



Listeners are sensitive to the speech breathing time series: Evidence from a gap detection task

Alexis Deighton MacIntyre^{*,1}, Sophie K. Scott

University College London, United Kingdom

ARTICLE INFO

Keywords:

Breathing
Speech
Rhythm
Entrainment
Gap detection

ABSTRACT

The effect of non-speech sounds, such as breathing noise, on the perception of speech timing is currently unclear. In this paper we report the results of three studies investigating participants' ability to detect a silent gap located adjacent to breath sounds during naturalistic speech. Experiment 1 ($n = 24$, in-person) asked whether participants could either detect or locate a silent gap that was added adjacent to breath sounds during speech. In Experiment 2 ($n = 182$; online), we investigated whether different placements within an utterance were more likely to elicit successful detection of gaps. In Experiment 3 ($n = 102$; online), we manipulated the breath sounds themselves to examine the effect of breath-specific characteristics on gap identification. Across the study, we document consistent effects of gap duration, as well as gap placement. Moreover, in Experiment 2, whether a gap was positioned before or after an interjected breath significantly predicted accuracy as well as the duration threshold at which gaps were detected, suggesting that nonverbal aspects of audible speech production specifically shape listeners' temporal expectations. We also describe the influences of the breath sounds themselves, as well as the surrounding speech context, that can disrupt objective gap detection performance. We conclude by contextualising our findings within the literature, arguing that the verbal acoustic signal is not "speech itself" per se, but rather one part of an integrated percept that includes speech-related respiration, which could be more fully explored in speech perception studies.

1. Introduction

Breathing provides both flow and form to speech. The development of the intercostal muscular control required to sustain vocal air pressure was a turning point in human evolution (MacLarnon & Hewitt, 1999), and David Abercrombie, an early proponent of one influential theory of speech rhythm, considered breathing as its basis, stating that speech rhythm is "essentially a muscular rhythm, and the muscles concerned are the breathing muscles" (Abercrombie, 1967).² Despite the vital contribution of breathing to vocalisation, its role in determining the temporal-dynamic structure of speech remains scientifically enigmatic, in part because the percept of breathing is rarely investigated experimentally. Indeed, the speech research community has been subject to recent criticism concerning its perceived over-reliance on unnatural, laboratory-produced stimuli, but even these commentaries neglect to engage with breathing as a necessary and ubiquitous, yet overlooked

component to speech production and perception in the wild (see, for example, Alexandrou, Saarinen, Kujala, & Salmelin, 2020; Hamilton & Huth, 2020; Hitczenko, Mazuka, Elsner, & Feldman, 2020). That speech breathing has not received widespread scientific attention may be attributable to its very banality: we tend not to consciously notice respiration—our own, or another person's—at all, unless some irregularity or breakdown has occurred. The addition of breath sounds (but not acoustically similar, non-breath sounds) to synthesized speech, however, improves listeners' recall (Elmers, Werner, Muhlack, Möbius, & Trouvain, 2021; Whalen, Hoequist, & Sheffert, 1995), and a speaking virtual character is perceived as more human and trustworthy when it audibly breathes (Bernardet, Kang, Feng, DiPaola, & Shapiro, 2019). Multi-sentence synthesized speech (e.g., audiobook narration) with breath sounds modelled on human respiratory patterns is strongly preferred for naturalism over the equivalent with silent pauses (Braunschweiler & Chen, 2013). Although little is known concerning

* Corresponding author at: 17 Queen Square, London WC1N 3AR, United Kingdom.

E-mail address: alexisdeighton.macintyre@mrc-cbu.cam.ac.uk (A.D. MacIntyre).

¹ Now at MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer, Rd, Cambridge CB2 7EF, United Kingdom.

² Unlike Abercrombie (1967), who presumed the existence of isochronous units in speech, we define rhythm simply as the non-arbitrary configuration of events in time, which need not be characterised by equal inter-event intervals.

<https://doi.org/10.1016/j.cognition.2022.105171>

Received 24 July 2021; Received in revised form 29 April 2022; Accepted 11 May 2022

Available online 19 May 2022

0010-0277/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

how speech breathing is planned, analysis so far suggests that breaths are predominantly taken at syntactic boundaries (Fuchs, Petrone, Krivokapić, & Hoole, 2013; Grosjean & Collins, 1979; Henderson, Goldman-Eisler, & Skarbek, 1965; Winkworth, Davis, Ellis, & Adams, 1994), and that longer phrases are usually preceded by longer or larger inhalations (Rochet-Capellan & Fuchs, 2013; Whalen & Kinsella-Shaw, 1997; Winkworth et al., 1994). There is also a strong social-pragmatic function for speech breathing, with dyadic and group experimental data revealing systematic relationships between respiratory and turn-taking behaviours in conversation (Aare, Gilmartin, Włodarczak, Lipus, & Heldner, 2020; Rochet-Capellan & Fuchs, 2014; Torreira, Bögels, & Levinson, 2015; Włodarczak & Heldner, 2016a, 2016b). Taken together, these preliminary data suggest that we expect to hear breath sounds in accompaniment of speech; that the duration of vocal inhalation may offer some cues as to the upcoming content being communicated; and that seamless interaction is possibly facilitated by interlocutors' awareness of each other's respiratory movements. What is unclear, however, is how breathing is perceptually integrated on the sub-second timescale as a component of speech prosody or rhythm. Do listeners, consciously or non-consciously, impose temporal expectations upon speech-related inhalation, for example, by expecting speech to occur at a certain point after hearing a breath sound? If so, are they sensitive to perturbations of the natural speech breathing time series?

Towards an initial understanding of listeners' sensitivity to the timing of speech breath sounds, we devised two modified gap detection tasks. In its most basic form, a gap detection task consists of the serial presentation of auditory stimuli, such as tones or noise, in which a very brief silent (i.e., acoustically empty) gap may or may not occur. The participant's job is to report whether or not they noticed a gap, and the task can incorporate adaptive procedures to determine an individual's unique perceptual threshold in terms of the Gap Duration of the gap, or some other property of the stimulus. The results are taken as an indication of auditory temporal resolution (Phillips, 1999), a dimension that predicts future reading problems in preschoolers (Boets et al., 2011), and shows decline in the early stages of dementia (Jalaei et al., 2019). This form of perceptual acuity is considered essential to speech processing, which by its nature relies on the ability to track a rapidly changing signal in potentially challenging listening environments, an effort that demands efficiency. This need for efficiency may explain why sensitivity to gaps does not appear to be uniformly distributed in time: Henry and Obleser (2012) employed a gap detection task in conjunction with electroencephalography (EEG) to show that attending power—that is, the likelihood of successful gap detection—is concentrated at particular instants in the phase-relationship between a rhythmic stimulus (a frequency modulating tone) and neural oscillations in the delta band (3 Hz). Given the purported social-linguistic salience of breath sounds, it is possible that the interlocked rhythms of speech and speech breathing could also influence listeners' discrimination of gap targets; for instance, listeners may be more likely to detect a gap that follows, rather than anticipates, a breath sound associated with speech.

In the current paper, we ask how listeners temporally assimilate a speaker's breaths, and the extent to which such temporal assimilation is flexible, by manipulating the respiratory time series in the context of varied and naturalistic speech stimuli. Across three experiments, we alter the placement and duration of an artificially added silent gap adjacent to breath sounds and speech in order to assess the approximate threshold at which a gap can be reliably detected or located by listeners. Using two tasks in the first experiment, we asked participants to 1) report whether or not they perceived a gap that could occur between a breath sound and speech; and 2) identify after which of two breaths a gap occurred within a single utterance. In the second experiment, we followed up on the first task, but added multiple positions in which the gap could occur relative to two breath sounds. In the third and final experiment, we extended the results of the second task in Experiment 1 by manipulating the breath sounds themselves, in order to interrogate breath-specific influences when determining where the gap occurred.

These paradigms allow us to interrogate listeners' expectations of the timing of speech breathing from slightly different angles, and at differing levels of naturalism. Before describing the experiments in more detail, we make some broad predictions concerning the perceptual thresholds at which listeners may detect the manipulation of the natural speech breathing time series and outline their justifications in the following section.

1.1.1. Predictions

1.1.1.1. Gap duration. Speech breathing is physiologically (Mckay, Evans, Frackowiak, & Corfield, 2003; Von Euler, 1982), behaviourally (Winkworth et al., 1994), and acoustically (Trouvain, Möbius, & Werner, 2019) distinct from metabolic breathing, and it follows that a listener should plausibly expect to hear speech following a speech-associated inhalation sound. What is less obvious is how flexibly listeners treat the variable pause between the cessation of air intake, and the ensuing speech onset. Although this question appears to be scientifically unexplored thus far, previous physiological analyses of speakers' respiratory activity provide some a priori rationale to expect that listeners will be sensitive to temporal manipulations in this domain. Specifically, it is reported that phonation usually begins between 100 and 300 ms after the post-inspiration period, although nontrivial interspeaker differences are also recorded (Atkinson, 1973; Lieberman, 1967; Slifka, 2003). Other researchers have focused on the acoustic presentation of breathing, observing that audible breath sounds tend to be flanked by silent "edges", which tend to last for tens of seconds (Fukuda, Ichikawa, & Nishimura, 2018; Trouvain et al., 2019). In view of the recognisable acoustic profile of speech breathing, its necessary association with imminent speech, and the normative transition period between the end of a breath and vocalisation, we speculate that perturbations as short as an extra 200 ms between breath and speech may be sufficient for detection; however, it is likely that performance will increase with increasing gap length, as would be expected in the traditional gap detection task.

1.1.1.2. Gap position. We also expect that the verbal context surrounding speech breathing should play a role in how respiratory sounds are perceived, as rhythmic priming has been shown to modulate temporal expectations (McAuley & Jones, 2003; Pashler, 2001; Yee, Holleran, & Jones, 1994). Moreover, recent data indicate that adjusting the pace of an ongoing speech stream leads to changes in subsequent phonemic awareness, potentially demonstrating a causal influence on fine-grain temporal processing in the speech domain (Kösem et al., 2018). One mechanism proposed to underlie these effects is a process known as *entrainment*. Broadly speaking, entrainment is the idea that endogenous neural and/or behavioural processes can become temporally coupled with other time series in the environment (Jones & Boltz, 1989; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Large & Jones, 1999), thereby supporting the optimal distribution of attending and action-planning resources across time. Entrainment is associated with enhanced behavioural responses, such as faster reaction times (e.g., Kunert and Jongman (2017)) or lower perceptual thresholds (e.g., Chang, Bosnyak, and Trainor (2019)). In the case of speech, entrainment may allow listeners to not only track, but anticipate upcoming speech, thereby supporting the real time processing of a complex and highly changeable signal (Ding & Simon, 2014; Peelle & Davis, 2012; Zoefel, 2018). The specifics of speech entrainment are currently the focus of an active and contentious literature (Ding & Simon, 2014; Doelling & Assaneo, 2021). For our purposes, we assume that with increased exposure to speech, participants will form stricter temporal expectations, and therefore a silent gap inserted close to the interjected, rather

than initial, breath should be more likely to register as a violation of natural speech breathing timing. This idea is congruent with evidence that, at least in nonverbal neural entrainment, the stimulus-brain phase relationship consolidates over several seconds of rhythmic stimulation (Bauer, Bleichner, Jaeger, Thorne, & Debener, 2018). It is also possible that the acoustic profile of interjected speech breaths maybe make them more amenable to temporal discrimination: by comparison, initial breaths tend to be longer lasting and larger in volume (Winkworth et al., 1994; Winkworth, Davis, Adams, & Ellis, 1995), and are hypothetically freer to vary than interjected breaths, which presumably conform to the surrounding speech rhythm. In sum, we predict that participants should be more likely to correctly identify silent gaps associated with interjected than initial breaths.

1.1.1.3. Rhythm sensitivity. In addition to the speech breathing gap detection tasks, we also administer a nonverbal rhythm discrimination task. Musicians achieve comparatively lower perceptual thresholds in traditional auditory gap detection tasks (Donai & Jennings, 2016; Mishra, Panda, & Herbert, 2014), and demonstrate enhanced abilities in auditory processing more generally (see Strait and Kraus (2014); Vasuki, Sharma, Demuth, and Arciuli (2016) for review), including speech-in-noise perception (see Coffey, Mogilever, and Zatorre (2017) for review). It is therefore likely that musical aptitude, especially in the domain of rhythm, confers an advantage in the present experiment; however, rather than rely on self-reported musician status, we opt to measure sensitivity to rhythm directly, as rhythmic aptitude is known to vary across musicians by genre and training (Bailey & Penhune, 2010; Matthews, Thibodeau, Gunther, & Penhune, 2016). The rhythm task employed consists of same-different judgements following the presentation of rhythmic drum loops. We predict that performance in this rhythm discrimination task will positively correlate with accuracy in gap detection.

To summarise, we expect that listeners will be more likely to detect silent gaps adjacent to breath sounds that occur within, rather than before, the speech stream. We also expect that increasing the gap duration will facilitate gap detection, and that participants with an enhanced sense of timing in the nonverbal domain will also perform relatively better in making judgements about the timing of speech breath sounds. Finally, in Experiment 2, we contrast the detection of gaps added either before or after interjected breaths. If gap detection is superior for gaps occurring after breath sounds, this may suggest that breath sounds are informative concerning the timing of upcoming speech. In practice, temporally asymmetric sensitivity to breath sounds could serve as a sort of phase reset, thereby enhancing early entrainment to the speech stream. We discuss these and other findings in view of embodied speech perception and the recent uptake of interest in rhythmic speech entrainment. We close in making our case for speech breathing as a consequential, if hitherto latent, factor in auditory speech processing more generally.

2. Experiment 1

2.1. Methods and materials

2.1.1. Participants

Participants were locally recruited from the University College London subject pool and were paid for taking part in the experiment, which was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki. All participants had normal or corrected to normal hearing, and gave informed consent to participate. 25 participants took part in this experiment, of which one participant had to be excluded due to a data saving problem, resulting in a total $N = 24$ (Male 14, Female 10; ages 18–35).

2.1.2. Design

2.1.2.1. Gap detection tasks. Experiment 1 consisted of two gap detection tasks, the first of which (Experiment 1A) simply asked participants to report whether or not they detected a silent gap that may have been inserted between a breath sound and speech. Half of trials contained a gap, and the other half did not. Each utterance repeated once across the pseudo-random trial order, such that participants heard both the gap and no-gap versions of each unique utterance. The gaps ranged in duration from [1,001,600 ms], the exact values of which were determined by generating a random gamma distribution with 95% CI [325,595 ms] using the `gamrnd` function in MATLAB. Each duration value occurred once per speaker. The gamma distribution was chosen to give participants more chances to respond to trials with shorter, and presumably more difficult, gaps.

In the second task (Experiment 1B), participants heard an utterance containing two breath sounds, an initial and an interjected breath, one of which would be followed by a silent gap before speech started or resumed. The participant's job was to answer after which breath they thought the gap had occurred. The gap followed the initial breath (Breath 1) in half of trials, and the interjected breath (Breath 2) in the other half. The utterances did not repeat. The gap duration consisted of discrete levels, resulting in a 2 (Gap Position: Breath 1, Breath 2) \times 3 (Gap Duration: 200, 400, 800 ms) within-subjects factorial design. The counts of gap duration values followed a ratio of 3:2:1, with more gaps having the shortest duration, and fewer gaps having the longest duration. The trial structures are depicted in Fig. 1.

2.1.2.2. Nonverbal rhythm discrimination task. To gauge participants' nonverbal musical rhythm processing skills, we administered a forced-choice discrimination task. Each trial consisted of the presentation of a short drum loop, followed by a pause, before a second drum loop played. Participants' job was to answer whether the second drum loop was the same or different from the first. The drum rhythm loop stimuli consisted of 3.2 s rhythmic sequences introduced by Tierney and Kraus (2015), adapted from Povel and Essens (1985), which were made up of nine conga drum sounds separated by the following inter-onset intervals: 5×200 ms; 2×400 ms; 1×600 ms; and 1×800 ms, the re-organization of which generate distinct rhythmic patterns. There were 40 trials, which were equally distributed between same/different correct responses. Each participant's percentage correct (Rhythm Task) was used as a covariate when modelling gap detection performance.

2.1.3. Stimuli

2.1.3.1. Gap detection task. The corpus from which we produced all speech stimuli used in the current study was balanced across four speakers (2 male, 2 female, ages 18–35), comprising of approximately 50% reading and 50% spontaneous speaking styles. Texts consisted of: adaptations of popular science articles chosen for an accessible reading level and neutral tone; four poems characterised as typical of traditional English rhyming verse; and a variety of prompts, characterised as small talk-style questions, to produce the spontaneous speech. The full-length recordings, consisting of approximately one hour of speech per speaker, were combed through for fluent utterances (i.e., without hesitations or silent pauses) where there was a discrete but minimal transition between the breath sound and the speech. In Experiment 1A, the gap detection task stimuli consisted of 108 unique utterances containing one breath followed by a complete phrase or sentence. Each speech stimulus was repeated (i.e., once with and without a gap), resulting in 216 trials in total. The mean duration of breath sounds was 578 ms (SD 260 ms) and the mean duration of the speech was 4.46 s (SD 1.58 s). In Experiment 1B, the stimuli consisted of 150 unique utterances that naturally consisted of a breath-speech-breath-speech structure. The speech stimuli did not repeat within the task. The mean duration of the first breath sound

A. Experiment 1A



B. Experiment 1B



Fig. 1. Graphic representations of the stimulus format and prompt to participants in Experiments 1A (Panel A) and 1B (Panel B).

was 578 ms (SD 280 ms), and the second breath sound was 418 ms (SD 167 ms). Mean Speech 1 duration was 3.00 s (SD 1.17 s), and mean Speech 2 was 2.49 s (SD 0.96 s). In both Experiments 1A and 1B, the stimuli were roughly balanced across all four speakers (2 male, 2 female) and speaking styles.

The speakers were recruited from the broader UCL research community and were unaware of the specific aims of the experiment. They were given no instructions nor direction regarding their breathing during speaking. Source recordings were made in a sound-attenuated studio using SM58 cardioid dynamic microphones (Shure Inc., Niles, IL), positioned via a mic stand in front of the speaker's mouth and sampled at 44,100 Hz. The speakers wore transducer plethysmography belts (MLT1132, ADInstruments, Castle Hill, Australia) to monitor breathing and aid in breath event identification. The acoustic stimuli were segmented and preprocessed in Audacity (Audacity Development Team, 2020), and the beginning and ending of breaths were labelled following visual inspection of the spectrogram with corroboration from the respiratory belt data. Speech and individual breath sounds were root mean square normalised for intensity, which was verified by the authors and adjusted where needed to ensure an approximate perceptual balance within and across excerpts. Onset and offset ramping (10 ms) was applied to the boundaries of speech and breath sounds, and low-intensity background pink noise was added throughout each trial, including during the "silent" gaps, to mask any residual noise or artifacts caused by splicing the audio.

2.1.3.2. Nonverbal rhythm discrimination task. The conga sounds used to generate the rhythmic drum loop stimuli are freely available from MusicRadar (Music radar drum samples, 2022).

2.1.4. Apparatus

The complete experiment was run in PsychoPy (Peirce, 2007) using a laptop computer connected to studio headphones. Participants' responses were recorded via the built-in laptop keyboard.

2.1.5. Procedure

The components of Experiment 1 were performed in the following order: 1. Administration of information sheet and informed consent; 2. Experiment 1A: "Was there a gap?"; 3. Experiment 1B: "Where was the gap?"; 4. Nonverbal rhythm discrimination task; Verbal debrief on the aims and background of the study. The tasks all contained introductory practice sessions, each consisting of three example trials, with otherwise unused stimuli. In the practice sessions, automatic accuracy feedback was given and participants could repeat the practice sessions as many times as they liked. Both of the gap detection tasks and the rhythm discrimination task incorporated a visual warning stimulus (a headphones icon) to alert participants that an upcoming trial was about to begin in 500 ms. There was a variable delay of 500–1000 ms between each trial presentation. All testing took place in a private, quiet space at the Institute of Cognitive Neuroscience. An experimenter was on hand to answer any questions and discreetly ensure task compliance. Throughout the tasks, participants were periodically given the option to take a self-paced rest.

2.2. Data processing and analysis

All analyses in this study were performed using R 3.5.0 (R Core Team, 2013). To explore the factors affecting correct/incorrect responses in the two gap detection tasks, we ran logistic regressions with a binomial distribution and logit link function, using a generalised linear mixed effect model with the glmer function of the lme4 package with bobyqa optimisation (Bates et al., 2012). As a secondary analysis, reaction time data were also analysed using a linear mixed effect model with the lmer function, estimated with restricted maximum likelihood and optimx optimisation (Bates et al., 2012). In Experiment 1A ("Was there a gap?"), the explanatory variables were Gap Duration (continuous) and Rhythm Task (continuous) with a random intercept for Participant and Speech Excerpt, and a random slope for Gap Duration within Participant. Breath Duration and Speech Duration were also fit as continuous terms to control for natural variability in the speech excerpts. Only responses to trials truly containing gaps were analysed in the modelling, but signal detection analyses were also conducted to

examine sensitivity. Hit rate was defined as the proportion of trials on which participants correctly detected the gap, and the False Alarm rate was defined as the proportion of trials in which participants incorrectly responded that a gap was present, with the loglinear correction applied where rates were 1 or 0 (Stanislaw & Todorov, 1999). Z-values were computed for Hit and False Alarm rates and d-prime was calculated. In Experiment 1B (“Where was the gap?”), explanatory variables were Gap Position (Breath 1/Breath 2), Gap Duration (200/400/800 ms), and Rhythm Task (continuous) with random intercepts for Participant and Speech Excerpt, as well as random slopes for Gap Duration and Gap Position within Participant. We also fit Breath 1 Duration, Breath 2 Duration, Speech 1 Duration, and Speech 2 Duration as covariates. Model selection was performed using a step-wise additive approach, beginning first with random intercepts, before including fixed main effects and interactions, then random slopes, with side-by-side model comparisons reported. Model residuals were visually inspected for normality. The significance of fixed effects was tested using Wald chi-square tests (Anova from the car package (Fox et al., 2012)). Statistical variation explained by fixed effects was estimated as semi-partial R^2 (R_{sp}^2) calculated in the r2glmm package (Jaeger, 2017). Post hoc contrasts were conducted using estimated marginal means using the package emmeans (Lenth, 2018). Bonferroni correction was employed for multiple comparisons and adjusted P -values are reported where applicable. Unless otherwise noted, all descriptive statistics (e.g., mean, standard deviation) are calculated within Participant before aggregation at the group level.

2.3. Results

2.3.1. Experiment 1A: “was there a gap?”

2.3.1.1. Accuracy. Descriptive statistics for both tasks are provided in Table 1. <1% (count 22) of Experiment 1A reaction times exceeded 5 s, and these observations were removed from further analysis. We found that the length of gap was strongly positively correlated with successful

Table 1
Summary statistics for Experiment 1.

Descriptive Statistics	
Count	24
Experiment 1A: “Was there a gap?”	
Mean	68%
Median	67%
SD	10%
Min	52%
Max	87%
Experiment 1B: “Where was the gap?”	
Mean	68%
Median	60%
SD	17%
Min	48%
Max	96%
Nonverbal Rhythm Discrimination Task	
Mean	78%
Median	83%
SD	14%
Min	55%
Max	100%
Spearman Rank Correlations	
Experiment 1A × Nonverbal Rhythm Discrimination Task	
$r_s = 0.71$	$p = 0.003$
Experiment 1B × Nonverbal Rhythm Discrimination Task	
$r_s = 0.59$	$p = 0.003$
Experiment 1A × Experiment 1B	
$r_s = 0.89$	$p < 0.001$

gap detection (Beta = 0.64, SE = 0.10, $z = 6.53$, $p < 0.001$, $R_{sp}^2 = 0.060$). Participants were able to achieve $\geq 70\%$ correct at gap duration values as short as ~ 440 ms (Fig. 2); however, mean accuracy of trials with gaps <200 ms duration hovered around chance (48%, SD 15%). By contrast, ceiling performance (approximately 80%) was reached by 700 ms, with negligible improvement as gap duration increased beyond that. We found no association between Rhythm Task and correct responses to trials containing gaps (Appendix A, Table A2), despite Rhythm Task correlating very highly with overall performance in the task ($r_s = 0.71$, $p = 0.003$). Signal detection analysis revealed that this relationship was driven by the False Alarm rate. Summarising by median-split Rhythm Task, participants with higher nonverbal rhythm perception ability had a much lower False Alarm rate (Mean 0.19, SD 0.11) in comparison to those with lower nonverbal rhythm perception (Mean 0.41, SD 0.11). The Hit rate, however, was more similar between participants with higher (Mean 0.71, SD 0.11) and lower (Mean 0.64, SD 0.09) nonverbal rhythm. In other words, although doing well in the Rhythm Task did not confer an advantage to gap detection, higher nonverbal rhythm-skilled participants were less likely to report gaps erroneously (Fig. 3).

Turning to stimulus-specific effects, we found Speech Duration did not improve model fit, but increases in Breath Duration were significantly associated with higher likelihood of gap detection (Beta = 0.18, SE = 0.06, $z = 3.17$, $p = 0.002$, $R_{sp}^2 = 0.005$). To test whether the opposite effect was found in trials that truly contained no gap, we fit a model with Breath Duration and Speech Duration, as well as Rhythm Task. Indeed, participants were more likely to incorrectly report hearing gaps with increasing Breath Duration (Beta = -0.22 , SE = 0.07, $z = -3.07$, $p = 0.002$, $R_{sp}^2 = 0.000$) and Speech Duration (Beta = -0.20 , SE = 0.06, $z = -3.29$, $p < 0.001$, $R_{sp}^2 = 0.001$). These effects were of a smaller magnitude than the benefit of Rhythm Task (Beta = 0.65, SE = 0.15, $z = 4.37$, $p < 0.001$, $R_{sp}^2 = 0.003$), and there were no significant interactions between these predictors ($p \geq 0.24$). Descriptive statistics for d-prime and other details are provided in Appendix A, Tables A1, A2, A3.

2.3.1.2. Reaction times. We found an effect of response, with Correct

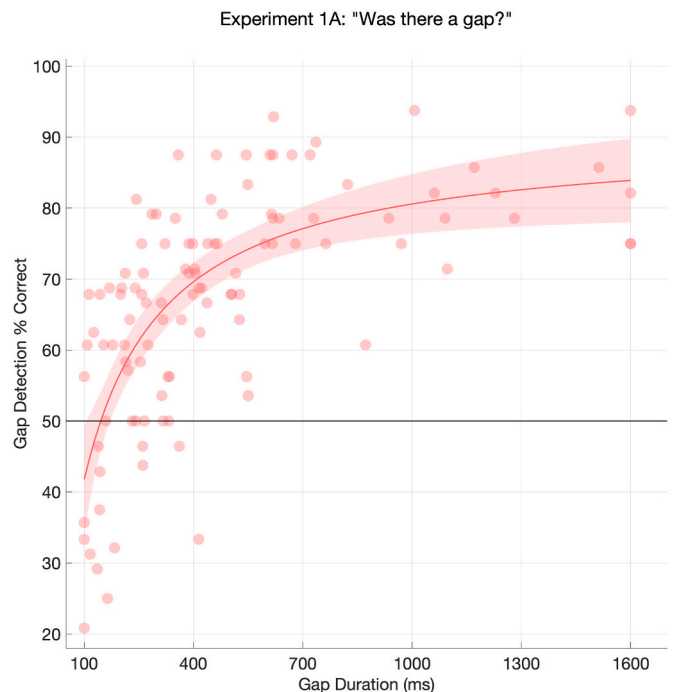


Fig. 2. Results for the Gap Detection Task in Experiment 1A showing effect of Gap Duration on percent correct (calculated within trials). The solid horizontal line indicates 50% performance. Shaded regions represent 95% confidence bounds of the fit curve.

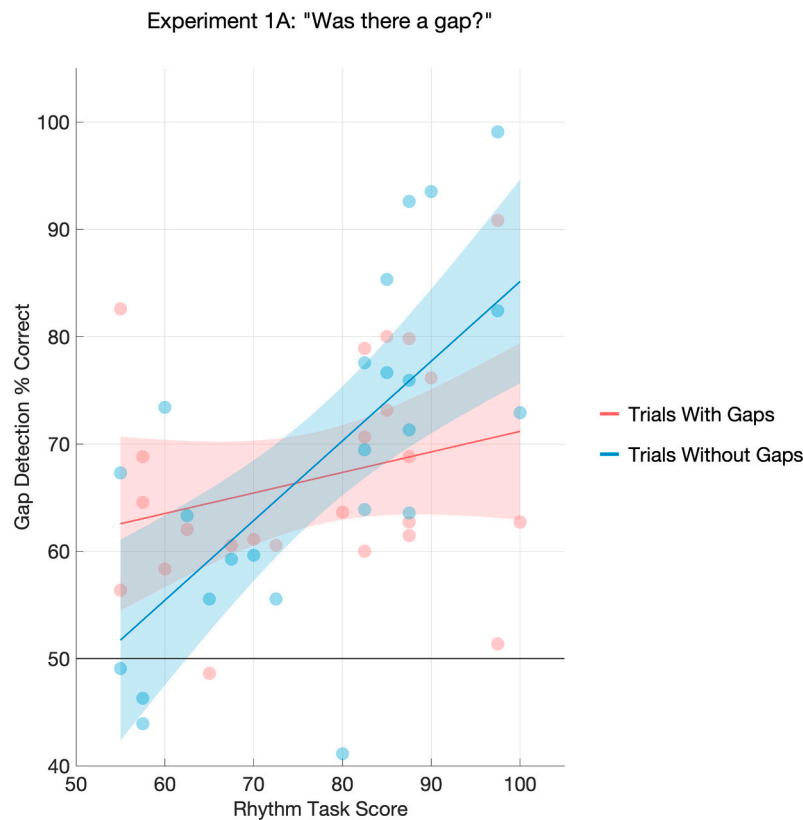


Fig. 3. Results for the Gap Detection Task in Experiment 1A showing effect of Rhythm Task score on percent correct (calculated within participants). The solid horizontal line indicates 50% performance. Shaded regions represent 95% confidence bounds of the linear fit.

answers (Mean 594 ms, SD 335 ms) associated with shorter reaction times than Incorrect responses (Mean 719 ms, SD 402 ms; Beta = -0.13 , SE = 0.02, $t = -5.86$, $p < 0.001$, $R_{sp}^2 = 0.009$). There was no effect of Breath Duration ($p = 0.85$). Increasing Gap Duration was associated with slower reaction times, but only when participants answered incorrectly (Estimate = 0.0002, SE = 0.00008, $t = 3.08$, $p = 0.006$). There was no relationship between Gap Duration and Correct response time ($p = 1.00$).

2.3.2. Experiment 1B: "where was the gap?"

2.3.2.1. Accuracy. <1% (count 23) of Experiment 1B reaction times exceeded 5 s, and these observations were removed from further analysis. Participants were more likely to correctly identify gaps that occurred after Breath 2 (Mean 76%, SD 18%) in comparison to gaps that followed Breath 1 (Mean 61%, SD 18%; Beta = 0.61, SE = 0.22, $z = 2.74$, $p = 0.006$, $R_{sp}^2 = 0.007$). Moreover, Bonferroni-corrected post hoc tests found no effect of Gap Duration in trials where the gap occurred after Breath 1 ($p = 1.00$; Fig. 4). For gaps that followed Breath 2, however, we observed that gaps with 400 ms (Estimate = 0.52, SE = 0.08, $z = -4.15$, $p < 0.001$) and 800 ms duration (Estimate = 0.31, SE = 0.07, $z = -4.85$, $p < 0.001$) were more likely to be correctly identified than 200 ms-long gaps. The contrast between 800 ms and 400 ms did not reach significance ($p = 0.12$).

Unlike in Experiment 1A, in this task there was an overall benefit to scoring higher in the Rhythm Task (Beta = 0.39, SE = 0.16, $z = 2.37$, $p = 0.02$, $R_{sp}^2 = 0.006$). Summarised by median-split group, participants with higher nonverbal rhythm sensitivity were better able to identify the gap's location (Mean 79%, SD 14%) than participants with lower rhythm sensitivity (Mean 55%, SD 6%). The interaction analysis with Gap Position and Duration showed that higher Rhythm Task scores predicted accuracy if the gap occurred after Breath 1 and was at least

400 ms long (Trend = 0.05, SE = 0.01, $z = 3.93$, $p < 0.001$). In comparison, Rhythm Task conferred an advantage for gaps following Breath 2 with duration as short as 200 ms (Trend = 0.05, SE = 0.02, $z = 3.20$, $p = 0.008$), but had ceased to do so once the gap was 800 ms long ($p = 0.72$), possibly due to a ceiling effect. Finally, though we observed a deleterious effect of Breath 2 Duration on accuracy regardless of Gap Position (Beta = -0.13 , SE = 0.05, $z = 2.72$, $p = 0.007$, $R_{sp}^2 = 0.002$), there was no influence of Breath 1 Duration ($p = 0.30$). Details for the model selection and final model are provided in Appendix B, Tables B1, B2.

2.3.2.2. Reaction times. There was a main effect of Correct (Beta = -0.15 , SE = 0.03, $t = -4.59$, $p < 0.001$, $R_{sp}^2 = 0.004$) showing that participants' correct responses tended to be faster (Mean 615 ms, SD 382 ms) than their incorrect responses (Mean 774 ms, SD 555 ms). Post hoc contrasts furthermore revealed that correct responses were made more quickly when the gap occurred after Breath 2 and had a duration of 400 ms (Mean 556 ms, SD 361 ms; Estimate = 0.12, SE = 0.03, $t = 3.92$, $p = 0.002$) or 800 ms (Mean 463 ms, SD 306 ms; Estimate = 0.15, SE = 0.04, $t = 3.55$, $p = 0.01$), in comparison to 200 ms (Mean 700 ms, SD 452 ms). There were no main effects of Gap Position ($p = 0.36$) nor of Rhythm Task ($p = 0.90$), and none of the other pairwise contrasts reached significance ($p \geq 0.13$).

2.4. Interim discussion

Few experiments directly examine the perception of speech breathing, and to our knowledge, none so far have investigated its timing on the sub-second scale. We therefore relied on data from speech production to form our expectation that listeners would be sensitive to violations of the natural breathing time series at the level of 200 ms. We instead found in Experiment 1A that, as a group, participants did not

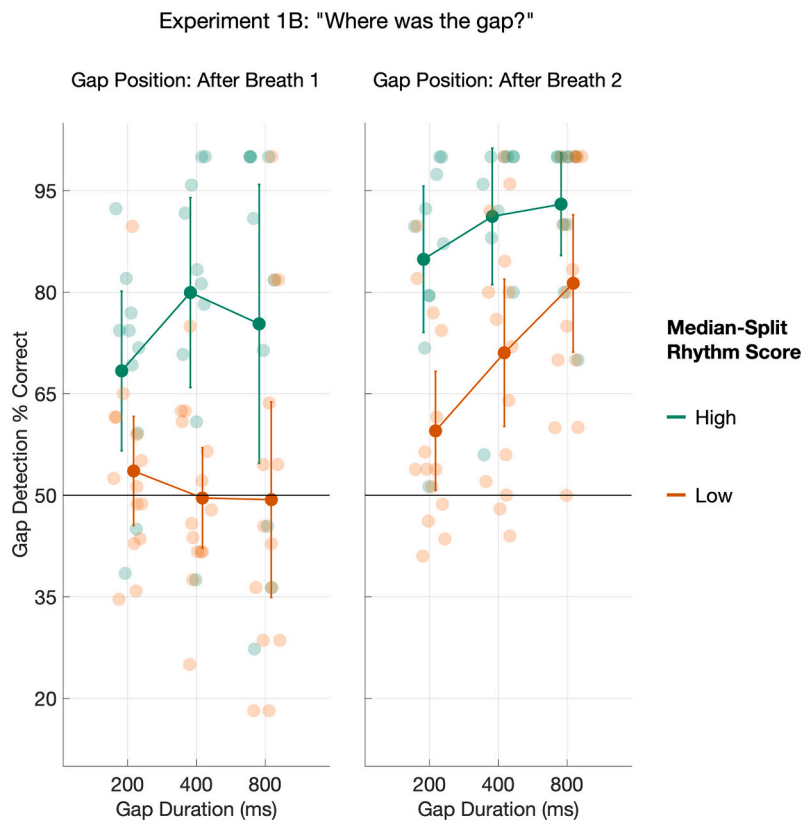


Fig. 4. Results for the Gap Detection Task in Experiment 1B showing effect of Gap Duration and Gap Position on percent correct (calculated within participants). For visual simplicity, the covariate Rhythm Task is plotted as median-split group (marker colour). The solid horizontal line indicates 50% performance. Bold markers show the group mean and error bars represent 95% confidence intervals of the mean.

reliably detect gaps (i.e., at >70% accuracy) until the gaps reached approximately 440 ms long. The covariate Rhythm Task, which was a measure of participants' nonverbal rhythm processing ability, failed to predict successful gap detection, but it was associated with the lower likelihood of incorrectly reporting gaps. By comparison, we observed a significant benefit of higher Rhythm Task score in Experiment 1B, which forced participants to choose whether they thought a gap had occurred after Breath 1 or after Breath 2. It is possible that participants adopted diverging strategies across the two tasks. In Experiment 1A, for instance, listeners might have relied on their sense of absolute or interval timing when deciding if a gap had occurred. In choosing between two potential gap locations, however, listeners could instead make a relative judgement, in which case the precise duration of the gap need not be known. It is therefore possible that the Rhythm Task, which involves making same-different comparisons between time series, draws on similar cognitive resources to Experiment 1B. Performance was not uniform between the two gap positions, however. Gaps that occurred after Breath 2, the breath that was interjected partway through the utterance, were more often correctly identified than gaps occurring after Breath 1, at the beginning of the utterance. Finally, in addition to the experimental manipulations, we observed that longer values of Breath Duration led participants to report having heard a gap, correctly or not (Experiment 1A). When asked after which breath a gap occurred, however, only the duration of the second, interjected breath influenced participants (Experiment 1B). We were somewhat surprised to see that increases in Breath 2 Duration resulted in fewer correct responses, regardless of whether the gap occurred after Breath 1 or Breath 2. We speculate that longer Breath 2 sounds may have been distracting, or could have possibly taxed participants' working memory. Unfortunately, because each speech utterance was paired to the same experimental condition across participants, the opportunity to tease apart speech-specific

influences is limited in Experiment 1.

As a first exploration into the perception of the speech breathing time series, Experiment 1 provides early indication of the thresholds at which participants perceive an artificially introduced silent gap. We can moreover confirm that listeners' acuity when making judgements about the locations of gaps varies between breaths within an utterance, with stronger performance associated with breaths that interject, in comparison to breaths that initiate, ongoing speech. On the other hand, the role of individual differences in nonverbal rhythm perception was unclear due to task-related differences. The sample ($n = 24$) is not large, although we did not estimate power a priori, given the exploratory nature of this study. It is nonetheless possible that we were unable to estimate the true underlying relationship between speech breathing gap detection and inter-individual differences in nonverbal rhythm processing. Another consideration is that the naturalistic speech stimuli do not vary across conditions, which is problematic given the observed stimulus-specific effects. Finally, the advantage of Breath 2 in Experiment 1B led us to wonder whether participants were simply better able to identify a gap within speech; that is, to succeed in the task, listeners could ignore the breath sounds and attend to interruptions of speech timing only. Gap Position is therefore confounded with exposure to speech.

To confirm our in-person findings, extend the current results, and address the limitations discussed above, we planned two larger, online studies. Experiment 2 is, like Experiment 1A, a "true" gap detection task, in that we asked participants to report whether or not they heard a gap somewhere in the utterance; however, we introduced multiple gap positions to better understand the contribution of the breath sounds and to disentangle the role of speech exposure. For instance, if breath sounds contribute to listeners' expectations concerning the timing of upcoming speech, we should expect that gap detection is better after, rather than

before a breath sound. If, on the other hand, the advantage of later gap positions is due to speech exposure alone, we won't find a difference between the two interjected breath gap positions. By comparison, Experiment 3 is a more faithful replication of Experiment 1B, as both share the same trial structure. The key development is that we manipulate the breath sounds themselves in Experiment 3. Specifically, Experiment 3 asks participants to judge after which breath a gap occurs, but the naturally occurring breath sounds have been inverted in half of trials. This allows us to dissociate Gap Position and inherent differences between initial and interjected breath sounds in naturalistic speech. We continue in Section 3 by describing Experiment 2 and its implementation in more detail.

3. Experiment 2

3.1. Methods and materials

3.1.1. Participants

Participants were recruited using either the professional online subject pool Prolific (Palan & Schitter, 2018), or University College London's internal online undergraduate subject pool. The former group were paid approximately £8 per hour for their time, and the latter were awarded course credits in exchange for participating. A total of 123 participants on Prolific and 76 participants from the university completed the experiment, resulting in a data set of $N = 199$ (gender was not confirmed; ages 18–35). Participants were prescreened for normal hearing, and all reported English as a primary language. On the basis of quality control, we excluded 15 participants from the Prolific set, and 2 participants from the university set, resulting in a final $N = 182$.

3.1.2. Design

3.1.2.1. Gap detection task. In Experiment 2, we examine silent gap detection using a similar paradigm to Experiment 1A, except that the gap can occur in one of three Gap Positions: After Breath 1, Before Breath 2, or After Breath 2 (Fig. 5). Participants were asked to report whether or not they heard a gap within each breath-speech-breath-speech trial, but the exact positions were not identified, and speech

breathing was not mentioned in the task instructions. This experiment was designed as a 3 (Gap Position: After Breath 1, Before Breath 2, After Breath 2) \times 5 (Gap Duration: 200, 325, 450, 575, 700 ms) within-subjects design, with Rhythm Task as covariate. To address variability in the naturalistic speech stimuli, we shuffled each utterance across the factorial combinations, meaning that participants all heard the same stimuli, but in different experimental conditions. The trial structure is depicted in Fig. 5, Panel A.

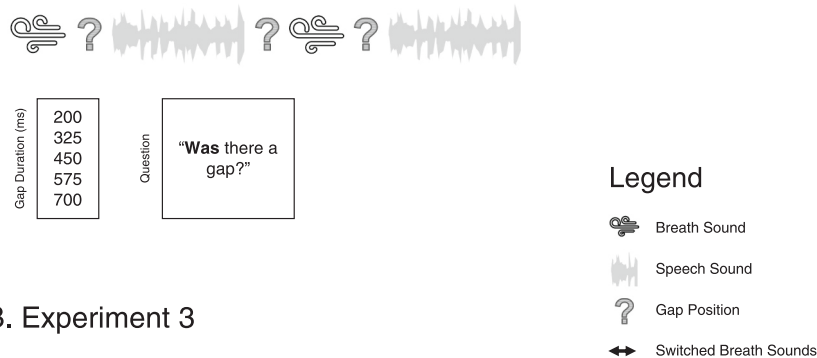
3.2. Stimuli

To enhance experimental control, we employed a subset from the full corpus of speech excerpts produced by the two speakers (1 male and 1 female) who happened to produce the highest number of matched read excerpts that naturally followed the same breath-speech-breath-speech structure. In the case of spontaneous speech, we attempted to pair the excerpts as closely as possible between the speakers. This resulted in a balanced data set of 80 utterances, which consisted of 50% reading and 50% spontaneous speech. There were a total of 80 trials, of which 25% truly contained no added silent gap. The remaining 75% of trials were divided up evenly among the three possible gap positions. There were $3 \times$ the shortest two durations (200 and 325 ms) and $2 \times$ the middle duration (450 ms) for every $1 \times$ the longest two durations (575 and 700 ms), or a ratio of 3:2:1. Trial order was pseudorandom, with seven different order conditions: four for the Prolific group, and three for the university group, with a projected 25 participants per order. The speech excerpts were shuffled across experimental conditions by trial order, meaning that a single excerpt would be associated with multiple trials across different experimental conditions, but only heard once per participant.

3.3. Apparatus

The experimental components and participant experience were, in general, unchanged from Experiment 1, except that the user interface was delivered online using Gorilla experiment builder (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020).

A. Experiment 2



B. Experiment 3



Fig. 5. Graphic representations of the stimulus format and prompt to participants in Experiments 2 and 3.

3.4. Procedure

The components of the study were performed in the following order: 1. Administration of participant information and informed consent; 2. Gap detection task; 3. Nonverbal rhythm discrimination task; 3. Text-based debrief on the aims and background of the study.

3.4.1.1. Gap detection task. At the start of the task, participants were instructed that they would hear clips of speech wherein a brief silent gap may or may not have been added, and that this gap could occur in different locations within the speech. No mention was made of the gap's proximity to the breath sounds, nor of breathing more generally. Participants were given a practice session of 4 trials, consisting of no gap and one of each of the possible gap positions, with feedback on their accuracy. The task was otherwise implemented the same as in Experiment 1A.

3.4.1.2. Nonverbal rhythm discrimination task. The task procedure was virtually identical to that of Experiment 1.

3.5. Data processing and analysis

The analysis pipeline was in general similar to Experiment 1A. The explanatory variables were Gap Position (After Breath 1/Before Breath 2/After Breath 2), Gap Duration (200/325/450/575/700 ms), and Rhythm Task (continuous) with random intercepts for Participant and Speech Excerpt, as well as random slopes for Gap Duration and Gap Position within Participant. Breath 1 Duration, Breath 2 Duration, Speech 1 Duration, and Speech 2 Duration were also fit as covariates. Signal detection and reaction time analyses were also undertaken. In addition, we performed an exploratory analysis of the speech stimuli, using a data-driven random forest classification of responses to examine the role of linguistic and acoustic factors in gap detection. The details of this analysis are provided in [Appendix D](#).

3.6. Results

3.6.1. Accuracy

Descriptive statistics are provided in [Table 2](#). ~4% (count = 475) of reaction times exceeded 5 s, and these responses were removed from further analysis. Accuracy varied by Gap Position, most strikingly in that

Table 2
Summary statistics for Experiment 2.

Descriptive Statistics	
Count	182
Gap Detection Task	
Mean	57%
Median	57%
SD	15%
Min	13%
Max	87%
Nonverbal Rhythm Discrimination Task	
Mean	83.3%
Median	85%
SD	11.6%
Min	52.5%
Max	100%
Spearman Rank Correlation	
Gap Detection Task × Nonverbal Rhythm Discrimination Task	
$r_s = 0.12$	$p = 0.20$
SD Standard Deviation	

participants were much more likely to detect gaps that occurred Before Breath 2 (Mean 68%, SD 20%; Beta = 1.73, SE = 0.14, $z = 12.07$, $p < 0.001$, $R_{sp}^2 = 0.013$) or After Breath 2 (Mean 78%, SD 15%; Beta = 2.18, SE = 0.14, $z = 15.28$, $p < 0.001$, $R_{sp}^2 = 0.020$), when compared to After Breath 1 (Mean 27%, SD 19%). Bonferroni-corrected post hoc tests between the three levels of Gap Position were significant at nearly every Gap Duration ($p \leq 0.001$), such that After Breath 2 was most likely to predict correct answers, followed by Before Breath 2, and then After Breath 1. The exception was the Before versus After Breath 2 contrast, which failed to reach significance at Gap Duration 700 ms ($p = 0.35$). Closer examination of Gap Duration revealed a shift to lower gap detection thresholds from gaps placed Before Breath 2 to After Breath 2 ([Fig. 6](#)). Specifically, gaps as short as 450 ms yielded a mean accuracy of 93% (SD 6%) when they were placed After Breath 2, with a maximum performance of 96% (SD 4%) at 700 ms. None of the pairwise tests between levels of Gap Duration that were longer than 450 ms met significance ($p \geq 0.63$). By contrast, when gaps occurred Before Breath 2, responses to 450 ms-long gaps were 77% (SD 28%) correct on average. Ceiling performance was not approached until 575 ms (Mean 89%, SD 11%) with a maximum mean accuracy of 91% (SD 7%) at 700 ms. In other words, the perceptual threshold of Gap Duration was lower for gaps placed after, rather than before the interjected breath. Participants were very unlikely to report a gap that followed Breath 1, even when it was 700 ms long (Mean 38%, SD 18%). The addition of Breath 1 Duration, Speech 1 Duration, and Breath 2 Duration did not improve model fit ([Appendix C, Table C1](#)). We did, however, find a global negative effect of Speech 2 Duration on gap detection, regardless of Gap Position (Beta = -0.45, SE = 0.07, $z = -6.64$, $p < 0.001$, $R_{sp}^2 = 0.018$). Taking trials without gaps into consideration, signal detection analysis showed that, whereas the mean Hit rate was 0.57 (SD 0.15, Range [0.14 0.86]), the mean False Alarm rate was just 0.18 (SD 0.18, Range [0.01 0.88]). Overall, d-prime ranged from a minimum of -0.29 to a maximum of 3.09, with a mean of 1.36 (SD 0.70). 1% of participants ($n = 2$) produced negative d-prime values, meaning that their False Alarm rate (i.e., incorrectly reported gaps) was higher than their Hit rate (i.e., correctly reported gaps). On the whole, however, it appears that participants adopted a generally conservative strategy when reporting gaps in Experiment 2. As we had also seen in Experiment 1A, Rhythm Task predicted correct answer in trials without gaps ($r_s = 0.21$, $p = 0.03$), but not trials with gaps ($r_s = -0.01$, $p = 0.90$; [Fig. 7](#)). Details concerning the modelling and post hoc tests are given in [Appendix C, Tables C1, C2, C3, C4](#).

3.6.2. Reaction times

Trials where the gap occurred After Breath 1 produced the slowest reaction times (Mean 1.42 s, SD 0.42 s), followed by Before Breath 2 (Mean 1.30, SD 0.41; Beta = 0.10, SE = 0.04, $t = 2.15$, $p = 0.03$, $R_{sp}^2 < 0.001$), followed by After Breath 2 (Mean 1.22, SD 0.40; Beta = 0.14, SE = 0.05, $t = 2.83$, $p = 0.005$, $R_{sp}^2 = 0.001$; [Fig. 8](#)). There were multiple interactions between Gap Position, Gap Duration, and Correct answer, however. At midpoint Gap Duration value (450 ms), when participants successfully detected the gap, they were fastest in responding to trials where the gap occurred After Breath 2 (Mean 1.03 s, SD 0.68 s) in comparison to After Breath 1 (Mean 1.48 s, SD 1.03 s; Estimate = 0.41, SE = 0.06, $t = 6.26$, $p < 0.001$), as well as Before Breath 2 (Mean 1.12 s, SD 0.77 s; Estimate = 0.14, SE = 0.05, $t = 2.97$, $p = 0.01$). For missed 450 ms-long gaps, however, reaction times were faster After Breath 1 (Mean 1.40 s, SD 0.89 s) than Before Breath 2 (Mean 1.57 s, SD 1.08 s; Estimate = -0.20, SE = 0.08, $t = -2.68$, $p = 0.02$) and After Breath 2 (Mean 1.89 s, SD 1.07 s; Estimate = -0.35, SE = 0.13, $t = -2.72$, $p = 0.02$). Taking reaction time as a proxy of confidence, we can thus infer that participants rejected true gaps with the most certainty After Breath 1, and correctly reported gaps with the most certainty After Breath 2. Finally, there was also a positive association between increasing Speech 2 Duration and faster reaction times (Beta = -0.12, SE = 0.02, $t = -7.58$, $p < 0.001$, $R_{sp}^2 = 0.019$).

Experiment 2: "Was there a gap?"



Fig. 6. Results for the Gap Detection Task in Experiment 2 showing effect of Gap Duration and Gap Position on percent correct (calculated within participants). For visual simplicity, the covariate Rhythm Task is plotted as median-split group (marker colour). The solid horizontal line indicates 50% performance. Bold markers show the group mean and error bars represent 95% confidence intervals of the mean.

3.7. Interim discussion

Experiment 2 built on Experiment 1A, by asking participants to detect whether or not a silent gap had been added between a breath sound and speech; however, we found that participants were far more likely to correctly report the gap when it occurred adjacent to the second of the two breaths, perhaps due to speech entrainment-related enhancement in temporal processing. We moreover found evidence of a strong bias *against* the initial gap position. It is unclear why participants were so unlikely to report perceiving a gap in that instance, given participants in Experiment 1A could reliably detect gaps as short as 440 ms that followed initial breaths. If speech entrainment indeed facilitates temporal processing, it may be that the obviousness of gaps inserted later within the speech stream (i.e., after entrainment has begun) had the unintended effect of suppressing awareness of the gaps that preceded speech. Put otherwise, perhaps because participants were so confident when recognising gaps occurring in the two later gap positions, they may have overlooked or second-guessed the more difficult After Breath 1 trials. They were unlikely to realise that there were more trials with than without gaps, rather than a 50/50 split, which likely compounded the bias against gaps After Breath 1. On the other hand, we also found that the Before Breath 2 and After Breath 2 gap positions differed from another. Specifically, gaps following Breath 2 elicited higher rates of detection, and at lower thresholds in duration, suggesting that respiratory sounds may indeed contribute to listeners' temporal expectations concerning the onset of speech. Reaction times corroborated these results, indicating that participants responded more slowly even when correctly identifying gaps that occurred Before Breath 2, in comparison to After Breath 2. Finally, we replicated the pattern of results seen in Experiment 1A concerning Rhythm Task; namely, nonverbal rhythm skills did not predict Hit rates, but they were associated with lower False

Alarm rates.

Up to this point in the current study, the natural locations of the breath sounds within the speech stimuli have been preserved. This means that the placement of gaps is hitherto confounded with potential breath-specific latent factors. For example, in our corpus and as reported more widely in the literature, interjected breaths tend to be shorter than breaths that initiate utterances (Winkworth et al., 1995, 1994). Although we did not find an effect of either Breath 1 or Breath 2 Duration in Experiment 2, the latter of these was associated with poorer performance in Experiment 1B, which instead asked participants where the gap occurred. In planning Experiment 3, we therefore determined to manipulate the breath sounds themselves directly. A relatively naturalistic change is to simply re-arrange the natural breath sounds, such that the initial breath would instead be heard partway through the utterance, and the formerly interjected breath would then take place at the beginning of the utterance. This step disentangles gap position and unique characteristics of the natural breath sounds. We wondered whether properties of the respiratory sounds would have any bearing on listeners' perception of ensuing speech, in particular, if their ability to correctly identify the gap position would change depending on if the speaker's natural breathing was either preserved or manipulated. Although the exploratory acoustic analysis of Experiment 2 (Appendix D) did not indicate that acoustic properties of the breath sounds played a strong role in gap detection accuracy, inverting the breath sounds provides the opportunity to explore this possibility further.

Another question concerns the role of nonverbal rhythm perception skills. So far, we have observed that participants with higher Rhythm Task scores produce lower False Alarm rates during gap detection, and are also better at determining where a gap has occurred when given the choice between possible locations. It is possible, however, that the benefit of Rhythm Task merely reflects superior auditory processing, or

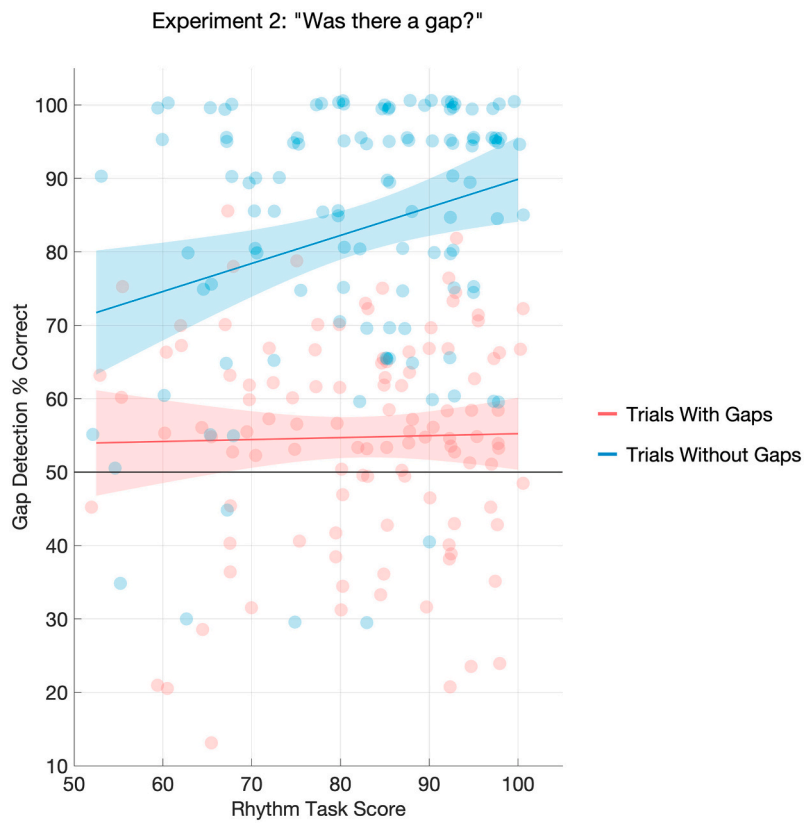


Fig. 7. Results for the Gap Detection Task in Experiment 2 showing effect of Rhythm Task score on percent correct (calculated within participants). The solid horizontal line indicates 50% performance. Shaded regions represent 95% confidence bounds of the linear fit.

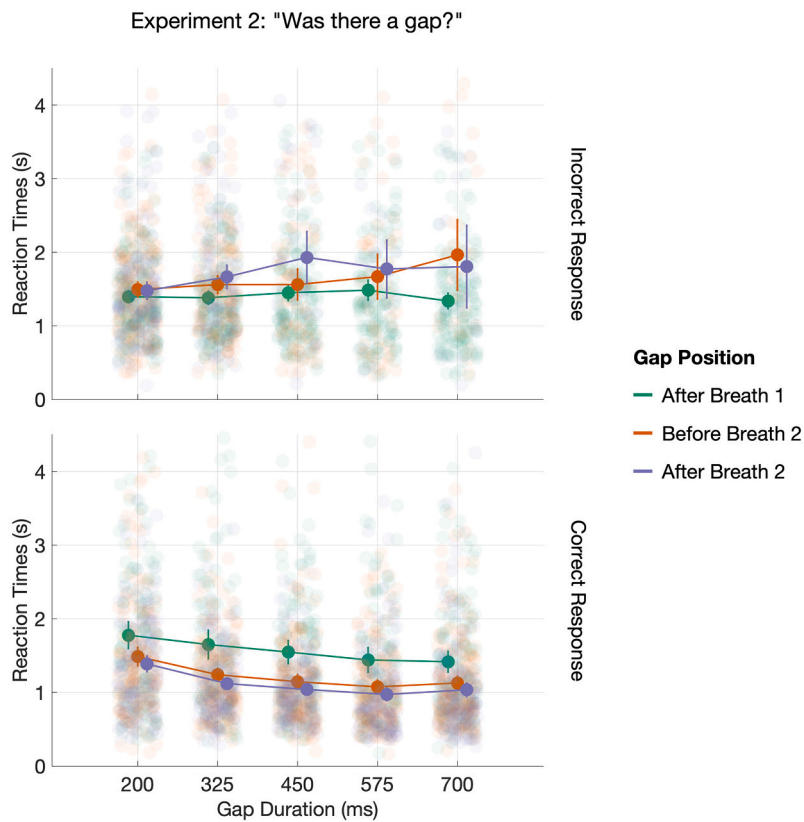


Fig. 8. Results for the Gap Detection Task in Experiment 2 showing effect of Gap Position and Gap Duration on reaction times (calculated within participants). Shaded regions represent 95% confidence bounds of the linear fit.

even forced-choice judgement, rather than timing or rhythm perception per se. We therefore developed a new task that could both serve as a control to the nonverbal rhythm discrimination task, as well as help ascertain the possible influence of conscious awareness of switched breath sounds. The solution was a switched breath recognition task. Briefly summarised, participants were presented with speech stimuli, half of which contained switched breaths, and the remaining half consisted of the unaltered utterances. The participant's objective was to report whether the breaths have been switched or not.

In planning Experiment 3 (Fig. 5, Panel B), we chose the same paradigm as in Experiment 1B, which asked participants to locate, rather than detect, a gap between breath sounds and speech. We were firstly motivated to replicate our in-person findings in a larger, online sample. The second reason was that the overall effect of Rhythm Task was stronger in Experiment 1B than in the other experiments, and we wished to confirm whether this relationship could be attributed to task-specific demands, rather than a spurious correlation arising from the small sample size. Finally, we found no effect of any breath duration in Experiment 2, but Breath 2 Duration was associated with poorer accuracy in Experiment 1B. It therefore seemed reasonable to select the paradigm that had, so far, elicited the strongest breath sound-specific effects. This final experiment is more exploratory than the others, and we formed three hypotheses, the first being the null hypothesis that altering the natural order of the breath sounds will not affect gap detection accuracy. The second hypothesis is that participants will in fact be more likely to correctly locate a gap if the breath sounds are manipulated, possibly due to the stimulus' unnaturalness, leading in turn to conscious awareness and/or heightened attention. The third hypothesis is that accuracy will be *degraded* in manipulated breath sound trials, potentially because of misleading information about upcoming speech timing conveyed by the breath sounds themselves.

4. Experiment 3

4.1. Methods and materials

4.1.1. Participants

Participants were recruited using Prolific (Palan & Schitter, 2018) and were paid £8 per hour for their time. A total of 109 participants completed Experiment 3 (ages 18–35). Participants were screened in the same manner as Experiment 2, and had not taken part in our earlier experiments. On the basis of data quality measures, we were forced to exclude 7 participants, resulting in a final $N = 102$.

4.1.2. Design

4.1.2.1. Gap detection task. Experiment 3 was similar to Experiment 1B, in that participants were asked to choose which of two gap positions (After Breath 1 / After Breath 2) contained a silent gap. To explore breath-specific effects, the natural initial and interjected breath sounds were inverted in half the trials. This allowed us to test whether unique characteristics of natural speech-related breathing would moderate successful gap detection. Experiment 3 was formed as a 2 (Switched Breath: Yes, No) \times 2 (Gap Position: Breath 1, Breath 2) \times 3 (Gap Duration: 200, 450, 700 ms) within-subjects design. The primary covariate terms were Rhythm Task, as well as score correct in the switched breath recognition task (Switched Breath Task), but we also fit terms for Breath 1 and Breath 2 Duration (i.e., the duration of the first and second breaths heard by the participant), and Speech 1 and Speech 2 Duration.

4.1.2.2. Switched breath recognition task. To determine whether conscious awareness of switched breaths exerted an influence on gap detection performance, we administered a short 2-forced choice task wherein participants answered whether they thought the breaths had been switched. To make the task as easy as possible, utterances were

chosen as the 4 with the most contrasting breaths by duration, per speaker, in the corpus, resulting in 16 trials in total, 8 of which contained truly switched breaths. Participants were distributed across two pseudorandom trial orders, with the switched and non-switched versions of each utterance counterbalanced across trial orders.

4.1.3. Stimuli

From the full corpus of four speakers, we selected, for each speaker, the twenty utterances with the greatest contrast in duration between the first breath and second breath. Within this subset, the mean duration was 705 ms (SD 295 ms) for initial breaths, and 371 ms (SD 108 ms) for interjected breaths. Although we did not explicitly balance the three speech types, there was nonetheless good representation of each, consisting of 38.75% Articles Reading, 28.75% Poems Reading, and 32.5% Spontaneous Speech. The stimuli were generated using the same audio processing pipeline as in Experiments 1 and 2. 50% of trials belonged to the Switched Breath condition, and the remaining 50% were unaltered in this respect. Gap position and duration were balanced across the 80 utterances, with the exception of the 700 ms duration condition, of which half as many occurred as the other two duration values. This approach was determined given the relatively high accuracy for longer gaps in Experiment 1B. Trial order was pseudorandom, with eight different order conditions across participants. The speech excerpts were shuffled across experimental condition by trial order, ensuring each utterance appeared across multiple conditions, but was heard only once per participant.

4.1.4. Apparatus

The experiment was again run via the Gorilla online testing platform (Anwyl-Irvine et al., 2020), using the same participant interface as in Experiment 2.

4.1.5. Procedure

The components of the study were performed in the following order: 1. Administration of information sheet and informed consent; 2. Gap detection task; 3. Switched breath recognition task; 4. Nonverbal rhythm discrimination task; 5. Text-based debrief on the aims and background of the study.

4.1.5.1. Gap detection task. The task instructions were the same as Experiment 1B and the user interface was essentially similar to Experiment 2. Although we did not describe or refer to any inversion of natural breath sounds in the task instructions, the practice session contained trials from both switched and non-switched conditions.

4.1.5.2. Switched breath recognition task. We implemented this new task similarly to the other components in the experiment, with the same inter-trial timing and visual cues prompting participants at the beginning of trials. Participants were explicitly informed that, in some trials, the breath at the beginning and the breath in the middle of the speech had been switch around. Their job was to guess whether the breaths were switched in the speech they heard. Due to time constraints, there was no practice session with feedback for this task.

4.1.5.3. Nonverbal rhythm discrimination task. The task procedure was the same as in the previous experiments.

4.2. Data processing and analysis

Analysis was performed using the same pipeline, in general, as in Experiments 1 and 2. The factorial explanatory variables of interest were Switched Breath (Yes/No), Gap Position (Breath 1/Breath 2), Gap Duration (200/450/700 ms). The covariate terms were Rhythm Task, Switched Breath Task, Breath 1 Duration, Breath 2 Duration, Speech 1 Duration, and Speech 2 Duration. There were random intercepts for

Participant and Speech Excerpt, and random slopes for Gap Duration and Gap Position within Participant.

4.3. Results

4.3.1. Accuracy

Descriptive statistics are provided in Table 3. ~5% (count = 380) of reaction times exceeded 5 s and were removed from further analysis. We found that the inclusion of the Switched Breath term did not improve the model fit (Appendix E, Table E1), and nor did we find any correlation between participants' ability to discern switched breaths and their gap detection ($r_s = 0.05, p = 0.64$; Fig. 10). There was, however, a main effect of Rhythm Task (Beta = 0.33, SE = 0.11, $z = 2.86, p = 0.004, R_{sp}^2 = 0.008$). Summarising by median-split group, low Rhythm Task scores were associated with a mean accuracy of 79% (SD 41%), and high scores 84% (SD 37%). Similar to our findings from the in-person Experiment 1B, gaps following Breath 2 were more likely to be correctly identified (Mean 83%, SD 38%) than those that occurred after Breath 1 (Mean 77%, SD 42%), even after controlling for the duration of the breath and speech segments (Beta = 0.57, SE = 0.17, $z = 3.47, p < 0.001, R_{sp}^2 = 0.002$; Fig. 9). Unlike Experiment 1B, however, we found that increasing Gap Duration predicted higher accuracy, regardless of Gap Position, and that all pairwise contrasts between the factorial predictors were significant ($p < 0.04$). Hence, the current data are, in short, less variable, and we see a comparatively subtler difference driven by the placement of the gap. Closer inspection of the complex interactions between Gap Position and the speech-specific covariates showed that increase in Breath 1 Duration was associated with improved accuracy, but only when the gap followed Breath 1 (Trend = 0.0008, SE = 0.0003, $z = 2.43, p = 0.03$; Fig. 11). Similarly, we found a positive correlation between Breath 2 Duration and accuracy when the gap occurred after Breath 2 (Trend = 0.0.002, SE = 0.0004, $z = 5.19, p < 0.001$); on the other hand, when the gap was in fact after Breath 1, then Breath 2 Duration had a negative influence on accuracy (Trend = -0.003, SE = 0.0003, $z = -11.13, p$

Table 3
Summary statistics for Experiment 3.

Descriptive Statistics	
Count	102
Gap Detection Task	
Mean	79.9%
Median	85%
SD	14.12%
Min	48.75%
Max	100%
Nonverbal Rhythm Discrimination Task	
Mean	82.23%
Median	82.5%
SD	11.18%
Min	50%
Max	100%
Switched Breath Recognition Task	
Mean	66.36%
Median	68.75%
SD	17.19%
Min	6.25%
Max	100%
Spearman Rank Correlation	
Gap Detection Task × Nonverbal Rhythm Discrimination Task	
$r_s = 0.30$	$p = 0.003$
Gap Detection Task × Switched Breath Recognition Task	
$r_s = 0.05$	$p = 0.64$
Switched Breath Recognition Task × Nonverbal Rhythm Discrimination Task	
$r_s = 0.09$	$p = 0.38$
SD Standard Deviation	

<0.001). Speech 1 Duration was associated with correct answers, but only when Breath 2 Duration <400 ms, and the gap followed Breath 2 (Trend = 0.22, SE = 0.08, $z = 2.65, p = 0.008$). None of the other post hoc covariate terms reached statistical significance (Appendix E, Table E4).

4.3.2. Reaction times

Consistent with the previous experiments, Correct answers were associated with faster reaction times (Beta = -132.83, SE = 46.46, $t = -2.86, p = 0.004, R_{sp}^2 = 0.001$). Although the main effect of Gap Position was marginal ($p = 0.06$), post hoc analysis revealed that Correct responses to trials with gaps after Breath 2 were faster than those with gaps after Breath 1 at every level of Gap Duration ($p \leq 0.008$). Reaction times did not differ as a function of Breath 1 Duration ($p = 0.21$), nor of Breath 2 Duration ($p = 0.91$), but increases both in Speech 1 Duration (Beta = -49.97, SE = 15.80, $t = -3.16, p = 0.002, R_{sp}^2 = 0.002$) and Speech 2 Duration (Beta = -96.73, SE = 15.85, $t = -6.10, p < 0.001, R_{sp}^2 = 0.008$) were associated with faster responses overall. Finally, Rhythm Task also predicted shorter reaction times (Beta = -119.61, SE = 42.87, $t = -2.79, p = 0.006, R_{sp}^2 = 0.017$).

4.4. Interim discussion

In Experiment 3, we confirmed the association between Rhythm Task and the ability to locate a gap in the speech breathing time series, which we had first observed in Experiment 1B. Given that Rhythm Task only predicted False Alarm rates in Experiments 1A and 2, this suggests that the role of nonverbal rhythm sensitivity may indeed be task-specific, although the correlation was weaker in Experiment 3 ($r_s = 0.30, p = 0.003$) in comparison to Experiment 1B ($r_s = 0.59, p = 0.003$). This difference may be attributable to the smaller sample size of Experiment 1B, or the demographic character of our in-person versus online samples, the latter of which had a slightly higher Rhythm Task score (Mean 82%, SD 12%) than in the in-person group (Mean 78%, SD 14%). Similarly, participants in Experiment 3 were less troubled by gaps that occurred after Breath 1, although the results followed the same overall pattern as in Experiment 1B, with greater accuracy for gaps following Breath 2. Because we switched true initial and interjected breaths across our stimuli, we were moreover able to experimentally decouple breath duration from Gap Position, and show that the latter was not entirely dependent on differences between initial and interjected breaths.

The influences of Breath 1 and especially Breath 2 Duration were, nonetheless, appreciable, along with their interactions with Gap Position. Specifically, we saw that increase in breath duration was associated with greater propensity for participants to choose that breath as the gap's location. In particular, longer Breath 2 Duration appears to have dissuaded participants from correctly identifying gaps occurring after Breath 1. As a factor, however, Switched Breath status did not significantly explain any variability above and beyond Breath 1 and Breath 2 Duration. These results do not support our second hypothesis, which was that listeners would be more likely to correctly locate the gap in Switched Breath trials, perhaps having been alerted by the unnaturalness of switched breath sounds. Indeed, despite reasonable group performance in the switched breath recognition task (Mean 66%, SD 17%), participants' ability to explicitly identify whether or not breaths had been switched did not at all correspond with their success in the gap detection task. Given that we deliberately made the switched breath recognition task as easy as possible, this result may indicate that the identification of gaps indeed draws upon temporal acuity (as measured in the Rhythm Task), rather than auditory processing skills more generally. In any case, although we did not find definitive evidence for the alternative hypotheses that Switched Breath trials would confer either an advantage or disadvantage when identifying gap positions, we can establish that the duration of breath sounds affects listeners' impression of the timing of speech onsets.

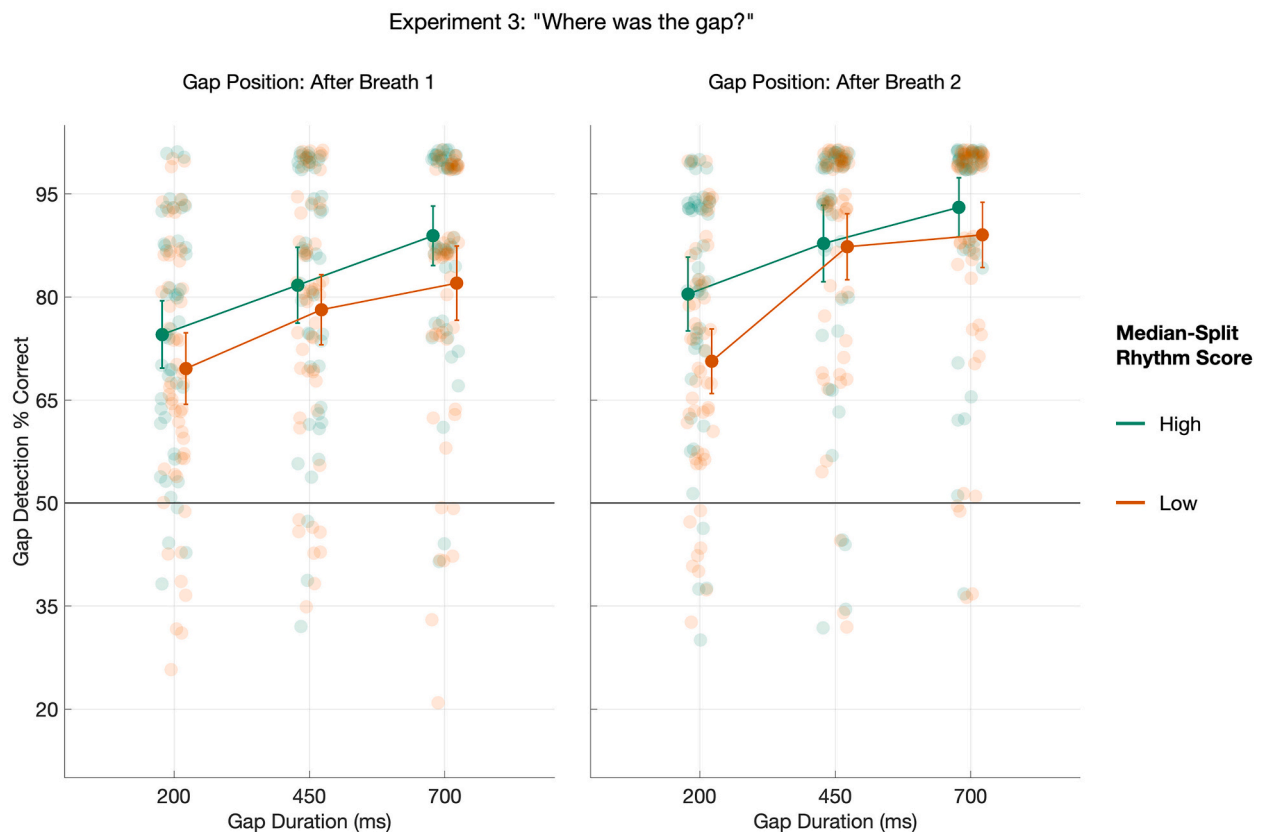


Fig. 9. Results for the Gap Detection Task in Experiment 3 showing effect of Gap Duration and Gap Position on percent correct (calculated within participants). For visual simplicity, the covariate Rhythm Task is plotted as median-split group (marker colour). The solid horizontal line indicates 50% performance. Bold markers show the group mean and error bars represent 95% confidence intervals of the mean.

5. General discussion

5.1. Experimental results

5.1.1. Gap position and duration

The primary motivation of this study was to determine whether listeners impose strict temporal expectations upon the speech breathing time series. We explored this question using a more implicit, classic gap detection paradigm (Experiments 1A and 2), as well as an explicit gap location paradigm (Experiments 1B and 3). Across the three experiments (total $n = 308$), we report consistent results concerning the effect of gap position and its mediation of gap duration; that is, listeners are sensitive to sub-second perturbations in the speech and speech breathing time series, but this sensitivity is not uniformly distributed across time. Rather, short silent gaps are more likely to be detected and identified when they occur following an interjected breath, rather than an initial breath. But we also found greater accuracy for gaps occurring *after*, rather than just before, interjected breaths (Experiment 2). In other words, the amount of prior speech exposure alone does not explain the perceptual benefits associated with later gaps. Although, as we discuss in the following section, participants were also influenced by speech stimulus-specific characteristics, on the whole, the effect of gap position held when controlling for natural differences between initial and interjected breaths, as well as the surrounding speech context. The gap's position also determined at what duration threshold it would be noticed. In the constrained paradigm used in Experiments 1B and 3, participants were able to locate gaps as short as 200 ms, especially when the gaps followed the interjected breath. Experiments 1A and 2, however, provided a more nuanced look at gap duration. For instance, when the gap followed the interjected breath in Experiment 2, there was a large jump in accuracy from about 55% for gap duration 200 ms, to over 75% at

325 ms. Ceiling performance, >90% correct, was reached by 450 ms. By comparison, ceiling performance was lower and not achieved until gap duration 575 ms when the gap occurred *before* the interjected breath. Surprisingly, gap duration was relatively ineffective in trials where the gap occurred after the initial breath in Experiment 2, with participants displaying a strong bias against gaps even as long as 700 ms. This stands in sharp contrast to Experiment 1A, where the gap could only ever follow Breath 1, in which case, participants reached a 70% detection rate by about 440 ms. A major difference between Experiment 2 and the other experiments is that we did not explicitly describe or direct participants' attention to breath sounds in the task instructions; hence, the gap duration at which participants can reliably detect or identify gaps may scale with attention and their expectations as listeners.

5.1.2. Nonverbal rhythm sensitivity

We also observed that individual differences in nonverbal rhythmic auditory processing affected performance in Experiments 1B and 3, wherein listeners were forced to choose between gap positions. Additionally, Rhythm Task in Experiment 3 was associated with faster reaction times, providing indirect evidence that participants with enhanced rhythm processing skills may have also been more confident when judging where the gap occurred. Individual differences in awareness of manipulated breath sounds were *not* correlated with that task, suggesting a specific role for sensitivity to rhythm, and not enhanced auditory processing more generally. For Experiments 1A and 2, in which participants were asked to report the presence of a gap, participants with higher Rhythm Task scores were less likely to incorrectly report having heard a gap in trials when there truly was no gap, meaning that they performed gap detection task with greater specificity. But their hit rate was not improved as a function of sensitivity to nonverbal rhythm. We can again speculate that the nature of the task

Experiment 3: "Where was the gap?"

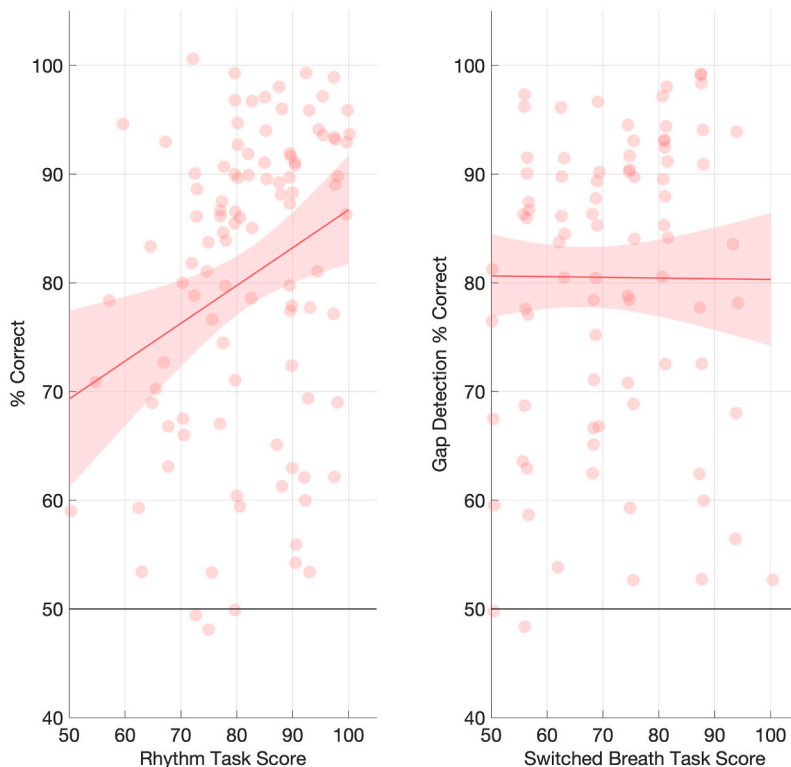


Fig. 10. Results for the Gap Detection Task in Experiment 3 showing effects of Rhythm Task score and Switched Breath Recognition Task score on percent correct (calculated within participants). The solid horizontal line indicates 50% performance. Shaded regions represent 95% confidence bounds of the linear fit.

Experiment 3: "Where was the gap?"

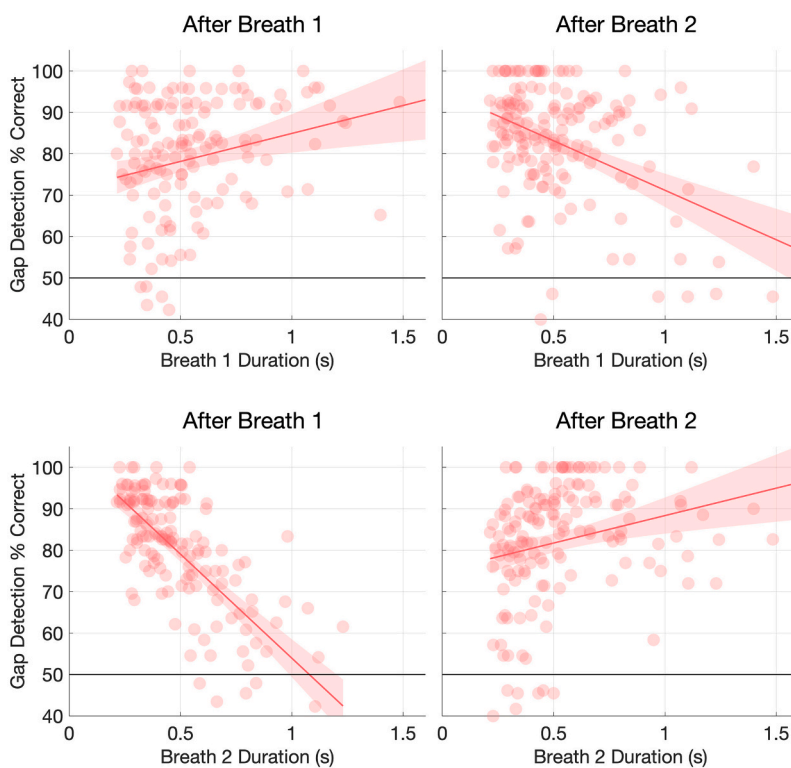


Fig. 11. Results for the Gap Detection Task in Experiment 3 showing effects of Breath 1 Duration and Breath 2 Duration on percent correct (calculated within speech stimuli). The solid horizontal line indicates 50% performance. Shaded regions represent 95% confidence bounds of the linear fit.

and listeners' beliefs may underlie the changing influence of nonverbal rhythm perception skills between experiments, a result that speaks to the idea of rhythm processing as a form of active sensing (Morillon, Hackett, Kajikawa, & Schroeder, 2015). Namely, the forced choice between possible *locations* of the gap may have pushed listeners to adopt an active listening strategy more similar to that used in the Rhythm Task. Specifically, the participants may have taken on a listening approach optimised to facilitate the later comparison between events in a time series. By contrast, in Experiments 1A and 2, participants could have approached gap detection differently; for instance, they may have relied on their sense of duration or interval timing, which is behaviourally and neurally distinct from event timing (Teki, Grube, Kumar, & Griffiths, 2011; Tierney & Kraus, 2015). Put differently, participants paid attention to *how long*, rather than *when*, a gap happened. There are possible implications here for daily life, in that optimal listening strategies probably also vary across different social contexts. For example, during spontaneous conversation, we may be sensitive to the timing of our partner's breathing to facilitate turn-taking and conversational flow (Rochet-Capellan & Fuchs, 2014; Torreira et al., 2015). In other situations (e.g., audiobook listening), wherein respiratory cues need not be acted upon, we do not need to engage with breath sounds in the same way. In any case, we can qualify our results concerning Rhythm Task, and suggest that there is indeed a specific relationship between nonverbal rhythm perception and speech breathing gap detection, but this correspondence is not universal across listening conditions.

5.2. Speech entrainment

The current work contributes new evidence that exposure to speech modulates temporal processing (Bosker & Ghitz, 2018; Kösem et al., 2018), by demonstrating that listeners were substantially more likely to both detect and locate gaps that occurred within an ongoing speech stream, rather than before speech begins. We ensured that the breath sounds had a standardised intensity and even manipulated them directly in Experiment 3, in addition to visually priming participants at the onset of each trial. As such, the effect of gap position cannot be attributed to breath sound-specific effects (e.g., louder interjected than initial breath sounds), nor subjects' having been caught unawares by the first breath sound in the case of Experiment 2. Although this is an interesting finding in the context of speech perception generally, a primary question motivating our study concerned the anticipatory temporal relationship between speech breathing and speech production. Given the relevance of breath sounds to social interaction (Rochet-Capellan & Fuchs, 2014; Włodarczyk, Heldner, & Edlund, 2015), does a listener's perception of a speaker's respiratory activity help them in forming predictions about forthcoming speech timing? Experiment 1A provides preliminary evidence that listeners require just a few hundred milliseconds of silence before they are sure of a gap's presence, even before having heard any speech. Moreover, our results in Experiment 2 show that gaps occurring after, rather than before, a breath sound are detected at a lower threshold of duration, and that listeners also respond comparatively more quickly to gaps that follow the breath. If breath sounds did not interact with ongoing speech entrainment processes, we wouldn't necessarily expect to see this asymmetry. Taking these findings into account, what is it that breath sounds contribute to temporal expectations? Research into visuomotor synchronisation suggests that humans are highly sensitive to changes in velocity when perceiving biological motion (Su, 2014; Varlet et al., 2014). Similarly, the instance of peak velocity of the stimulus envelope has also been shown to be the primary target during sensorimotor synchronisation to speech (Rathcke, Lin, Falk, & Bella, 2021; Scott, 1998). In short, moments of rapid acceleration within a given signal may provide a relatively discrete cue or anchor by which listeners can anticipate future events. When measured using inductance plethysmography, the kinematic profile of speech-related inhalation can be characterised as having a relatively shallow early slope, before quickly rising to terminate in a sharp peak as speech

begins. To our knowledge, the acoustic correlates of this speech breathing "shape" are not established, but it seems intuitively likely that breath sounds convey kinematic information about speech preparation, thereby helping listeners to form temporal predictions. Some support for this idea can be taken from the switched breath recognition task (Experiment 3). Although it was primarily used as a control in the current study, performance in the task was above chance on average (61%), and some participants achieved a perfect or near-perfect score. If listeners can detect manipulations of the natural breath-utterance pairings, it suggests that unique properties of breath sounds may well carry some meaning concerning forthcoming speech, even though it was not associated with judgement of gap position in that experiment. Future work could explore this idea further by replacing natural breath sounds with spectrally matched noise or time-reversed breaths, for example.

5.3. Gap detection within and between utterances

Although our focus was on listeners' perception of the speech breathing time series, the data may also be of interest to pause perception, an area of research that explores sensitivity to interruptions during speech timing more generally. Previous work investigating silent pauses inserted between words suggests above-chance detection rates with ~200 ms gap duration (Lovgren & Doorn, 2005; Warner, Whalen, Harel, & Jackson, 2022). Given that naturally occurring gaps—if present at all—normally last for just tens of milliseconds in connected speech, this approximate threshold seems plausible, with the caveat that the studies by Lovgren and Doorn (2005) and Warner et al. (2022) each used just a small number of tightly controlled speech excerpts. In any case, the threshold for interruptions to speech breathing boundaries seems to be about twice this duration (~400 ms). We speculate that, whereas within-speech gap detection may operate at the faster, syllabic time scale (e.g., ~4 – 8 Hz), speech breathing gap detection could instead be more comparable to stressed syllable timing (e.g., ~2 – 3 Hz). Cross-linguistic studies could investigate this possibility further, for example, by comparing speech breathing gap detection between speakers of languages known for their differing rhythmic structure.

Indeed, multiple cues appear to influence listeners when asked to segment speech or detect pauses in real time, which extend beyond the absolute duration of silences (Cole, Mo, & Baek, 2010; Duez, 1985, 1993; Lundholm Fors, 2015). These cues can include intelligibility (Butcher, 1981; Duez, 1985), prosodic cues such as vowel lengthening (Duez, 1993), and syntactical constraints, in that listeners may be primed to expect a pause at grammatical junctures (Cole et al., 2010). Moreover, evidence from cross-linguistic studies suggests that listeners of unfamiliar foreign speech are more objective than native listeners when estimating the length of pauses, at least for English listeners of Italian (Chiappetta, Monti, & O'Connell, 1987). Finally, the rapidity with which participants categorise pauses may also be influenced by online lexical processing, such that reaction times to pauses are slower when the pause occurs following a plausible word, in comparison to a nonsense word in short utterances (Mattys & Clark, 2002). In sum, temporal expectation in pause perception is subject to both acoustic and linguistic influences. With regards to speech breathing, we observed more mixed effects of the speech and breath sounds. Specifically, in Experiments 1A and 3, increasing breath duration was associated with the likelihood of listeners reporting or identifying a gap, correctly or otherwise. We did not, however, find strong modulation of performance by breath duration in Experiment 2, and longer interjected breaths in Experiment 1B tended to make participants perform worse, no matter where the gap was.

There was also a generally deleterious effect of longer speech that followed both breath sounds in Experiment 2. Given that Speech 2 by definition occurred after the presentation of any gap, this finding could potentially be linked to cognitive load. Cognitive load is typically studied with dual-task paradigms, which the current study did not incorporate. On the other hand, speech perception is itself an active,

cognitive process, rather than passive and automatic (Heald & Nusbaum, 2014). It is therefore possible that, if making temporal judgements draws upon the same attending resources as ongoing speech perception, increasing the amount of speech participants heard could have inhibited their retrospective awareness of gaps. In this case, the effect of Speech 2 Duration may be related to previous work showing that higher levels of cognitive load interfere with listeners' ability to make fine acoustic-phonetic judgements, such as vowel duration estimation (Chiu, Rakusen, & Mattys, 2019; Mattys, Barden, & Samuel, 2014). (Bosker, Reinisch, & Sjerps, 2017) reported that increasing cognitive load led participants to overestimate the preceding speech rate and subjectively perceive a subsequent ambiguous word as longer. Hence, if increasing Speech 2 Duration did in fact introduce additional cognitive load, it could be that participants' recollection of the gap "shrunk", along with their sense of the timing of speech more generally. An alternative, but related explanation for how Speech 2 Duration could have affected gap perception is based on working memory constraints: (Teki & Griffiths, 2014) found in a nonverbal temporal memory task that accuracy decayed with an increasing number of sequential intervals to be memorised. The relationship between cognitive load and working memory is beyond the scope of the present discussion, but our incidental finding here suggests that gap detection tasks present a potentially useful paradigm in which to explore these ideas further.

5.4. Limitations

Having used naturalistic speech stimuli, we forfeited a level of control, which we have addressed via counter-balancing across experimental conditions in large sample sizes, as well as incorporating terms like breath duration in the modelling. We moreover interrogated speaker and other speech-specific predictors in the exploratory analyses (Appendix D), finding that, although present, the influence of unique characteristics of the speech excerpts did not in general distort our experimental results. Future work should incorporate stimuli more systematically produced to interrogate these features. For example, Werner, Fuchs, Trouvain, and Möbius (2021) found that utterance-interjected breath sounds tend to have higher intensity values than utterance-initial breath sounds. We standardised the intensity of breath sounds when producing the stimuli, and so may have artificially subdued qualities of the breath sounds that would otherwise act as cues. With regards to syntactical aspects of the speech, we did not perform an in-depth linguistic analysis and cannot speak to these factors, such as the likely interaction between between grammar and gap detection. We did employ automatically calculated measures of pace and regularity in speech timing in the exploratory analysis, but saw no obvious effect of these terms. That said, algorithmically obtained parameters may not have been precise enough to reflect potentially subtle effects of speech timing (MacIntyre, Cai, & Scott, 2022). In light of the pause perception literature covered earlier, we expect that systematically varying the timing of breath sounds relative to local grammatical context could influence listeners' awareness of artificially imposed silent gaps, although breathing and grammar are not easily decoupled in naturalistic speech (Fuchs et al., 2013; Grosjean & Collins, 1979; Henderson et al., 1965; Winkworth et al., 1994). Another limitation is that we only tested English speech on English primary language listeners, so the impact of intelligibility and other cross-linguistic factors remains unknown. Finally, there is of course, the artificial nature of the gap detection tasks themselves, which required participants to scrutinise speech in a manner unlikely to arise outside of a laboratory. Conversational speech is often conducted in live, face-to-face settings, but visual and other multimodal

aspects of speech breathing are not addressed in the present experiments. Future studies could investigate how the visual and auditory percepts of breathing interact by experimentally decoupling them.

5.5. Conclusion

Speech breathing is not well explored by the current speech perception literature. Yet behavioural studies of social interaction (McFarland, 2001; Rochet-Capellan & Fuchs, 2014; Torreira et al., 2015) imply a facilitating role, and perhaps even an adaptive advantage, for sensitivity to speech breathing. Taken together, results in the present study indicate that the verbal acoustic signal alone does not comprise the "speech itself", but rather can be understood as part of an integrated percept that also includes nonverbal traces. Hence, instead of scrubbing the breath sounds from naturalistic recordings, or neglecting to account for respiration in synthesized stimuli, researchers might consider an active role for speech breathing in their own experimental work. For one hypothetical example, the distinctive sound of speech-related inhalations could serve as a rhythmic "phase reset", thereby readying listeners to entrain to an upcoming speech stream. In which case, altering the placement of breath sounds relative to speech may afford a unique opportunity to manipulate the perception of speech rhythm with a greater degree of naturalism than is normally possible, leading to new avenues in entrainment research. Indeed, engaging directly with speech breathing could open the door to generating new hypotheses using many paradigms already familiar to speech science, as well as help build towards a more naturalistic and complete understanding of speech perception in the wild. But beyond speech, the current work underscores the dynamic nature of auditory attention, and in particular the potential for embodied traces to exert a persuasive influence on temporal perception more generally. If confirmed in future work, these results may be of consequence for clinical applications, such as in the development of smart hearing devices that are sensitive to speech-related respiration. After all, if humans are specially attuned to social signals, few signs affirm the presence—nor communicative intent—of another so well as breathing does.

Data availability

Data associated with the gap detection tasks are available to download from <https://osf.io/3ecj7/>

Data from the other tasks, the stimuli, and the scripts used to generate stimuli and run the analyses are available by request from the corresponding author.

Funding

Funding was provided by University College London Graduate and Overseas Research Scholarships to Alexis Deighton MacIntyre.

Declarations of interest

None.

CRediT authorship contribution statement

Alexis Deighton MacIntyre: Conceptualization, Methodology, Software, Investigation, Data Curation, Formal analysis, Visualization, Writing – original draft. **Sophie K. Scott:** Supervision, Resources, Writing – review & editing.

Appendix A

A.1. Experiment 1A

Table A1
Descriptive statistics for the signal detection analysis of Experiment 1A.

D-Prime			
Mean	SD	Min	Max
1.01	0.65	0.10	2.24
Hit Rate			
Mean	SD	Min	Max
0.67	0.10	0.49	0.90
False Alarm Rate			
Mean	SD	Min	Max
0.32	0.16	0.01	0.59

Table A2
Model selection results are shown for the generalised linear mixed modelling of Correct Answer in Experiment 1A.

Model	Fixed Effects	Random Effects		Model Fit				Likelihood Ratio Test				
		Intercept	Slope	NP	AIC	BIC	LL	Dev.	χ^2	DF	p	
1		Participant		2	3244.7	3256.4	-1620.3	3240.7				
2		Trial Order, Participant		3	3246.2	3263.8	-1620.1	3240.2	0.53	1	0.47	
3		Excerpt + Participant		3	3164.1	3181.7	-1579.0	3158.1	82.61	1	<0.001	
4	Rhythm Task	Excerpt + Participant		4	3164.3	3187.7	-1578.1	3156.3	1.81	1	0.18	
5	Gap Duration	Excerpt + Participant		4	3115.4	3138.9	-1553.7	3107.4	50.66	1	<0.001	
6	Gap Duration	Excerpt + Participant	Gap Duration Participant	6	3105.4	3140.7	-1546.8	3093.5	13.91	2	0.001	
7	Gap Duration + Breath 1 Duration	Excerpt + Participant	Gap Duration Participant	7	3097.9	3138.9	-1542.0	3083.9	9.63	1	0.002	
8	Gap Duration × Breath Duration	Excerpt + Participant	Gap Duration Participant	8	3097.5	3144.4	-1540.7	3081.5	2.43	1	0.12	
9	Gap Duration + Breath Duration + Speech Duration	Excerpt + Participant	Gap Duration Participant	8	3099.8	3146.7	-1541.9	3083.8	0.06	1	0.81	

Response: Correct Answer, Count Observations: 2597, Groups: Participants (24), Excerpts (108).

Key: + Main Effects, × Interaction, |Grouped by, NP Number of Parameters, AIC Aikake Information Criterion, BIC Bayesian Information Criterion, LL Log Likelihood, Dev. Deviance, DF Degrees of Freedom, χ^2 Chi-square.

Table A3
Details for the generalised linear mixed modelling of Correct Answer in Experiment 1A.

Fixed Effects						
	Beta	SE	95% CI		z	p
			Lower	Upper		
Intercept	0.85	0.12	0.60	1.09	6.89	<0.001
Gap Duration	0.64	0.10	0.45	0.83	6.53	<0.001
Breath Duration	0.18	0.06	0.07	0.29	3.17	0.002
Random Effects						
		Variance	SD	Corr.		
Excerpt (Intercept)		0.11	0.33			
Participant (Intercept)		0.28	0.53			
Gap Duration, Participant (Slope)		0.11	0.32	0.44		
Model Fit						
R ²		Marginal	Conditional			
		0.09	0.21			

Model equation: Correct Answer ~ Gap Duration + Breath Duration + (1|Excerpt) + (1 + Gap Duration|Participant).

Key: SE Standard Error, CI Confidence Intervals, SD Standard Deviation, Corr. Correlation.

Confidence Intervals are calculated using the Wald method.

Table A4
Details for the linear mixed modelling of Reaction Times in Experiment 1A.

Fixed Effects							
	Beta	SE	95% CI		DF	t	p
			Lower	Upper			
Intercept	0.73	0.07	0.59	0.88	26.02	9.98	<0.001
Correct	-0.13	0.02	-0.18	-0.09	2569	-5.86	<0.001
Gap Duration	0.09	0.03	0.03	0.14	87.71	3.09	0.003
Breath Duration	0.00	0.01	-0.03	0.02	100.8	-0.19	0.85
Correct × Gap Duration	-0.10	0.02	-0.15	-0.05	2490	-4.21	<0.001
Random Effects							
		Variance	SD	Corr.			
Excerpt (Intercept)		0.01	0.09				
Participant (Intercept)		0.12	0.35				
Gap Duration, Participant (Slope)		0.01	0.08	-0.37			
Residual		0.25	0.49				
Model Fit							
R ²		Marginal	Conditional				
		0.01	0.36				

Model equation: Reaction Times ~ Correct × Gap Duration + Breath Duration + (1|Excerpt) + (1 + Gap Duration|Participant).

Key: SE Standard Error, CI Confidence Intervals, DF Degrees of Freedom, SD Standard Deviation, Corr. Correlation.

Confidence Intervals are calculated using the Wald method.

Appendix B

B.1. Experiment 1B

Table B1

Model selection results are shown for the generalised linear mixed modelling of Correct Answer in Experiment 1B.

Model	Fixed Effects	Random Effects		Model Fit				Likelihood Ratio Test				
		Intercept	Slope	NP	AIC	BIC	LL	Dev.	χ ²	DF	p	
1		Participant		2	4048.5	4060.9	-2022.3	4044.5				
2		Excerpt + Participant		3	3970.1	3988.7	-1982.1	3964.1	80.40	1		<0.001
3	Rhythm Task	Excerpt + Participant	4	3961.4	3986.1	-1976.7	3953.4	10.77	1		0.001	
4	Rhythm Task + Gap Position	Excerpt + Participant		5	3915.4	3946.3	-1952.7	3905.4	47.95	1		<0.001
5	Rhythm Task + Gap Position	Excerpt + Participant	Gap Position Participant	7	3860.2	3903.4	-1923.1	3846.2	59.19	2		<0.001
6	Rhythm Task + Gap Position + Gap Duration	Excerpt + Participant	Gap Position Participant	9	3848.6	3904.1	-1915.3	3830.6	15.68	2		<0.001
7	Rhythm Task + Gap Position × Gap Duration	Excerpt + Participant	Gap Position Participant	11	3838.2	3906.1	-1908.1	3816.2	14.35	2		<0.001
8	Rhythm Task + Gap Position × Gap Duration	Excerpt + Participant	Gap Position × Gap Duration Participant	29	3846.9	3788.9	-1894.4	3819.4	27.35	18		0.07
9	Rhythm Task × Gap Position × Gap Duration	Excerpt + Participant	Gap Position Participant	16	3840.6	3805.1	-1902.5	3808.6	11.16	5		0.05
10	Rhythm Task × Gap Position × Gap Duration + Breath 1 Duration	Excerpt + Participant	Gap Position Participant	17	3837.1	3941.9	-1901.5	3803.1	5.53	1		0.02
11	Rhythm Task × Gap Position × Gap Duration + Breath 1 Duration + Speech 1 Duration	Excerpt + Participant	Gap Position Participant	18	3838.3	3939.4	-1901.2	3802.3	0.72	1		0.40
12	Rhythm Task × Gap Position × Gap Duration + Breath 1 Duration + Breath 2 Duration	Excerpt + Participant	Gap Position Participant	18	3831.8	3942.8	-1897.9	3795.8	7.26	1		0.007
13	Rhythm Task × Gap Position × Gap Duration + Breath 1 Duration + Breath 2 Duration + Speech 2 Duration	Excerpt + Participant	Gap Position Participant	19	3833.8	3951.0	-1897.9	3795.8	0.03	1		0.86
14	Rhythm Task × Gap Position × Gap Duration + Breath 1 Duration × Gap Position + Breath 2 Duration	Excerpt + Participant	Gap Position Participant	19	3833.7	3950.9	-1897.9	3795.7	0.06	1		0.80
15	Rhythm Task × Gap Position × Gap Duration + Breath 1 Duration + Breath 2 Duration × Gap Position	Excerpt + Participant	Gap Position Participant	19	3833.6	3958.8	-1897.8	3795.6	0.15	1		0.69

Response: Correct Answer, Count Observations: 3529, Groups: Participants (24), Excerpts (150).

Key: + Main Effects, × Interaction, |Grouped by, NP Number of Parameters, AIC Aikake Information Criterion, BIC Bayesian Information Criterion, LL Log Likelihood, Dev. Deviance, DF Degrees of Freedom, χ² Chi-square.

Table B2
 Details for the generalised linear mixed modelling of Correct Answer in Experiment 1B.

Fixed Effects						
	Beta	SE	95% CI		z	p
			Lower	Upper		
Intercept	0.54	0.17	0.20	0.88	3.15	0.002
Rhythm Task	0.39	0.16	0.07	0.71	2.37	0.02
Gap Position	0.61	0.22	0.17	1.04	2.74	0.006
Gap Duration (400)	0.06	0.14	-0.21	0.34	0.48	0.63
Gap Duration (800)	0.01	0.19	-0.36	0.38	0.06	0.96
Breath 1 Duration	-0.05	0.05	-0.15	0.04	-1.02	0.30
Breath 2 Duration	-0.13	0.05	-0.22	-0.04	-2.72	0.007
Rhythm Task × Gap Position	0.38	0.22	-0.04	0.80	1.75	0.08
Rhythm Task × Gap Duration (400)	0.31	0.12	0.07	0.55	2.57	0.01
Rhythm Task × Gap Duration (800)	0.29	0.16	-0.01	0.61	1.87	0.06
Gap Position × Gap Duration (400)	0.57	0.21	0.17	0.99	2.77	0.006
Gap Position × Gap Duration (800)	1.15	0.30	0.57	1.74	3.85	<0.001
Rhythm Task × Gap Position × Gap Duration (400)	-0.29	0.19	-0.66	0.08	-1.53	0.13
Rhythm Task × Gap Position × Gap Duration (800)	-0.58	0.28	-1.13	-0.03	-2.06	0.04
Random Effects						
		Variance	SD	Corr.		
Excerpt (Intercept)		0.09	0.29			
Participant (Intercept)		0.53	0.72			
Gap Position, Participant (Slope)		0.77	0.88	-0.11		
Model Fit						
R ²		Marginal	Conditional			
		0.15	0.35			

Model equation: Correct Answer ~ Rhythm Task × Gap Position.
 × Gap Duration + Breath 1 Duration + Breath 2 Duration + (1|Excerpt) + (1 + Gap Position|Participant).
 Key: SE *Standard Error*, CI *Confidence Intervals*, SD *Standard Deviation*, Corr. *Correlation*.
 Confidence Intervals are calculated using the Wald method.

Table B3
 The results of post hoc contrasts using estimated marginal means for Gap Position, Gap Duration, and Rhythm Task in Experiment 1B.

Gap Position × Gap Duration						
Contrast	Condition	Estimate	SE	z	p	
Gap Position (Breath 1) / (Breath 2)	Gap Duration (200)	0.55	0.12	-2.74	0.006	
	Gap Duration (400)	0.31	0.08	-4.69	<0.001	
	Gap Duration (800)	0.17	0.06	-5.33	<0.001	
Gap Duration (200) / (400)	Gap Position (Breath 1)	0.94	0.13	-0.48	1.00	
	Gap Position (Breath 2)	0.52	0.08	-4.15	<0.001	
Gap Duration (200) / (800)	Gap Position (Breath 1)	0.99	0.19	-0.06	1.00	
	Gap Position (Breath 2)	0.31	0.07	-4.85	<0.001	
Gap Duration (400) / (800)	Gap Position (Breath 1)	1.06	0.21	0.29	1.00	
	Gap Position (Breath 2)	0.60	0.15	-2.05	0.12	
Rhythm Task						
Condition	Trend	SE	z	p		
Gap Position (Breath 1)	Gap Duration (200)	0.03	0.01	2.37	0.11	
	Gap Duration (400)	0.05	0.01	3.93	<0.001	
	Gap Duration (800)	0.05	0.01	3.33	0.005	
Gap Position (Breath 2)	Gap Duration (200)	0.05	0.02	3.20	0.008	
	Gap Duration (400)	0.06	0.02	3.08	0.01	
	Gap Duration (800)	0.03	0.02	1.56	0.72	

P-values are adjusted using the Bonferroni method.

Table B4
 Details for the linear mixed modelling of Reaction Times in Experiment 1B.

Fixed Effects							
	Beta	SE	95% CI		DF	t	p
			Lower	Upper			
Intercept	0.75	0.09	0.58	0.92	26.08	8.65	<0.001
Breath 1 Duration	-0.02	0.01	-0.04	0.00	310.70	-2.01	0.04
Breath 2 Duration	0.00	0.01	-0.02	0.02	695.40	0.16	0.88

(continued on next page)

Table B4 (continued)

Fixed Effects							
	Beta	SE	95% CI		DF	t	p
			Lower	Upper			
Rhythm Task	-0.01	0.09	-0.18	0.16	25.08	-0.13	0.90
Gap Position	-0.04	0.05	-0.13	0.05	267.90	-0.93	0.36
Gap Duration (400)	-0.09	0.05	-0.18	0.00	2256.00	-2.03	0.04
Gap Duration (800)	-0.12	0.06	-0.24	0.01	1394.00	-1.88	0.06
Correct	-0.15	0.03	-0.22	-0.09	3423.00	-4.59	<0.001
Rhythm Task × Gap Position	-0.02	0.04	-0.11	0.06	263.10	-0.56	0.57
Rhythm Task × Gap Duration (400)	-0.04	0.04	-0.13	0.04	3462.00	-1.05	0.30
Rhythm Task × Gap Duration (800)	-0.10	0.05	-0.21	0.01	3466.00	-1.77	0.08
Gap Position × Gap Duration (400)	0.02	0.08	-0.13	0.17	2722.00	0.29	0.77
Gap Position × Gap Duration (800)	0.16	0.11	-0.06	0.38	3040.00	1.39	0.16
Rhythm Task × Correct	-0.08	0.03	-0.14	-0.02	3422.00	-2.46	0.01
Gap Position × Correct	0.08	0.05	-0.02	0.18	3089.00	1.62	0.10
Gap Duration (400) × Correct	0.10	0.06	-0.01	0.21	3479.00	1.75	0.08
Gap Duration (800) × Correct	0.18	0.07	0.04	0.33	3481.00	2.43	0.02
Rhythm Task × Gap Duration (400) × Gap Position	0.01	0.08	-0.14	0.15	3453.00	0.07	0.94
Rhythm Task × Gap Duration (800) × Gap Position	-0.03	0.12	-0.26	0.19	3456.00	-0.30	0.76
Rhythm Task × Gap Position × Correct	0.01	0.05	-0.09	0.11	3227.00	0.24	0.81
Rhythm Task × Gap Duration (400) × Correct	-0.01	0.06	-0.12	0.09	3465.00	-0.25	0.80
Rhythm Task × Gap Duration (400) × Correct	0.02	0.07	-0.12	0.17	3481.00	0.31	0.75
Gap Position × Gap Duration (400) × Correct	-0.15	0.09	-0.32	0.02	3478.00	-1.75	0.08
Gap Position × Gap Duration (800) × Correct	-0.37	0.12	-0.62	-0.13	3470.00	-3.00	0.003
Rhythm Task × Gap Position × Gap Duration (400) × Correct	0.05	0.09	-0.12	0.23	3458.00	0.61	0.54
Rhythm Task × Gap Position × Gap Duration (800) × Correct	0.11	0.13	-0.14	0.37	3468.00	0.86	0.39
Random Effects							
		Variance	SD	Corr.			
Excerpt (Intercept)		0.00	0.06				
Participant (Intercept)		0.16	0.40				
Gap Duration, Participant (Slope)		0.01	0.08	-0.62			
Residual		0.23	0.48				
Model Fit							
R ²	Marginal	Conditional					
	0.04	0.40					

Model equation: Reaction Times ~ Correct × Rhythm Task × Gap Position.

× Gap Duration + Breath 1 Duration + Breath 2 Duration + (1|Excerpt) + (1 + Gap Position|Participant).

Key: SE Standard Error, CI Confidence Intervals, DF Degrees of Freedom, SD Standard Deviation, Corr. Correlation.

Confidence Intervals are calculated using the Wald method.

Appendix C. Appendic C

C.1. Experiment 2

Table C1

Model selection results are shown for the generalised linear mixed modelling of Correct Answer in Experiment 2.

Model	Fixed Effects	Random Effects		Model Fit				Likelihood Ratio Test				
		Intercept	Slope	NP	AIC	BIC	LL	Dev.	χ ²	DF	p	
1		Participant		2	15,406	15,420	-7700.9	15,402				
2		Participant + Excerpt		3	14,631	14,653	-7312.3	14,625	777.14	1	<0.001	
3	Rhythm Task	Participant + Excerpt		4	14,632	14,662	-7312.2	14,624	0.24	1	0.62	
4	Gap Position	Participant + Excerpt		5	11,868	11,905	-5928.9	11,858	2766.73	2	<0.001	
5	Gap Position	Participant + Excerpt	Gap Position Participant	10	11,724	11,905	-5852.2	11,704	153.51	5	<0.001	
6	Gap Position + Gap Duration	Participant + Excerpt	Gap Position Participant	14	10,957	11,798	-5464.4	10,929	775.65	4	<0.001	
7	Gap Position × Gap Duration	Participant + Excerpt	Gap Position Participant	22	10,786	11,060	-5370.8	10,742	187.13	8	<0.001	
8	Gap Position × Gap Duration	Participant + Excerpt	Gap Position × Gap Duration Participant	136	10,891	10,948	-5309.4	10,619	122.89	114	0.27	
9	Gap Position × Gap Duration + Breath 1 Duration	Participant + Excerpt	Gap Position Participant	23	10,786	10,955	-5369.9	10,740	1.82	1	0.18	
10	Gap Position × Gap Duration + Breath 2 Duration	Participant + Excerpt	Gap Position Participant	23	10,786	10,955	-5369.8	10,740	2.02	1	0.16	
11	Gap Position × Gap Duration + Speech 1 Duration	Participant + Excerpt	Gap Position Participant	23	10,787	20,956	-5370.3	10,741	0.94	1	0.33	

(continued on next page)

Table C1 (continued)

Model	Fixed Effects	Random Effects		Model Fit				Likelihood Ratio Test			
		Intercept	Slope	NP	AIC	BIC	LL	Dev.	χ^2	DF	p
12	Gap Position × Gap Duration + Speech 2 Duration	Participant + Excerpt	Gap Position Participant	23	10,752	10,922	-5353.3	10,706	35.08	1	<0.001
13	Gap Position × Gap Duration × Speech 2 Duration	Participant + Excerpt	Gap Position Participant	23	10,752	10,936	-5350.9	10,702	4.80	2	0.09

Response: Correct Answer, Count Observations: 11,677, Groups: Participants (182), Excerpts (80).

Key: + Main Effects, × Interaction, |Grouped by, NP Number of Parameters, AIC Aikake Information Criterion, BIC Bayesian Information Criterion, LL Log Likelihood, Dev. Deviance, DF Degrees of Freedom, χ^2 Chi-square.

Table C2

Details for the generalised linear mixed modelling of Correct Answer in Experiment 2.

Fixed Effects						
	Beta	SE	95% CI		z	p
			Lower	Upper		
Intercept	-1.82	0.14	-2.09	-1.54	-12.98	<0.001
Before Breath 2	1.73	0.14	1.45	2.00	12.07	<0.001
After Breath 2	2.18	0.14	1.90	2.46	15.28	<0.001
Gap Duration (325)	0.14	0.13	-0.11	0.39	1.08	0.28
Gap Duration (450)	0.62	0.14	0.35	0.88	4.48	<0.001
Gap Duration (575)	0.56	0.15	0.27	0.86	3.74	<0.001
Gap Duration (700)	1.11	0.13	0.86	1.37	8.65	<0.001
Speech 2 Duration	-0.45	0.07	-0.58	-0.32	-6.64	<0.001
Gap Duration (325) × Gap Position (Before Breath 2)	0.84	0.17	0.51	1.17	5.00	<0.001
Gap Duration (450) × Gap Position (Before Breath 2)	1.34	0.19	0.96	1.71	7.00	<0.001
Gap Duration (575) × Gap Position (Before Breath 2)	1.92	0.23	1.47	2.37	8.35	<0.001
Gap Duration (700) × Gap Position (Before Breath 2)	1.95	0.24	1.47	2.41	8.10	<0.001
Gap Duration (325) × Gap Position (After Breath 2)	1.21	0.18	0.86	1.55	6.87	<0.001
Gap Duration (450) × Gap Position (After Breath 2)	2.06	0.24	1.59	2.53	8.64	<0.001
Gap Duration (575) × Gap Position (After Breath 2)	2.66	0.31	2.06	3.26	8.71	<0.001
Gap Duration (700) × Gap Position (After Breath 2)	2.00	0.30	1.42	2.58	6.74	<0.001
Random Effects						
		Variance	SD	Corr.		
Excerpt (Intercept)		0.35	0.59			
Participant (Intercept)		1.22	1.05			
Gap Position (Before Breath 2), Participant (Slope)		1.12	1.06	-0.44		
Gap Position (After Breath 2), Participant (Slope)	0.80	0.89	-0.58	0.72		
Model Fit						
R ²		Marginal	Conditional			
		0.40	0.59			

Model equation: Correct Answer ~ Gap Duration × Gap Position + Speech 2 Duration + (1|Excerpt) + (1 + Gap Position|Participant).

Key: SE Standard Error, CI Confidence Intervals, SD Standard Deviation, Corr. Correlation.

Confidence Intervals are calculated using the Wald method.

Table C3

The results of post hoc contrasts using estimated marginal means for Gap Position by Gap Duration in Experiment 2.

Gap Position × Gap Duration					
Contrast	Condition	Estimate	SE	z	p
Gap Position (After Breath 1) / (Before Breath 2)	Gap Duration (200)	-0.34	0.03	-11.40	<0.001
	Gap Duration (325)	-0.55	0.03	-21.20	<0.001
	Gap Duration (450)	-0.72	0.03	-24.18	<0.001
	Gap Duration (575)	-0.69	0.03	-25.00	<0.001
	Gap Duration (700)	-0.62	0.03	-20.37	<0.001
Gap Position (Before Breath 2) / (After Breath 2)	Gap Duration (200)	-0.11	0.03	-3.58	0.001
	Gap Duration (325)	-0.14	0.02	-5.86	0.001
	Gap Duration (450)	-0.09	0.02	-5.07	<0.001
	Gap Duration (575)	-0.06	0.02	-3.72	<0.001
	Gap Duration (700)	-0.02	0.01	-1.58	0.35
Gap Position (After Breath 1) / (After Breath 2)	Gap Duration (200)	-0.45	0.02	-16.15	<0.001
	Gap Duration (325)	-0.69	0.02	-34.82	<0.001
	Gap Duration (450)	-0.72	0.03	-28.00	<0.001
	Gap Duration (575)	-0.75	0.03	-27.30	<0.001
	Gap Duration (700)	-0.64	0.03	-20.67	<0.001

P-values are adjusted using the Bonferroni method.

Table C4
The results of post hoc contrasts using estimated marginal means for Gap Duration by Gap Position in Experiment 2.

Gap Position × Gap Duration					
Contrast	Condition	Estimate	SE	z	p
Gap Duration (200) / (325)	Gap Position (After Breath 1)	-0.02	0.02	-1.07	1.00
	Gap Position (Before Breath 2)	-0.23	0.03	-8.87	<0.001
Gap Duration (200) / (450)	Gap Position (After Breath 2)	-0.26	0.03	-10.25	<0.001
	Gap Position (After Breath 1)	-0.09	0.02	-4.10	<0.001
	Gap Position (Before Breath 2)	-0.39	0.03	-13.91	<0.001
Gap Duration (200) / (575)	Gap Position (After Breath 2)	-0.36	0.03	-12.76	<0.001
	Gap Position (After Breath 1)	-0.08	0.02	-3.40	0.006
	Gap Position (Before Breath 2)	-0.44	0.03	-14.82	<0.001
Gap Duration (200) / (700)	Gap Position (After Breath 2)	-0.38	0.03	-13.03	<0.001
	Gap Position (After Breath 1)	-0.19	0.03	-7.46	<0.001
	Gap Position (Before Breath 2)	-0.47	0.03	-15.47	<0.001
Gap Duration (325) / (450)	Gap Position (After Breath 2)	-0.38	0.03	-12.88	<0.001
	Gap Position (After Breath 1)	-0.07	0.02	-3.24	0.01
	Gap Position (Before Breath 2)	-0.16	0.02	-6.97	<0.001
Gap Duration (325) / (575)	Gap Position (After Breath 2)	-0.11	0.02	-6.70	<0.001
	Gap Position (After Breath 1)	-0.06	0.02	-2.64	0.08
	Gap Position (Before Breath 2)	-0.21	0.02	-8.48	<0.001
Gap Duration (325) / (700)	Gap Position (After Breath 2)	-0.13	0.02	-7.58	<0.001
	Gap Position (After Breath 1)	-0.17	0.03	-6.61	<0.001
	Gap Position (Before Breath 2)	-0.24	0.02	-9.60	<0.001
Gap Duration (450) / (575)	Gap Position (After Breath 2)	-0.12	0.02	-7.22	<0.001
	Gap Position (After Breath 1)	0.00	0.03	0.34	1.00
	Gap Position (Before Breath 2)	-0.05	0.02	-2.77	0.06
Gap Duration (450) / (700)	Gap Position (After Breath 2)	-0.02	0.01	-1.86	0.63
	Gap Position (After Breath 1)	-0.10	0.03	-3.58	0.004
	Gap Position (Before Breath 2)	-0.09	0.02	-4.99	<0.001
Gap Duration (575) / (700)	Gap Position (After Breath 2)	-0.02	0.01	-1.48	1.00
	Gap Position (After Breath 1)	-0.11	0.03	-3.81	0.001
	Gap Position (Before Breath 2)	-0.04	0.02	-2.31	0.2
	Gap Position (After Breath 2)	-0.00	0.00	0.32	1.00

P-values are adjusted using the Bonferroni method.

Table C5
Details for the linear mixed modelling of Reaction Times in Experiment 2.

Fixed Effects							
	Beta	SE	95% CI		DF	t	p
			Lower	Upper			
Intercept	1.38	0.04	1.30	1.46	626.80	33.46	<0.001
Correct	0.16	0.06	0.05	0.28	11,500.00	2.74	0.006
Before Breath 2	0.10	0.04	0.01	0.18	11,120.00	2.15	0.03
After Breath 2	0.14	0.05	0.04	0.24	10,760.00	2.83	0.005
Gap Dur. (325)	0.01	0.04	-0.07	0.09	10,810.00	0.26	0.79
Gap Dur. (450)	0.09	0.05	-0.01	0.18	10,520.00	1.83	0.07
Gap Dur. (575)	0.08	0.05	-0.03	0.18	10,940.00	1.48	0.14
Gap Dur. (700)	0.04	0.05	-0.05	0.13	11,280.00	0.89	0.37
Speech 2 Duration	-0.12	0.02	-0.16	-0.09	84.59	-7.58	<0.001
Correct × Before Breath 2	-0.29	0.08	-0.44	-0.14	11,480.00	-3.75	<0.001
Correct × After Breath 2	-0.33	0.08	-0.48	-0.17	11,490.00	-4.15	<0.001
Correct × Gap Dur. (325)	0.01	0.09	-0.16	0.17	11,470.00	0.07	0.94
Correct × Gap Dur. (450)	-0.23	0.09	-0.41	-0.06	11,470.00	-2.63	0.008
Correct × Gap Dur. (575)	-0.30	0.10	-0.49	-0.12	11,480.00	-3.16	0.002
Correct × Gap Dur. (700)	-0.25	0.08	-0.41	-0.09	11,480.00	-3.08	0.002
Before Breath 2 × Gap Dur. (325)	0.01	0.07	-0.12	0.14	11,150.00	0.09	0.93
After Breath 2 × Gap Dur. (325)	0.10	0.08	-0.05	0.26	11,290.00	1.35	0.18
Before Breath 2 × Gap Dur. (450)	0.11	0.09	-0.07	0.28	11,340.00	1.21	0.23
After Breath 2 × Gap Dur. (450)	0.21	0.14	-0.06	0.48	11,460.00	1.53	0.13
Before Breath 2 × Gap Dur. (575)	0.17	0.12	-0.06	0.40	11,460.00	1.44	0.15
After Breath 2 × Gap Dur. (575)	0.31	0.19	-0.07	0.69	11,470.00	1.58	0.11
Before Breath 2 × Gap Dur. (700)	0.43	0.14	0.15	0.71	11,480.00	3.02	0.003
After Breath 2 × Gap Dur. (700)	0.15	0.19	-0.23	0.53	11,500.00	0.79	0.43
Correct × Before Breath 2 × Gap Dur. (325)	-0.18	0.11	-0.40	0.03	11,470.00	-1.66	0.10
Correct × After Breath 2 × Gap Dur. (325)	-0.32	0.12	-0.55	-0.09	11,470.00	-2.77	0.006
Correct × Before Breath 2 × Gap Dur. (450)	-0.18	0.12	-0.42	0.06	11,480.00	-1.44	0.15
Correct × After Breath 2 × Gap Dur. (450)	-0.43	0.16	-0.74	-0.11	11,460.00	-2.63	0.009
Correct × Before Breath 2 × Gap Dur. (575)	-0.28	0.15	-0.58	0.02	11,470.00	-1.83	0.07
Correct × After Breath 2 × Gap Dur. (575)	-0.40	0.22	-0.82	0.03	11,450.00	-1.84	0.07
Correct × Before Breath 2 × Gap Dur. (700)	-0.45	0.17	-0.77	-0.12	11,460.00	-2.69	0.007
Correct × After Breath 2 × Gap Dur. (700)	-0.25	0.21	-0.66	0.17	11,460.00	-1.17	0.24

(continued on next page)

Table C5 (continued)

Fixed Effects							
	Beta	SE	95% CI		DF	t	p
			Lower	Upper			
Random Effects							
		Variance	SD				
-5 Excerpt (Intercept)		0.02	0.14				
Participant (Intercept)		0.13	0.26				
Residual		0.62	0.79				
Model Fit							
R ²		Marginal	Conditional				
		0.06	0.24				

Model equation: Reaction Times ~ Correct × Gap Duration × Gap Position + Speech 2 Duration + (1|Excerpt) + (1|Participant).

Key: SE Standard Error, CI Confidence Intervals, DF Degrees of Freedom, SD Standard Deviation, Corr. Correlation.

Confidence Intervals are calculated using the Wald method.

Appendix D

D.1. Exploratory acoustic analysis of speech stimuli

D.1.1. Introduction

The speech stimuli employed in the current study were naturally produced by non-professional speakers reading from a variety of texts and speaking spontaneously. The reason for including a variety of speaking styles was to enhance the generalisability of our results; however, we wondered whether some acoustic or linguistic aspects of the stimuli may have influenced accuracy in the gap detection task. We therefore determined to extract a set of acoustic features to explore this question further, using data from Experiment 2, which employed the most naturalistic paradigm in the current study (participants were not informed of the relevance of breathing during task instructions). First, we estimated acoustic features from the breath sounds, the speech, and short segments (250 ms) of the speech immediately adjacent to (i.e., before and/or after) the breath sounds, which we term “Post-Breath 1”, “Pre-Breath 2”, etc. The extracted acoustic features included root mean square, spectral centroid, and F0 values. In addition, we calculated duration values of Breath 1 and Breath 2, as well as of Speech 1 and Speech 2, which refer to the complete vocalisations following Breath 1 and Breath 2, respectively. Finally, in the case of the speech, we generated the amplitude envelope following the approach described by [Oganian and Chang \(2019\)](#) and [MacIntyre et al. \(2022\)](#). From this envelope, we used the peaks in its first derivative to estimate vocalic onsets, and took the inter-vocalic onset as an approximation of the syllabic unit ([MacIntyre et al., 2022](#); [Oganian & Chang, 2019](#)). This time series allows us to derive the mean and coefficient of variation of inter-event intervals as metrics of speech timing. We performed this process twice, varying the peak-finding algorithm (islocalmax in MATLAB) to be less or more stringent (i.e., by adjusting the threshold of peak prominence) in order to estimate the syllable- and stressed syllable-levels of timing, respectively. Finally, the estimated counts of syllables and stressed syllables were also divided by the durations of Speech 1/ Speech 2 to produce the measure of speech rate. Gap Position, Gap Duration, Rhythm Task score, Speech Type (Articles/Poems/Spontaneous), and Speaker Identity were also included as predictors.

D.1.2. Feature selection with random forest classification

With this many potential predictors, there is a strong danger of over-parameterising, which may lead to unreliable estimates and overfitting in models, especially where there is likely to be collinearity between features ([Harrison et al., 2018](#); [Zuur, Ieno, & Elphick, 2010](#)). We therefore performed a feature selection analysis with a random forest classifier using the caret package in R ([Kuhn, 2008](#); [Kuhn & Johnson, 2019](#)). A random forest is an ensemble machine learning method that consists of many decision trees—logical nodes—that classify new observations by majority vote. During training, each decision tree applies a subset of randomly selected predictors to uniquely bootstrapped samples, making this technique highly robust to over-fitting ([Biau & Scornet, 2016](#)). Random forests are also sensitive to interactions between variables (provided they are of a sufficient magnitude, see discussion in [Darst, Malecki, & Engelman, 2018](#); [Inglis, Parnell, & Hurley, 2022](#)). The random forest was conducted using 5-fold cross validation repeated 10 times, with 10% of the data held out for testing. The class to be predicted was “correct” or “incorrect” response, and we also included data from the trials where no gap had occurred (Total number of observations 15,022).

The classifier accuracy for test data was 75% (95% CI [73% 77%]), and this overall accuracy rate was found to be significantly greater than the rate of the majority class, which was “correct” (No Information Rate 62%; $p < 0.001$). Having established that the model performed reasonably well, we extracted the variable importance weights ([Kuhn, 2012](#)), with the complete list of weights given in [Appendix D, Table D3](#). As can be seen, Gap Position, Rhythm Task, and Gap Duration are weighted most highly, with a substantial drop-off from Gap Duration (weight 39.03) to the highest-ranked acoustic feature, which was Speech 2 Duration (weight 6.11). No single acoustic predictor has a comparatively large weight, and the decision of which ones to retain was therefore somewhat arbitrary. As 29/36 (81%) of the variables were awarded a weight < 3.00 , we opted to take forward the predictors with weight ≥ 3.00 , each of which was duration-based: Speech 2 Duration (weight 6.11), Breath 2 Duration (weight 4.75), and Speech 1 Duration (weight 3.43). Breath 1 Duration, though not weighted as highly, was also included for completeness. These speech excerpt-specific, continuous predictors were fit as fixed effect terms in the modelling throughout the current study.

Table D1

Pearson’s two-tailed correlations between Breath 1 Duration, Speech 1 Duration, Breath 2 Duration, and Speech 2 Duration, calculated within the subset of speech excerpts that formed the stimuli of Experiment 2.

Correlations Within Speech Excerpts			
Experiment 2			
Breath 1 Duration	× Speech 1 Duration	0.28	0.05
	× Breath 2 Duration	0.30	0.03
	× Speech 2 Duration	0.32	0.02
Speech 1 Duration	× Breath 2 Duration	-0.01	1.00
	× Speech 2 Duration	0.04	1.00
Breath 2 Duration	× Speech 2 Duration	0.13	0.69

P-values are adjusted using the Bonferroni method.

Table D2

Descriptive statistics for the duration and speech rate (syllables per second) of segments within utterances by Speech Type.

Utterance Duration and Rate by Speech Type			
Mean (SD)	Articles	Poems	Spontaneous
Speech 1 Duration	3.19 (0.95)	2.97 (0.79)	3.02 (1.40)
Speech 1 Syllable Rate	5.78 (0.19)	5.81 (0.13)	5.74 (0.27)
Speech 2 Duration	2.50 (0.87)	2.96 (0.71)	2.23 (1.00)
Speech 2 Syllable Rate	5.76 (0.16)	5.77 (0.29)	5.69 (0.27)

Table D3

Variable weights for the experimental factors and covariate and acoustic features extracted in the exploratory acoustic analysis of Experiment 2. The weights are produced by fitting a random forest to 90% of observations, with the remaining 10% held out to verify the accuracy of the classifier, which was 75% (95% CI [73% 77%]) and significantly above chance (No Information Rate 62%; $p < 0.001$). Weights are scaled within range [0100].

Feature	Weight
Gap Position	100.00
Rhythm Task	76.27
Gap Duration	39.03
Speech 2 Duration	6.11
Breath 2 Duration	4.75
Speech 1 Duration	3.43
Breath 2 RMS	2.97
Pre-Breath 2 RMS	2.97
Speech 2 Inter-Stressed Syllable CV	2.85
Post-Breath 1 RMS	2.46
Breath 1 RMS	2.25
Pre-Breath 2 SpC	2.23
Breath 1 SpC	2.16
Speech 1 Inter-Syllable Mean	2.11
Speech 2 Inter-Stressed Syllable Mean	1.92
Speech 2 Inter-Syllable CV	1.89
Speech 2 Syllable Rate	1.85
Post-Breath 2 RMS	1.84
Speech 1 Stressed Syllable Rate	1.82
Speech 1 Inter-Stressed Syllable CV	1.80
Pre-Breath 2 F0	1.79
Speech 1 Inter-Syllable CV	1.78
Post-Breath 1 SpC	1.74
Breath 1 Duration	1.71
Breath 2 F0	1.69
Breath 1 F0	1.68
Breath 2 SpC	1.64
Speech 1 Inter-Stressed Syllable Mean	1.63
Post-Breath 2 SpC	1.61
Speech 2 Stressed Syllable Rate	1.53
Speech 2 Inter-Syllable Mean	1.48
Post-Breath 2 F0	1.47
Speech 1 Syllable Rate	1.47

(continued on next page)

Table D3 (continued)

Feature	Weight
Post-Breath 1 F0	1.37
Speech Type	1.07
Speaker	0.00

Key: CV Coefficient of Variation, F0 Fundamental Frequency, RMS Root Mean Square, SpC Spectral Centroid.

Appendix E

E.1. Experiment 3

Table E1

Model selection results are shown for the generalised linear mixed modelling of Correct Answer in Experiment 3.

Model	Fixed Effects	Random Effects		Model Fit				Likelihood Ratio Test				
		Intercept	Slope	NP	AIC	BIC	LL	Dev.	χ^2	DF	p	
1		Participant		2	6909.1	6923	-3452.5	6905.1				
2		Participant + Excerpt		3	6841.3	6862.2	-3417.7	6835.3	69.72	1	<0.001	
3	Gap Position	Participant + Excerpt		4	6809.1	6836.9	-3400.5	6801.1	34.26	1	<0.001	
4	Gap Position	Participant + Excerpt	Gap Position Participant	6	6746.2	6788	-3367.1	6734.2	66.83	2	<0.001	
5	Gap Position + Gap Duration	Participant + Excerpt	Gap Position Participant	8	6527.4	6583.1	-3255.7	6511.4	222.80	2	<0.001	
6	Gap Position × Gap Duration	Participant + Excerpt	Gap Position Participant	10	6514	6583.6	-3247	6494	17.43	2	<0.001	
7	Gap Position × Gap Duration	Participant + Excerpt	Gap Position × Gap Duration Participant	28	Did not converge.							
8	Gap Position × Gap Duration + B1	Participant + Excerpt	Gap Position Participant	11	6506.9	6583.4	-3242.4	6484.9	9.15	1	0.002	
9	Gap Position × Gap Duration + B1:Gap Position	Participant + Excerpt	Gap Position Participant	12	6435.5	6519	-3205.7	6411.5	73.38	1	<0.001	
10	Gap Position × Gap Duration + B1 + B1:Gap Position + B2	Participant + Excerpt	Gap Position Participant	13	6418.4	6508.9	-3196.2	6392.4	19.11	1	<0.001	
11	Gap Position × Gap Duration + B1 + B1:Gap Position + B2 + B2:Gap Position	Participant + Excerpt	Gap Position Participant	14	6166.3	6263.7	-3069.2	6138.3	254.07	1	<0.001	
12	Gap Position × Gap Duration + B1 + B1:Gap Position + B2 + B2:Gap Position + SP1	Participant + Excerpt	Gap Position Participant	15	6165	6269.3	-3067.5	6135	3.36	1	0.07	
13	Gap Position × Gap Duration + B1 + B1:Gap Position + B2 + B2:Gap Position + SP1:Gap Position	Participant + Excerpt	Gap Position Participant	16	6155.2	6266.6	-3061.6	6123.2	15.10	2	<0.001	
14	Gap Position × Gap Duration + B1 + B1:Gap Position + B2 + B2:Gap Position + SP1:Gap Position + SP2	Participant + Excerpt	Gap Position Participant	17	6157.2	6275.5	-3061.6	6123.2	0.00	1	0.96	
15	Gap Position × Gap Duration + B1 + B1:Gap Position + B2 + B2:Gap Position + SP1:Gap Position + SP2:GapPosition	Participant + Excerpt	Gap Position Participant	18	6147.3	6272.5	-3055.6	6111.3	11.94	2	0.002	
16	Gap Position × Gap Duration + B1 + B1:Gap Position + B2 + B2:Gap Position + SP1:Gap Position + SP2:GapPosition + Switched	Participant + Excerpt	Gap Position Participant	19	6149.3	6281.5	-3055.6	6111.3	0.01	1	0.92	
17	Gap Position × Gap Duration + B1 + B1:Gap Position + B2 + B2:Gap Position + SP1:Gap Position + SP2:GapPosition + Switched:Gap Position	Participant + Excerpt	Gap Position Participant	20	6151.1	6290.3	-3055.6	6111.1	0.16	2	0.93	
18	Gap Position × Gap Duration + B1 × B2 × Gap Position	Participant + Excerpt	Gap Position Participant	20	6139.8	6279	-3049.9	6099.8	11.44	2	0.003	
19	Gap Position × Gap Duration + B1 × B2 × SP1 × Gap Position	Participant + Excerpt	Gap Position Participant	26	6132.7	6313.6	-3040.3	6080.7	19.15	6	0.004	
20	Gap Position × Gap Duration + B1 × B2 × SP1 × SP2 × Gap Position	Participant + Excerpt	Gap Position Participant	40	6132.5	6410.9	-3026.2	6052.5	28.20	14	0.01	
21	Gap Position × Gap Duration + B1 × B2 × SP1 × SP2 × Gap Position + Switched Breath Task	Participant + Excerpt	Gap Position Participant	17	6502.7	6621	-3234.3	6468.7	0.03	1	0.85	
22	Gap Position × Gap Duration + B1 × B2 × SP1 × SP2 × Gap Position + Switched Breath Task:Gap Position	Participant + Excerpt	Gap Position Participant	18	6502.2	6627.5	-3233.1	6466.2	2.50	2	0.29	
23	Gap Position × Gap Duration + B1 × B2 × SP1 × SP2 × Gap Position + Rhythm Task	Participant + Excerpt	Gap Position Participant	17	6494.2	6612.6	-3230.1	6460.2	8.48	1	0.003	
24	Gap Position × Gap Duration + B1 × B2 × SP1 × SP2 × Gap Position + Rhythm Task:Gap Position	Participant + Excerpt	Gap Position Participant	42	6127.8	6420.1	-3021.9	6043.8	0.77	1	0.38	

Response: Correct Answer, Count Observations: 7780, Groups: Participants (102), Excerpts (80).

Key: B1 Breath 1 Duration, B2 Breath 2 Duration, SP1 Speech 1 Duration, SP2 Speech 2 Duration, + Main Effects, × Interaction, |Grouped by, NP Number of Parameters, AIC Aikake Information Criterion, BIC Bayesian Information Criterion, LL Log Likelihood, Dev. Deviance, DF Degrees of Freedom, χ^2 Chi-square.

Table E2
Details for the generalised linear mixed modelling of Correct Answer in Experiment 3.

Fixed Effects						
	Beta	SE	95% CI		z	p
			Lower	Upper		
Intercept	1.27	0.14	0.99	1.55	8.98	<0.001
Gap Duration (450)	0.60	0.10	0.40	0.80	5.88	<0.001
Gap Duration (700)	1.02	0.14	0.75	1.29	7.47	<0.001
Gap Position	0.57	0.17	0.25	0.90	3.47	<0.001
B1	0.21	0.09	0.04	0.38	2.43	0.02
B2	-0.93	0.08	-1.10	-0.77	-11.13	<0.001
SP1	0.08	0.08	-0.07	0.23	1.01	0.31
SP2	0.12	0.07	-0.03	0.26	1.61	0.11
Rhythm Task	0.33	0.11	0.10	0.55	2.86	0.004
Gap Duration (450) × Gap Position	0.42	0.15	0.11	0.72	2.69	0.007
Gap Duration (700) × Gap Position	0.41	0.21	-0.01	0.83	1.92	0.06
Gap Position × SP1	-0.04	0.11	-0.25	0.18	-0.32	0.75
Gap Position × SP2	-0.30	0.11	-0.51	-0.09	-2.77	0.005
B1 × B2	0.26	0.14	-0.01	0.53	1.87	0.06
B1 × SP1	-0.04	0.09	-0.21	0.14	-0.41	0.69
B2 × SP1	0.13	0.09	-0.05	0.31	1.38	0.17
B1 × SP2	-0.13	0.11	-0.34	0.09	-1.15	0.25
B2 × SP2	-0.05	0.11	-0.26	0.16	-0.46	0.65
SP1 × SP2	0.14	0.08	-0.01	0.29	1.86	0.06
Gap Position × B1	-0.32	0.13	-0.57	-0.06	-2.45	0.01
Gap Position × B2	1.52	0.13	1.26	1.78	11.40	<0.001
B1 × B2 × SP1	-0.19	0.19	-0.57	0.19	-0.97	0.33
B1 × B2 × SP2	-0.24	0.18	-0.60	0.12	-1.33	0.19
B1 × SP1 × SP2	0.03	0.11	-0.18	0.24	0.29	0.77
B2 × SP1 × SP2	0.25	0.12	0.00	0.49	2.00	0.05
Gap Position × B1 × B2	0.29	0.20	-0.11	0.68	1.43	0.15
Gap Position × B1 × SP1	-0.26	0.13	-0.51	-0.01	-2.00	0.05
Gap Position × B2 × SP1	-0.46	0.13	-0.71	-0.21	-3.55	<0.001
Gap Position × B1 × SP2	0.04	0.16	-0.27	0.35	0.25	0.8
Gap Position × B2 × SP2	-0.18	0.16	-0.49	0.12	-1.17	0.24
Gap Position × SP1 × SP2	-0.21	0.11	-0.42	0.00	-1.96	0.05
B1 × B2 × SP1 × SP2	-0.05	0.22	-0.48	0.38	-0.23	0.82
Gap Position × B1 × B2 × SP1	0.10	0.26	-0.40	0.60	0.40	0.69
Gap Position × B1 × B2 × SP2	0.37	0.25	-0.13	0.86	1.44	0.15
Gap Position × B1 × SP1 × SP2	-0.24	0.15	-0.53	0.06	-1.55	0.12
Gap Position × B2 × SP1 × SP2	-0.38	0.17	-0.71	-0.05	-2.26	0.02
Gap Position × B1 × B2 × SP1 × SP2	-0.22	0.30	-0.79	0.36	-0.73	0.46
Random Effects						
		Variance	SD	Corr.		
-5 Excerpt (Intercept)		0.04	0.19			
Participant (Intercept)		1.18	1.08			
Gap Position, Participant (Slope)		0.81	0.90	-0.11		
Model Fit						
R ²		Marginal	Conditional			
		0.17	0.43			

Model equation: Correct Answer ~ Gap Duration × Gap Position +

Breath 1 Duration × Breath 2 Duration × Speech 1 Duration × Speech 2 Duration × Gap Position + Rhythm Task + (1|Excerpt) + (1 + Gap Position|Participant).

Key: B1 Breath 1 Duration, B2 Breath 2 Duration, SP1 Speech 1 Duration, SE Standard Error, CI Confidence Intervals, SD Standard Deviation, Corr. Correlation.

Confidence Intervals are calculated using the Wald method.

Table E3
The results of post hoc contrasts using estimated marginal means for Gap Position and Gap Duration in Experiment 3.

Gap Position × Gap Duration					
Contrast	Condition	Estimate	SE	z	p
Gap Position (Breath 1) / (Breath 2)	Gap Duration (200)	-0.08	0.02	-3.54	0.004
	Gap Duration (450)	-0.08	0.01	-5.19	<0.001
	Gap Duration (700)	-0.06	0.01	-3.97	<0.001

(continued on next page)

Table E3 (continued)

Gap Position × Gap Duration					
Contrast	Condition	Estimate	SE	z	p
Gap Duration (200) / (450)	Gap Position (Breath 1)	-0.09	0.02	-5.30	<0.001
	Gap Position (Breath 2)	-0.08	0.01	-5.75	<0.001
Gap Duration (200) / (700)	Gap Position (Breath 1)	-0.13	0.02	-6.78	<0.001
	Gap Position (Breath 2)	-0.10	0.02	-6.06	<0.001
Gap Duration (450) / (700)	Gap Position (Breath 1)	-0.04	0.01	-3.08	0.006
	Gap Position (Breath 2)	-0.02	0.01	-2.46	0.04

P-values are adjusted using the Bonferroni method.

Table E4

The results of post hoc contrasts using estimated marginal means for Breath 1 Duration, Breath 2 Duration, Speech 1 Duration, and Speech 2 Duration with Gap Position in Experiment 3.

Breath 1 Duration					
Condition	Trend	SE	z	p	
Gap Position	Breath 1	0.0008	0.0003	2.43	0.03
	Breath 2	-0.0004	0.0004	-1.00	0.63
Breath 2 Duration					
Condition	Trend	SE	z	p	
Gap Position	Breath 1	-0.003	0.0003	-11.13	<0.001
	Breath 2	0.002	0.0004	5.19	<0.001
Speech 1 Duration					
Condition	Trend	SE	z	p	
Gap Position	Breath 1	0.08	0.08	1.01	0.62
	Breath 2	0.05	0.10	0.47	1.00
Speech 2 Duration					
Condition	Trend	SE	z	p	
Gap Position	Breath 1	0.14	0.08	1.61	0.22
	Breath 2	-0.22	0.11	-2.00	0.09

P-values are adjusted using the Bonferroni method.

Table E5

Details for the linear mixed modelling of Reaction Times in Experiment 3.

Fixed Effects							
	Beta	SE	95% CI		DF	t	p
			Lower	Upper			
(Intercept)	1420.35	58.32	1306.03	1534.66	314.65	24.35	<0.001
Correct	-132.83	46.46	-223.89	-41.77	7679.89	-2.86	0.004
Gap Duration (450)	153.88	60.40	35.50	272.25	7651.73	2.55	0.01
Gap Duration (700)	227.94	86.20	58.98	396.90	7635.02	2.64	0.008
Gap Position	108.36	57.32	-3.98	220.71	7324.41	1.89	0.06
B1	-18.61	14.73	-47.49	10.26	231.74	-1.26	0.21
B2	1.57	13.39	-24.67	27.81	113.95	0.12	0.91
SP1	-49.97	15.80	-80.94	-19.00	75.05	-3.16	0.002
SP2	-96.73	15.85	-127.80	-65.66	74.80	-6.10	<0.001
Rhythm Task	-119.61	42.87	-203.63	-35.59	95.32	-2.79	0.006
Correct × Gap Duration (450)	-214.15	68.55	-348.50	-79.79	7649.60	-3.12	0.002
Correct × Gap Duration (700)	-277.14	94.33	-462.02	-92.27	7657.44	-2.94	0.003
Correct × Gap Position	-199.77	66.50	-330.10	-69.44	7666.76	-3.00	0.003
Gap Position × Gap Duration (450)	-126.24	94.90	-312.23	59.75	7618.19	-1.33	0.18
Gap Position × Gap Duration (700)	-188.51	136.83	-456.70	79.67	7508.42	-1.38	0.17
Gap Position × B1	26.34	19.43	-11.73	64.42	5309.32	1.36	0.18
Correct × Gap Position × Gap Duration (450)	56.75	104.80	-148.66	262.16	7662.82	0.54	0.59
Correct × Gap Position × Gap Duration (700)	108.19	146.08	-178.11	394.50	7662.77	0.74	0.46
Random Effects							
		Variance	SD				
Excerpt (Intercept)		7523	86.74				
Participant (Intercept)		180,380	424.71				
Residual		625,017	790.58				

Model Fit

(continued on next page)

Table E5 (continued)

Fixed Effects	Beta	SE	95% CI		DF	t	p
			Lower	Upper			
			Marginal 0.06	Conditional 0.27			
R^2							

Model equation: Reaction Time ~ Correct × Gap Duration × Gap Position + Breath 1 Duration ×

Breath 2 Duration × Speech 1 Duration × Speech 2 Duration × Gap Position + Rhythm Task + (1|Excerpt) + (1|Participant).

Key: B1 *Breath 1 Duration*, B2 *Breath 2 Duration*, SP1 *Speech 1 Duration*, SP2 *Speech 2 Duration*, SE *Standard Error*, CI *Confidence Intervals*, DF *Degrees of Freedom*, SD *Standard Deviation*, Corr. *Correlation*.

Confidence Intervals are calculated using the Wald method.

References

- Aare, K., Gilmartin, E., Włodarczak, M., Lippus, P., & Heldner, M. (2020). Breath holds in chat and chunk phases of multiparty casual conversation. In *Proceedings of speech prosody* (pp. 779–783).
- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.
- Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2020). Cortical entrainment: What we can learn from studying naturalistic speech perception. *Language, Cognition and Neuroscience*, 35(6), 681–693.
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407.
- Atkinson, J. E. (1973). *Aspects of intonation in speech: Implications from an experimental study of fundamental frequency*. Unpublished doctoral dissertation. University of Connecticut Doctoral dissertation.
- Audacity Development Team. (2020). Audacity 2.4 [software]. Retrieved 2020-26-01, from <https://www.audacityteam.org/>.
- Bailey, J. A., & Penhune, V. B. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Experimental Brain Research*, 204(1), 91–101.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... Scheipl, F. (2012). *Package 'lme4'*. CRAN. Vienna, Austria: R Foundation for Statistical Computing.
- Bauer, A.-K. R., Bleichner, M. G., Jaeger, M., Thorne, J. D., & Debener, S. (2018). Dynamic phase alignment of ongoing auditory cortex oscillations. *NeuroImage*, 167, 396–407.
- Bernardet, U., Kang, S.-H., Feng, A., DiPaola, S., & Shapiro, A. (2019). Speech breathing in virtual humans: An interactive model and empirical study. In *2019 IEEE Virtual Humans and Crowds for Immersive Environments (vhcie)* (pp. 1–9).
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227.
- Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquiere, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, 32(2), 560–570.
- Bosker, H. R., & Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalisation. *Language, Cognition and Neuroscience*, 33(8), 955–967.
- Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, 94, 166–176.
- Braunschweiler, N., & Chen, L. (2013). Automatic detection of inhalation breath pauses for improved pause modelling in hhm-tts. In *Eighth isca workshop on speech synthesis*. Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communication functions. *Arbeitsberichte Kiel*, 15, 1–233.
- Chang, A., Bosnyak, D. J., & Trainor, L. J. (2019). Rhythmicity facilitates pitch discrimination: Differential roles of low and high frequency neural oscillations. *NeuroImage*, 198, 31–43.
- Chiappetta, J., Monti, L. A., & O'Connell, D. C. (1987). Pause perception: Some cross-linguistic comparisons. *Bulletin of the Psychonomic Society*, 25(2), 103–105.
- Chiu, F., Rakusen, L. L., & Mattys, S. L. (2019). Cognitive load elevates discrimination thresholds of duration, intensity, and f0 for a synthesized vowel. *The Journal of the Acoustical Society of America*, 146(2), 1077–1084.
- Coffey, E. B., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69.
- Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language & Cognitive Processes*, 25(7–9), 1141–1177.
- Darst, B. F., Malecki, K. C., & Engelman, C. D. (2018). Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genetics*, 19(1), 1–6.
- Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, 8, 311.
- Doelling, K. B., & Assaneo, M. F. (2021). Neural oscillations are a start toward understanding brain activity rather than the end. *PLoS Biology*, 19(5), Article e3001234.
- Donai, J. J., & Jennings, M. B. (2016). Gaps-in-noise detection and gender identification from noise-vocoded vowel segments: Comparing performance of active musicians to non-musicians. *The Journal of the Acoustical Society of America*, 139(5), EL128–EL134.
- Duez, D. (1985). Perception of silent pauses in continuous speech. *Language and Speech*, 28(4), 377–389.
- Duez, D. (1993). Acoustic correlates of subjective pauses. *Journal of Psycholinguistic Research*, 22(1), 21–39.
- Elmers, M., Werner, R., Muhlack, B., Möbius, B., & Trouvain, J. (2021). Take a breath: Respiratory sounds improve recollection in synthetic speech. In *Proc. Interspeech 2021* (pp. 3196–3200).
- Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., ... others. (2012). *Package 'car'*. Vienna: R Foundation for Statistical Computing.
- Fuchs, S., Petrone, C., Krivokapić, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, 41(1), 29–47.
- Fukuda, T., Ichikawa, O., & Nishimura, M. (2018). Detecting breathing sounds in realistic Japanese telephone conversations and its application to automatic speech recognition. *Speech Communication*, 98, 95–103.
- Grosjean, F., & Collins, M. (1979). Breathing, pausing and reading. *Phonetica*, 36(2), 98–114.
- Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, 35(5), 573–582.
- Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., ... Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6, Article e4794.
- Heald, S., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, 8, 35.
- Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1965). Temporal patterns of cognitive activity and breath control in speech. *Language and Speech*, 8(4), 236–242.
- Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences*, 109(49), 20095–20100.
- Hitzenko, K., Mazuka, R., Elsner, M., & Feldman, N. H. (2020). When context is and isn't helpful: A corpus study of naturalistic speech. *Psychonomic Bulletin & Review*, 1–37.
- Inglis, A., Parnell, A., & Hurley, C. B. (2022). Visualizing variable importance and variable interaction effects in machine learning models. *Journal of Computational and Graphical Statistics*, 1–13.
- Jaeger, B. (2017). *Package 'r2glmm'*. Vienna: R Found Stat Comput. available CRAN R-project org/package=R2glmm.
- Jalaei, B., Valadbeigi, A., Panahi, R., Nahrani, M. H., Arefi, H. N., Zia, M., & Ranjbar, N. (2019). Central auditory processing tests as diagnostic tools for the early identification of elderly individuals with mild cognitive impairment. *Journal of Audiology & Otology*, 23(2), 83.
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96(3), 459.
- Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, 28(18), 2867–2875.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28, 1–26.
- Kuhn, M. (2012). Variable importance using the caret package. *Journal of Statistical Software*, 6.
- Kuhn, M., & Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press.
- Kunert, R., & Jongman, S. R. (2017). Entrainment to an auditory signal: Is attention involved? *Journal of Experimental Psychology: General*, 146(1), 77.
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320(5872), 110–113.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1), 119.
- Lenth, R. (2018). Package 'lsmeans'. *The American Statistician*, 34(4), 216–221.
- Lieberman, P. (1967). Intonation, perception, and language. In *MIT research monograph*.
- Lovgren, T., & Doorn, J. V. (2005). Influence of manipulation of short silent pause duration on speech fluency. In *Disfluency in spontaneous speech*.
- Lundholm Fors, K. (2015). *Production and perception of pauses in speech* (Unpublished doctoral dissertation). University of Gothenburg.

- MacIntyre, A. D., Cai, C. Q., & Scott, S. K. (2022). Pushing the envelope: Evaluating speech rhythm with different envelope extraction techniques. *The Journal of the Acoustical Society of America*, 151(3), 2002–2026.
- MacLarnon, A. M., & Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 109(3), 341–363.
- Matthews, T. E., Thibodeau, J. N., Gunther, B. P., & Penhune, V. B. (2016). The impact of instrument-specific musical training on rhythm perception and production. *Frontiers in Psychology*, 7, 69.
- Mattys, S. L., Barden, K., & Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin & Review*, 21(3), 748–754.
- Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language*, 47(3), 343–359.
- McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(6), 1102.
- McFarland, D. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*, 44(1), 128–143.
- Mckay, L. C., Evans, K. C., Frackowiak, R. S., & Corfield, D. R. (2003). Neural correlates of voluntary breathing in humans. *Journal of Applied Physiology*, 95(3), 1170–1178.
- Mishra, S. K., Panda, M. R., & Herbert, C. (2014). Enhanced auditory temporal gap detection in listeners with musical training. *The Journal of the Acoustical Society of America*, 136(2), EL173–EL178.
- Morillon, B., Hackett, T. A., Kajikawa, Y., & Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Current Opinion in Neurobiology*, 31, 230–238.
- Music radar drum samples. (2022, April). http://cdn.mos.musicradar.com/audio/sample_s/musicradar-drum-samples.zip (Accessed: 2022-04-24).
- Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*, 5(11), eaay6279.
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27.
- Pashler, H. (2001). Perception and production of brief durations: Beat-based versus interval-based timing. *Journal of Experimental Psychology: Human Perception and Performance*, 27(2), 485.
- Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, 3, 320.
- Peirce, J. W. (2007). Psychopy—Psychophysics software in python. *Journal of Neuroscience Methods*, 162(1–2), 8–13.
- Phillips, D. P. (1999). Auditory gap detection, perceptual channels, and temporal resolution in speech perception. *Journal of the American Academy of Audiology*, 10(6), 343.
- Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2(4), 411–440.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria.
- Rathcke, T., Lin, C.-Y., Falk, S., & Bella, S. D. (2021). Tapping into linguistic rhythm. *Laboratory Phonology*, 12(1).
- Rochet-Capellan, A., & Fuchs, S. (2013). The interplay of linguistic structure and breathing in german spontaneous speech. In *14th annual conference of the international speech communication association (interspeech 2013)* (p. 1228).
- Rochet-Capellan, A., & Fuchs, S. (2014). Take a breath and take the turn: How breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 369(1658), 20130399.
- Scott, S. K. (1998). The point of p-centres. *Psychological Research*, 61(1), 4–11.
- Slifka, J. (2003). Respiratory constraints on speech production: Starting an utterance. *The Journal of the Acoustical Society of America*, 114(6), 3343–3353.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149.
- Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: Musicians as a model of auditory learning. *Hearing Research*, 308, 109–121.
- Su, Y.-H. (2014). Peak velocity as a cue in audiovisual synchrony perception of rhythmic stimuli. *Cognition*, 131(3), 330–344.
- Teki, S., & Griffiths, T. D. (2014). Working memory for time intervals in auditory rhythmic sequences. *Frontiers in Psychology*, 5, 1329.
- Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *Journal of Neuroscience*, 31(10), 3805–3812.
- Tierney, N., & Kraus, N. (2015). Evidence for multiple rhythmic skills. *PLoS One*, 10(9), Article e0136645.
- Torreira, F., Bögel, S., & Levinson, S. C. (2015). Breathing for answering. The time course of response planning in conversation. *Frontiers in Psychology*, 6.
- Trouvain, J., Möbius, B., & Werner, R. (2019). On acoustic features of inhalation noises in read and spontaneous speech. In *1st international seminar on the foundations of speech: Breathing, pausing, and voice*.
- Varlet, M., Coey, C. A., Schmidt, R. C., Marin, L., Bardy, B. G., & Richardson, M. J. (2014). Influence of stimulus velocity profile on rhythmic visuomotor coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 40(5), 1849.
- Vasuki, P. R. M., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research*, 342, 112–123.
- Von Euler, C. (1982). Some aspects of speech breathing physiology. In *Speech motor control* (pp. 95–103). Elsevier.
- Warner, H. J., Whalen, D., Harel, D., & Jackson, E. S. (2022). The effect of gap duration on the perception of fluent versus disfluent speech. *Journal of Fluency Disorders*, 105896.
- Werner, R., Fuchs, S., Trouvain, J., & Möbius, B. (2021). Inhalations in speech: Acoustic and physiological characteristics. In *Proc. Interspeech 2021* (pp. 3186–3190).
- Whalen, D. H., Hoquist, C. E., & Sheffert, S. M. (1995). The effects of breath sounds on the perception of synthetic speech. *The Journal of the Acoustical Society of America*, 97(5), 3147–3153.
- Whalen, D. H., & Kinsella-Shaw, J. M. (1997). Exploring the relationship of inspiration duration to utterance duration. *Phonetica*, 54(3–4), 138–152.
- Winkworth, A. L., Davis, P. J., Adams, R. D., & Ellis, E. (1995). Breathing patterns during spontaneous speech. *Journal of Speech, Language, and Hearing Research*, 38(1), 124–144.
- Winkworth, A. L., Davis, P. J., Ellis, E., & Adams, R. D. (1994). Variability and consistency in speech breathing during reading: Lung volumes, speech intensity, and linguistic factors. *Journal of Speech, Language, and Hearing Research*, 37(3), 535–556.
- Włodarczak, M., & Heldner, M. (2016a). Respiratory belts and whistles: A preliminary study of breathing acoustics for turn-taking. In *Interspeech 2016, San Francisco, USA, September 8–12, 2016* (pp. 510–514).
- Włodarczak, M., & Heldner, M. (2016b). Respiratory turn-taking cues. In *Interspeech 2016, San Francisco, USA, September 8–12, 2016* (pp. 1275–1279).
- Włodarczak, M., Heldner, M., & Edlund, J. (2015). Breathing in conversation: An unwritten history. In *2nd european and the 5th nordic symposium on multimodal communication, Tartu, Estonia, August 6–8, 2014* (pp. 107–112).
- Yee, W., Holleran, S., & Jones, M. R. (1994). Sensitivity to event timing in regular and irregular sequences: Influences of musical skill. *Perception & Psychophysics*, 56(4), 461–471.
- Zoefel, B. (2018). Speech entrainment: Rhythmic predictions carried by neural oscillations. *Current Biology*, 28(18), R1102–R1104.
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14.