adams, R.D., and Ellis, E. (1995). Breathing patterns during spontaneous speech. J. Speech Hear. Res. *38*, 124–144. https://doi.org/10.1044/jshr.3801.124.
29. McFarland, D.H., and Smith, A. (1992). Effects of vocal task and respiratory phase on prephonatory chest wall movements. J. Speech Hear. Res. *35*, 971–982. https://doi.org/10.1044/jshr.3505.971.
30. Huber, J.E. (2008). Effects of utterance length and vocal loudness on speech breathing in older adults. Respir. Physiol. Neurobiol. *164*, 323–330. https://doi.org/10.1016/j.resp.2008.08.007.
31. Barsalou, L.W. (2008). Grounded cognition. Annu. Rev. Psychol. *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.
32. Rochet-Capellan, A., and Fuchs, S. (2013). Changes in breathing while listening to read speech: the effect of reader and speech mode. Front. Psychol. *4*, 906. https://doi.org/10.3389/fpsyg.2013.00906.
33. Stuldreher, I.V., Thammasan, N., van Erp, J.B.F., and Brouwer, A.-M. (2020). Physiological synchrony in EEG, electrodermal activity and heart rate reflects shared selective auditory attention. J. Neural.

Eng. *17*, 046028. https://doi.org/10.1088/1741-2552/aba87d.

34. Madsen, J., Júlio, S.U., Gucik, P.J., Steinberg, R., and Parra, L.C. (2021). Synchronized eye movements predict test scores in online video education. Proc. Natl. Acad. Sci. USA *118*, e2016980118. https://doi.org/10.1073/pnas.2016980118.

35. Codrons, E., Bernardi, N.F., Vandoni, M., and Bernardi, L. (2014). Spontaneous group synchronization of movements and respiratory rhythms. PLoS One *9*, e107538. https://doi.org/10.1371/journal.pone.0107538.

36. Paccalin, C., and Jeannerod, M. (2000). Changes in breathing during observation of effortful actions. Brain Res. *862*, 194–200. https://doi.org/10.1016/s0006-8993(00)02145-4.

37. Bartlett, D., and Leiter, J.C. (2012). Coordination of breathing with nonrespiratory activities. Compr. Physiol. *2*, 1387–1415. https://doi.org/10.1002/cphy.c110004.

38. Rassler, B., and Raabe, J. (2003). Co-ordination of breathing with rhythmic head and eye movements and with passive turnings of the body. Eur. J. Appl. Physiol. *90*, 125–130. https://doi.org/10.1007/s00421-003-0876-5.

39. Ebert, D., Hefter, H., Binkofski, F., and Freund, H.-J. (2002). Coordination between breathing and mental grouping of pianistic finger movements. Percept. Mot. Skills *95*, 339–353. https://doi.org/10.2466/pms.2002.95.2.339.

40. Garssen, B. (1979). Synchronization of respiration. Biol. Psychol. *8*, 311–315. https://doi.org/10.1016/0301-0511(79)90013-9.

41. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869. https://doi.org/10.1155/2011/156869.

42. Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., et al. (2013). Good practice for conducting and reporting MEG research. Neuroimage *65*, 349–363. https://doi.org/10.1016/j.neuroimage.2012.10.001.

43. Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. PLoS Comput. Biol. *5*, e1000436. https://doi.org/10.1371/journal.pcbi.1000436.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Preprocessed MEG + respiration data	This study	osf.io/pxguq
Software and algorithms		
Matlab code for present analyses	This study	osf.io/pxguq
FieldTrip	Oostenveld et al., 2011[41]	fieldtriptoolbox.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Daniel Kluger (daniel.kluger@wwu.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Preprocessed data have been deposited at the Open Science Framework and are publicly available as of the date of publication. The access link is listed in the key resources table.
- All original code has been deposited at the Open Science Framework and is publicly available as of the date of publication. The access link is listed in the key resources table.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We recruited thirty native German-speaking participants (15 female, mean age 25.1 ± 2.8 years [M ± SD], range 20–32 years). The study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki. Prior written informed consent was obtained before the measurement and participants received monetary compensation after their participation.

METHOD DETAILS

Data acquisition

Respiratory and speech signals were recorded simultaneously. The speech recording had a sampling rate of 44.1 kHz. Audio data was captured with a microphone which was placed at a distance of 155 cm from the participant's mouth in order not to cause any artefacts by the microphone itself. The respiratory signal was measured as thoracic circumference by means of a respiration belt transducer (BIOPAC Systems, Goleta, USA) placed around the participant's chest. Individual respiration time courses were visually inspected for irregular breathing patterns such as breath holds or unusual breathing frequencies, but no such artefacts were detected.

Paradigm

Participants were asked to sit relaxed while performing the given tasks and to keep their eyes focused on a white fixation cross. This study consisted of four separate recordings: i) speech production, ii) speech production while perception of their own speech was masked, iii) speech perception while listening to their own normal speech generated in the first recording, and iv) speech perception while listening to their own masked speech generated in the second recording. For the speech production recording, there were seven 60-second trials for overt speech. During each trial, participants answered a given question such as 'What does a typical weekend look like for you?'. A colour change of the fixation cross from white to blue indicated the beginning of the time period in which participants should speak and the end was marked by a colour change back to white. In the second recording, participants were asked to perform the same task as in the first recording while they heard white noise, leaving them largely unable to hear their own voice. The questions were different to the prior recording in order to prevent repetition of prefabricated answers. Questions covering neutral topics were chosen to avoid emotional confounds.

In the third and fourth recording sessions participants listened to audio-recordings of their own voice which were collected in the first and second recordings respectively.

Preprocessing and data analysis

In the preprocessing and data analysis steps, custom-made scripts in Matlab R2020 (The Mathworks, Natick, MA, USA) in combination with the Matlab-based FieldTrip toolbox[41] were used in accord with current guidelines.[42] Three participants where respiration recording failed were excluded from analysis.

The wideband amplitude envelope of the speech signal was computed using the method presented in.[43] Nine logarithmically spaced frequency bands between 100-10000 Hz were constructed by bandpass filtering (third-order Butterworth filters). Then, we computed the amplitude envelope for each frequency band as the absolute value of the Hilbert transform and downsampled them to 1200 Hz. Finally, we averaged them across bands and used this computed wideband amplitude envelope for all further analysis. In all respiration signals obtained during speaking and listening we identified the time points corresponding to peak inhalation. To this end, we employed the findpeaks.m function in matlab on the z-scored time series after smoothing (Savitzky-Golay filter of order 3 and frame length 1591). Three participants were excluded because peak inhalation could not be reliably identified in their data. Results were validated by visual inspection. The temporal distance between peak inhalation times yielded the respiration cycle durations (and their variability) presented in Figure 1.

QUANTIFICATION AND STATISTICAL ANALYSIS

These data were subjected to a linear mixed effects model (LMEM) using the equation in Wilkinson notation:

$$\text{data} \sim \text{condition} + \text{masking} + \text{condition} * \text{masking} + (1|\text{subject}) \qquad \text{(Equation 1)}$$

The model fit was obtained with the fitlme function in Matlab R2022a (Mathworks).

Time points of peak inhalation during speaking were used to study the relationship between peak inhalation amplitude and the summed speech envelope of the subsequent breath group (i.e. the speech envelope in the time window until next inhalation). The LMEM used the equation in Wilkinson notation:

$$\text{summed speech envelope} \sim \text{peak inhalation amplitude} + (1|\text{subject}) \qquad \text{(Equation 2)}$$

The main analysis was based on the temporal distance of inhalation peaks between listening and speaking. Recall that participants were listening to the same speech that they had themselves produced in an earlier recording session. We aimed to test if inhalation during listening was more likely to occur at time points where inhalation occurred in the speaking condition. Therefore, we identified for each inhalation peak during speaking the temporally closest inhalation peak in the listening condition. To improve statistical sensitivity we pooled the normal speech and masked speech condition leading to 14 min of data (7 min speaking, 7 min masked speaking). Finally, we statistically compared the distribution of temporal distances to two other control distributions. The first control distribution was constructed by identifying temporal distances between non-matching stimuli. While the original distribution was constructed from matched stimuli ([natural speaking, natural listening] and [masked speaking, masked listening]), the first control distribution was constructed from the non-matched stimuli pairings ([natural speaking, masked listening] and [masked speaking, natural listening]). This control distribution therefore represents the distribution of delays that can be expected by chance. The second control distribution was constructed artificially: For each individual participant, we computed a new vector of peak inhalation times by picking a random start time for the first inhalation and then successively adding to this time point randomly picked respiration cycle durations from the individual real respiration cycle durations (see Figure 4A). Therefore, this surrogate list had the same distribution of respiration cycle durations as the original individual listening condition - but not in the right order. This procedure would destroy any temporal alignment of inhalation between speaking and listening while preserving the overall statistics of the respiration cycle duration.

A test on the relationship between peak inhalation amplitude in speaking and listening was conducted with an LMEM using the equation:

$$\text{peak inhalation amplitude (listening)} \sim \text{peak inhalation amplitude (speaking)} + (1|\text{subject}) \quad \text{(Equation 3)}$$

Bayes factors were computed with the bayesFactor toolbox (github.com/klabhub/bayesFactor).