Synchronized Neural Activity Indexes Engagement with Spoken Stories under Acoustic Masking

Vanessa C. Irsik[1,*], Ingrid S. Johnsrude[1,2], and Björn Herrmann[1,3,4]

[1] Department of Psychology & The Brain and Mind Institute,
The University of Western Ontario, London, ON, Canada N6A 3K7

[2] School of Communication and Speech Disorders,
The University of Western Ontario, London, ON, N6A 5B7, Canada

[3] Rotman Research Institute,
Baycrest, M6A 2E1, Toronto, ON, Canada

[4] Department of Psychology,
University of Toronto, M5S 1A1, Toronto, ON, Canada

* Correspondence concerning this article should be addressed to Vanessa C. Irsik, The Brain and Mind Institute, The University of Western Ontario, London, Ontario, N6A 5B7, Canada. E-mail: virsik@uwo.ca

Acknowledgements:
This research was supported by the Canadian Institutes of Health Research (MOP133450 to I.S. Johnsrude). BH was supported by a BrainsCAN Tier I postdoctoral fellowship (Canada First Research Excellence Fund; CFREF) and the Canada Research Chair program.

Pages: 38

Figures: 6

Abstract: 233

Introduction: 649

Discussion: 1410

Abstract

Older people with hearing problems often experience difficulties understanding speech in the presence of background sound. As a result, they may disengage in social situations, which has been associated with negative psychosocial health outcomes. Measuring listening (dis-)engagement during challenging listening has received little attention thus far. We recruit normal-hearing human adults (both sexes) and investigate how speech intelligibility and engagement during naturalistic story listening is affected by the level of acoustic masking (12-talker babble). In Experiment 1, we observed that word-report scores were above 80% for all but the lowest SNR (-3 dB SNR) we tested, at which performance dropped to 54%. In Experiment 2, we calculated inter-subject correlation (ISC) in electroencephalography (EEG) data to identify dynamic spatial patterns of shared neural activity evoked by the stories. ISC was stronger during story listening compared to rest. The magnitude of ISC was high and stable across all SNRs except for the lowest one, at which it dropped substantially. Comparing ISC and intelligibility directly demonstrated that word-report performance declined more strongly with decreasing SNR compared to ISC. Observing significant ISC despite the presence of background noise suggests that participants were able to remain engaged despite missing segments of the story during especially difficult SNRs. Our work provides a novel approach to observe speech intelligibility and listener engagement using ecologically valid spoken materials, which can be used to investigate (dis)engagement in older adults with hearing impairment.

Keywords: naturalistic stimuli; hearing; speech intelligibility; engagement; inter-subject correlation

Significance Statement

Individuals who frequently experiencing listening challenges may disengage in social situations, which puts them at risk for social isolation and negative health outcomes. Measuring the extent to which listeners engage and disengage during listening when background sound is present may improve our ability to detect people who are at risk of social isolation before it manifests. We utilize ecologically valid spoken stories to investigate how acoustic masking reduces intelligibility and listening engagement, using inter-subject correlation as a neural signature of engagement. We found that healthy listeners continue to engage with listening material even under adverse listening conditions and reduced speech intelligibility, although engagement is less when the level of background sound is high.

Introduction

Many listening environments we encounter contain background masking sounds (Olsen, 1998; Smeds et al., 2015) that can lead to listening challenges for a significant proportion of people aged 50 and older (Frisina and Frisina, 1997; Gordon-Salant, 2006). People who frequently experience listening difficulties may avoid environments with increased levels of background noise, resulting in social isolation (Dawes et al., 2015; Palmer et al., 2016) and negative psychosocial and physical health outcomes (Pichora-Fuller et al., 2015; Wayne and Johnsrude, 2015). However, social isolation is likely subsequent to within-situation disengagement; a coping mechanism to reduce cognitive demand. A person may temporarily "zone out" in conversational situations because continuous listening is too difficult (Heffernan et al., 2016; Herrmann and Johnsrude, 2020a, 2020b). A greater understanding of the listening conditions under which a person may disengage, and development of quantitative measures of within-situation disengagement, may improve our ability to diagnose hearing problems before social isolation manifests.

Two factors are likely critical for listening engagement (Herrmann and Johnsrude, 2020b). First, engagement depends on degree to which a listener judges that, with deliberate deployment of cognitive resources, some of the masked speech will be understandable (Brehm and Self, 1989; Richter, 2013; Wright, 2014; Pichora-Fuller et al., 2016; Herrmann and Johnsrude, 2020b). An individual may disengage if they believe speech comprehension will be impossible (Eckert et al., 2016; Richter, 2016; Peelle, 2018). A second very important factor is motivation. If the individual is motivated to listen, for example, because they find what they are hearing to be enjoyable and rewarding (Matthen, 2016), they may engage in listening despite background sound (Pichora-Fuller et al., 2016; Herrmann and Johnsrude, 2020b). However, typical investigations of speech comprehension involve listening to isolated sentences (Davis and Johnsrude, 2003; Duncan and Aarts, 2006; Obleser et al., 2007) that lack a topical thread, are not very interesting (e.g., He buttoned his shirt), and may not therefore foster volition to listen (i.e., conation; Reitan and Wolfson, 2000). Further, listening engagement also likely develops more slowly, at time scales beyond individual sentences (Mandler and Goodman, 1982). Spoken stories, in contrast, involve event descriptions along a topical

thread, are intrinsically motivating to a listener, and are common in everyday life (Dunlop and Walker, 2013). Utilizing engaging spoken stories under varying levels of acoustic masking may thus provide a fruitful avenue to investigate intelligibility, and listening engagement and disengagement.

Engagement (and disengagement) can be measured behaviorally through assessment of an individual's experience (Busselle and Bilandzic, 2008, 2009; Dmochowski et al., 2014; Kuijpers et al., 2014; Herrmann and Johnsrude, 2020a). However, behavioral assessment of engagement typically requires the listener to retroactively introspect about their experiences (Herrmann and Johnsrude, 2020a). In contrast, neural activity recorded using functional imaging or electroencephalography (EEG) may provide a real-time window on engagement. Naturalistic materials such as a movie or a spoken story lead to synchronized patterns of neural activity across individuals that scale with the degree of engagement (Hasson et al., 2010; Yeshurun et al., 2017; Nastase et al., 2019; Nguyen et al., 2019). The strength of synchrony across individuals, quantified as inter-subject correlation (ISC; Hasson et al., 2004, 2008; Dmochowski et al., 2012, 2014), is stronger when stimuli are captivating or exciting (Hasson et al., 2010; Schmälzle et al., 2015), and is predictive of behavioral measures reflecting engagement (Dmochowski et al., 2014). Conversely, ISC is reduced when individuals do not attend to naturalistic materials (Ki et al., 2016) or when stimuli are unstructured or temporally scrambled (Hasson et al., 2008b; Dmochowski et al., 2012).

In the current study, we utilize spoken stories to investigate how challenging listening situations affect speech intelligibility and story engagement. In Experiment 1, we assess intelligibility of spoken stories presented with different levels of 12-talker babble masker. In Experiment 2, we record EEG in another group of individuals while they listen to the same stories used in Experiment 1 and investigate ISC as a neural signature of engagement.

Experiment 1: Examining story intelligibility and engagement

In Experiment 1, we investigate the extent to which participants understand spoken stories when they are masked with 12-talker babble at different signal-to-noise ratios. We further assess overall listening

engagement behaviorally using a modified narrative absorption scale (Kuijpers et al., 2014) adapted for use with spoken stories (Herrmann and Johnsrude, 2020). We will use speech intelligibility ratings from Experiment 1 to analyze with data from Experiment 2, in which we investigate neural signatures of story engagement, to assess the extent to which story engagement depends on speech intelligibility.

Materials and Methods

Participants

Eighty-two individuals (mean: 28.8 years; age-range: 18-36 years; 51 males 31 females) without self-reported hearing loss, neurological or psychiatric disorders, participated in Experiment 1. All participants were recruited from the Amazon Mechanical Turk online participant pool (MTurk; https://www.mturk.com/) via the participant sourcing platform Cloud Research (previously TurkPrime; Litman et al., 2017). Participants provided informed consent and the study protocol was approved by the University of Western Ontario's Non-medical Research Ethics Board (REB #112574). Each individual received financial compensation of $6 USD following completion of the study ($10 USD hourly rate).

Online research can be subject to increased levels of random responders as experimenters have limited control over the testing environment. However, the evidence indicates that online studies generally replicate findings of in-person data collection (Gosling et al., 2004; Buhrmester et al., 2011; Mason and Suri, 2012; Berinsky et al., 2014; Thomas and Clifford, 2017; Buchanan and Scofield, 2018). Twenty-one additional individuals participated in the study but were not included, due to a technical error during data recording (N=2); reporting hearing aid usage or constant ringing in at least one ear (N=9); not wearing headphones during the study (N=3); or submitting random one-word answers to all questions on the intelligibility task (N=7).

Acoustic stimulation and procedure

Four stories were selected from the story-telling podcast The Moth (https://themoth.org): *Reach for the Stars One Small Step at a Time* (by Richard Garriott, ~13 min), *The Bounds of Comedy* (by Colm O'Regan, ~10 min), *Nacho Challenge* (by Omar Qureshi, ~11 min), and *Discussing Family Trees in School Can Be Dangerous* (by Paul Nurse, ~10 min). Each story had 12-talker babble noise added as a masker. Babble noise was derived using the masker materials from the Revised Speech in Noise (R-SPIN) test (Bilger, 1984). Individual 5-s babble noise snippets were randomly selected from a total set of 100 and concatenated to equal the length of each story. The signal-to-noise ratio (SNR) was pseudo-randomly varied approximately every 30 to 33 s throughout each story (32, 30, 33, and 30-s, respectively, for the four stories). Five SNRs were chosen: clear, +12, +7, +2, and -3 dB signal-to-noise ratio (SNR) relative to a 12-talker babble masker. SNR was manipulated by adjusting the dB level of both the story and masker. This ensured that the overall sound level remained constant throughout a story and was similar for all stories. Three versions of SNR condition order were generated for each story in order to ensure that specific parts of a story were not confounded with a specific SNR. Within each version, SNR was varied pseudo-randomly such that a particular SNR could not be heard twice in succession. Four 30, 33, or 30-s segments per SNR were presented for stories by O'Regan, Qureshi, and Nurse, respectively, and five 32-s segments per SNR were presented for the longest story by Garriott.

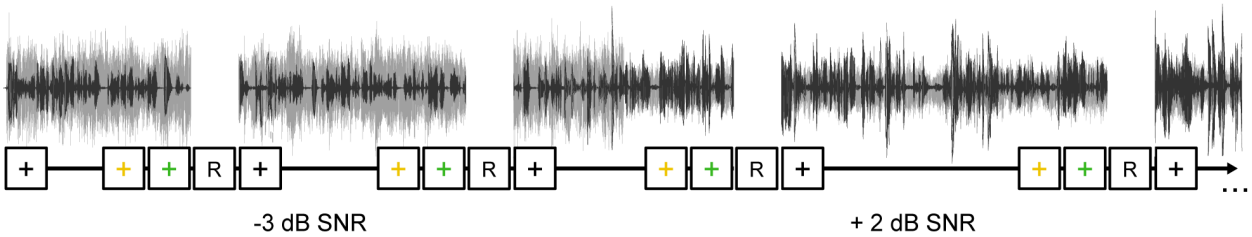

Figure 1. Schematic representation of the speech intelligibility task. Participants listened to a spoken story (black) masked with different levels of 12-talker babble noise (grey). A fixation cross was displayed on the computer screen throughout the study and changed colors to communicate which parts of the spoken story participants would need to recall. The fixation cross turned yellow 2-s prior to the beginning of a test phrase/sentence, cueing the participant to prepare for intelligibility testing. The fixation cross turned green during the phrase/sentence participants would be asked to report back. The story stopped with the offset of

the test phrase/sentence, at which point participants would report back the phrase/sentence. The story resumed once a response was submitted.

For each story, phrases/sentences ranging from 4 to 8 words (range of durations: 0.62–3.6 s) were selected for intelligibility (word report) testing. These test phrases/sentences did not occur during the transition period from one SNR to the next (for approximately 5-s before, and 1-s after, the SNR transition). Four phrases per 30–33-s segment were selected, resulting in 100 phrases for the story by Garriott, and 80 for each of the other three stories. Two of the four selected phrases per 30–33-s segment were used as one intelligibility test set, whereas the other two of the four selected phrases were used as a second intelligibility test set (5 or 4 segments × 5 SNRs × 2 phrases/sentences × 2 intelligibility test sets = 100 or 80 phrases). Having two test sets ensured that the observed intelligibility effects were not confounded by specific phrases/sentences.

The experiment was conducted online, using custom written JavaScript/html and jsPsych code (Version 6.1.0, a high-level JavaScript library used for precise stimulus control; de Leeuw, 2015). The experiment code was stored at an online Gitlab repository (https://gitlab.pavlovia.org) and hosted via Pavlovia (https://pavlovia.org/). During the main experiment, each participant listened to one pseudo-randomly selected story (Garriott: N = 21; O'Regan: N = 22; Qureshi: N = 20; Nurse: N = 19). SNR order and intelligibility test set were randomly assigned. A black fixation cross was presented at the center of the screen throughout the story. The fixation cross turned yellow two seconds prior to the beginning of a test phrase/sentence, cueing the participant to prepare for intelligibility testing (see Figure 1). The fixation cross then turned green for the duration of the test phrase in the story, indicating to the participant the phrase they would be asked to report back. The story stopped with the offset of the test phrase, and an input text box appeared on the screen. Participants were asked to type their answer into the text box, after which the story resumed from the beginning of the sentence most recently heard (allowing for story continuation). After the story ended, participants answered questions that assessed comprehension and rated statements about their story listening experiences (see Table 1).

In order to familiarize participants with the intelligibility task, a brief practice block was presented prior to the main experiment. Participants heard a ~3-min story (a shortened version of *A Shoulder Bag to Cry On* by Laura Zimmerman), without added babble noise, and performed 12 trials of the intelligibility task (2 trials per 30-s segment).

Online research quality assurance measures

Participants completed two initial listening tasks at the very beginning of the online session. First, participants listened to a 15-s stream of pink noise normalized to the same root-mean-square amplitude as the stories and were instructed to adjust their volume to a comfortable listening level. Participants had the option to replay the noise if they needed additional time to adjust their volume. This task ensured that participants had an opportunity to adjust their volume to a comfortable level prior to the intelligibility task, after which they were instructed to not make further adjustments.

Second, participants completed a headphone-check procedure to determine whether participants were wearing headphones as instructed (cf. Woods et al., 2017). Participants performed a tone discrimination task (6 trials; ~2 minutes total duration), in which they determined which of three consecutive 200-Hz sine tones was the quietest. The three tones differed such that one was presented at the comfortable listening level, one at –6 dB relative to the other two tones, and one at the comfortable listening level with a 180° phase difference between the left and right headphone channels (anti-phase tone). This task is straightforward over headphones, but difficult over loudspeakers, because the pressure waves generated from an anti-phase tone interfere if heard through loudspeakers (Woods et al 2017). If they were listening through loudspeakers, they would likely falsely select the anti-phase tone as the quietest tone. No participants were excluded solely on the basis of performance on this test, however, it did provide a metric which could flag potentially non-compliant online subjects.

Assessment of intelligibility

We calculated the proportion of correctly reported words for each SNR (Clear, +12, +7, +2, -3), across the three versions of the four stories. Different or omitted words were counted as errors, but minor misspellings and incorrect grammatical number (singular vs. plural) were not. Word-report performance was assessed using a repeated-measures analysis of variance (rmANOVA) with SNR (Clear, +12, +7, +2, -3 dB SNR) as a within-subjects factor. In addition, we fit a quadratic function to the proportion of correctly reported words as a function of SNR (treating them as equally spaced) for each participant. The estimated linear and quadratic coefficients from the fit were tested against zero using one-sample *t*-tests. A significant difference from zero would indicate that word-report performance changed linearly and/or quadratically as a function of SNR.

Assessment of story comprehension

Story comprehension was assessed using 8 statements, which either correctly or incorrectly described an element from the story the participant heard. Participants were asked to categorize each statement as true or false. Comprehension performance was calculated as the proportion of statements categorized correctly. Comprehension was statistically examined using a one-sample *t*-test which tested scores against chance-level performance of 0.5.

Assessment of story engagement, enjoyment, and motivation

Following the assessment of story comprehension, we also assessed how much participants were engaged with the story, how much they enjoyed the story, and how motivated they were to listen. In order to assess story engagement, we utilized a narrative absorption scale (NAS; Kuijpers et al., 2014) adapted previously for spoken stories (Herrmann and Johnsrude, 2020a). The NAS contains statements along four dimensions (attention, emotional engagement, mental imagery, and transportation; Kuijpers et al., 2014; Herrmann and Johnsrude, 2020), which participants rated on a 7-point scale, where ‘1’ referred to completely disagree and ‘7’ referred to completely agree (Table 1). Participants also rated statements about enjoyment and listening

motivation on the same 7-point scale (Table 1). Rating scores for each statement were averaged separately for narrative absorption (NAS) and enjoyment. Ratings for NAS and enjoyment were statistically examined in separate one-sample *t*-tests which tested ratings against a neutral response (test value: 4). Given that motivation was assessed using a single question (Table 1), the data remain ordinal rather than continuous. Motivation was therefore assessed using a one-sample Wilcoxon signed-rank test, against a neutral response (test value: 4).

Table 1. Statements of the narrative absorption scale (NAS), enjoyment, and motivation (Herrmann and Johnsrude, 2020a)

Dimension	Statement
Attention (NAS)	When I finished listening I was surprised to see that time had gone by so fast.
Attention (NAS)	When I was listening I was focused on what happened in the story.
Attention (NAS)	I felt absorbed in the story.
Attention (NAS)	The story gripped me in such a way that I could close myself off for things that were happening around me.
Attention (NAS)	I was listening in such a concentrated way that I had forgotten the world around me.
Emotional engagement (NAS)	When I listened to the story I could imagine what it must be like to be in the shoes of the main character(s).
Emotional engagement (NAS)	I felt sympathy for the main character(s).
Emotional engagement (NAS)	I felt connected with the main character(s) of the story.
Emotional engagement (NAS)	I felt how the main character(s) was/were feeling.
Emotional engagement (NAS)	I felt for what happened in the story.
Mental imagery (NAS)	When I was listening to the story I had an image of the main character(s) in mind.
Mental imagery (NAS)	When I was listening to the story I could see the situations happening in the story being played out before my eyes.
Mental imagery (NAS)	I could imagine what the world in which the story took place looked like.
Transportation (NAS)	When I was listening to the story it sometimes seemed as if I were in the story world too.
Transportation (NAS)	When listening to the story there were moments in which I felt that the story world overlapped with my own world.
Transportation (NAS)	The world of the story sometimes felt closer to me than the world around me.
Transportation (NAS)	When I was finished with listening to the story it felt like I had taken a trip to the world of the story.
Transportation (NAS)	Because all of my attention went into the story, I sometimes felt as if I could not exist separate from the story.
Enjoyment	I thought it was an exciting story.

Enjoyment	I thought it was an enthralling story.
Enjoyment	I listened to the story with great interest.
Enjoyment	I thought the story was beautiful.
Enjoyment	I thought the story was presented well.
Motivation	I was motivated to listen to the story.

Assessment of the relationship between measures of engagement, enjoyment, and motivation

To characterize the relationship between behavioral ratings of absorption (NAS), enjoyment, motivation, and story comprehension, we calculated correlations (Pearson) between pairs of behavioral measures (6 pairs).

Experimental design and statistical analysis

Statistical analyses were conducted using IBM SPSS Statistics for Windows (v24) and MATLAB. Details of the specific variables and statistical tests for each analysis can be found in analysis subsections for each measure. In general, effects were examined either using a rmANOVA, paired-samples *t*-tests, or one-sample *t*-tests. Behavioral ratings that were considered were ordinal, not continuous, were analyzed using non-parametric tests for either one sample (Wilcoxon signed rank test) or independent sample comparisons (Wilcoxon rank sum). Significant effects were followed up using *t*-tests, with multiple comparisons corrected using the false discovery rate (FDR; Benjamini and Hochberg, 2016). FDR corrected p-values are reported as p_{FDR}. Effect sizes are reported as partial eta squared (η^2_p) for rmANOVAs and $r_{equivalent}$ (r_e; Rosenthal and Rubin, 2003), for *t*-tests. Greenhouse-Geisser corrected p-values are reported when sphericity assumptions have not been met (reported as p_{GG}). This experiment was not preregistered. Data are available upon reasonable request.

Results and Discussion

Story comprehension, motivation, and enjoyment are high despite speech masking

Performance on comprehension questions was significantly above chance (M = .80, se = .02; t_{81} = 14.79, p = 1 × 10^{-24}, r_e = .85), which suggests participants were able to grasp details from each story despite the varying SNRs and performing the speech intelligibility task while listening to each story.

We also examined participants' level of story engagement by testing mean ratings for NAS, enjoyment, and motivation against a rating of 4 (neutral response). Scores on the NAS were not significantly different from a neutral response (p = .832), but enjoyment (t_{81} = 2.87, p = .005, r_e = .3) and motivation (V = 2103, p = 5 × 10^{-11}) ratings were higher than the neutral point (see Figure 2a). This suggests that although participants were ambivalent regarding whether they were fully immersed in the story they heard, they appeared to enjoy listening to the story and were motivated to do so. These findings are somewhat inconsistent with previous research on story engagement (Herrmann and Johnsrude, 2020a), which found engagement to be high and largely unaffected by moderate noise levels. The presence of the intelligibility task during story listening may have altered the listening experience such that engagement was reduced to enable detail-oriented listening to ensure participants could report back specific words when prompted.

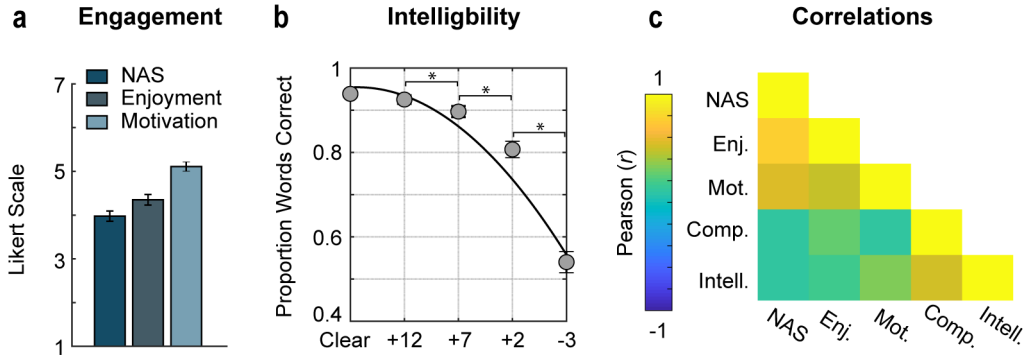

Figure 2. Behavioral measures of story comprehension, engagement, and intelligibility. (a) Mean ratings of the statements comprising the narrative absorption scale (NAS), and those assessing enjoyment and motivation. Ratings were provided on a 7-point Likert scale, where '1' refers to completely disagree and '7' refers to completely agree. (b) Mean proportion of correctly reported words plotted as a function of SNR (clear, +12, +7, +2, -3 dB). The black line corresponds to a quadratic function fit to the proportion of correct words reported. (c) Correlation matrix depicting the correlation (Pearson) between behavioral

measures of story comprehension, engagement, and intelligibility. Error bars reflect the standard error of the mean. *p < 0.05.

Story intelligibility declines quadratically as SNR decreases

For the intelligibility task, the proportion of correctly reported words declined with decreasing SNR ($F_{4,324}$ = 176.5, $p = 2.98 \times 10^{-38}$, $\eta^2_p = .69$; see Figure 2b), as predicted (Bronkhorst, 2000; Brungart, 2001; Brungart et al., 2001; Duncan and Aarts, 2006). We further characterized the effect of SNR on story intelligibility by fitting a quadratic function to the word-report data. Both the linear (t_{81} = -15.07, $p = 3.28 \times 10^{-25}$, $r_e = .86$) and the quadratic coefficient (t_{81} = -13.65, $p = 1.05 \times 10^{-22}$, $r_e = .83$) were significantly smaller than zero, suggesting that intelligibility was relatively stable at most SNRs, and then fell off quadratically: intelligibility declined by only about 4 % between clear speech and +7dB SNR; a further 10% between +7 and +2 dB SNR, and 26% between +2 and -3 dB SNR. Comparing intelligibility at successive pairs of SNRs, in order to identify the SNRs at which a significant drop in performance occurred, revealed that intelligibility was comparable between the Clear and +12 dB SNR conditions ($p_{FDR} = .077$), but declined from +12 to +7 dB SNR ($t_{81} = 3.33$, $p_{FDR} = .002$, $r_e = .35$), +7 to +2 dB SNR ($t_{81} = 5.7$, $p_{FDR} = 3.71 \times 10^{-7}$, $r_e = .54$), and +2 to -3 dB SNR ($t_{81} = 14.94$, $p_{FDR} = 2.22 \times 10^{-24}$, $r_e = .86$).

Relating intelligibility and behavioral measures of engagement and comprehension

In an effort to quantify how intelligibility may relate to engagement and comprehension measures, we calculated correlations between intelligibility (averaged across SNRs), NAS, motivation, enjoyment, and comprehension scores (see Figure 2c). We found that intelligibility was significantly correlated with motivation ($r_{80} = .33$, $p_{FDR} = .004$) and comprehension scores ($r_{80} = .50$, $p_{FDR} = 6.7 \times 10^{-6}$), such that better word-report performance was associated with higher levels of motivation to listen and better comprehension of the stories. We did not find a significant correlation between intelligibility and NAS ($r_{82} = .10$, $p_{FDR} = .381$) or intelligibility and enjoyment scores ($r_{80} = .16$, $p_{FDR} = .223$), suggesting that enjoyment/engagement with the story and being able to report the specific words spoken during the story may be independent (cf. Herrmann and Johnsrude, 2020).

We also found a significant correlation between NAS and enjoyment (r_{80} = .77, p_{FDR} = 3.9 ×10^{-16}), and NAS and motivation (r_{80} = .55, p_{FDR} = 3.4 ×10^{-6}), suggesting that higher levels of engagement were associated with greater enjoyment of the stories and motivation to listen. We also found a significant relationship between enjoyment and motivation (r_{80} = .48, p_{FDR} = 1.6 ×10^{-5}), and enjoyment and comprehension (r_{80} = .26, p_{FDR} = .029). This indicates that enjoyment was associated with greater motivation to listen and better comprehension. No relationship was observed between NAS and comprehension (r_{80} = .11, p_{FDR} = .381), or between motivation and comprehension (r_{80} = .12, p_{FDR} = .367).

The results of Experiment 1 suggest that participants were able to maintain a relatively high level of story comprehension and were generally motivated and enjoyed listening to the stories despite the presence of the babble masker. Further, speech intelligibility was relatively good at all SNRs except the lowest. The intelligibilty and engagement data obtained in this experiment will be used to interpret the results of Experiment 2, which investigates neural signatures of engagement.

Experiment 2: Examining inter-subject correlation during story listening

In Experiment 2, we investigate how neural signatures of engagement with spoken stories are affected by 12-talker background babble noise at different SNRs. We use inter-subject correlation (ISC), a neural measure of synchronized neural activity across participants, which is thought to be driven by engagement and shared experience (Hasson et al., 2010; Yeshurun et al., 2017; Nastase et al., 2019; Nguyen et al., 2019), to identify and characterize different neural components of engagement (Hasson et al., 2008; Dmochowski et al., 2012).

Materials and Methods

Participants

Thirty-nine individuals (mean: 20.3 years; age-range: 18-32 years; 19 males 20 females) without hearing loss, neurological issues, or psychiatric disorders participated in Experiment 2. All participants were recruited from Western University or the surrounding community of London, Canada. All participants provided written informed consent and were financially compensated with $10 CAD per hour. The study was conducted in accordance with the Declaration of Helsinki, the Canadian Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2-2014), and approved by the local Health Sciences Research Ethics Board of the University of Western Ontario (REB #112015). Five additional individuals participated but were not included either due to an issue with sound delivery (N =2), or a technical error during data recording (N=3).

Acoustic stimulation and procedure

The experiment was conducted in a single-walled sound-attenuating booth (Eckel Industries). Sounds were delivered through Sennheiser (HD 25 Light) headphones at a comfortable listening level, using a Focusrite Scarlett 2i4 external soundcard controlled by a PC (Windows 10) and Psychtoolbox (Version 3) in MATLAB (R2017b).

Stimuli were the same four stories used Experiment 1, with the same SNRs (Clear, +12, +7, +2, -3) and the same three versions of SNR condition order per story. EEG was recorded while participants listened to each of the four stories consecutively (see Figure 3) Story version and story order were counterbalanced across participants. After each story, comprehension, story engagement, motivation, and enjoyment were assessed, and these data analyzed, as in Experiment 1. After listening to all four stories, participants completed a resting block, where they sat quietly for six minutes with their eyes open while EEG was recorded. This block was used as a baseline for ISC analyses of story listening.

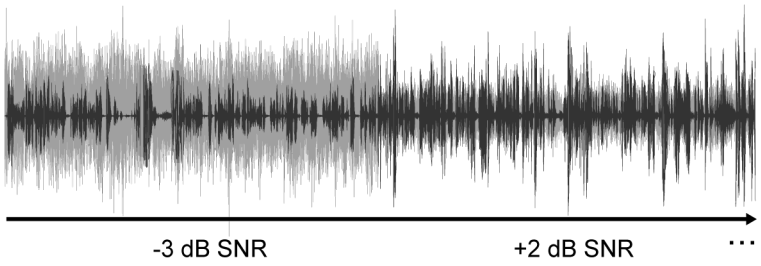

Figure 3. Stimulus design for Experiment 2. Participants listened to four consecutive spoken stories (black) masked with varied levels of 12-talker babble noise (grey); two of the SNRs are shown here. Each SNR lasts for ~30-33 sec (see text for details).

EEG recording and preprocessing

EEG was recorded from 64 active electrodes (Ag/AgCl) placed on the scalp using an electrode cap and two additional electrodes placed on both mastoids, with spacing on the scalp according to the 10/20 system (Biosemi ActiveTwo system). During data recording, all electrodes were referenced to a feedback loop of two electrodes, a common mode sense (CMS) active electrode and a driven passive electrode (see www.biosemi.com/faq/cms&drl.htm). EEG was recorded at 1024 Hz with an online low-pass filter of 208 Hz to focus on cortical sources.

All pre-processing was carried out offline using MATLAB software, the Fieldtrip toolbox (Oostenveld et al., 2011), and custom scripts. EEG data were re-referenced to the average of the signal from both mastoids. EEG data were then notch-filtered at 60 Hz to attenuate line noise, and then high-pass (0.5 Hz, 3429 points, Hann window) and low-pass filtered (22 Hz, 211 points, Kaiser window). EEG data were also downsampled to 256 Hz. Artifacts due to eye movements (saccades, blinks) and muscle activity were removed using independent components analysis (Makeig et al., 1996). To further exclude additional artifacts from subsequent analyses, data segments in which the EEG signal changed by more than 80 µV within a 0.2-s period in any channel were set to 0 µV (cf. Dmochowski et al., 2012, 2014; Cohen and Parra, 2016).

Inter-subject correlation

We quantified inter-subject correlation (ISC) using correlated component analysis (CCA; Dmochowski et al., 2012; Parra et al., 2019), a signal decomposition method that identifies a set of electrode weights (i.e., spatial filters) yielding a linear combination of electrode activation (components) maximally correlated across participants. CCA can uncover patterns of neural activity that would not be possible with an electrode-to-electrode correlation method (Dmochowski et al., 2012; Cohen and Parra, 2016; Ki et al., 2016). For mathematical details of the method see (Parra et al., 2019), for available MATLAB scripts see (https://www.parralab.org/isc/), and for the adapted version of the MATLAB scripts used in the current study see (https://osf.io/tv7kg/). We calculated the components for each story individually to enable leave-one-out analyses (see below), such that spatial filter calculations are independent from subsequent analyses (see also Crosse et al., 2016; Herrmann et al., 2018; Broderick et al., 2019). In line with previous work, we restrict our analysis to the three components with highest overall ISC for each story (cf. Dmochowski et al., 2012, 2014; Cohen and Parra, 2016; Ki et al., 2016), because they show consistent spatial projections onto the scalp across stories (see Figure 4, left panel).

Separately for each of the three components (i.e., spatial filters), component weights for one story were multiplied with participants' EEG data for the other three stories. Spatially filtered data for the three stories were then concatenated, leading to one response time course which reflects the component's underlying sources (cf. Crosse et al., 2016; Broderick et al., 2019). This was repeated for all 4 stories, resulting in four spatially filtered EEG time courses for each component, per participant (4 concatenated datasets × 3 components =12 spatially filtered time courses). All ISC analyses described in subsequent paragraphs were calculated after using this leave-one-out approach. ISC values were subsequently averaged across the four leave-one out iterations for each component.

For an overall ISC analysis independent of SNR conditions, each participant's time course was correlated with the time course of each of the other participants, resulting in N-1 correlations per participant. ISC for a unique participant was calculated as the mean across these N-1 correlation values. In order to test

whether story listening led to higher ISC than resting state activity in these 'story-listening' networks, the EEG data from the resting block were projected through each story's spatial filters using the leave one-out approach described above. Resting ISC was calculated similarly to story-listening ISC. Mean story-listening ISC was compared to resting ISC using a paired sample t-test, separately for each of the three components. Note that results from using surrogate-data testing (Lancaster et al., 2018), by circularly shifting response time courses of participants, mirrored results from the resting-state contrasts.

To examine ISC as a function of SNR, we concatenated EEG segments with the same SNR and spatially filtered using the above leave-one-out approach. This resulted in a spatially-filtered EEG time series for each participant, SNR condition, and spatial-filter component. Separately for each SNR, each participant's time course was correlated with the time course of each of the other participants who heard the same story version (i.e., same order of SNR conditions), and subsequently averaged across the resulting N-1 correlations for each subject. We analyzed ISC using rmANOVAs with SNR (Clear, +12, +7, +2, -3 dB) as one within-subjects factor, separately for each of the three components. We also fit a quadratic function to ISC values, separately for each participant, to assess whether the effect of SNR on ISC could be quantified as linear or quadratic. Linear and quadratic coefficients were against zero using a one-sample *t*-test.

Assessment of the relationship between inter-subject correlation and behavioral measures

To characterize the relationship between inter-subject correlation and behavioral ratings of absorption (NAS), enjoyment, motivation, and story comprehension, we calculated correlations (Pearson) between overall ISC for each of the three components and these behavioral measures.

Comparison of ISC and speech intelligibility

In order to directly compare word-report performance in Experiment 1 with inter-subject correlation during story listening in Experiment 2, we transformed both word-report scores and ISC values to z-scores. We restricted our analysis to the component with the highest ISC (Component 1, see Figure 4,

right panel). We also fit a quadratic function to z-scored word-report and ISC values, separately for each participant. To assess the extent to which linear and quadratic coefficients differed for word-report and ISC data, we conducted two separate independent t-tests, with experiment measure (intelligibility [Exp 1], ISC [Exp 2]) as the grouping variable. To capture how each non-clear SNR condition (+12, +7, +2, -3 dB) changed relative to the clear condition, we calculated difference scores between each SNR level and the clear condition, separately for word-report scores and ISC values. Resulting values were submitted to a mixed design ANOVA with measure (intelligibility, ISC) as the between-subjects variable and SNR (+12, +7, +2, -3 dB) as the within-subjects variable.

Results and Discussion

Comprehension and engagement are high despite story masking

For story comprehension, similar to Experiment 1, participants' performance was significantly higher than chance level performance (M = .86, se = .01; t_{38} = 28.77, p = 2.1 × 10^{-27}, r_e =.98), indicating that they attended and grasped details from the stories despite the varying babble masker.

Table 2. Mean engagement (NAS), enjoyment, and motivation ratings for Experiment 1 and 2

	NAS	Enjoyment	Motivation
	M ± s.e.	M ± s.e.	M ± s.e.
Experiment 1	3.98 ± 0.12	4.35 ± 0.12	5.11 ± 0.11
Experiment 2	4.5 ± 0.14	5.23 ± 0.17	5.59 ± 0.18

We examined participants' level of story engagement by testing mean ratings for each engagement category (NAS, enjoyment, motivation) against a rating of 4 (neutral response). In contrast to Experiment 1, scores on the NAS (t_{38} = 3.62, p = .001, r_e = .51), enjoyment (t_{38} =7.28, p = 1.02 × 10^{-8}, r_e = .76), and motivation (t_{38} = 8.9, p = 7.6 × 10^{-11}, r_e = .82), were all significantly higher than a neutral response (see Table 2). This indicates that participants were engaged, enjoyed the stories, and felt motivated to listen. As

a follow-up, we directly compared behavioral ratings between Experiment 1 and 2 using either separate independent t-tests (NAS, enjoyment) or a Wilcoxon rank sum test (motivation), with Experiment as the grouping factor (Experiment 1, Experiment 2). We found that narrative absorption (t_{119} = 2.7, p_{FDR} = .008, r_e= .24), enjoyment (t_{119} = 4.14, p_{FDR} = 2 × 10^{-4}, r_e= .35), and motivation (Z = 2.77, p_{FDR} = .008) were higher in Experiment 2 compared to Experiment 1 (Table 2), perhaps because Experiment 2 was in person, whereas Experiment 1 was on line.

Inter-subject correlation is higher during active listening compared to rest

Figure 4 (left panel) shows the scalp projections of the first three components (Parra et al., 2005): topographical distributions are consistent with previous work (Dmochowski et al., 2012, 2014; Cohen and Parra, 2016; Ki et al., 2016). The first component has a fronto-central scalp distribution, which is consistent with a source in auditory cortex (Näätänen and Picton, 1987; Picton et al., 2003), but may also reflect contributions from frontal cortex. The second component has a parietal scalp distribution, suggesting a source in parietal cortex or posterior auditory cortex. The third component shows a parieto-occipital peak with frontal polarity reversal. For each component we observed ISC that, while appearing small, were within an expected range reported previously (Dmochowski et al., 2012, 2014; Ki et al., 2016). ISC was stronger during story listening compared to rest for all three components (Component 1: t38 = 31.6, p = 7.3 × 10-29, re = .98; Component 2: t38 =19.6, p = 1.7 × 10-21, re = .95; Component 3: t38 =16.7, p = 4.6 × 10-19, re = .94; see, right panel). ISC for surrogate data (generated by circularly shifting participants' time courses during the stories) also yielded significantly lower ISC values compared to un-shifted (story listening) data for each component. That the ISC was stronger during story listening compared to rest for the three strongest spatial components suggests that it is driven by patterns of shared neural activity evoked by the stories.

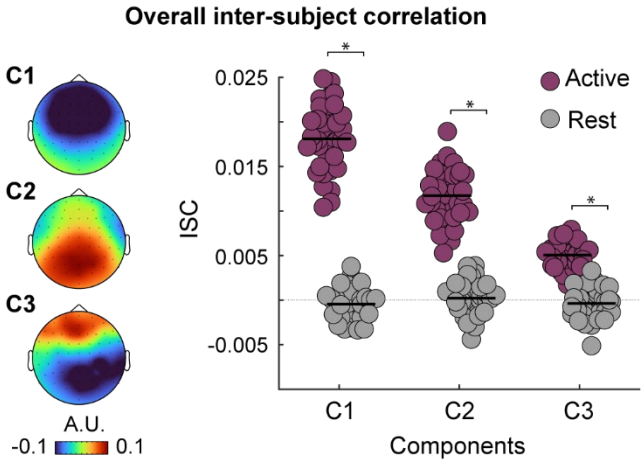

Figure 4. Overall inter-subject correlation. Left panel: Scalp projections of the three most correlated components. Right panel: Overall ISC is plotted as a function of component number (Component 1, Component 2, Component 3) and listening condition (active: listening to stories; rest: relaxing with eyes open). Dots represent data points for individual participants. Black solid lines indicate means. The thin, dashed line marks ISC of zero. *p < 0.05.

Inter-subject correlation declines quadratically as SNR decreases

In order to examine whether ISC is affected by SNR, we conducted an rmANOVA. Consistent with behavioral performance from Experiment 1, we observed that ISC declined with decreasing SNR for all three components (Component 1: $F_{4,152} = 20.64$, $p = 1.3 \times 10^{-13}$, $\eta^2_p = .35$; Component 2: $F_{4,152} = 18.66$, $p = 1.7 \times 10^{-12}$, $\eta^2_p = .33$; Component 3: $F_{4,152} = 3.22$, $p = .014$, $\eta^2_p = .08$; Figure 5a). We further characterized the effect of SNR on ISC by fitting a quadratic function to ISC values across SNRs and analyzing the resulting coefficients. For Component 1, the linear ($t_{38} = -6.9$, $p = 3.2 \times 10^{-8}$, $r_e = .75$) and quadratic ($t_{38} = -4.86$, $p = 2 \times 10^{-5}$, $r_e = .62$) coefficients were significant (Figure 5b). For Component 2, we also found a significant linear ($t_{38} = -7.14$, $p = 1.6 \times 10^{-8}$, $r_e = .76$) and quadratic ($t_{38} = -5.53$, $p = 2.5 \times 10^{-6}$, $r_e = .67$) trend. Finally, we only found a significant linear trend for Component 3 ($t_{38} = -3.2$, $p = .003$, $r_e = .46$). The presence of quadratic trends in Components 1 and 2 suggests that ISC remained fairly stable at higher SNRs, but decreased most substantially at the lowest SNRs, consistent with the intelligibility data.

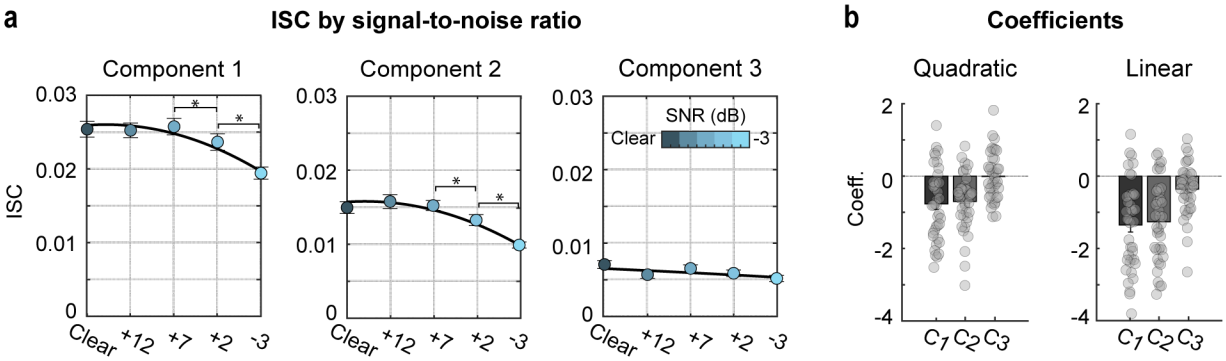

Figure 5. Inter-subject correlation as a function of SNR. (a) Mean ISC plotted as a function of SNR (clear, +12,+7,+2,-3 dB) and component (Component 1, Component 2, Component 3). (b) Mean quadratic (left panel) and linear coefficients (right panel) are plotted as a function of component. Grey dots represent data for individual participants. Error bars reflect the standard error of the mean. *p < 0.05.

We then compared ISC at successive pairs of SNRs to identify the point at which a significant drop in ISC occurred. For Component 1, ISC strength did not differ between Clear and +12 dB SNR (p_{FDR} = .859), or between +12 and +7 dB SNR (p_{FDR} = .703), but decreased significantly from +7 to +2 dB SNR (t_{38} = 2.7, p_{FDR} = .02, r_e = .4), and from +2 to -3 dB SNR (t_{38} = 5.7, p_{FDR} = 5.2×10^{-6}, r_e = .68). An identical pattern was observed for Component 2, such that ISC strength did not differ between Clear and +12 dB SNR (p_{FDR} = .452), or between +12 and +7 dB SNR (p_{FDR} = .498), but decreased significantly from +7 to +2 dB SNR (t_{38} = 2.54, p_{FDR} = .03, r_e = .38), and from +2 to -3 dB SNR (t_{38} = 4.72, p_{FDR} = .0001, r_e = .61). No differences were observed between successive SNRs for Component 3 after correcting for multiple comparisons (p_{FDR} >.14). Neural ISC is thought to index the degree of engagement with naturalistic materials (Hasson et al., 2010; Yeshurun et al., 2017; Nastase et al., 2019; Nguyen et al., 2019), and our data may thus suggest that a listener remains engaged during naturalistic listening when about 10% of words are heard incorrectly (+7 dB SNR) but starts disengaging when approximately 20% of words are missed (+2 dB SNR).

Finally, we compared the SNR condition with the lowest ISC (-3 dB SNR) to ISC during rest to determine whether the reduced synchronization observed for the most challenging listening condition was still greater than average synchronization during rest. We observed stronger ISC during the -3 dB SNR

condition for Component 1 (t_{38} = -21.44, p_{FDR} = 2.4 × 10^{-22}, r_e = .96), Component 2 (t_{38} = -17.74, p_{FDR} = 8.4 × 10^{-20}, r_e = .94), and Component 3 (t_{38} =-11.49, p_{FDR} = 6.2 × 10^{-14}, r_e =.88). Therefore, while engagement was significantly reduced at -3 dB relative to the other less challenging SNRs, observing stronger ISC at -3 dB compared to rest may suggest participants were still somewhat engaged. These results are consistent with previous behavioral work suggesting that listeners continue to engage with spoken stories despite the presence of masking and reduced speech intelligibility (Herrmann and Johnsrude, 2020.

Relating inter-subject correlation and behavioral measures of engagement and comprehension

Correlations between overall ISC for each component (Component 1, Component 2, Component 3) and behavioral ratings of absorption (NAS), enjoyment, motivation, and story comprehension were not significant after correcting for multiple comparisons (r_{37} < .3, p_{FDR} > .22).

Consistent with Experiment 1, we found a significant correlation between NAS and enjoyment (r_{37} = .81, p_{FDR} = 6.6 × 10^{-9}), and NAS and motivation (r_{37} = .70, p_{FDR} = 4 × 10^{-6}). Higher levels of absorption were associated with greater enjoyment of the stories and motivation to listen. We also found a significant relationship between enjoyment and motivation (r_{37} = .86, p_{FDR} = 5.8 × 10^{-11}): enjoyment was associated with greater motivation to listen. NAS did not correlate with story comprehension (r_{37} =.21, p_{FDR} = .192), and no other significant correlations were found (r_{37} <.23, p_{FDR} >.36).

Assessing the relative effect of SNR on intelligibility and inter-subject correlation

We directly compared the relative effect of SNR on intelligibility and inter-subject correlation by fitting a quadratic function to mean-normalized (z-score) word-report scores and inter-subject correlation values for the component with the strongest ISC (Component 1; see Figure 6a), and comparing the resulting coefficients. We observed no difference between quadratic coefficients for word-report scores and ISC for Component 1 (p = .21), suggesting similar quadratic patterns are present in both measures. We utilized a Bayesian independent t-test to determine the likelihood that the quadratic coefficients are indeed equal, but found weak evidence for this (B_{10} = 0.42). Additionally, we observed larger linear coefficients for the

intelligibility task relative to ISC ($t_{119} = 4.71$, $p = 7 \times 10^{-6}$), suggesting that speech intelligibility declined more strongly with SNR than neural story engagement measured as ISC (Figure 6b.)

In order to capture how SNR may differentially affect intelligibility and ISC we calculated difference scores, separately for mean-normalized word-report and ISC, between each SNR and the clear condition and analyzed the resulting z-scores using an ANOVA. The relative change from clear for both word-report and ISC decreased significantly as SNR decreased (effect of SNR: $F_{3,357} = 197.5$, $p = 1.8 \times 10^{-75}$), but the decline from 2 to -3 dB SNR was larger for word report than for ISC (Task × SNR interaction: $F_{3,357} = 5.99$, $p = .001$, $\eta^2_p = .05$; WS contrast: $F_{1,119} = 7.76$, $p_{FDR} = .01$, $\eta^2_p = .06$; see Figure 6c).

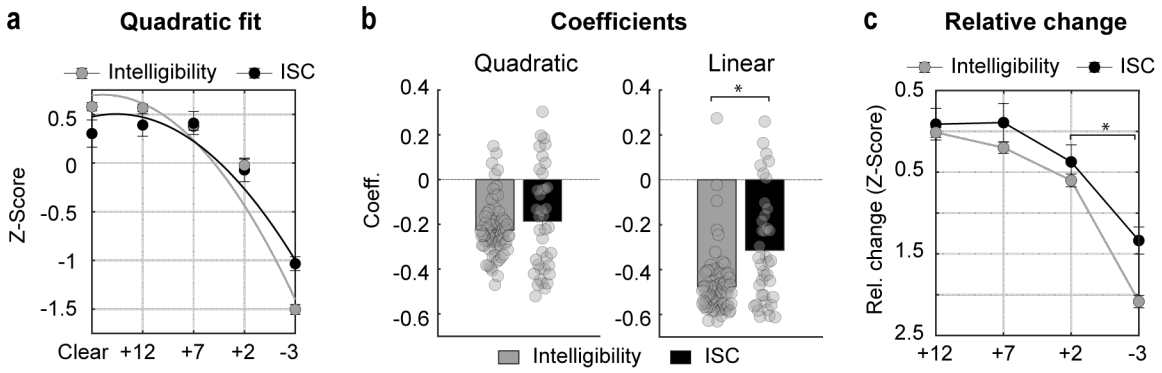

Figure 6. Normalized measures of intelligibility and inter-subject correlation. (a) Mean normalized word report (Exp 1) and ISC (Exp 2: Component 1) are plotted as a function of SNR (clear, +12, +7, +2, -3 dB). Solid lines correspond to a quadratic function fit to normalized means. (b) Mean quadratic (left panel) and linear coefficients (right panel) are plotted as a function of experiment measure (intelligibility, ISC). Grey dots represent data for individual participants. (c) The change in normalized word-report (z-score) and ISC for each SNR relative to clear. Error bars reflect standard error of the mean. *p < 0.05.

General Discussion

We investigated how acoustic masking affects speech intelligibility, engagement, and neural activity during listening to spoken stories. In Experiment 1, we observed that that speech intelligibility quadratically declined with decreasing SNR, such that intelligibility was high (80% or higher) at all but the lowest SNR, at which word report accuracy declined to ~55%. In Experiment 2, inter-subject correlation (ISC), a neural signature of engagement, exhibited a fairly similar quadratic pattern, decreasing most substantially for the lowest SNR. However, engagement (ISC) was actually more stable across all but the lowest SNRs, whereas

speech intelligibility declined more rapidly with SNR. Our study suggests that engagement with naturalistic, spoken stories does not change even under adverse listening conditions that reduce intelligibility by approximately 10%.

Story comprehension and overall engagement remain high despite background noise

In both experiments, participants scored well above chance on the comprehension questions, indicating that they were able to encode details from the stories despite regular changes in the level of background noise. Participants' self-rated story absorption in Experiment 1 (our behavioral measure of engagement) was neutral, but they reported being motivated and enjoyed listening to the stories. In Experiment 2 participants reported increased levels of absorption, enjoyment, and motivation relative to neutral, which is consistent with previous research suggesting engagement and motivation for interesting narratives are largely resistant to moderate levels of noise (Herrmann and Johnsrude, 2020a). In addition, participants in Experiment 2 reported higher levels of absorption, enjoyment, and motivation relative to Experiment 1 (see Table 2).

The differences in behavioral engagement metrics between experiments may have several explanations. Engagement with narrative materials requires individuals to generate and maintain mental models of events, situations, and characters (Zwaan et al., 1995; Mar and Oatley, 2008; Oatley, 2016; Zwaan, 2016). Disruptions to mental model generation and maintenance may reduce engagement with a narrative (Busselle and Bilandzic, 2009). In Experiment 1, stories were briefly discontinued about every 10-20 seconds and participants performed the intelligibility task, whereas stories were played without interruption in Experiment 2. Moreover, Experiment 1 was conducted as online, whereas Experiment 2 was conducted in-person in the lab, with possibly fewer distractions for participants. Generation and maintenance of a mental model may thus have been more disrupted in Experiment 1 than 2, leading to lower narrative absorption scores in Experiment 1 (Table 2). Alternatively, the different levels of reported engagement between the two experiments may reflect differences in listening strategy. The word-report task in Experiment 1 may have led listeners to adopt a detail-oriented strategy to ensure they could report

back specific words when prompted. In contrast, the absence of a dual task in Experiment 2 may have motivated a 'gist' or 'global' strategy (Harding et al., 2007), in which individuals focus on understanding the message but are not remembering words verbatim. The conditions of Experiment 2, in which people listened without a secondary intelligibility task, is more similar to real life, and suggests that, consistent with previous work (Herrmann and Johnsrude, 2020a), engagement with spoken stories – which probably relies on gist perception - is unaffected even if a listener misses 10% of words.

Speech intelligibility declines quadratically with decreasing signal to noise ratio

We utilized spoken stories masked with varied levels of background noise to approximate naturalistic listening conditions and found that speech intelligibility decreased as the SNR decreased. This is consistent with previous studies investigating how speech intelligibility changes with increasing speech degradation or with masking by short speech utterances, such as disconnected sentences (Davis and Johnsrude, 2003; Duncan and Aarts, 2006; Obleser et al., 2007; Akeroyd, 2008; Wild et al., 2012b, 2012a). We further demonstrate that the relation between acoustic masking and speech intelligibility was quadratic, such that intelligibility was relatively high at moderate SNRs (>80% words reported for +2 dB SNR and higher) but was particularly poor when the SNR was low (~55% words reported for -3 dB SNR). Performance at moderate SNRs may have benefited from the influence of contextual information, as semantic or lexical context can facilitate intelligibility when the speech signal is degraded or an individual has impaired hearing (Pichora-Fuller et al., 1995; Dubno et al., 2000; Obleser et al., 2007; Desjardins and Doherty, 2014; Holmes et al., 2018). This strategy was perhaps not possible for the lowest SNR, at which only about half of the words could be understood – the context gleaned from this proportion of sentences may not be sufficient to enhance comprehension.

Inter-subject correlation is stronger when listening to stories compared to rest

We observed that ISC was stronger during story listening compared to rest for all three spatial components. This is consistent with previous studies showing that ISC is stronger when attending to a coherent narrative

compared to when attention is directed elsewhere (Ki et al., 2016) or compared to rest (Hasson et al., 2004; Wilson et al., 2008). Observing significant ISC despite the presence of background noise suggests that participants were able to remain engaged despite presumably missing bits of the story during periods when the SNR was low.

Inter-subject correlation declines as signal to noise ratio decreases

Consistent with word-report performance in Experiment 1, we found that ISC declined with decreasing SNRs in a quadratic fashion. ISC did not differ between clear, +12, and +7 dB SNR despite missing approximately 10% of words for +7 dB SNR, but declined from +7 to +2, and more drastically from +2 dB to -3 dB SNR. Despite similar overall patterns, comparing ISC and intelligibility directly demonstrated that word-report performance declined more strongly with decreasing SNR compared to ISC (Figures 6a and 6c). Additionally, while ISC was lowest at -3 dB SNR, ISC for all three components was still stronger than during rest, which may suggest participants were still somewhat engaged when speech intelligibility was ~55%. This is consistent with a recent behavioral study demonstrating that listeners were as absorbed during a story that was continuously masked by babble at +4 dB SNR as they were during a clear story, despite much higher rated listening effort for the former (Herrmann and Johnsrude, 2020a). The current data provide neural evidence that engagement with spoken stories is unaffected by masking that reduces intelligibility to about 90%, but begins to decrease when intelligibility reduces to 80%. We speculate that sustained engagement despite acoustic masking may have resulted from participants accurately perceiving the story gist, and enjoying it, which perpetuated their intrinsic motivation to continue listening (Eckert et al., 2016; Matthen, 2016; Richter, 2016; Herrmann and Johnsrude, 2020b).

Listening and trying to understand masked speech is supported by processes such as sustained and selective attention (Davis and Johnsrude, 2007; Obleser et al., 2007; Wild et al., 2012b) which, when recruited, result in effortful listening (Pichora-Fuller et al., 2016). When the SNR is so poor that the gist is lost, or attentional resources become depleted (Ivarsson and Arlinger, 1994; Hornsby et al., 2016), this may

lead to temporary disengagement or attentional lapses (Hallberg and Carlsson, 1991; Heffernan et al., 2016). This is consistent with our observation of lowest, but still significant, ISC at the most difficult SNR (-3 dB SNR) for which participants missed about ~45% of words on average. Our study suggests that a listener can engage and follow a story's thread as long as intelligibility is over 80%.

Using narratives to approximate realistic listening scenarios

Everyday listening situations involve speech material comprised of sentences that relate to each other, and that are interesting to the listener. Such listening situations are commonly subject to increased levels of background noise, but are also rich with many positive aspects of listening that may drive motivation to listen (Matthen, 2016; Herrmann and Johnsrude, 2020b). Similarly, the spoken stories used here are intrinsically motivating to a listener, reflecting such everyday listening situations (Dunlop and Walker, 2013). Our work shows, in line with recent studies (Broderick et al., 2018, 2019, 2021; Brodbeck et al., 2020; Erb et al., 2020; Polonenko and Maddox, 2021), that utilizing naturalistic, spoken stories to investigate speech listening provide a useful avenue to investigate listening in ecologically valid conditions.

Conclusions

We utilized naturalistic spoken stories to investigate how challenging listening situations affect intelligibility and listener engagement. Listener engagement, measured via inter-subject correlation of electroencephalographic activity, was unaffected by moderately challenging SNRs, despite listeners missing approximately 10-20% of words. Our work provides a unique approach to observe speech intelligibility and listener engagement using ecologically valid, engaging stories, and suggests that listeners continue to engage with spoken stories even under adverse listening conditions when intelligibility is reduced.

References

Akeroyd MA (2008) Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. Int J Audiol 47:S53–S71.

Benjamini Y, Hochberg Y (2016) Controlling the false discovery rate : A practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300.

Berinsky AJ, Margolis MF, Sances MW (2014) Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. Am J Pol Sci 58:739–753.

Bilger RC (1984) Manual for the clinical use of the revised SPIN test. Champaign, IL: University of Illinois Press.

Brehm JW, Self EA (1989) The intensity of motivation. Annu Rev Psychol 40:109–131.

Brodbeck C, Jiao A, Hong LE, Simon JZ (2020) Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. PLoS Biol 18:1–22.

Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC (2018) Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Curr Biol 28:803–809.

Broderick MP, Anderson AJ, Lalor EC (2019) Semantic context enhances the early auditory encoding of natural speech. J Neurosci 39:7564–7575.

Broderick MP, Di Liberto GM, Anderson AJ, Rofes A, Lalor EC (2021) Dissociable electrophysiological measures of natural language processing reveal differences in speech comprehension strategy in healthy ageing. Sci Rep 11:1–12.

Bronkhorst AW (2000) The cocktail party phenomenon: A review of research on speech intelligibility in

multiple-talker conditions. Acta Acust United With Acust 86:117–128.

Brungart DS (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. J Acoust Soc Am 109:1101–1109.

Brungart DS, Simpson BD, Ericson MA, Scott KR (2001) Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am 110:2527–2538.

Buchanan EM, Scofield JE (2018) Methods to detect low quality data and its implication for psychological research. Behav Res Methods 50:2586–2596.

Buhrmester M, Kwang T, Gosling SD (2011) Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 6:3–5.

Busselle R, Bilandzic H (2008) Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement. Commun Theory 18:255–280.

Busselle R, Bilandzic H (2009) Measuring narrative engagement. Media Psychol 12:321–347.

Cohen SS, Parra LC (2016) Memorable audiovisual narratives synchronize sensory and supramodal neural responses. eNeuro 3:1–11.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016) The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. Front Hum Neurosci 10:1–14.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Davis MH, Johnsrude IS (2007) Hearing speech sounds: Top-down influences on the interface between audition and speech perception. Hear Res 229:132–147.

Dawes P, Emsley R, Cruickshanks KJ, Moore DR, Fortnum H, Edmondson-Jones M, McCormack A, Munro KJ (2015) Hearing loss and cognition: The role of hearing aids, social isolation and depression. PLoS One 10:1–9.

de Leeuw JR (2015) jsPsych: A JavaScript library for creating behavioral experiments in a web browser. Behav Res Methods 47:1–12.

Desjardins JL, Doherty KA (2014) The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. Ear Hear 35:600–610.

Dmochowski JP, Bezdek MA, Abelson BP, Johnson JS, Schumacher EH, Parra LC (2014) Audience preferences are predicted by temporal reliability of neural processing. Nat Commun 5:1–9.

Dmochowski JP, Sajda P, Dias J, Parra LC (2012) Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? Front Hum Neurosci 6:1–9.

Dubno JR, Ahlstrom JB, Horwitz AR (2000) Use of context by young and aged adults with normal hearing. J Acoust Soc Am 107:538–546.

Duncan KR, Aarts NL (2006) A comparison of the HINT and Quick SIN tests. J Speech-Language Pathol Audiol 30:86–94.

Dunlop WL, Walker LJ (2013) The life story: Its development and relation to narration and personal identity. Int J Behav Dev 37:235–247.

Eckert MA, Teubner-Rhodes S, Vaden KI (2016) Is listening in noise worth it? The neurobiology of speech recognition in challenging listening conditions. Ear Hear 37:101S-110S.

Erb J, Schmitt LM, Obleser J (2020) Temporal selectivity declines in the aging human auditory cortex. Elife 9:1–21.

Frisina DR, Frisina RD (1997) Speech recognition in noise and presbycusis: Relations to possible neural

mechanisms. Hear Res 106:95–104.

Gordon-Salant S (2006) Speech perception and auditory temporal processing performance by older listeners : implications for real-world communication. Semin Hear 27:264–268.

Gosling SD, Vazire S, Srivastava S, John OP (2004) Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. Am Psychol 59:93–104.

Hallberg LR, Carlsson SG (1991) A qualitative study of strategies for managing a hearing impairment. Br Soc Audiol 25:201–211.

Harding S, Cooke M, König P (2007) Auditory gist perception: An alternative to attentional selection of auditory streams? In: Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint (Paletta L, Rome E, eds), pp 399–416. Berlin, Heidelberg: Springer.

Hasson U, Furman O, Clark D, Dudai Y, Davachi L (2008a) Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron 57:452–462.

Hasson U, Malach R, Heeger DJ (2010) Reliability of cortical activity during natural stimulation. Trends Cogn Sci 14:40–48.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. Sci 303:1634–1640.

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008b) A hierarchy of temporal receptive windows in human cortex. J Neurosci 28:2539–2550.

Heffernan E, Coulson NS, Henshaw H, Barry JG, Ferguson MA (2016) Understanding the psychosocial experiences of adults with mild-moderate hearing loss: An application of Leventhal's self-regulatory model. Int J Audiol 55:S3–S12.

Herrmann B, Johnsrude I (2020a) Absorption and enjoyment during listening to acoustically masked

stories. Trends Hear 24:1–18.

Herrmann B, Johnsrude IS (2020b) A model of listening engagement (MoLE). Hear Res 397:108016.

Herrmann B, Maess B, Johnsrude IS (2018) Aging affects adaptation to sound-level statistics in human auditory cortex. J Neurosci 38:1989–1999.

Holmes E, Folkeard P, Johnsrude IS, Scollie S (2018) Semantic context improves speech intelligibility and reduces listening effort for listeners with hearing impairment. Int J Audiol 57:483–492.

Hornsby BWY, Naylor G, Bess FH (2016) A taxonomy of fatigue concepts and their relation to hearing loss. Ear Hear 37:136S-144S.

Ivarsson US, Arlinger SD (1994) Speech recognition in noise before and after a work-day's noise exposure. Scand Audiol 23:159–163.

Ki JJ, Kelly SP, Parra LC (2016) Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. J Neurosci 36:3092–3101.

Kuijpers MM, Hakemulder F, Tan ES, Doicaru MM (2014) Exploring absorbing reading experiences: Developing and validating a self-report scale to measure story world absorption. Sci Study Lit 4:89–122.

Lancaster G, Iatsenko D, Pidde A, Ticcinelli V, Stefanovska A (2018) Surrogate data for hypothesis testing of physical systems. Phys Rep 748:1–60.

Litman L, Robinson J, Abberbock T (2017) TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. Behav Res Methods 49:433–442.

Makeig S, Bell AJ, Jung T-P, Sejnowski TJ (1996) Independent component analysis of electroencephalographic signals. In: Advances in neural information processing systems 8 (D. Touretzky MM and MH, ed), pp 145–151. Cambridge, MA: MIT Press.

Mandler JM, Goodman MS (1982) On the psychological validity of story structure. J Verbal Learning Verbal Behav 21:507–523.

Mar RA, Oatley K (2008) The function of fiction is the abstraction and simulation of social experience. Perspect Psychol Sci 3:173–192.

Mason W, Suri S (2012) Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 44:1–23.

Matthen M (2016) Effort and displeasure in people who are hard of hearing. Ear Hear 37:28S-34S.

Näätänen R, Picton TW (1987) The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. Psychophysiology 24:375–425.

Nastase SA, Gazzola V, Hasson U, Keysers C (2019) Measuring shared responses across subjects using intersubject correlation. Soc Cogn Affect Neurosci 14:669–687.

Nguyen M, Vanderwal T, Hasson U (2019) Shared understanding of narratives is correlated with shared neural responses. Neuroimage 184:161–170.

Oatley K (2016) Fiction: Simulation of social worlds. Trends Cogn Sci 20:618–628.

Obleser J, Wise RJS, Dresner MA, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci 27:2283–2289.

Olsen WO (1998) Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. Am J Audiol 7:21–25.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011.

Palmer AD, Newsom JT, Rook KS (2016) How does difficulty communicating affect the social

relationships of older adults? An exploration using data from a national survey. J Commun Disord 62:131–146.

Parra LC, Haufe S, Dmochowski JP (2019) Correlated components analysis - Extracting reliable dimensions in multivariate data. arXiv Available at: http://arxiv.org/abs/1801.08881.

Parra LC, Spence CD, Gerson AD, Sajda P (2005) Recipes for the linear analysis of EEG. Neuroimage 28:326–341.

Peelle JE (2018) Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. Ear Hear 39:204–214.

Pichora-Fuller KM, Schneider BA, Daneman M (1995) How young and old adults listen to and remember speech in noise. J Acoust Soc Am 97:593–608.

Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BWY, Humes LE, Lemke U, Lunner T, Matthen M, Mackersie CL, Naylor G, Phillips NA, Richter M, Rudner M, Sommers MS, Tremblay KL, Wingfield A (2016) Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). Ear Hear 37:5S-27S.

Pichora-Fuller MK, Mick P, Reed M (2015) Hearing, cognition, and healthy aging: Social and public health implications of the links between age-related declines in hearing and cognition. Semin Hear 36:122–139.

Picton TW, John MS, Dimitrijevic A, Purcell D (2003) Human auditory steady-state responses. Int J Audiol 42:177–219.

Polonenko MJ, Maddox RK (2021) Exposing distinct subcortical components of the auditory brainstem response evoked by continuous naturalistic speech. Elife 10:1–37.

Reitan RM, Wolfson D (2000) Conation: A neglected aspect of neuropsychological functioning. Arch

Clin Neuropsychol 15:443–453.

Richter M (2013) A closer look into the multi-layer structure of motivational intensity theory. Soc Personal Psychol Compass 7:1–12.

Richter M (2016) The moderating effect of success importance on the relationship between listening demand and listening effort. Ear Hear 37:111S-117S.

Schmälzle R, Häcker FEK, Honey CJ, Hasson U (2015) Engaged listeners: Shared neural processing of powerful political speeches. Soc Cogn Affect Neurosci 10:1137–1143.

Smeds K, Wolters F, Rung M (2015) Estimation of signal-to-noise ratios in realistic sound scenarios. J Am Acad Audiol 26:183–196.

Thomas KA, Clifford S (2017) Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. Comput Human Behav 77:184–197.

Wayne R V, Johnsrude IS (2015) A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. Ageing Res Rev 23:154–166.

Wild CJ, Davis MH, Johnsrude IS (2012a) Human auditory cortex is sensitive to the perceived clarity of speech. Neuroimage 60:1490–1502.

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012b) Effortful listening: The processing of degraded speech depends critically on attention. J Neurosci 32:14010–14021.

Wilson SM, Molnar-Szakacs I, Iacoboni M (2008) Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. Cereb Cortex 18:230–242.

Woods KJP, Siegel MH, Traer J, McDermott JH (2017) Headphone screening to facilitate web-based auditory experiments. Attention, Perception, Psychophys 79:2064–2072.

Wright RA (2014) Presidential address 2013: Fatigue influence on effort-considering implications for self-regulatory restraint. Motiv Emot 38:183–195.

Yeshurun Y, Swanson S, Simony E, Chen J, Lazaridi C, Honey CJ, Hasson U (2017) Same story, different story: The neural representation ofiInterpretive frameworks. Psychol Sci 28:307–319.

Zwaan RA (2016) Situation models, mental simulations, and abstract concepts in discourse comprehension. Psychon Bull Rev 23:1028–1034.

Zwaan RA, Langston MC, Graesser AC (1995) The construction of situation models in narrative comprehension: An event-indexing model. Psychol Sci 6:292–297.