Neural activity during story listening is synchronized across individuals despite 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

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 situations has received little attention thus far. We recruit young, 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) at different signal-to-noise ratios (SNRs). 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) using electroencephalography (EEG) data to identify dynamic spatial patterns of shared neural activity evoked by the stories. ISC has previously been used as a neural measure of participants' engagement with naturalistic materials. Our results show that ISC was stable across all but the lowest SNRs, despite reduced speech intelligibility. Comparing ISC and intelligibility demonstrated that word-report performance declined more strongly with decreasing SNR compared to ISC. Our measure of neural engagement suggests that individuals remain engaged in story listening despite missing words due to background noise. Our work provides a potentially fruitful approach to investigate listener engagement with naturalistic, spoken stories that may be used to investigate (dis)engagement in older adults with hearing impairment.

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

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 & 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 & 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, which we define as "the (automatic or volitional) recruitment of executive and other cognitive resources, when speech comprehension serves a valued communication goal" (Herrmann & Johnsrude, 2020a). 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 & Self, 1989; Herrmann & Johnsrude, 2020a; Pichora-Fuller et al., 2016; Richter, 2013; Wright, 2014). An individual may disengage if they believe speech comprehension will be impossible (Eckert et al., 2016; Peelle, 2018; Richter, 2016). A second important factor is motivation. An individual who is motivated to listen, for example, because they find what they are hearing to be enjoyable and rewarding (Matthen, 2016), may engage in listening despite the presence of background sounds (Herrmann & Johnsrude, 2020a; Pichora-Fuller et al., 2016). However, typical investigations of speech comprehension involve listening to isolated sentences (Davis & Johnsrude, 2003; Duncan & 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

1   engagement also likely develops more slowly, at time scales beyond individual sentences (Mandler &

2   Goodman, 1982). Spoken stories, in contrast, involve event descriptions along a topical thread, are

3   intrinsically motivating to a listener, and are common in everyday life (Dunlop & Walker, 2013). Utilizing

4   engaging spoken stories under varying levels of acoustic masking may thus provide a fruitful avenue to

5   investigate intelligibility, and listening engagement and disengagement.

6   Engagement (and disengagement) can be measured behaviorally through assessment of an

7   individual's experience (Busselle & Bilandzic, 2008, 2009; Dmochowski et al., 2014; Herrmann &

8   Johnsrude, 2020b; Kuijpers et al., 2014). However, behavioral assessment of engagement typically requires

9   the listener to retroactively introspect about their experiences (Herrmann & Johnsrude, 2020b). In contrast,

10   neural activity recorded using functional imaging or electroencephalography (EEG) may provide a real-

11   time window on engagement. Naturalistic materials such as movies or spoken stories lead to synchronized

12   patterns of neural activity across individuals that scale with the degree of engagement (Hasson et al., 2010;

13   Nastase et al., 2019; Nguyen et al., 2019; Yeshurun et al., 2017). The strength of synchrony across

14   individuals, quantified as inter-subject correlation (ISC; Hasson et al., 2004, 2008; Dmochowski et al.,

15   2012, 2014), is stronger when stimuli are captivating or exciting (Hasson et al., 2010; Schmälzle et al.,

16   2015), and is predictive of behavioral measures reflecting engagement (S. S. Cohen et al., 2017; Dikker et

17   al., 2017; Dmochowski et al., 2014; Poulsen et al., 2017; Song et al., 2021) and recall of the materials (Chan

18   et al., 2019; S. S. Cohen et al., 2018; S. S. Cohen & Parra, 2016; Davidesco et al., 2019; Hasson, Furman,

19   et al., 2008; Piazza et al., 2021; Song et al., 2021; Stephens et al., 2010). Conversely, ISC is reduced when

20   individuals do not attend to naturalistic materials (S. S. Cohen et al., 2018; Ki et al., 2016; Kuhlen et al.,

21   2012; Rosenkranz et al., 2021) or when stimuli are unstructured or temporally scrambled (Dmochowski et

22   al., 2012; Hasson, Yang, et al., 2008).

23   ISC is conceptually similar to measures which estimate the correlation between specific stimulus

24   features (such as the amplitude envelope) and neural responses (cf. Ding & Simon, 2013; Iotzov & Parra,

25   2019; Peelle et al., 2013). However, critically, neural synchronization to stimulus features is only one

1   potential source that may drive shared neural patterns across subjects during naturalistic listening. Engaging

2   narratives are designed to provide a shared conscious experience to listeners — this is driven, in part, by

3   the recruitment of processes that enable each listener to integrate incoming information with existing

4   schemas, refine their understanding, infer what is unsaid, and anticipate what may be coming next, while

5   filtering out distractions (Naci et al., 2014). This manifests as ongoing engagement in the story. These

6   cognitive processes are executive in nature since they are coordinating all mental activity as the story

7   unfolds, enabling the listener to follow along and understand. ISC is sensitive to such experiential aspects

8   of engaging materials (Dmochowski et al., 2014; Naci et al., 2014; Nastase et al., 2019; Nummenmaa et

9   al., 2012; Schmälzle et al., 2015; Yeshurun et al., 2017).

10   In the current study, we use spoken stories to investigate how challenging listening situations affect

11   speech intelligibility and story engagement in a group of young, normal-hearing adults. We opted to

12   measure intelligibility and ISC in separate experiments because we suspected the inclusion of the

13   intelligibility task may influence participants' engagement levels during story listening. Therefore, in

14   Experiment 1, we assess intelligibility of spoken stories presented with different levels of 12-talker babble

15   masker. In Experiment 2, we record EEG in another group of individuals while they listen to the same

16   stories used in Experiment 1 and investigate how synchronized activity (inter-subject correlation) is affected

17   by background masking during story listening.

Experiment 1: Examining story intelligibility and engagement

19   In Experiment 1, we investigate the extent to which participants understand spoken stories when they are

20   masked with 12-talker babble at different signal-to-noise ratios (SNRs). We further assess overall listening

21   engagement behaviorally using a modified narrative absorption scale (Kuijpers et al., 2014) adapted for use

22   with spoken stories (Herrmann & Johnsrude, 2020b). Speech intelligibility ratings from Experiment 1 will

23   be related to data from Experiment 2, in which we investigate neural signatures of story engagement, to

24   assess the extent to which story engagement relates to 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, previous work indicates that online studies generally replicate findings of in-person data collection (Berinsky et al., 2014; Buchanan & Scofield, 2018; Buhrmester et al., 2011; Gosling et al., 2004; Mason & Suri, 2012; Thomas & Clifford, 2017). 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

1   to equal the length of each story. The signal-to-noise ratio (SNR) was pseudo-randomly varied

2   approximately every 30 to 33 s throughout each story (32, 30, 33, and 30-s, respectively, for the four

3   stories). Five SNRs were chosen: clear, +12, +7, +2, and -3 dB signal-to-noise ratio (SNR) relative to a 12-

4   talker babble masker. SNR was manipulated by adjusting the dB level of both the story and masker. This

5   ensured that the overall sound level remained constant throughout a story and was similar for all stories.

6   Three versions of SNR condition order were generated for each story in order to ensure that specific parts

7   of a story were not confounded with a specific SNR. Within each version, SNR was varied pseudo-randomly

8   such that a particular SNR could not be heard twice in succession. Four 30, 33, or 30-s segments per SNR

9   were presented for stories by O'Regan, Qureshi, and Nurse, respectively, and five 32-s segments per SNR

10   were presented for the longest story by Garriott.

11   Insert Figure 1

12   For each story, phrases/sentences ranging from 4 to 8 words (range of durations: 0.62–3.6 s) were

13   selected for intelligibility (word report) testing. These test phrases/sentences did not occur during the

14   transition period from one SNR to the next (for approximately 5-s before, and 1-s after, the SNR transition).

15   Four phrases per 30–33-s segment were selected, resulting in 100 phrases for the story by Garriott, and 80

16   for each of the other three stories. Two of the four selected phrases per 30–33-s segment were used as one

17   intelligibility test set, whereas the other two of the four selected phrases were used as a second intelligibility

18   test set (5 or 4 segments × 5 SNRs × 2 phrases/sentences × 2 intelligibility test sets = 100 or 80 phrases).

19   Having two test sets ensured that the observed intelligibility effects were not confounded by specific

20   phrases/sentences.

21   The experiment was conducted online, using custom written JavaScript/html and jsPsych code

22   (Version 6.1.0, a high-level JavaScript library used for precise stimulus control; de Leeuw, 2015). The

23   experiment code was stored at an online Gitlab repository (https://gitlab.pavlovia.org) and hosted via

24   Pavlovia (https://pavlovia.org/). During the main experiment, each participant listened to one pseudo-

25   randomly selected story (Garriott: N = 21; O'Regan: N = 22; Qureshi: N = 20; Nurse: N = 19). SNR order

1   and intelligibility test set were randomly assigned. A black fixation cross was presented at the center of the

2   screen throughout the story. The fixation cross turned yellow two seconds prior to the beginning of a test

3   phrase/sentence, cueing the participant to prepare for intelligibility testing (see Figure 1). The fixation cross

4   then turned green for the duration of the test phrase in the story, indicating to the participant the phrase they

5   would be asked to report back. The story stopped with the offset of the test phrase, and an input text box

6   appeared on the screen. Participants were asked to type their answer into the text box, after which the story

7   resumed from the beginning of the sentence most recently heard (allowing for story continuation). After

8   the story ended, participants answered questions that assessed comprehension and rated statements about

9   their story listening experiences (see Table 1).

10   In order to familiarize participants with the intelligibility task, a brief practice block was presented

11   prior to the main experiment. Participants heard a ~3-min story (a shortened version of *A Shoulder Bag to*

12   *Cry On* by Laura Zimmerman), without added babble noise, and performed 12 trials of the intelligibility

13   task (2 trials per 30-s segment).

14   **Online research quality assurance measures**

15   Participants completed two initial listening tasks at the very beginning of the online session. First,

16   participants listened to a 15-s stream of pink noise normalized to the same root-mean-square amplitude as

17   the stories and were instructed to adjust their volume to a comfortable listening level. Participants had the

18   option to replay the noise if they needed additional time to adjust their volume. This task ensured that

19   participants had an opportunity to adjust their volume to a comfortable level prior to the intelligibility task,

20   after which they were instructed to not make further adjustments.

21   Second, participants completed a headphone-check procedure to determine whether participants

22   were wearing headphones as instructed (cf. Woods et al., 2017). Participants performed a tone

23   discrimination task (6 trials; ~2 minutes total duration), in which they determined which of three

24   consecutive 200-Hz sine tones was the quietest. The three tones differed such that one was presented at the

1   comfortable listening level, one at –6 dB relative to the other two tones, and one at the comfortable listening

2   level with a 180° phase difference between the left and right headphone channels (anti-phase tone). This

3   task is straightforward over headphones, but difficult over loudspeakers, because the pressure waves

4   generated from an anti-phase tone interfere if heard through loudspeakers (Woods et al 2017). If they were

5   listening through loudspeakers, they would likely falsely select the anti-phase tone as the quietest tone. No

6   participants were excluded solely on the basis of performance on this test, however, it did provide a metric

7   which could flag potentially non-compliant online subjects.

8   **Assessment of intelligibility**

9   We calculated the proportion of correctly reported words for each SNR (Clear, +12, +7, +2, -3), across the

10 three versions of the four stories. Different or omitted words were counted as errors, but minor misspellings

11 and incorrect grammatical number (singular vs. plural) were not. Word-report performance was assessed

12 using a repeated-measures analysis of variance (rmANOVA) with SNR (Clear, +12, +7, +2, -3 dB SNR) as

13 the within-subjects factor.

14 **Assessment of story comprehension**

15 Story comprehension was assessed using 8 statements, which either correctly or incorrectly described an

16 element from the story the participant heard. Participants were asked to categorize each statement as true

17 or false. Comprehension performance was calculated as the proportion of statements categorized correctly.

18 Comprehension was statistically examined using a one-sample *t*-test which tested scores against chance-

19 level performance of 0.5.

20 **Assessment of story engagement, enjoyment, and motivation**

21 Following the assessment of story comprehension, we also assessed how much participants were engaged

22 with the story, how much they enjoyed the story, and how motivated they were to listen. In order to assess

23 story engagement, we utilized a narrative absorption scale (NAS; Kuijpers et al., 2014) adapted previously

1   for spoken stories (Herrmann & Johnsrude, 2020b). The NAS contains statements along four dimensions

2   (attention, emotional engagement, mental imagery, and transportation; Herrmann & Johnsrude, 2020b;

3   Kuijpers et al., 2014) , which participants rated on a 7-point scale, where '1' referred to completely disagree

4   and '7' referred to completely agree (Table 1). Participants also rated statements about enjoyment and

5   listening motivation on the same 7-point scale (Table 1). Rating scores for each statement were averaged

6   separately for narrative absorption (NAS) and enjoyment. Ratings for NAS and enjoyment were statistically

7   examined in separate one-sample t-tests which tested ratings against a neutral response (test value: 4). Given

8   that motivation was assessed using a single question (Table 1), the data remain ordinal rather than

9   continuous. Motivation was therefore assessed using a one-sample Wilcoxon signed-rank test, against a

10   neutral response (test value: 4).

11   **Table 1. Statements of the narrative absorption scale (NAS), enjoyment, and motivation (Herrmann**
12   **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 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 varying the SNR 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 & Johnsrude, 2020b), 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.

Insert Figure 2

Story intelligibility declines 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 × 10^{-38}, η^2_p = .69; see Figure 2b), as predicted (Bronkhorst, 2000; Brungart, 2001; Brungart et al., 2001; Duncan & Aarts, 2006). 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 × 10^{-7}, r_e = .54), and +2 to -3 dB SNR (t_{81} =14.94, p_{FDR} = 2.22 × 10^{-24}, r_e = .86).

1   **Relating intelligibility and behavioral measures of engagement and comprehension**

2   In an effort to quantify how intelligibility may relate to engagement and comprehension measures, we

3   calculated correlations between intelligibility (averaged across SNRs), NAS, motivation, enjoyment, and

4   comprehension scores (see Figure 2c). We found that intelligibility was significantly correlated with

5   motivation (r_{80} = .33, p_{FDR} = .004) and comprehension scores (r_{80} = .50, p_{FDR} = 6.7 ×10^{-6}), such that better

6   word-report performance was associated with higher levels of motivation to listen and better comprehension

7   of the stories. We did not find a significant correlation between intelligibility and NAS (r_{82} = .10, p_{FDR} =

8   .381) or intelligibility and enjoyment scores (r_{80} = .16, p_{FDR} =.223), suggesting that enjoyment/engagement

9   with the story and being able to report the specific words spoken during the story may be independent (cf.

10   Herrmann & Johnsrude, 2020b).

11   We also found a significant correlation between NAS and enjoyment (r_{80} = .77, p_{FDR} = 3.9 ×10^{-16}),

12   and NAS and motivation (r_{80} = .55, p_{FDR} = 3.4 ×10^{-6}), suggesting that higher levels of engagement were

13   associated with greater enjoyment of the stories and motivation to listen. We further observed a significant

14   relationship between enjoyment and motivation (r_{80} = .48, p_{FDR} = 1.6 ×10^{-5}), and enjoyment and

15   comprehension (r_{80} = .26, p_{FDR} = .029). This indicates that enjoyment was associated with greater

16   motivation to listen and better comprehension. No relationship was observed between NAS and

17   comprehension (r_{80} = .11, p_{FDR} = .381), or between motivation and comprehension (r_{80} = .12, p_{FDR} = .367).

18   The results of Experiment 1 suggest that participants were able to maintain a relatively high level

19   of story comprehension, were generally motivated, and enjoyed listening to the stories, despite the presence

20   of the babble masker. Further, speech intelligibility was relatively good for SNRs up to +2 dB (~80% correct

21   words report), but substantially declined for a higher masking level (-3 dB SNR; ~54%). The intelligibilty

22   and engagement data obtained in this experiment will be related to inter-subject correlation during story

23   listening in Experiment 2.

1   **Experiment 2: Examining inter-subject correlation during story listening**

2   Synchronization of neural activity across different individuals – measured as inter-subject correlation – has

3   previously been suggested to provide a window onto engagement and shared experience with naturalistic

4   materials (Hasson et al., 2010; Nastase et al., 2019; Nguyen et al., 2019; Yeshurun et al., 2017). In

5   Experiment 2, we investigate how inter-subject correlation during story listening is affected by 12-talker

6   background babble noise at different SNRs.

7   **Materials and Methods**

8   **Participants**

9   Thirty-nine individuals (mean: 20.3 years; age-range: 18-32 years; 19 males 20 females) without hearing

10   loss, neurological issues, or psychiatric disorders participated in Experiment 2. All participants were

11   recruited from Western University or the surrounding community of London, Canada. All participants

12   provided written informed consent and were financially compensated with $10 CAD per hour. The study

13   was conducted in accordance with the Declaration of Helsinki, the Canadian Tri-Council Policy Statement

14   on Ethical Conduct for Research Involving Humans (TCPS2-2014), and approved by the local Health

15   Sciences Research Ethics Board of the University of Western Ontario (REB #112015). Five additional

16   individuals participated but were not included either due to an issue with sound delivery (N =2), or a

17   technical error during data recording (N=3).

18   **Acoustic stimulation and procedure**

19   The experiment was conducted in a single-walled sound-attenuating booth (Eckel Industries). Sounds were

20   delivered through Sennheiser (HD 25 Light) headphones at a comfortable listening level, using a Focusrite

21   Scarlett 2i4 external soundcard controlled by a PC (Windows 10) and Psychtoolbox (Version 3) in

22   MATLAB (R2017b).

1   Stimuli were the same four stories used Experiment 1, with the same SNRs (Clear, +12, +7, +2, -

2   3) and the same three versions of SNR condition order per story. EEG was recorded while participants

3   listened to each of the four stories consecutively (see Figure 3). Story version and story order were

4   counterbalanced across participants. After each story, comprehension, story engagement, motivation, and

5   enjoyment were assessed, and these data analyzed, as in Experiment 1. After listening to all four stories,

6   participants completed a resting block, where they sat quietly for six minutes with their eyes open while

7   EEG was recorded. This block was used as a baseline for ISC analyses of story listening.

8   Insert Figure 3

9   **EEG recording and preprocessing**

10   EEG was recorded from 64 active electrodes (Ag/AgCl) placed on the scalp using an electrode cap and two

11   additional electrodes placed on both mastoids, with spacing on the scalp according to the 10/20 system

12   (Biosemi ActiveTwo system). During data recording, all electrodes were referenced to a feedback loop of

13   two electrodes, a common mode sense (CMS) active electrode and a driven passive electrode (see

14   www.biosemi.com/faq/cms&drl.htm). EEG was recorded at 1024 Hz with an online low-pass filter of 208

15   Hz to focus on cortical sources.

16   All pre-processing was carried out offline using MATLAB software, the Fieldtrip toolbox

17   (Oostenveld et al., 2011), and custom scripts. EEG data were re-referenced to the average of the signal from

18   both mastoids. EEG data were then notch-filtered at 60 Hz to attenuate line noise, and then high-pass (0.5

19   Hz, 3429 points, Hann window) and low-pass filtered (22 Hz, 211 points, Kaiser window). EEG data were

20   also downsampled to 256 Hz. Artifacts due to eye movements (saccades, blinks) and muscle activity were

21   removed using independent components analysis (Makeig et al., 1996). To further exclude additional

22   artifacts from subsequent analyses, data segments in which the EEG signal changed by more than 80 µV

23   within a 0.2-s period in any channel were set to 0 µV (cf. Dmochowski et al., 2012, 2014; Cohen and Parra,

24   2016).

1 **Inter-subject correlation**

2 We quantified inter-subject correlation (ISC) using correlated component analysis (CCA; Dmochowski et

3 al., 2012; Parra et al., 2019), a signal decomposition method that identifies a set of electrode weights (i.e.,

4 spatial filters) yielding a linear combination of electrode activation (components) maximally correlated

5 across participants. CCA can uncover patterns of neural activity that would not be possible with an

6 electrode-to-electrode correlation method (S. S. Cohen & Parra, 2016; Dmochowski et al., 2012; Ki et al.,

7 2016). For mathematical details of the method see (Parra et al., 2019), for available MATLAB scripts see

8 (https://www.parralab.org/isc/), and for the adapted version of the MATLAB scripts used in the current

9 study see (https://osf.io/tv7kg/). We calculated the components for each story individually to enable leave-

10 one-out analyses (see below), such that spatial filter calculations are independent from subsequent analyses

11 (see also Crosse et al., 2016; Herrmann et al., 2018; Broderick et al., 2019). In line with previous work, we

12 restrict our analysis to the three components with highest overall ISC for each story (cf. Dmochowski et al.,

13 2012, 2014; Cohen and Parra, 2016; Ki et al., 2016), because they show consistent spatial projections onto

14 the scalp across stories (see Figure 4, left panel).

15 Separately for each of the three components (i.e., spatial filters), component weights for one story

16 were multiplied with participants' EEG data for the other three stories. Spatially filtered data for the three

17 stories were then concatenated, leading to one response time course which reflects the component's

18 underlying sources (cf. Crosse et al., 2016; Broderick et al., 2019). This was repeated for all 4 stories,

19 resulting in four spatially filtered EEG time courses for each component, per participant (4 concatenated

20 datasets × 3 components = 12 spatially filtered time courses). All ISC analyses described in subsequent

21 paragraphs were calculated after using this leave-one-out approach. ISC values were subsequently averaged

22 across the four leave-one out iterations for each component.

23 For an overall ISC analysis independent of SNR conditions, each participant's time course was

24 correlated with the time course of each of the other participants, resulting in N-1 correlations per participant.

25 ISC for a unique participant was calculated as the mean across these N-1 correlation values. In order to test

1   whether story listening led to higher ISC than resting state activity in these 'story-listening' networks, the

2   EEG data from the resting block were projected through each story's spatial filters using the leave one-out

3   approach described above. Resting ISC was calculated similarly to story-listening ISC. Mean story-

4   listening ISC was compared to resting ISC using a paired sample t-test, separately for each of the three

5   components. Note that results from using surrogate-data testing (Lancaster et al., 2018), by circularly

6   shifting response time courses of participants, mirrored results from the resting-state contrasts.

7   To examine ISC as a function of SNR, we concatenated EEG segments with the same SNR and

8   spatially filtered using the above leave-one-out approach. This resulted in a spatially-filtered EEG time

9   series for each participant, SNR condition, and spatial-filter component. Separately for each SNR, each

10   participant's time course was correlated with the time course of each of the other participants who heard

11   the same story version (i.e., same order of SNR conditions), and subsequently averaged across the resulting

12   N-1 correlations for each subject. We analyzed ISC using rmANOVAs with SNR (Clear, +12, +7, +2, -3

13   dB) as one within-subjects factor, separately for each of the three components.

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

15   To characterize the relationship between inter-subject correlation and behavioral ratings of absorption

16   (NAS), enjoyment, motivation, and story comprehension, we calculated correlations (Pearson) between

17   overall ISC for each of the three components and these behavioral measures.

18   **Comparison of ISC and speech intelligibility**

19   In order to directly compare word-report performance in Experiment 1 with inter-subject correlation during

20   story listening in Experiment 2, we transformed both word-report scores and ISC values to z-scores. We

21   restricted our analysis to the component with the highest ISC (Component 1, see Figure 4, right panel). To

22   assess the extent to which the effect of SNR differentially impacts intelligibility and ISC, we fit an

23   exponential function to the z-scored word-report and ISC values using the following equation:

24   $$y = a * e^{(-x*\tau)} + s \qquad\qquad\qquad \text{Eq. 1}$$

1   where y is the exponential function, *a* is the parameter for amplitude, *x* is a vector of linearly spaced values

2   (1 to 5) corresponding to each SNR (clear, +12, +7, +2, -3 dB SNR), τ (tau) controls the decay strength,

3   and the *s* parameter which allows the function to shift along the y-axis. Note that the inclusion of *s* was

4   essential to fit the exponential function to z-scored data. To detect differences in how strongly intelligibility

5   and ISC declined with SNR, we analyzed the τ (tau) parameter using an independent t-test, with experiment

6   measure (intelligibility [Exp 1], ISC [Exp 2]) as the grouping variable. To make sure our results are robust

7   to the type of function fit, we also fit a quadratic function to the z-scored word-report and ISC values, and

8   compared quadratic and linear coefficients using two separate independent t-tests, with experiment measure

9   (intelligibility [Exp 1], ISC [Exp 2]) as the grouping variable.

10   Finally, to capture how each non-clear SNR condition (+12, +7, +2, -3 dB) changed relative to the

11   clear condition, we calculated difference scores between each SNR level and the clear condition, separately

12   for word-report scores and ISC values. Resulting values were submitted to a mixed design ANOVA with

13   measure (intelligibility, ISC) as the between-subjects variable and SNR (+12, +7, +2, -3 dB) as the within-

14   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 (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

1   We examined participants' level of story engagement by testing mean ratings for each engagement

2   category (NAS, enjoyment, motivation) against a rating of 4 (neutral response). In contrast to Experiment

3   1, scores on the NAS ($t_{38} = 3.62$, $p = .001$, $r_e = .51$), enjoyment ($t_{38} = 7.28$, $p = 1.02 \times 10^{-8}$, $r_e = .76$), and

4   motivation ($t_{38} = 8.9$, $p = 7.6 \times 10^{-11}$, $r_e = .82$), were all significantly higher than a neutral response (see

5   Table 2). This indicates that participants were engaged, enjoyed the stories, and felt motivated to listen. As

6   a follow-up, we directly compared behavioral ratings between Experiment 1 and 2 using either separate

7   independent t-tests (NAS, enjoyment) or a Wilcoxon rank sum test (motivation), with Experiment as the

8   grouping factor (Experiment 1, Experiment 2). We found that narrative absorption ($t_{119} = 2.7$, $p_{FDR} = .008$,

9    $r_e = .24$), enjoyment ($t_{119} = 4.14$, $p_{FDR} = 2 \times 10^{-4}$, $r_e = .35$), and motivation ($Z = 2.77$, $p_{FDR} = .008$) were higher

10   in Experiment 2 compared to Experiment 1 (Table 2), perhaps because Experiment 2 was in person, whereas

11   Experiment 1 was online and Experiment 2 did not require participants to provide word reports during the

12   story.

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

14   Figure 4 (left panel) shows the scalp projections of the first three components (Parra et al., 2005):

15   topographical distributions are consistent with previous work (S. S. Cohen & Parra, 2016; Dmochowski et

16   al., 2012, 2014; Ki et al., 2016). The first component has a fronto-central scalp distribution, which is

17   consistent with a source in auditory cortex (Näätänen & Picton, 1987; Picton et al., 2003), but may also

18   reflect contributions from frontal cortex. The second component has a parietal scalp distribution, suggesting

19   a source in parietal cortex or posterior auditory cortex. The third component shows a parieto-occipital peak

20   with frontal polarity reversal.

21   Insert Figure 4

22   For each component we observed ISC values that, while appearing small, were within an expected

23   range reported previously (Dmochowski et al., 2012, 2014; Ki et al., 2016). ISC was stronger during story

1   listening compared to rest for all three components (Component 1: t_{38} = 31.6, p = 7.3 × 10^{-29}, r_e = .98;

2   Component 2: t_{38} =19.6, p = 1.7 × 10^{-21}, r_e = .95; Component 3: t_{38} =16.7, p = 4.6 × 10^{-19}, r_e = .94; see, right

3   panel). ISC for surrogate data (generated by circularly shifting participants' time courses during the stories)

4   also yielded significantly lower ISC values compared to un-shifted (story listening) data for each

5   component. That the ISC was stronger during story listening compared to rest for the three strongest spatial

6   components suggests that it is driven by patterns of shared neural activity evoked by the stories.

7   **Inter-subject correlation declines more strongly with challenging SNRs**

8   In order to examine whether ISC is affected by SNR, we conducted an rmANOVA. Consistent with

9   behavioral performance from Experiment 1, we observed that ISC declined with decreasing SNR for all

10   three components (Component 1: $F_{4,152}$ = 20.64, p = 1.3 × 10^{-13}, η^2_p = .35; Component 2: $F_{4,152}$ = 18.66, p =

11   1.7 × 10^{-12}, η^2_p = .33; Component 3: $F_{4,152}$ = 3.22, p = .014, η^2_p = .08; Figure 5).

12   Insert Figure 5

13   We then compared ISC at successive pairs of SNRs to identify the point at which a significant drop

14   in ISC occurred. For Component 1, ISC strength did not differ between Clear and +12 dB SNR (p_{FDR} =

15   .859), or between +12 and +7 dB SNR (p_{FDR} = .703), but decreased significantly from +7 to +2 dB SNR

16   (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

17   pattern was observed for Component 2, such that ISC strength did not differ between Clear and +12 dB

18   SNR (p_{FDR} = .452), or between +12 and +7 dB SNR (p_{FDR} = .498), but decreased significantly from +7 to

19   +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).

20   No differences were observed between successive SNRs for Component 3 after correcting for multiple

21   comparisons (p_{FDR} >.14). Together, this suggests that ISC for Component 1 and 2 remained fairly stable at

22   moderate SNRs, but decreased most substantially at the lowest SNRs. Neural ISC correlates with the degree

23   of engagement with naturalistic materials (Hasson et al., 2010; Nastase et al., 2019; Nguyen et al., 2019;

24   Yeshurun et al., 2017), and our data may thus suggest that listeners remained engaged during naturalistic

1   listening even when about 10% of words were missed (+7 dB SNR; see Figure 2 for intelligibility data),

2   but started disengaging when approximately 20% of words were missed (+2 dB SNR).

3   Finally, we compared the SNR condition with the lowest ISC (-3 dB SNR) to ISC during rest to

4   determine whether the reduced synchronization observed for the most challenging listening condition was

5   still greater than average synchronization during rest. We observed stronger ISC during the -3 dB SNR

6   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

7   × 10^{-20}, r_e = .94), and Component 3 (t_{38} =-11.49, p_{FDR} = 6.2 × 10^{-14}, r_e =.88). Therefore, while ISC was

8   significantly reduced at -3 dB relative to the other less challenging SNRs, observing stronger ISC at -3 dB

9   compared to rest may suggest participants were still somewhat engaged. These results are consistent with

10   previous behavioral work suggesting that listeners continue to engage with spoken stories despite the

11   presence of masking and reduced speech intelligibility (Herrmann & Johnsrude, 2020b).

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

13   Correlations between overall ISC for each component (Component 1, Component 2, Component 3) and

14   behavioral ratings of absorption (NAS), enjoyment, motivation, and story comprehension were not

15   significant after correcting for multiple comparisons (r_{37} < .3, p_{FDR} > .22).

16   Consistent with Experiment 1, we found a significant correlation between NAS and enjoyment (r_{37}

17   = .81, p_{FDR} = 6.6 × 10^{-9}), and NAS and motivation (r_{37} = .70, p_{FDR} = 4 × 10^{-6}). Higher levels of absorption

18   were associated with greater enjoyment of the stories and motivation to listen. We also found a significant

19   relationship between enjoyment and motivation (r_{37} = .86, p_{FDR} = 5.8 × 10^{-11}): enjoyment was associated

20   with greater motivation to listen. NAS did not correlate with story comprehension (r_{37} =.21, p_{FDR} = .192),

21   and no other significant correlations were found (r_{37} <.23, p_{FDR} >.36).

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

23   We directly compared the effect of SNR on intelligibility and inter-subject correlation by fitting an

24   exponential function z-scored word-report values and z-scored ISC values for the component with the

1   strongest ISC (Component 1), and compared the resulting decay coefficients between word report and ISC.

2   We observed a larger τ coefficient for the intelligibility task relative to ISC (t_{119} = -21.44, p = .001, r_e =

3   .29), suggesting that speech intelligibility declined more strongly with SNR than ISC (Figure 6a). To

4   confirm that our results are robust to the type of function used to describe the data, we also fit a quadratic

5   function and compared the resulting quadratic and linear coefficients between word report and ISC (Figure

6   6b). The quadratic coefficients did not differ between word report and ISC (p = .21). However, we observed

7   more negative linear coefficients for the intelligibility task relative to ISC (t_{119} = 4.71, p = 7×10^{-6}, r_e = .4),

8   further confirming that speech intelligibility declined more strongly with SNR than ISC.

9   In order to capture at which SNR intelligibility and ISC diverged, we calculated difference scores,

10   separately for mean-normalized word-report and ISC, between each SNR and the clear condition and

11   analyzed the resulting z-scores using an ANOVA. The relative change from clear for both word-report and

12   ISC decreased significantly as SNR decreased (effect of SNR: $F_{3,357}$ = 197.5, p = 1.8×10^{-75}), but the decline

13   from 2 to -3 dB SNR was larger for word report than for ISC (Task × SNR interaction: $F_{3,357}$ = 5.99, p =

14   .001, η^2_p = .05; WS contrast: $F_{1,119}$ =7.76, p_{FDR} = .01, η^2_p = .06; see Figure 6c).

15   Insert Figure 6

16   **General Discussion**

17   We investigated how acoustic masking affects speech intelligibility, engagement, and neural activity during

18   story listening. In Experiment 1, we observed that speech intelligibility declined with decreasing SNR, such

19   that intelligibility was relatively high (80% or higher) at all but the lowest SNR (-3 dB), at which word

20   report accuracy declined to 54%. In Experiment 2, we recorded EEG while individuals listened to spoken

21   stories and investigated inter-subject correlation (ISC) as a possible window into engagement. We observed

22   that ISC remained stable across all but the lowest SNRs, whereas speech intelligibility measured in

23   Experiment 1 declined more rapidly with decreasing SNR. Our analyses suggest that speech intelligibility

24   is less robust to changes in SNR than synchronized neural activity (ISC). Speech intelligibility may thus be

1   an insufficient predictor of naturalistic speech listening. Our study suggests that individuals may continue

2   to engage with naturalistic, spoken stories even under adverse listening conditions that reduce intelligibility

3   by 10%.

4   **Story comprehension and overall engagement remain high despite background noise**

5   In both experiments, participants scored well above chance on the comprehension questions, indicating that

6   they were able to encode details from the stories despite regular changes in the level of background noise.

7   Participants' self-rated story absorption in Experiment 1 (our behavioral measure of engagement) was

8   neutral, but they reported being motivated and enjoyed listening to the stories. In Experiment 2, participants

9   reported increased levels of absorption, enjoyment, and motivation relative to neutral, which is consistent

10   with previous research suggesting engagement and motivation for interesting narratives are largely resistant

11   to moderate levels of noise (Herrmann & Johnsrude, 2020b). In addition, participants in Experiment 2

12   reported higher levels of absorption, enjoyment, and motivation relative to Experiment 1 (see Table 2).

13   The differences in behavioral engagement metrics between experiments may have several

14   explanations. Engagement with narrative materials requires individuals to generate and maintain mental

15   models of events, situations, and characters (Mar & Oatley, 2008; Oatley, 2016; Zwaan, 2016; Zwaan et

16   al., 1995). Disruptions to mental model generation and maintenance may reduce engagement with a

17   narrative (Busselle & Bilandzic, 2009). In Experiment 1, stories were briefly discontinued about every 10-

18   20 seconds and participants performed the intelligibility task, whereas stories were played without

19   interruption in Experiment 2. Moreover, Experiment 1 was conducted online, whereas Experiment 2 was

20   conducted in-person in the lab, with possibly fewer distractions for participants. Generation and

21   maintenance of a mental model may thus have been more disrupted in Experiment 1 than 2, leading to lower

22   narrative absorption scores in Experiment 1 (Table 2). Alternatively, the different levels of reported

23   engagement between the two experiments may reflect differences in listening strategy. The word-report

24   task in Experiment 1 may have led listeners to adopt a detail-oriented strategy to ensure they could report

1   back specific words when prompted. In contrast, the absence of a dual task in Experiment 2 may have

2   motivated a 'gist' or 'global' strategy (Harding et al., 2007), in which individuals focus on understanding

3   the message but are not remembering words verbatim. The conditions of Experiment 2, in which people

4   listened without a secondary intelligibility task, is more similar to real life, and suggests that, consistent

5   with previous work (Herrmann & Johnsrude, 2020b), engagement with spoken stories – which likely relies

6   on gist perception – is unaffected even if a listener misses 10% of words.

7   **Speech intelligibility declines with decreasing signal to noise ratio**

8   We utilized spoken stories masked with varied levels of background noise to approximate naturalistic

9   listening conditions and found that speech intelligibility decreased as the SNR decreased. This is consistent

10   with previous studies investigating how speech intelligibility changes with increasing speech degradation

11   or with masking by short speech utterances, such as disconnected sentences (Akeroyd, 2008; Davis &

12   Johnsrude, 2003; Duncan & Aarts, 2006; Obleser et al., 2007; Wild, Davis, et al., 2012; Wild, Yusuf, et al.,

13   2012). This is also consistent with other studies using indirect measures of speech intelligibility during

14   naturalistic listening, such as using comprehension or fill-in-the-blank questions (Best et al., 2016; Humes

15   & Dubno, 2010; MacPherson & Akeroyd, 2013; Power et al., 2012; Xia et al., 2017). One critical difference

16   is that our paradigm directly measures intelligibility by asking participants to report back phrases just after

17   being heard, which allows for a more direct comparison with previous psychoacoustic work using isolated

18   speech utterances.

19   We further demonstrate that intelligibility was relatively high at moderate SNRs (>80% words

20   reported for +2 dB SNR and higher) but was particularly poor when the SNR was low (~55% words reported

21   for -3 dB SNR). Performance at moderate SNRs may have benefited from the influence of contextual

22   information, as semantic or lexical context can facilitate intelligibility when the speech signal is degraded

23   or an individual has impaired hearing (Desjardins & Doherty, 2014; Dubno et al., 2000; Holmes et al.,

24   2018; Obleser et al., 2007; Pichora-Fuller et al., 1995). This strategy was perhaps not possible for the lowest

1   SNR, at which only about half of the words could be understood – the context gleaned from this proportion

2   of sentences may not be sufficient to enhance comprehension.

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

4   Inter-subject correlation (ISC) captures the aspects of neural activity that are synchronized across

5   participants, typically while they watch a movie or listen to a spoken story (Dmochowski et al., 2012, 2014;

6   Hasson et al., 2004; Hasson, Furman, et al., 2008). We observed that ISC was stronger during story listening

7   compared to rest for all three spatial components. This is consistent with previous studies showing that ISC

8   is stronger when attending to a coherent narrative compared to when attention is directed elsewhere (S. S.

9   Cohen et al., 2018; Ki et al., 2016; Kuhlen et al., 2012; Rosenkranz et al., 2021) or compared to rest (Hasson

10   et al., 2004; Wilson et al., 2008). Inter-subject correlation is often thought of as a neural signature of

11   engagement, because ISC is stronger for captivating or exciting stimuli (Hasson et al., 2010; Schmälzle et

12   al., 2015) and predicts behavioral engagement measures (S. S. Cohen et al., 2017; Dikker et al., 2017;

13   Dmochowski et al., 2014; Poulsen et al., 2017; Song et al., 2021). Observing significant ISC despite the

14   presence of background noise in the current study may suggest that participants were able to remain engaged

15   despite missing bits of the story during periods when the SNR was low.

16   Critically, the term engagement is not always used in the same manner. In literature and media

17   studies, engagement is typically defined as a multidimensional construct that includes attention, emotional

18   engagement, mental imagery, and "transportation", the experience of being absorbed in the world of the

19   story (Busselle & Bilandzic, 2008, 2009; J. Cohen, 2001; Green et al., 2004; Herrmann & Johnsrude, 2020b;

20   Kuijpers et al., 2014; see Table 1). Engagement measured as synchronized neural activity during movie

21   watching or story listening is thought to be driven by a shared experience across individuals (Dmochowski

22   et al., 2012; Hasson et al., 2010; Nastase et al., 2019; Nguyen et al., 2019; Yeshurun et al., 2017) and thus

23   may converge with more experiential, cognitive engagement dimensions. Our finding that activity is less

24   synchronized across participants when they rest and their minds are not focused on a story (Figure 4) may

1   suggest that EEG-recorded ISC is perhaps sensitive to one or more of the aforementioned experiential

2   dimensions of engagement. However, this remains to be more directly addressed in future research.

3   **Inter-subject correlation is more robust to changes in signal to noise ratio than speech intelligibility**

4   We observed that ISC was reduced for speech listening during challenging SNRs, but not for moderate

5   SNRs. Specifically, ISC did not differ between clear, +12, and +7 dB SNR, despite our separate behavioral

6   experiment indicating participants missed approximately 10% of words for +7 dB SNR (Figure 2).

7   However, ISC significantly declined from +7 to +2 and from +2 dB to -3 dB SNR. Despite similar overall

8   patterns, comparing ISC and intelligibility directly demonstrated that word-report performance declined

9   more strongly with decreasing SNR compared to ISC (Figures 6a - 6c). This is consistent with related work

10   that investigated the relation between speech intelligibility and neural synchronization with the amplitude

11   envelope of speech. Synchronization with the speech envelope appears robust to moderate changes in SNR

12   or the presence of a competing speech stream, while intelligibility declines (Ding & Simon, 2012, 2013).

13   Additionally, while ISC was lowest at -3 dB SNR, ISC for all three components was still stronger than

14   during rest, which may suggest participants were still somewhat engaged when speech intelligibility was

15   ~55%. This is consistent with a recent behavioral study demonstrating that listeners were as absorbed during

16   a story that was continuously masked by babble at +4 dB SNR as they were during a clear story, despite

17   much higher rated listening effort for the former (Herrmann & Johnsrude, 2020b). The current data provide

18   neural evidence that across-participant synchronization during story listening is unaffected by masking that

19   reduces intelligibility to about 90%, but begins to decrease when intelligibility reduces to 80%. We

20   speculate that observing synchronized activity despite acoustic masking may have resulted from

21   participants accurately perceiving the story gist, and enjoying it, which perpetuated their intrinsic

22   motivation to continue listening and stay engaged (Eckert et al., 2016; Herrmann & Johnsrude, 2020a;

23   Matthen, 2016; Richter, 2016).

1   Listening and trying to understand masked speech is supported by processes such as sustained and

2   selective attention (Davis & Johnsrude, 2007; Obleser et al., 2007; Wild, Yusuf, et al., 2012) which, when

3   recruited, result in effortful listening (Pichora-Fuller et al., 2016). When the SNR is so poor that the gist is

4   lost, or attentional resources become depleted (Hornsby et al., 2016; Ivarsson & Arlinger, 1994), this may

5   lead to temporary disengagement or attentional lapses (Hallberg & Carlsson, 1991; Heffernan et al., 2016).

6   This is consistent with our observation of lowest, but still significant, ISC at the most difficult SNR (-3 dB

7   SNR) for which participants missed about ~45% of words on average. Our study suggests that a listener

8   can engage and follow a story's thread as long as intelligibility is over 80%. However, one limitation of the

9   current study is that intelligibility and EEG-measured ISC were assessed in separate groups of participants.

10   Future studies may consider using a within-subjects design to more directly assess how intelligibility relates

11   to synchronized activity across participants.

12   **Relating inter-subject correlation with behavioral measures of engagement**

13   To assess the relationship between the behavioral and neural measures of engagement we calculated

14   correlations between the overall ISC for each component (Component 1, Component 2, Component 3) and

15   behavioral ratings of absorption (NAS), enjoyment, and motivation, but did not observe any significant

16   relationships after correcting for multiple statistical comparisons ($r_{37} < .3$, $p_{FDR} > .22$). This is inconsistent

17   with previous literature which demonstrates that ISC strength is predictive of engagement (S. S. Cohen et

18   al., 2017; Dikker et al., 2017; Dmochowski et al., 2014; Poulsen et al., 2017; Song et al., 2021) as well as

19   other measures dependent on sustained attention, such as subsequent recall of the materials (Chan et al.,

20   2019; S. S. Cohen et al., 2018; S. S. Cohen & Parra, 2016; Davidesco et al., 2019; Hasson, Furman, et al.,

21   2008; Piazza et al., 2021; Song et al., 2021; Stephens et al., 2010). We speculate that the behavioral measure

22   of engagement did not correlate with ISC because we did not continuously assess changes in behavioral

23   engagement throughout each story, as has been done previously (S. S. Cohen et al., 2017; Dmochowski et

24   al., 2014; Song et al., 2021). Post-hoc measures of behavioral engagement (as we used here) have also been

1   used on occasion: for example, relating ISC with engagement in the classroom (Dikker et al., 2017; Poulsen

2   et al., 2017), but the listening conditions did not vary systematically, as in the current study. We changed

3   SNR every 30 seconds - this may have led to dynamic changes in engagement (especially at high masking

4   levels), which would be reflected in ISC, but which would be difficult to capture behaviorally with a single

5   post-hoc measure. Further, in contrast to previous work, we used a story absorption scale (Herrmann &

6   Johnsrude, 2020b; Kuijpers et al., 2014) as a behavioral engagement measure, but whether this accurately

7   reflects neural engagement is a matter for further research. It would be helpful to combine our story

8   paradigm with a continuous measure of behavioral engagement to further explore this topic (cf. Song et al.,

9   2021).

10   **Using narratives to approximate realistic listening scenarios**

11   Everyday listening situations involve speech material comprised of sentences that relate to each other, and

12   that are interesting to the listener. Such listening situations are commonly subject to increased levels of

13   background noise, but are also rich with many positive aspects of listening that may drive motivation to

14   listen (Herrmann & Johnsrude, 2020a; Matthen, 2016). Similarly, the spoken stories used here are

15   intrinsically motivating to a listener, reflecting such everyday listening situations (Dunlop & Walker, 2013).

16   Our work shows, in line with recent studies (Brodbeck et al., 2020; Broderick et al., 2018, 2019, 2020; Erb

17   et al., 2020; Polonenko & Maddox, 2021), that utilizing naturalistic, spoken stories to investigate speech

18   listening provide a useful avenue to investigate listening in ecologically valid conditions.

19   <div align="center">**Conclusions**</div>

20   We utilized naturalistic spoken stories to investigate how challenging listening affect intelligibility and

21   listener engagement. We used inter-subject correlation of electroencephalographic activity as a measure of

22   listener engagement. Inter-subject correlation was unaffected by moderately challenging SNRs, despite

23   speech-intelligibility scores suggesting listeners missed approximately 10% of words. Our data further

1   show that speech intelligibility declines faster than electrophysiologically measured listening engagement

2   under masked speech conditions. Our work provides a potentially fruitful approach with naturalistic, spoken

3   stories to investigate intelligibility and engagement during masked speech listening. Our approach may

4   importantly open future avenues to investigate (dis)engagement in older adults with hearing impairment.

References

6   Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in

7   cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults.

8   *International Journal of Audiology*, *47*, S53–S71. https://doi.org/10.1080/14992020802301142

9   Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful

10   approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*,

11   *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

12   Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers?

13   Making sure respondents pay attention on self-administered surveys. *American Journal of Political

14   Science*, *58*(3), 739–753. https://doi.org/10.1111/ajps.12081

15   Best, V., Keidser, G., Buchholz, J. M., & Freeston, K. (2016). Development and preliminary evaluation

16   of a new test of ongoing speech comprehension. *International Journal of Audiology*, *55*(1), 45–52.

17   https://doi.org/10.3109/14992027.2015.1055835

18   Bilger, R. C. (1984). *Manual for the clinical use of the Revised SPIN test*. University of Illinois Press.

19   Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*, 109–

20   131. https://doi.org/10.1146/annurev.ps.40.020189.000545

21   Brodbeck, C., Jiao, A., Hong, L. E., & Simon, J. Z. (2020). Neural speech restoration at the cocktail

22   party: Auditory cortex recovers masked speech of both attended and ignored speakers. *PLoS*

1   *Biology*, *18*(10), 1–22. https://doi.org/10.1371/journal.pbio.3000883

2   Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018).

3   Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural,

4   narrative speech. *Current Biology*, *28*(5), 803–809. https://doi.org/10.1016/j.cub.2018.01.080

5   Broderick, M. P., Anderson, A. J., & Lalor, E. C. (2019). Semantic context enhances the early auditory

6   encoding of natural speech. *Journal of Neuroscience*, *39*(38), 7564–7575.

7   https://doi.org/10.1523/JNEUROSCI.0584-19.2019

8   Broderick, M. P., Di Liberto, G., Anderson, A., Rofes, A., & Lalor, E. (2020). Dissociable

9   electrophysiological measures of natural language processing reveal differences in speech

10   comprehension strategy in healthy ageing. *Scientific Reports*, *11*(4963), 1–12.

11   https://doi.org/10.1101/2020.04.17.046201

12   Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility

13   in multiple-talker conditions. *Acta Acustica United With Acustica*, *86*(1), 117–128.

14   https://doi.org/10.3758/s13414-015-0882-9

15   Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous

16   talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109.

17   https://doi.org/10.1121/1.1345696

18   Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic

19   masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical*

20   *Society of America*, *110*(5), 2527–2538. https://doi.org/10.1121/1.1408946

21   Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for

22   psychological research. *Behavior Research Methods*, *50*(6), 2586–2596.

23   https://doi.org/10.3758/s13428-018-1035-6

1   Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of

2   inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5.

3   https://doi.org/10.1177/1745691610393980

4   Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing stories: A model

5   of narrative comprehension and engagement. *Communication Theory*, *18*(2), 255–280.

6   https://doi.org/10.1111/j.1468-2885.2008.00322.x

7   Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology*, *12*(4), 321–

8   347. https://doi.org/10.1080/15213260903287259

9   Chan, H. Y., Smidts, A., Schoots, V. C., Dietvorst, R. C., & Boksem, M. A. S. (2019). Neural similarity

10   at temporal lobe and cerebellum predicts out-of-sample preference and recall for video stimuli.

11   *NeuroImage*, *197*, 391–401. https://doi.org/10.1016/j.neuroimage.2019.04.076

12   Cohen, J. (2001). Defining Identification: A Theoretical Look at the Identification of Audiences With

13   Media Characters. *Mass Communication and Society*, *4*(3), 245–264.

14   http://www.tandfonline.com/doi/abs/10.1207/S15327825MCS0403_01

15   Cohen, S. S., Henin, S., & Parra, L. C. (2017). Engaging narratives evoke similar neural activity and lead

16   to similar time perception. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-

17   04402-4

18   Cohen, S. S., Madsen, J., Touchan, G., Robles, D., Lima, S. F. A., Henin, S., & Parra, L. C. (2018).

19   Neural engagement with online educational videos predicts learning performance for individual

20   students. *Neurobiology of Learning and Memory*, *155*(January), 60–64.

21   https://doi.org/10.1016/j.nlm.2018.06.011

22   Cohen, S. S., & Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and

23   supramodal neural responses. *ENeuro*, *3*(6), 1–11. https://doi.org/10.1523/ENEURO.0203-16.2016

1   Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response

2   function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli.

3   *Frontiers in Human Neuroscience*, *10*, 1–14. https://doi.org/10.3389/fnhum.2016.00604

4   Davidesco, I., Laurent, E., Valk, H., West, T., Dikker, S., Milne, C., & Poeppel, D. (2019). Brain-to-brain

5   synchrony between students and teachers predicts learning outcomes. *BioRxiv*, 644047.

6   https://doi.org/10.1101/644047

7   Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension.

8   *Journal of Neuroscience*, *23*(8), 3423–3431. https://doi.org/10.1523/jneurosci.23-08-03423.2003

9   Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface

10   between audition and speech perception. *Hearing Research*, *229*(1–2), 132–147.

11   https://doi.org/10.1016/j.heares.2007.01.014

12   Dawes, P., Emsley, R., Cruickshanks, K. J., Moore, D. R., Fortnum, H., Edmondson-Jones, M.,

13   McCormack, A., & Munro, K. J. (2015). Hearing loss and cognition: The role of hearing aids, social

14   isolation and depression. *PLoS ONE*, *10*(3), 1–9. https://doi.org/10.1371/journal.pone.0119616

15   de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web

16   browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

17   Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in

18   hearing-impaired adults. *Ear and Hearing*, *35*(6), 600–610.

19   https://doi.org/10.1097/AUD.0000000000000028

20   Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G.,

21   Van Bavel, J. J., Ding, M., & Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world

22   dynamic group interactions in the classroom. *Current Biology*, *27*(9), 1375–1380.

23   https://doi.org/10.1016/j.cub.2017.04.002

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. https://doi.org/10.1152/jn.00297.2011

Ding, N., & Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *Journal of Neuroscience*, *33*(13), 5728–5735. https://doi.org/10.1523/JNEUROSCI.5297-12.2013

Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., & Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications*, *5*, 1–9. https://doi.org/10.1038/ncomms5567

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Frontiers in Human Neuroscience*, *6*(112), 1–9. https://doi.org/10.3389/fnhum.2012.00112

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *The Journal of the Acoustical Society of America*, *107*(1), 538–546. https://doi.org/10.1121/1.428322

Duncan, K. R., & Aarts, N. L. (2006). A comparison of the HINT and Quick SIN tests. *Journal of Speech-Language Pathology and Audiology*, *30*(2), 86–94.

Dunlop, W. L., & Walker, L. J. (2013). The life story: Its development and relation to narration and personal identity. *International Journal of Behavioral Development*, *37*(3), 235–247. https://doi.org/10.1177/0165025413479475

Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. (2016). Is listening in noise worth it? The neurobiology of speech recognition in challenging listening conditions. *Ear and Hearing*, *37*, 101S-110S. https://doi.org/10.1097/AUD.0000000000000300

1   Erb, J., Schmitt, L. M., & Obleser, J. (2020). Temporal selectivity declines in the aging human auditory

2   cortex. *ELife*, *9*, e55300. https://doi.org/10.7554/eLife.55300

3   Frisina, D. R., & Frisina, R. D. (1997). Speech recognition in noise and presbycusis: Relations to possible

4   neural mechanisms. *Hearing Research*, *106*(1–2), 95–104. https://doi.org/10.1016/S0378-

5   5955(97)00006-3

6   Gordon-Salant, S. (2006). Speech perception and auditory temporal processing performance by older

7   listeners: Implications for real-world communication. *Seminars in Hearing*, *27*(4), 264–268.

8   https://doi.org/10.1055/s-2006-954852

9   Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A

10   comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*,

11   *59*(2), 93–104. https://doi.org/10.1037/0003-066X.59.2.93

12   Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of

13   transportation into narrative worlds. *Communication Theory*, *14*(4), 311–327.

14   https://doi.org/10.1111/j.1468-2885.2004.tb00317.x

15   Hallberg, L. R., & Carlsson, S. G. (1991). A qualitative study of strategies for managing a hearing

16   impairment. *British Society of Audiology*, *25*(3), 201–211.

17   https://doi.org/10.3109/03005369109079853

18   Harding, S., Cooke, M., & König, P. (2007). Auditory gist perception: An alternative to attentional

19   selection of auditory streams? In L. Paletta & E. Rome (Eds.), *Attention in Cognitive Systems.*

20   *Theories and Systems from an Interdisciplinary Viewpoint* (Vol. 4840, pp. 399–416). Springer.

21   Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008). Enhanced intersubject correlations

22   during movie viewing correlate with successful episodic encoding. *Neuron*, *57*(3), 452–462.

23   https://doi.org/10.1016/j.neuron.2007.12.009

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48. https://doi.org/10.1016/j.tics.2009.10.011

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*, 1634–1640. https://doi.org/doi: 10.1126/science.1089506

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, *28*(10), 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008

Heffernan, E., Coulson, N. S., Henshaw, H., Barry, J. G., & Ferguson, M. A. (2016). Understanding the psychosocial experiences of adults with mild-moderate hearing loss: An application of Leventhal's self-regulatory model. *International Journal of Audiology*, *55*, S3–S12. https://doi.org/10.3109/14992027.2015.1117663

Herrmann, B., & Johnsrude, I. S. (2020a). A model of listening engagement (MoLE). *Hearing Research*, *397*, 108016. https://doi.org/10.1016/j.heares.2020.108016

Herrmann, B., & Johnsrude, I. S. (2020b). Absorption and enjoyment during listening to acoustically masked stories. *Trends in Hearing*, *24*, 1–18. https://doi.org/10.1177/2331216520967850

Herrmann, B., Maess, B., & Johnsrude, I. S. (2018). Aging affects adaptation to sound-level statistics in human auditory cortex. *Journal of Neuroscience*, *38*(8), 1989–1999. https://doi.org/10.1523/JNEUROSCI.1489-17.2018

Holmes, E., Folkeard, P., Johnsrude, I. S., & Scollie, S. (2018). Semantic context improves speech intelligibility and reduces listening effort for listeners with hearing impairment. *International Journal of Audiology*, *57*(7), 483–492. https://doi.org/10.1080/14992027.2018.1432901

Hornsby, B. W. Y., Naylor, G., & Bess, F. H. (2016). A taxonomy of fatigue concepts and their relation to hearing loss. *Ear and Hearing*, *37*, 136S-144S. https://doi.org/10.1097/AUD.0000000000000289

Humes, L. E., & Dubno, J. R. (2010). Factors Affecting Speech Understanding in Older Adults. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.), *The Aging Auditory System* (pp. 211–257). Springer New York. https://doi.org/10.1007/978-1-4419-0993-0_8

Iotzov, I., & Parra, L. C. (2019). EEG can predict speech intelligibility. *Journal of Neural Engineering*, *16*(3). https://doi.org/10.1088/1741-2552/ab07fe

Ivarsson, U. S., & Arlinger, S. D. (1994). Speech recognition in noise before and after a work-day's noise exposure. *Scandinavian Audiology*, *23*(3), 159–163. https://doi.org/10.3109/01050399409047502

Ki, J. J., Kelly, S. P., & Parra, L. C. (2016). Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *Journal of Neuroscience*, *36*(10), 3092–3101. https://doi.org/10.1523/JNEUROSCI.2942-15.2016

Kuhlen, A. K., Allefeld, C., & Haynes, J. D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience*, *6*(SEPTEMBER), 1–15. https://doi.org/10.3389/fnhum.2012.00266

Kuijpers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. M. (2014). Exploring absorbing reading experiences: Developing and validating a self-report scale to measure story world absorption. *Scientific Study of Literature*, *4*(1), 89–122. https://doi.org/10.1075/ssol.4.1.05kui

Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., & Stefanovska, A. (2018). Surrogate data for hypothesis testing of physical systems. *Physics Reports*, *748*, 1–60. https://doi.org/10.1016/j.physrep.2018.06.001

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data

1   acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442.

2   https://doi.org/10.3758/s13428-016-0727-z

3   MacPherson, A., & Akeroyd, M. A. (2013). The Glasgow Monitoring of Uninterrupted Speech Task

4   (GMUST): A naturalistic measure of speech intelligibility in noise. *Proceedings of Meetings on*

5   *Acoustics*, *19*, 1–6. https://doi.org/10.1121/1.4799865

6   Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent component analysis of

7   electroencephalographic signals. In M. M. and M. H. D. Touretzky (Ed.), *Advances in neural*

8   *information processing systems 8* (pp. 145–151). MIT Press.

9   https://doi.org/10.1109/ICOSP.2002.1180091

10 Mandler, J. M., & Goodman, M. S. (1982). On the psychological validity of story structure. *Journal of*

11   *Verbal Learning and Verbal Behavior*, *21*(5), 507–523. https://doi.org/10.1016/S0022-

12   5371(82)90746-0

13 Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social

14   experience. *Perspectives on Psychological Science*, *3*(3), 173–192. https://doi.org/10.1111/j.1745-

15   6924.2008.00073.x

16 Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior*

17   *Research Methods*, *44*(1), 1–23. https://doi.org/10.3758/s13428-011-0124-6

18 Matthen, M. (2016). Effort and displeasure in people who are hard of hearing. *Ear and Hearing*, *37*, 28S-

19   34S. https://doi.org/10.1097/AUD.0000000000000292

20 Näätänen, R., & Picton, T. W. (1987). The N1 wave of the human electric and magnetic response to

21   sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425.

22   https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

1   Naci, L., Cusack, R., Anello, M., & Owen, A. M. (2014). A common neural code for similar conscious

2   experiences in different individuals. *Proceedings of the National Academy of Sciences of the United*

3   *States of America*, *111*(39), 14277–14282. https://doi.org/10.1073/pnas.1407007111

4   Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across

5   subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, *14*(6), 669–

6   687. https://doi.org/10.1093/scan/nsz037

7   Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with

8   shared neural responses. *NeuroImage*, *184*, 161–170.

9   https://doi.org/10.1016/j.neuroimage.2018.09.010

10   Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012).

11   Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings*

12   *of the National Academy of Sciences of the United States of America*, *109*(24), 9599–9604.

13   https://doi.org/10.1073/pnas.1206095109

14   Oatley, K. (2016). Fiction: Simulation of Social Worlds. *Trends in Cognitive Sciences*, *20*(8), 618–628.

15   https://doi.org/10.1016/j.tics.2016.06.002

16   Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain

17   regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*,

18   *27*(9), 2283–2289. https://doi.org/10.1523/JNEUROSCI.4663-06.2007

19   Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions: A

20   summary of the Pearson, Bennett, & Fidell (1977) report. *American Journal of Audiology*, *7*(2), 21–

21   25. https://doi.org/10.1044/1059-0889(1998/012)

22   Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for

23   advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational*

Intelligence and Neuroscience, *2011*. https://doi.org/10.1155/2011/156869

Palmer, A. D., Newsom, J. T., & Rook, K. S. (2016). How does difficulty communicating affect the social relationships of older adults? An exploration using data from a national survey. *Journal of Communication Disorders*, *62*, 131–146. https://doi.org/10.1016/j.jcomdis.2016.06.002

Parra, L. C., Haufe, S., & Dmochowski, J. P. (2019). Correlated components analysis - Extracting reliable dimensions in multivariate data. *ArXiv*. http://arxiv.org/abs/1801.08881

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage*, *28*(2), 326–341. https://doi.org/10.1016/j.neuroimage.2005.05.032

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Piazza, E. A., Cohen, A., Trach, J., & Lew-Williams, C. (2021). Neural synchrony predicts children's learning of novel words. *Cognition*, *214*(July 2020), 104752. https://doi.org/10.1016/j.cognition.2021.104752

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

1 Pichora-Fuller, M. K., Mick, P., & Reed, M. (2015). Hearing, cognition, and healthy aging: Social and

2   public health implications of the links between age-related declines in hearing and cognition.

3   *Seminars in Hearing*, *36*(3), 122–139. https://doi.org/10.1055/s-0035-1555116

4 Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and

5   remember speech in noise. *Journal of the Acoustical Society of America*, *97*(1), 593–608.

6   https://doi.org/10.1121/1.412282

7 Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses.

8   *International Journal of Audiology*, *42*(4), 177–219. https://doi.org/10.3109/14992020309101316

9 Polonenko, M. J., & Maddox, R. K. (2021). Exposing distinct subcortical components of the auditory

10   brainstem response evoked by continuous naturalistic speech. *ELife*, *10*, e62329.

11   https://doi.org/10.7554/eLife.62329

12 Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., & Hansen, L. K. (2017). EEG in the

13   classroom: Synchronised neural recordings during video presentation. *Scientific Reports*, *7*, 1–9.

14   https://doi.org/10.1038/srep43916

15 Power, A. J., Foxe, J. J., Forde, E. J., Reilly, R. B., & Lalor, E. C. (2012). At what time is the cocktail

16   party? A late locus of selective attention to natural speech. *European Journal of Neuroscience*,

17   *35*(9), 1497–1503. https://doi.org/10.1111/j.1460-9568.2012.08060.x

18 Reitan, R. M., & Wolfson, D. (2000). Conation: A neglected aspect of neuropsychological functioning.

19   *Archives of Clinical Neuropsychology*, *15*(5), 443–453. https://doi.org/10.1016/S0887-

20   6177(99)00043-8

21 Richter, M. (2013). A closer look into the multi-layer structure of motivational intensity theory. *Social*

22   *and Personality Psychology Compass*, *7*, 1–12. https://doi.org/10.1111/spc3.12007

1   Richter, M. (2016). The moderating effect of success importance on the relationship between listening

2   demand and listening effort. *Ear and Hearing*, *37*, 111S-117S.

3   https://doi.org/10.1097/AUD.0000000000000295

4   Rosenkranz, M., Holtze, B., Jaeger, M., & Debener, S. (2021). EEG-Based Intersubject Correlations

5   Reflect Selective Attention in a Competing Speaker Scenario. *Frontiers in Neuroscience*, *15*(June),

6   1–12. https://doi.org/10.3389/fnins.2021.685774

7   Schmälzle, R., Häcker, F. E. K., Honey, C. J., & Hasson, U. (2015). Engaged listeners: Shared neural

8   processing of powerful political speeches. *Social Cognitive and Affective Neuroscience*, *10*(8),

9   1137–1143. https://doi.org/10.1093/scan/nsu168

10   Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound

11   scenarios. *Journal of the American Academy of Audiology*, *26*(2), 183–196.

12   https://doi.org/10.3766/jaaa.26.2.7

13   Song, H., Finn, E. S., & Rosenberg, M. D. (2021). Neural signatures of attentional engagement during

14   narratives and its consequences for event memory. *Proceedings of the National Academy of*

15   *Sciences of the United States of America*, *118*(33). https://doi.org/10.1073/pnas.2021905118

16   Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful

17   communication. *Proceedings of the National Academy of Sciences of the United States of America*,

18   *107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107

19   Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion

20   methods and interactive experiments. *Computers in Human Behavior*, *77*, 184–197.

21   https://doi.org/10.1016/j.chb.2017.08.038

22   Wayne, R. V., & Johnsrude, I. S. (2015). A review of causal mechanisms underlying the link between

23   age-related hearing loss and cognitive decline. *Ageing Research Reviews*, *23*, 154–166.

1 https://doi.org/10.1016/j.arr.2015.06.002

2 Westgate, E. C., & Wilson, T. D. (2018). Boring thoughts and bored minds: The MAC model of boredom

3 and cognitive engagement. In *Psychological Review* (Vol. 125, Issue 5, pp. 689–713). American

4 Psychological Association. https://doi.org/10.1037/rev0000097

5 Wild, C. J., Davis, M. H., & Johnsrude, I. S. (2012). Human auditory cortex is sensitive to the perceived

6 clarity of speech. *NeuroImage*, *60*(2), 1490–1502. https://doi.org/10.1016/j.neuroimage.2012.01.035

7 Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful

8 listening: The processing of degraded speech depends critically on attention. *Journal of*

9 *Neuroscience*, *32*(40), 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

10 Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: Intersubject

11 correlations in narrative speech comprehension. *Cerebral Cortex*, *18*(1), 230–242.

12 https://doi.org/10.1093/cercor/bhm049

13 Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate

14 web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*(7), 2064–2072.

15 https://doi.org/10.3758/s13414-017-1361-2

16 Wright, R. A. (2014). Presidential address 2013: Fatigue influence on effort-considering implications for

17 self-regulatory restraint. *Motivation and Emotion*, *38*, 183–195. https://doi.org/10.1007/s11031-014-

18 9406-5

19 Xia, J., Kalluri, S., Micheyl, C., & Hafter, E. (2017). Continued search for better prediction of aided

20 speech understanding in multi-talker environments. *The Journal of the Acoustical Society of*

21 *America*, *142*(4), 2386–2399. https://doi.org/10.1121/1.5008498

22 Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same

1 story, different story: The neural representation ofiInterpretive frameworks. *Psychological Science*,

2 *28*(3), 307–319. https://doi.org/10.1177/0956797616682029

3 Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse

4 comprehension. *Psychonomic Bulletin and Review*, *23*(4), 1028–1034.

5 https://doi.org/10.3758/s13423-015-0864-x

6 Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in

7 narrative comprehension: An event-indexing model. *Psychological Science*, *6*(5), 292–297.

8 https://doi.org/10.1111/j.1467-9280.1995.tb00513.x

9   **Figure Legends**

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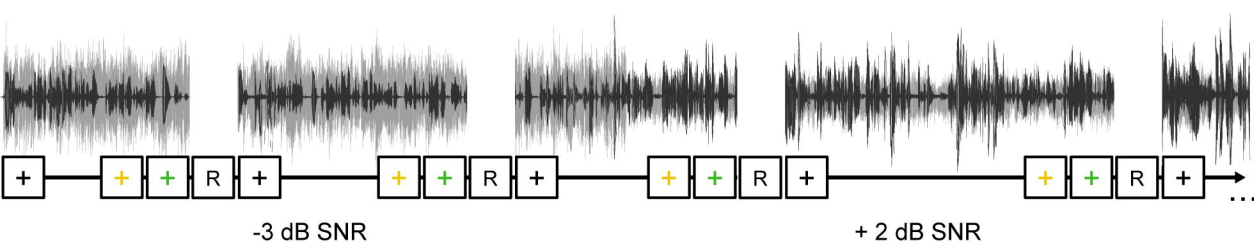17 resumed once a response was submitted.

18 **Figure 2. Behavioral measures of story comprehension, engagement, and intelligibility.** (a) Mean
19 ratings of the statements comprising the narrative absorption scale (NAS), and those assessing enjoyment
20 and motivation. Ratings were provided on a 7-point Likert scale, where '1' refers to completely disagree
21 and '7' refers to completely agree. (b) Mean proportion of correctly reported words plotted as a function of
22 SNR (clear, +12, +7, +2, -3 dB). (c) Correlation matrix depicting the correlation (Pearson) between
23 behavioral measures of story comprehension, engagement, and intelligibility. Error bars reflect the standard
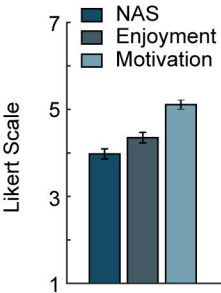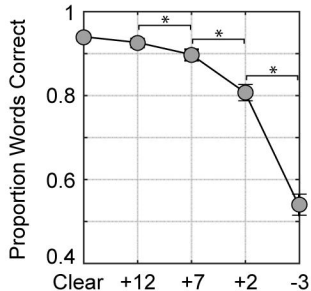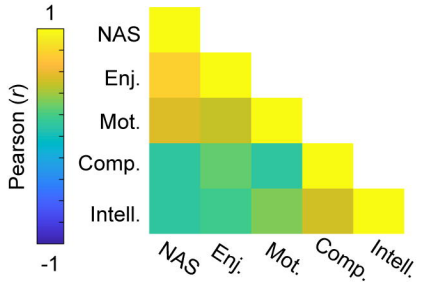24 error of the mean. *$p < 0.05$.

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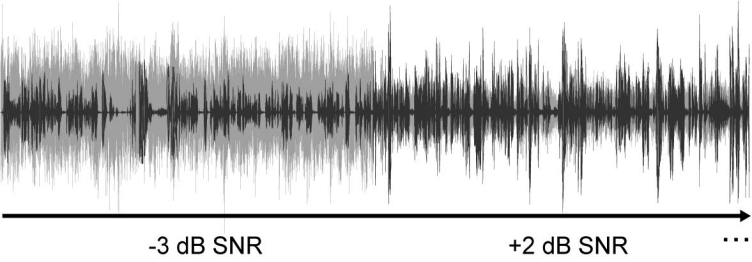27 SNR lasts for ~30-33 sec (see text for details).

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

1 **Figure 5. Inter-subject correlation as a function of SNR.** (a) Mean ISC is plotted as a function of SNR
2 (clear, +12,+7,+2,-3 dB) for each component (Component 1, Component 2, Component 3). Error bars
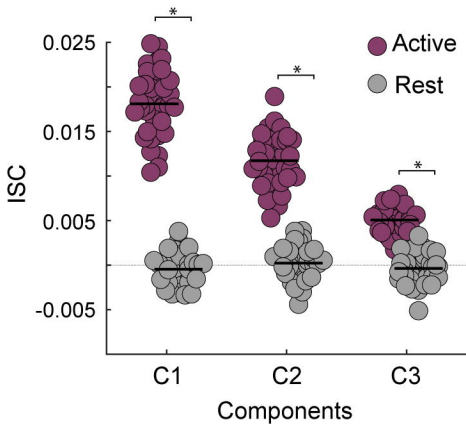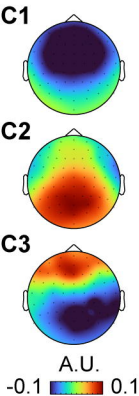3 reflect the standard error of the mean. *p < 0.05.

4 **Figure 6. Normalized measures of intelligibility and inter-subject correlation.** (a) Mean normalized
5 word report (Exp 1) and ISC (Exp 2: Component 1) are plotted as a function of SNR (clear, +12, +7, +2, -
6 3 dB). Solid lines correspond to an exponential function fit to normalized means. Mean tau coefficients
7 (decay strength) are plotted as a function of experiment measure (intelligibility, ISC). (b) Mean normalized
8 word report (Exp 1) and ISC (Exp 2: Component 1) are replotted as a function of SNR (clear, +12, +7, +2,
9 -3 dB). Solid lines correspond to a quadratic function fit to normalized means. Mean quadratic (left plot)
10 and linear coefficients (right plot) are plotted as a function of experiment measure (intelligibility, ISC). (c)
11 The change in normalized word-report (z-score) and ISC for each SNR relative to clear. Error bars reflect
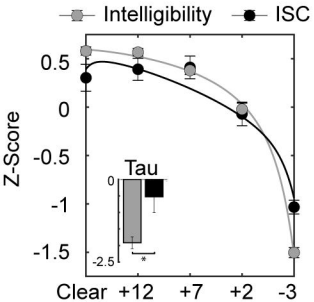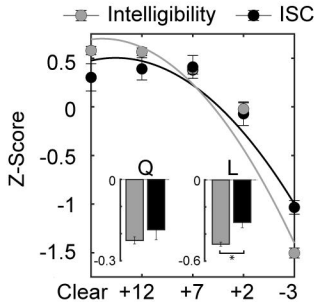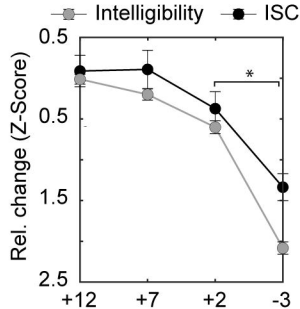12 standard error of the mean. *p < 0.05.

13

-3 dB SNR   + 2 dB SNR

a Engagement

Likert Scale

- NAS
- Enjoyment
- Motivation

b Intelligbility

Proportion Words Correct

Clear +12 +7 +2 -3

c Correlations

Pearson (r)

NAS
Enj.
Mot.
Comp.
Intell.

NAS Enj. Mot. Comp. Intell.

-3 dB SNR   +2 dB SNR   ...

Overall inter-subject correlation

a Exponential fit   **b** Quadratic fit   **c** Relative change

ISC by signal-to-noise ratio

Component 1 | Component 2 | Component 3