Absorption and Enjoyment During Listening to Acoustically Masked Stories

Björn Herrmann[1,2,3] (iD) and Ingrid S. Johnsrude[3,4]

Abstract

Comprehension of speech masked by background sound requires increased cognitive processing, which makes listening effortful. Research in hearing has focused on such challenging listening experiences, in part because they are thought to contribute to social withdrawal in people with hearing impairment. Research has focused less on positive listening experiences, such as enjoyment, despite their potential importance in motivating effortful listening. Moreover, the artificial speech materials—such as disconnected, brief sentences—commonly used to investigate speech intelligibility and listening effort may be ill-suited to capture positive experiences when listening is challenging. Here, we investigate how listening to naturalistic spoken stories under acoustic challenges influences the quality of listening experiences. We assess absorption (the feeling of being immersed/engaged in a story), enjoyment, and listening effort and show that (a) story absorption and enjoyment are only minimally affected by moderate speech masking although listening effort increases, (b) thematic knowledge increases absorption and enjoyment and reduces listening effort when listening to a story presented in multitalker babble, and (c) absorption and enjoyment increase and effort decreases over time as individuals listen to several stories successively in multitalker babble. Our research indicates that naturalistic, spoken stories can reveal several concurrent listening experiences and that expertise in a topic can increase engagement and reduce effort. Our work also demonstrates that, although listening effort may increase with speech masking, listeners may still find the experience both absorbing and enjoyable.

The speech that listeners encounter in everyday life—such as during conversations in trains, cars, schools, stores, restaurants, hospitals, or urban outdoor areas—is often masked by background sound (Hodgson et al., 2007; Olsen, 1998; Rusnock & Bush, 2012; Smeds et al., 2015). Background sound degrades the speech signal and requires a listener to recruit additional cognitive resources to understand what is being said (Johnsrude & Rodd, 2016; Mattys et al., 2012). Such cognitive investment—drawing on attention, memory, and prior knowledge—makes listening to masked speech effortful (Eckert et al., 2016; Peelle, 2018; Pichora-Fuller et al., 2016; Shenhav et al., 2017; Westbrook & Braver, 2015). Basic and applied auditory researchers increasingly recognize that effort experienced during speech listening has diagnostic value because it may capture interindividual differences

in behavior that are not explained by traditional hearing assessments such as pure-tone audiometry and speech-in-noise intelligibility testing (Herrmann & Johnsrude, 2020; Lemke & Besser, 2016; Pichora-Fuller et al., 2016; Strauss & Francis, 2017).

[1]Rotman Research Institute, Baycrest, Toronto, Ontario, Canada
[2]Department of Psychology, University of Toronto, Toronto, Ontario, Canada
[3]Department of Psychology, University of Western Ontario, London, Canada
[4]School of Communication Sciences & Disorders, University of Western Ontario, London, Canada

Corresponding Author:
Björn Herrmann, Rotman Research Institute, Baycrest, 3560 Bathurst St, North York, ON M6A 2E1, Canada.
Email: herrmann.b@gmail.com

In spite of enthusiasm in the hearing science community for measuring listening effort (Eckert et al., 2016; Peelle, 2018; Pichora-Fuller et al., 2016), progress to date has been limited: The materials used to investigate speech intelligibility and listening effort do not reflect what people listen to in their everyday lives (Herrmann & Johnsrude, 2020). In a typical study, participants listen to brief, isolated sentences, such as 'Big dogs can be dangerous', and respond behaviorally after each sentence, for example, reporting back the words heard or rating experienced effort (Alhanbali et al., 2017; Davis & Johnsrude, 2003; Duncan & Aarts, 2006; Lunner & Sundewall-Thorén, 2007; Wendt et al., 2016; Zekveld et al., 2010). Brief, isolated sentences may not reflect what people actually experience during listening in real life, including positive experiences (e.g., enjoyment, satisfaction; Matthen, 2016) and negative experiences (e.g., effort, fatigue; McGarrigle et al., 2014; Pichora-Fuller et al., 2016). Investigating speech processing with isolated, not very interesting sentences may thus not fully capture the processes recruited during speech listening in the real world.

In everyday life, people encounter spoken narratives and stories in form of gossip, anecdotes, and event descriptions (among other forms) that support understanding the world and ourselves, convey cultural history, and enable social connection (Bamberg, 2010; Dunlop & Walker, 2013; Graesser et al., 2002; Mar & Oatley, 2008; Ryan, 2007). In other words, normal speech is personally meaningful, embedded in a broader context, and follows some topical narrative (Dunlop & Walker, 2013). A listener is usually intrinsically motivated to understand and to follow a spoken narrative but may be less motivated to listen to the isolated sentences used in labs and clinics.

Progress in understanding why some people disengage while in challenging listening situations whereas others persist and continue to engage (Heffernan et al., 2016) has perhaps been limited because research has focused on listening effort and other aversive listening experiences and not on positive experiences, such as enjoyment (but see Matthen, 2016). People seek social interactions in bars, cafes, and restaurants because they enjoy them or experience other positive benefits (Matthen, 2016). A focus on listening effort may not do justice to the complexity of an individual's listening experience (Nabi & Krcmar, 2004; Wright et al., 2003). How other listening experiences besides effort are affected by the presence of background sound when listening to speech is currently unknown.

Literature and media-studies researchers have extensively investigated how individuals engage with (mostly written) narratives and stories, and what factors contribute to narrative enjoyment (Albrecht & O'Brien, 1993; Bilandzic & Busselle, 2017; Busselle & Bilandzic, 2008,

2009; Green et al., 2004; Kuijpers et al., 2014; Oatley, 1999). A recently developed scale captures engagement in reading experiences as *story world absorption* using psychological and folk-psychological constructs along four dimensions (Kuijpers et al., 2014). In this article, we use the dimensional definitions of Kuijpers et al. (2014): (a) *Attention* refers to the feeling of losing awareness and concentrating deeply; (b) *Mental imagery* refers to visualizing settings, characters, and situations in one's mind; (c) *Transportation* refers to the feeling of entering the story world and of being in the story (Green & Brock, 2000; Green et al., 2004); and (d) *Emotional engagement* captures feeling with and for characters: This dimension is related to empathy and to the identification with others (Cohen, 2001; Cohen & Tal-Or, 2017).

Engagement is a term used in psychology and cognitive neuroscience to describe the recruitment of resources for an activity (Herrmann & Johnsrude, 2020; Westgate & Wilson, 2018). Here, we use *engagement* as a superordinate term capturing a range of immersion experiences with a narrative or story, including the four dimensions captured by the Kuijpers et al. (2014) scale. We use *engagement* synonymously with *absorption* because we use the narrative absorption scale (NAS; Kuijpers et al., 2014) to investigate whether and how engagement with spoken narratives and stories is affected by masking background sound.

Narrative engagement has been described as involving the creation and updating of mental models that represent characters, goals, actions, situations, and so forth while a story unfolds (Albrecht & O'Brien, 1993; Busselle & Bilandzic, 2009; Zwaan, 2016; Zwaan et al., 1995). Creating and updating mental models depends on the extent to which incoming story information can be incorporated into the person's existing knowledge (Busselle & Bilandzic, 2009; Gerrig & Mumper, 2017; Green, 2004). Neural engagement measures (i.e., neural synchrony across individuals) derived from electroencephalography recordings indicate that repetition of audiovisual narratives reduces engagement (Dmochowski et al., 2012; Ki et al., 2016; Poulsen et al., 2017), suggesting that it is not exact knowledge about a narrative that supports engagement, but rather more general thematic knowledge. *Thematic knowledge* refers to broad knowledge of, and familiarity with, a circumscribed area, topic or theme that provide structure and meaning to experiences (see also DeSantis & Ugarriza, 2000; Hjørland, 2001). For example, prior knowledge about and experiences with homosexuality has been shown to facilitate transportation into a written story about a homosexual person (Green, 2004). That thematic knowledge and expertise also alleviates listening effort and increases listening engagement when

individuals listen to acoustically masked stories seems likely but has not been demonstrated.

Finally, differences in listening experiences between two people with similar performance on traditional hearing assessments may become particularly apparent when individuals listen to masked speech for a longer time (Phillips, 2016). The cognitive investment required to understand masked speech over an extended period may increase effort and lead to fatigue (McGarrigle et al., 2014), which, in turn, may lead to disengagement from listening. In fact, the degree to which an individual is able to engage persistently in a cognitively challenging activity may be a crucial factor for understanding listening behavior in real life (Phillips, 2016; Reitan & Wolfson, 2000, 2004). Yet, listening to masked speech over longer periods may not only have aversive effects. Individuals can adapt to some forms of degraded speech (i.e., noise-vocoded speech), leading to increased speech intelligibility as exposure to degraded speech progresses (Davis et al., 2005; Eisner et al., 2010; Erb et al., 2013; Erb & Obleser, 2013; Huyck & Johnsrude, 2012; Samuel & Kraljic, 2009). Whether and how prolonged listening to masked stories over a period of many minutes affects listening effort, absorption, enjoyment, and comprehension is unknown.

In the current study, we conduct a series of experiments using engaging, spoken stories under different degrees of masking to investigate the relationship between listening effort and story absorption. Experiment 1 investigates whether an NAS (Kuijpers et al., 2014) is sufficiently sensitive to engagement with spoken stories. Experiment 2 tests whether masking spoken stories with multitalker babble affects absorption. In Experiment 3, we examine whether thematic knowledge supports absorption and reduces effort during story listening. Experiment 4 seeks to answer whether absorption and listening effort change over time, when individuals listen to masked stories for a longer period.

Methods and Materials

Participants

Participants were recruited from the undergraduate student, graduate student, and postdoc population at the University of Western Ontario, Canada. They gave written informed consent prior to the experiment and received course credits or were paid $5 CAD per half hour for their participation. Demographic information about participants is provided separately for each of the following experiment. All participants self-reported normal hearing abilities. Participants were either native English speakers or nonnative speakers who were proficient English speakers—they all rated their English skills to be 5 or higher on a 7-point scale (note that we account for the nativeness of our participants during data analysis). The study was conducted in accordance with the Declaration of Helsinki, the Canadian Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2-2014), and was approved by the local Nonmedical Research Ethics Board of the University of Western Ontario (protocol ID: 106570).

Assessment of Story Engagement

In the current study, we administered an NAS that was developed to assess experiences with written narratives (Kuijpers et al., 2014) and that integrates other previous scales related to narrative engagement (Busselle & Bilandzic, 2009; Cohen, 2001; Green et al., 2004). As there is currently no engagement scale available that is specifically designed to use with spoken (as opposed to written or audiovisual) narratives, we opted to slightly adapt the scale to assess listening experiences. This required changing the wording of the statements (e.g., 'reading' to 'listening') and discarding two statements. The remaining 18 statements were used in the current study (Table 1).

In addition to the four dimensions of the NAS (described in the Introduction section), we also assessed the participants' story enjoyment ($N = 5$ items from Kuijpers et al., 2014), listening effort ($N = 3$), and story comprehension ($N = 3$). Enjoyment may correlate with the NAS but is thought to reflect a separate experience (Bilandzic & Busselle, 2017). All statements used in the study are listed in Table 1. Statements were rated on a 7-point scale, where 1 referred to *completely disagree* and 7 to *completely agree*. All 29 items were used in Experiments 2 to 4. Effort and comprehension were not assessed in Experiment 1 because all stories were presented under clear conditions.

General Procedure

Participants were seated in a quiet room or a sound-attenuated booth. Stories were presented via Grado SR225 headphones and a Focusrite Scarlett 2i4 external sound card. Stimulation was controlled by a Laptop running Psychtoolbox in MATLAB software (Mathworks Inc.).

Participants listened to spoken stories whose duration ranged between 5 and 13 min depending on the experiment. After listening to a story, participants rated each of the statements in Table 1 on a 7-point scale (1 = *completely disagree*; 7 = *completely agree*). Rating scores were averaged separately for NAS statements, enjoyment statements, effort statements, and comprehension statements (Table 1). All data analyses described

Table 1. Items of the Narrative Absorption Scale (NAS; Capturing Four Dimensions), Enjoyment, Effort, and Comprehension.

Dimension	Question
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.
Effort	I had to invest effort to understand what was said.
Effort	It was difficult to understand what was said.
Effort	Understanding the speaker was hard.
Comprehension	I got the gist of the story.
Comprehension	I have a good sense of what the story was about.
Comprehension	The story was understandable.

in the following were carried out in MATLAB and IBM SPSS software.

Data will be made available upon reasonable request.

Experiment 1: Sensitivity of the NAS

The purpose of Experiment 1 was to test the sensitivity of the NAS (Kuijpers et al., 2014) for engaging relative to less engaging spoken stories.

Participants

Thirty-four young, normal-hearing adults (mean age: 20.9 years; age range: 18–28 years; 26 females; 28 native English speakers) participated in Experiment 1.

Methods

We selected five ∼5- to 6-min stories with broad appeal from the storytelling podcast 'The Moth' (https://themoth.org/). The following were the stories: _Things I Knew For Sure_ (by Qing Zhao), _Fearless_ (by Lydia Velez), _A Shoulder Bag to Cry On_ (by Laura Zimmermann), _I Am Batman_ (by Paul Davis), and _A Monkey Meets a Seal Meets a Monkey_ (by Matthew McArthur). Each participant listened to two of these _The Moth_ stories. Each participant also listened to a third ∼5-min story, which was an excerpt from the _Sleep With Me_ podcast (https://www.sleepwithmepodcast.com/) intended to help people fall asleep: That is, its purpose is to help listeners disengage. The root-mean-square amplitude was matched across all six stories. We hypothesized that the _The Moth_ stories would be more

absorbing compared with the *Sleep* story. For each participant, the order of the three stories was randomized. After each story, participants rated the statements of the NAS and the enjoyment statements. Statements were presented in a different, randomized order for each participant.

We also assessed a participant's motivation to listen to a story while a story was presented to them. This involved presenting the statement 'I am keen to hear how the story evolves' on the screen at regular intervals (about every 1.5 min) throughout a story. Participants rated the statement on a 7-point scale, where 1 referred to *completely disagree* and 7 to *completely agree*. Rating scores for listening motivation were averaged across each whole story.

Paired-samples *t* tests were calculated to compare motivation, absorption, and enjoyment between the *Sleep* story and *The Moth* stories (averaged across the two The Moth stories to which each participant listened).

Results and Discussion

Figure 1A shows motivation, absorption, and enjoyment for the *Sleep* story and *The Moth* stories. Paired-samples *t* tests show that motivation, absorption, and enjoyment—all three tests: $t(33) > 9$, $p < 1 \times 10^{-10}$; Figure 1A—were rated higher for *The Moth* stories compared with the *Sleep* story. These differences were also observed when we limited our analysis to native English speakers—all $t(27) > 9$, $p < 1 \times 10^{-9}$. Figure 1B shows the difference score for each of the five *The Moth* stories relative to the *Sleep* story. Almost every participant rated *The Moth* stories to be more motivating,

absorbing, and enjoyable compared with the "Sleep" story.

We also calculated correlations among the three dependent measures: motivation correlated with absorption ($r = .549$, $p = .0008$) and enjoyment ($r = .815$, $p = 4.6 \times 10^{-9}$), and absorption correlated with enjoyment ($r = .730$, $p = 9.7 \times 10^{-7}$).

Experiment 1 demonstrates that the NAS (Kuijpers et al., 2014) and the motivation and enjoyment measures are sensitive to the difference between spoken The Moth stories and the Sleep story. Our results motivate using these measures in subsequent experiments (in which we focus on absorption and enjoyment).

Experiment 2: Effects of Acoustic Masking on Story Absorption

Here, we examine how acoustic masking of an engaging, spoken story affects absorption, enjoyment, listening effort, and comprehension. Experiment 2 comprised two subexperiments, in which we evaluate how perception of spoken materials is affected by masking with 12-talker background babble at different signal-to-noise ratios (SNRs). Experiment 2a mimics standard speech-in-noise testing (Duncan & Aarts, 2006), and we measure behavioral intelligibility levels (as word-report error) to short snippets from a story. In Experiment 2b, we investigate how absorption, enjoyment, effort, and comprehension of a coherent story are affected by masking.

Participants

Twenty-seven young, normal-hearing adults participated in Experiment 2a (mean age: 19.5 years;
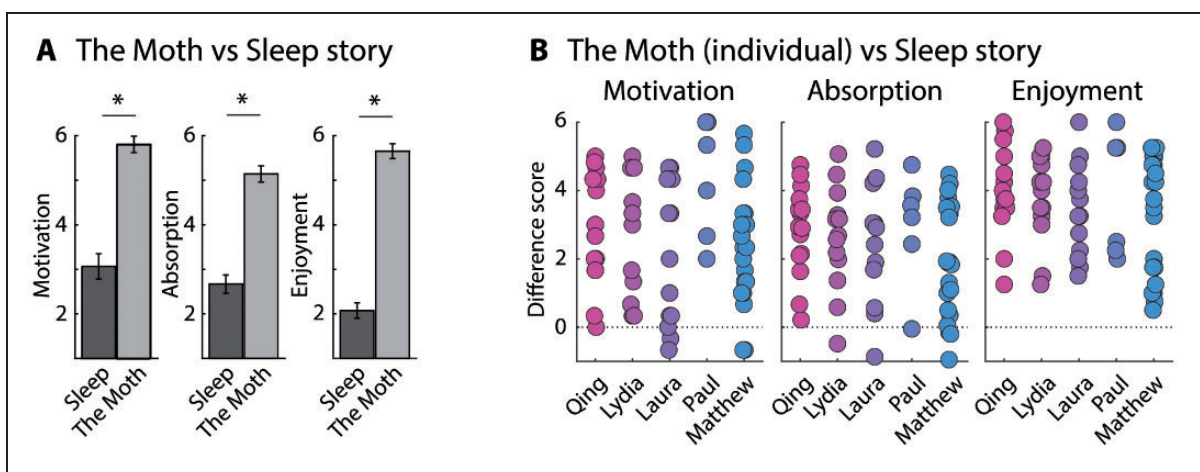

Figure 1. Rating Scores for The Moth Stories Versus the Sleep Story. A: Motivation, absorption, and enjoyment for *The Moth* stories (averaged across the two *The Moth* stories) relative to the *Sleep* story. All three measures were larger for *The Moth* stories than the *Sleep* story (*$p \leq .05$). B: Difference between scores for *The Moth* stories and scores for the *Sleep* story, separately for each *The Moth* story. Dots reflect data from individual participants. The dashed black line indicates no difference between the *The Moth* story and the *Sleep* story.

age range: 18–25 years; 10 females; 20 native English speakers). Eighty-eight different young, normal-hearing adults participated in Experiment 2b (mean age: 23.3 years; age range: 17–37 years; 55 females; 64 native English speakers). None of the participants who participated in Experiment 2a or 2b took part in Experiment 1.

Methods

The 5-min *The Moth* story by Laura Zimmermann (title: *A Shoulder Bag to Cry On*; see Figure 1B, third story) was selected for Experiments 2a and 2b: The story is about an American couple losing their passports in Portugal. Three story conditions were created: The story without background sound ('clear'), with added 12-talker babble at $+12$ dB SNR, and with added 12-talker babble at $+4$ dB SNR. The SNR levels chosen simulate conditions listeners might encounter in situations of everyday life, such as trains, schools, hospitals, and so forth (Olsen, 1998; Smeds et al., 2015). The root-mean-square amplitude was matched between story conditions.

For Experiment 2a, participants listened to 66 short snippets extracted from the story. Snippets consisted of words ($N = 1$), phrases ($N = 36$), or sentences ($N = 29$) that had a median duration of 1.407 s (minimum: 0.765 s; maximum: 2.982 s) and a median number of words of 5 (minimum: 1 word; maximum: 9 words). The speech snippets were assigned randomly to the 'clear', $+12$ dB SNR, and $+4$ dB SNR conditions (with equal proportions; $N = 22$ each). Assignment of snippets to individual speech conditions ('clear', $+12$ dB SNR, $+4$ dB SNR) was counterbalanced across participants such that each snippet was heard in each speech condition an equal number of times. Speech snippets for the different conditions were presented in random order (uniquely for each participant) that differed from the original order of the story (however, participants reported noticing that the snippets belonged to a coherent story). After each speech snippet, participants reported what they heard by typing it into the computer using the keyboard. Different or omitted words were counted as errors; misspellings, incorrect tenses, and incorrect grammatical number (singular vs. plural) were not. The proportion of errors was calculated. An analysis of variance (ANOVA) with the within-subject factor Condition (clear, $+12$ dB SNR, $+4$ dB SNR) and the between-subject factor Nativeness (native, nonnative English speaker) was calculated using the proportion of word-report errors as a dependent measure. The Greenhouse and Geisser correction was used when Mauchly's test of sphericity was violated (Greenhouse & Geisser, 1959). Paired-samples *t* tests were used to further resolve significant effects of Condition.

For Experiment 2b, participants were randomly assigned to one of three speech conditions. Thirty participants listened to the story under clear conditions, 29 participants listened to the story with added 12-talker babble at $+12$ dB SNR, and 29 participants listened to the story with added 12-talker babble at $+4$ dB SNR. After listening to the story, participants rated the absorption, enjoyment, effort, and comprehension statements listed in Table 1. Statements were presented in a different, randomized order for each participant. Linear regression models were calculated separately to predict absorption, enjoyment, effort, and comprehension. Predictors for each regression were Condition (clear, SNR12, SNR4), Nativeness (native, nonnative), Sex (female, male), and Age (in years). Sex and age were included in the analysis because differences in engagement between men and women have been reported previously (Oatley, 1999; but see Green, 2004) and experienced effort may be higher for older people, even in our young sample (age range: 17–37 years). Significant effects of Condition were resolved using linear models with the same predictors but with only two levels for Condition. The statistical results were similar for analyses without nuisance regressors (see Supplemental Materials).

To assess whether absorption and enjoyment ratings correlated with each other and/or with effort ratings, we calculated three additional regressions. Absorption ratings were used to predict enjoyment ratings and effort ratings, and enjoyment ratings were used to predict effort ratings. Nativeness (native, nonnative), Sex (female, male), and Age (in years) were used as additional predictors in the linear models. To ensure that these analyses are not biased by any mean differences among speech conditions (clear, $+12$ dB SNR, $+4$ dB SNR), the mean rating across participants for a given speech condition was subtracted from the rating of each participant for that speech condition. Mean subtraction for each of the three speech conditions was calculated separately for each measure (absorption, enjoyment, effort) prior to the regression analyses.

Results and Discussion

Word-report errors for snippets extracted from an engaging story with different degrees of masking are shown in Figure 2 (Experiment 2a). As expected, word-report errors increased with decreasing SNR—main effect of Condition: $F(2,50) = 26.084$, $p = 1.4 \times 10^{-5}$; errors were greater for $+12$ dB SNR than clear: $t(26) = 3.609$, $p = .0013$; greater for $+4$ dB SNR than $+12$ dB SNR: $t(26) = 5.509$, $p = 9 \times 10^{-6}$; and greater for $+4$ dB SNR than clear: $t(26) = 6.161$, $p = 2 \times 10^{-6}$—and were higher for nonnative compared with native English speakers—main effect of Nativeness:
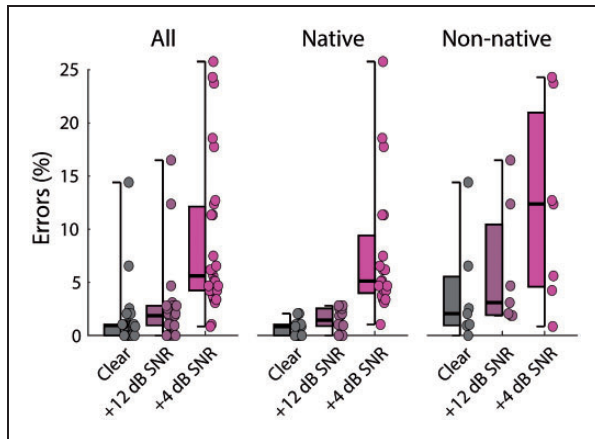
Figure 2. Word-Report Errors for Spoken Items at Different Signal-to-Noise Ratios (Experiment 2a). Box plots are displayed, and dots reflect data from individual participants.
SNR = signal-to-noise ratio.

$F(1,25) = 6.047$, $p = .021$. There was no interaction between Condition and Nativeness—$F(2,50) = 0.206$, $p = .676$. The absence of the interaction may in part be due to the low number of nonnative English speakers (7 out of 27). Note that the same pattern of results was observed using nonparametric statistics or data transformed using the rationalized arcsine transform (Studebaker, 1985; see Supplemental Materials). Test statistics for analyses limited to native English speakers show the same pattern (see Supplemental Materials).

Experiment 2a shows that for the most difficult speech condition (+4 dB SNR), the median word-report error was only 5.12% for native and 12.37% for nonnative English speakers (Figure 2, left). Thus, at this masking level and even for speech snippets with reduced speech context, native and nonnative English-speaking participants were able to understand more than 94% and 87% of the words, respectively (Figure 2, right). Clearly, the masker is not trivial—its presence degrades intelligibility—but these results set a lower bound on intelligibility for a story presented with a masker at +4 dB SNR. The enhanced context of a coherent story would probably result in higher intelligibility (Miller et al., 1951; Pickett & Pollack, 1963; Pollack & Pickett, 1963, 1964).

In Experiment 2b, different groups of people listened to the full story under one of the three masking conditions used in Experiment 2a (clear, +12 dB SNR, +4 dB SNR) and subsequently rated absorption, enjoyment, effort, and comprehension statements (Figure 3). Regression analyses did not reveal an effect of Condition on story absorption—$t(83) = 0.9303$, $p = .3549$; Figure 3, first column; none of the other predictors were significant, $p > .3$—suggesting that individuals are similarly absorbed by an engaging story under

clear conditions and moderate masking. Enjoyment was significantly affected by Condition—$t(83) = 2.1059$, $p = .0382$; Figure 3, second column; none of the other predictors were significant, $p > .6$. Enjoyment was rated lower for +12 dB SNR, $t(54) = 2.2676$, $p = .0274$, and +4 dB SNR, $t(54) = 2.1549$, $p = .0356$, compared with the clear condition. Enjoyment ratings for the +12 dB SNR and +4 dB SNR conditions did not differ, $t(53) = 0.042$, $p = .9667$. Effort significantly increased as SNR declined—$t(83) = 7.0985$, $p = 3.92 \times 10^{-10}$; Figure 3, third column; none of the other predictors were significant, $p > .4$—such that effort was rated higher for +12 dB SNR, $t(54) = 3.0977$, $p = .0031$, and +4 dB SNR, $t(54) = 7.389$, $p = 9.6 \times 10^{-10}$, than for clear, and higher for +4 dB SNR than +12 dB SNR, $t(53) = 3.5990$, $p = .0007$. Moreover, story comprehension was affected by Condition, $t(83) = 2.9322$, $p = .0043$; Figure 3, fourth column: Comprehension was rated lower for the +4 dB SNR condition compared with clear speech, $t(54) = 2.8242$, $p = .0066$. Comprehension was also rated lower by nonnative compared with native English speakers, $t(83) = 2.7953$, $p = .0064$; the other predictors were not significant, $p > .15$. The same pattern of results was observed when analyses were limited to native English speakers (see Supplemental Materials).

The results show that absorption is not affected, and enjoyment is only minimally affected, by masking of speech. Experienced effort, in contrast, increases strongly as SNR decreases. Our results also show that despite nonnative English speakers rating comprehension lower than native speakers (mirroring word-report errors in Experiment 2a), measures of absorption, enjoyment, and effort did not depend on Nativeness. Moreover, absorption and enjoyment did not differ between the +12 dB SNR and the +4 dB SNR conditions, whereas effort ratings were higher for the latter compared with the former. These results suggest that aversive experiences (listening effort) and positive experience (absorption, enjoyment) may be somewhat independent.

Figure 3 indicates high interindividual variability in rating scores, most prominently for absorption, enjoyment, and effort. To investigate whether absorption, enjoyment, and effort share variance that may explain some of the interindividual differences, correlations were calculated among these measures (after regressing out Nativeness, Sex, and Age). Figure 4 shows a strong correlation between absorption and enjoyment ($r = .758$, $p = 1.18 \times 10^{-17}$; mirroring Experiment 1), and moderate negative correlations between absorption and effort ($r = -.290$, $p = .0061$) and enjoyment and effort ($r = -.403$, $p = 9.88 \times 10^{-5}$). Note that these relations were also significant when the raw ratings were correlated or when analyses were limited to native English speakers (see Supplemental Materials).
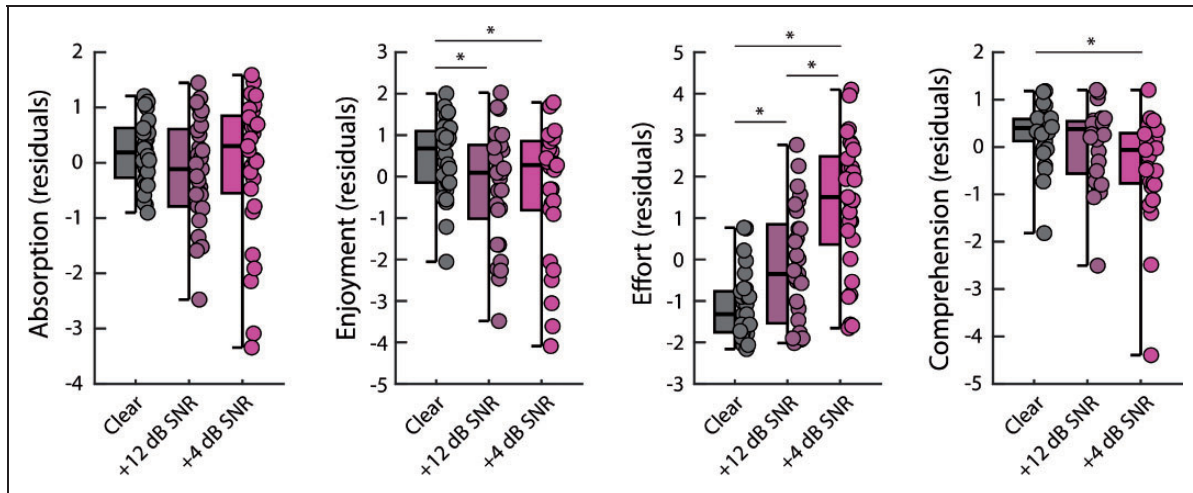
Figure 3. Ratings of Absorption, Enjoyment, Effort, and Comprehension for Stories Presented Under Different Degrees of Acoustic Degradation (Experiment 2b). Data reflect the residuals after Nativeness (native vs. nonnative English speaker), Sex (female, male), and Age were regressed out. That is, separately for each dependent variable of interest, we calculated regressions using predictors Nativeness, Sex, and Age and plotted the resulting residuals here. Note that analyses using the original ratings yielded the same statistical results because these regressors were mostly nonsignificant (see text). Box plots are displayed, and dots reflect data from individual participants. Significant differences are indicated by an asterisk. $*p \leq .05$.
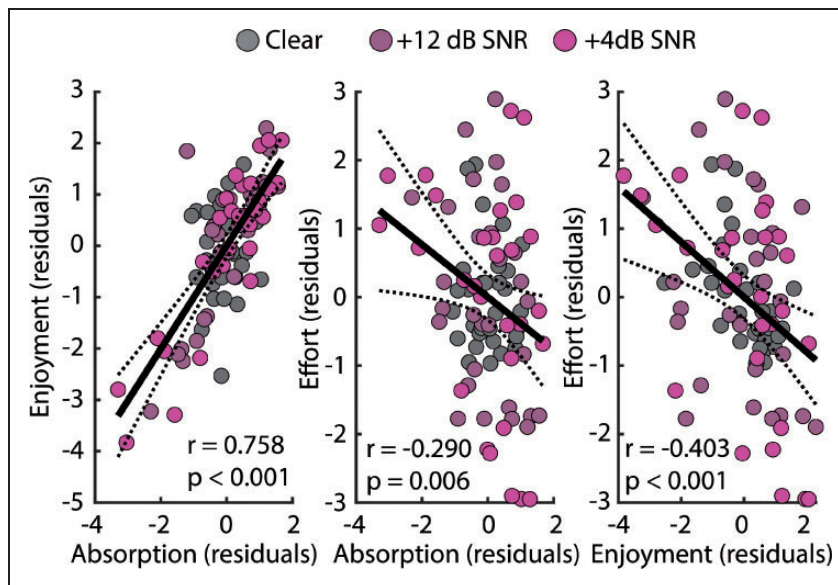SNR = signal-to-noise ratio.

Figure 4. Correlation Between Listening Experiences. Data reflect the residuals after regressing out Nativeness, Sex, and Age. That is, separately for each dependent variable of interest, we calculated regressions using predictors Nativeness, Sex, and Age and used the resulting residuals for the correlation plots. The solid line reflects the best fitting line. The dashed lines mark the 95% confidence intervals. All three correlations are significant, $p < .05$.
SNR = signal-to-noise ratio.

The correlation of effort ratings with absorption and enjoyment ratings may either indicate that absorption and enjoyment reduce listening effort or, alternatively, that less effort enables listeners to be more absorbed and enjoy listening. Alternatively, absorption and enjoyment may render a listener less likely to notice and report

effort. We cannot distinguish between these different potential relationships, but because only about 8% to 17% of variance in effort is explained by absorption or enjoyment, effort seems to be largely independent of these other dimensions. Moreover, given that absorption and enjoyment were not or only minimally affected by
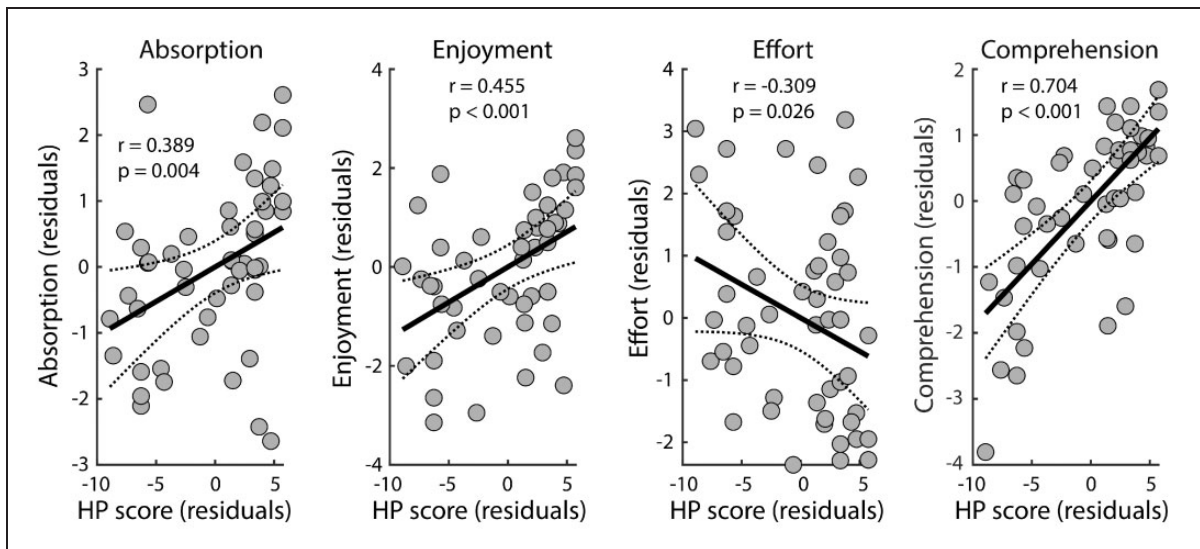
Figure 5. Correlation Between Harry Potter Score and Absorption, Enjoyment, Effort, and Comprehension. Data points reflect the residuals after regressing out Nativeness, Sex, and Age. The solid line reflects the best fitting line. The dashed lines mark the 95% confidence intervals. All correlations are significant, $p < .05$.
HP = Harry Potter.

speech masking (and were not affected by Nativeness), Experiment 2 results indicate that absorption and enjoyment are important dimensions of experience that cannot be reduced to changes in listening effort.

Experiment 3: The Effect of Thematic Knowledge on Story Absorption

Experiment 2b revealed that individuals find stories absorbing and enjoyable despite interference from masking and the experience of listening effort. The purpose of Experiment 3 was to explore whether thematic knowledge increases absorption and enjoyment and alleviates effort during story listening.

Participants

Fifty-two young, normal-hearing adults participated in Experiment 3 (mean age: 19.8 years; age range: 18–32 years; 28 females; 41 native English speakers). Fifteen of these participants also took part in Experiment 1, and 27 other participants also took part in Experiment 2a.

Methods

Participants listened to the audio of a 6-min audiovisual narrative summary of the first seven movies of the Harry Potter franchise (https://www.youtube.com/watch?v=TDnSdmznaTk). The summary is narrated by a male talker with an American accent, and the narrative incorporated short elements of the movies' soundtracks,

including sound effects, screams, speech, and music. Twelve-talker babble noise was added to the summary at +4 dB SNR (a level resulting in less than 6% or 13% word-report errors [native/nonnative English speakers] for speech snippets of a different story in Experiment 2a). After listening to the summary, participants rated the statements listed in Table 1 (absorption, enjoyment, effort, and comprehension). Participants also rated statements about liking Harry Potter ("I love the Harry Potter series.") and expertise with Harry Potter ("I have seen/watched most of the Harry Potter stories.") on a 7-point scale, where 1 referred to *completely disagree* and 7 to *completely agree*. The two rating scores we summed, leading to a "Harry Potter score" ranging from 2 to 14, with higher scores reflecting higher expertise with, and/or liking of, the Harry Potter series.

Four linear regression models were calculated to separately predict absorption, enjoyment, effort, and comprehension. Predictors were the Harry Potter score, Nativeness (native, nonnative), Sex (female, male), and Age (in years). To display correlations between the Harry Potter score and absorption, enjoyment, effort, and comprehension, we regressed out Nativeness, Sex, and Age, separately for each of the four metrics. That is, separately for each dependent variable of interest, we calculated regressions using predictors Nativeness, Sex, and Age and used the resulting residuals for correlation analyses. Correlations using the raw score/rating values revealed similar results compared with those incorporating nuisance regressors (see Supplemental Materials).

Results and Discussion

Regression analyses showed that absorption, $t(47) = 2.8948$, $p = .0057$; enjoyment, $t(47) = 3.5044$, $p = .001$; and comprehension, $t(47) = 6.7934$, $p = 1.68 \times 10^{-8}$, ratings significantly increased with increasing Harry Potter score and that effort ratings significantly decreased with increasing Harry Potter score, $t(47) = -2.2246$, $p = .0309$ (Figure 5). This pattern of results was also observed when we limited our analyses to data from native English speakers: absorption, $t(37) = 2.6977$, $p = .0105$; enjoyment, $t(37) = 3.4181$, $p = .0015$; and comprehension, $t(37) = 6.0876$, $p = 4.7 \times 10^{-7}$, ratings increased with increasing Harry Potter score and effort ratings decreased with increasing Harry Potter score, $t(37) = -2.8052$, $p = .008$.

These data show that thematic knowledge and expertise (without familiarity with the verbatim spoken narrative) can increase story absorption and enjoyment, reduce listening effort, and increase self-assessed comprehension.

Experiment 4: The Effect of Time on Story Absorption

Experiments 1 to 3 have shown that spoken stories are absorbing and enjoyable; that moderate interference from a masker increases listening effort but affects absorption and enjoyment only to a limited extent; and that thematic knowledge increases story absorption, enjoyment, and comprehension and reduces listening effort. In Experiments 2 and 3, listening times were quite short—each listener heard only one story, for a maximum of 6.5 min. In the real world, effortful listening can go on for much longer—over a meal in a restaurant for example. In Experiment 4, we investigated whether effort remains constant over multiple stories masked by multitalker babble, and, if so, whether this is related to declines in absorption and enjoyment.

Participants

Forty-eight young, normal-hearing adults participated in Experiment 4 (mean age: 22 years; age range: 18–34 years; 31 females; 34 native English speakers). Data from three additional participants were discarded due to technical problems during recording ($N = 2$) and due to insufficient demographic information to calculate regressions ($N = 1$). Participants did not take part in any of the other experiments.

Methods

We selected four stories from the storytelling podcast The Moth (https://themoth.org/). The following were the stories: *The Bounds of Comedy* (by Colm O'Regan; ~10 min), *Nacho Challenge* (by Omar Qureshi; ~11 min), *Microphone Uninhibited* (by Lydia Dubois; ~5.5 min), and *The Overview Effect* (by Richard Garriott; ~14 min). Two versions of each story were generated. Stories were used either in their original, clear version (no background sound) or with added 12-talker babble at +4 dB SNR (that resulted in less than 6% or 13% word-report errors [native/nonnative English speakers] for speech snippets of a different story in Experiment 2a). Clear stories and stories with added 12-talker babble were normalized to the same root-mean-square amplitude.

Participants were randomly assigned to the 'clear' group or the 'noise' group. Participants in the 'clear' group listened to all four stories under clear conditions. Participants in the 'noise' group listened to all four stories with added 12-talker babble at +4 dB SNR. Twelve-talker babble was only present while stories were played. The order of stories was counterbalanced across participants within each group such that each story was heard first, second, third, and fourth an equal number of times across both groups.

After each story, participants rated the absorption, enjoyment, effort, and comprehension statements listed in Table 1. We also assessed a participant's motivation to listen as each story progressed (similar to Experiment 1). This involved visually presenting the statement 'I am keen to hear how the story evolves' on the screen at regular intervals (about every 2 min) throughout a story, without stopping the story or the added babble in the 'noise' group. Participants rated the statement on a 7-point scale, where 1 referred to *completely disagree* and 7 to *completely agree*. Rating scores for listening motivation were averaged for each story. After each story, participants took a break of 2 to 3 min before initiating the next story.

For each measure (motivation, absorption, enjoyment, effort, and comprehension), an ANOVA was calculated using the within-subjects factor Story Number (first, second, third, fourth) and the between-subjects factor Group ('clear', 'noise'). Nativeness (native, nonnative), Sex (female, male), and Age were used as nuisance regressors. A significant Story Number × Group interaction was resolved using linear regressions with the predictor Group (clear, +4 dB SNR) and additional nuisance predictors Nativeness (native, nonnative), Sex (female, male), and Age. Regressions were calculated separately for each story number (first, second, third, fourth story). False discovery rate (FDR) was used to account for multiple comparisons (Benjamini & Hochberg, 1995; Genovese et al., 2002).

We further fit a linear function to rating scores as a function of story number (separately for each participant) to investigate whether rating scores changed over successive stories. The resulting slopes relating story

number (first, second, third, fourth story) to rating scores were tested against zero using a one-sample t test and FDR correction (Benjamini & Hochberg, 1995; Genovese et al., 2002). A positive slope means that rating scores increased over time, whereas a negative value means that rating scores decreased over time. To test group differences, a linear regression was calculated separately for each of the five metrics (motivation, absorption, enjoyment, effort, and comprehension) using the slope as a dependent measure. Predictors were Group ('clear', 'noise'), Nativeness (native, nonnative English speaker), Sex (female, male), and Age (in years).

Results and Discussion

Figure 6 shows motivation, absorption, enjoyment, effort, and comprehension ratings for the four stories presented successively. The ANOVAs revealed higher motivation, absorption, and enjoyment, and lower effort for participants listening to clear stories compared with those listening to stories with added babble—main effect of Group; for all: $F(1,43) > 4.5$, $p < .05$. There was no effect of Group for comprehension, $F(1,43) = 2.83$, $p = .1$. The Story Number × Group interaction was significant for motivation, absorption, and effort—for all: $F(3,129) > 2.9$; $p < .05$—but not for enjoyment and comprehension—for both: $F(3,129) < 2$, $p > .1$. There were no main effects of Story Number—for all: $F(3,129) < 2$, $p > .1$. The pattern of results was similar for analyses limited native English speakers (see Supplemental Materials).

Separate regression analyses for each story number revealed that motivation, $t(43) = -4.572$, $p = 4.06 \times 10^{-5}$; absorption, $t(43) = -4.051$, $p = 2.1 \times 10^{-4}$; and enjoyment, $t(43) = -3.904$, $p = 3.28 \times 10^{-4}$, were lower in the 'noise' group compared with the 'clear' group, but only for the first of four stories (FDR-thresholded). Moreover, listeners in the 'noise' group rated effort higher for the first three stories compared with listeners in the 'clear' group—first: $t(43) = 5.783$, $p = 7.57 \times 10^{-7}$; second: $t(43) = 2.633$, $p = .0117$; third: $t(43) = 2.823$, $p = .0072$; FDR-thresholded. Other tests were not significant. The data suggest that motivation, absorption, and enjoyment while listening to stories in multitalker babble increases over the course of four different stories (40 min), whereas effort decreases.

To examine changes in rating scores over time, a linear function was fit to ratings as a function of story number (first, second, third, and fourth story), independently for the 'clear' and 'noise' group. The slopes relating story number to motivation, $t(22) = 3.728$, $p = .001$; absorption, $t(22) = 2.568$, $p = .018$; and enjoyment, $t(22) = 2.601$, $p = .016$, ratings were significantly larger and effort ratings significantly smaller, $t(22) = -3.4151$, $p = .003$, than zero for the 'noise' group (FDR-thresholded). Slopes were not significantly different from zero for the 'clear' group and for story comprehension in either group (Figure 7). The slope was significantly more positive for motivation, $t(43) = 3.682$, $p = .0006$, and absorption, $t(43) = 2.704$, $p = .0098$; marginally significant for enjoyment, $t(43) = 1.807$, $p = .0777$; and more negative for effort, $t(43) = -3.1550$, $p = .0029$, for the 'noise' group compared with the 'clear' group
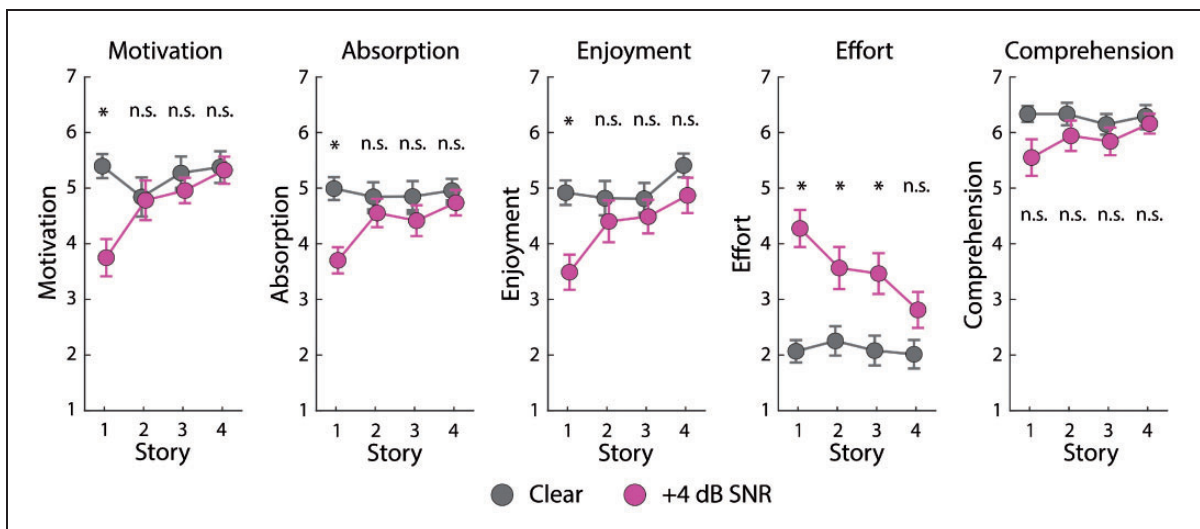

Figure 6. Rating Scores for Listening to Four Stories in Succession Under Clear or Background Babble Conditions. Between-group (clear vs. +4 dB SNR) effects were tested using regression analysis with Nativeness, Sex, and Age as additional regressors. FDR correction was used to account for multiple comparisons (Benjamini & Hochberg, 1995; Genovese et al., 2002). Error bars reflect the standard error of the mean. *$p \leq .05$.
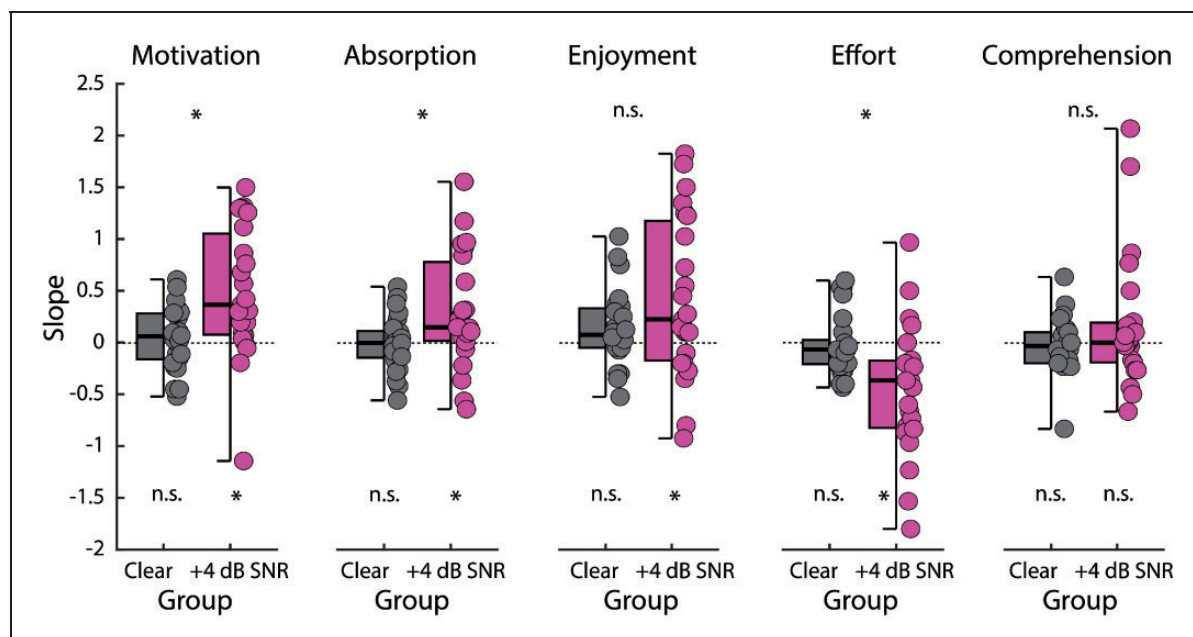SNR = signal-to-noise ratio; n.s. = not significant.

Figure 7. Slopes From Linear Fits Relating Story Number to Rating Scores. Positive values mean that rating scores increased with the number of stories, whereas negative values mean that rating scores decreased with the number of stories. Slopes were tested against zero using a one-sample *t* test (indicted below the box plots; FDR-thresholded; Benjamini & Hochberg, 1995; Genovese et al., 2002). Between-group effects were tested using regression analysis with Nativeness, Sex, and Age as additional regressors. Dotted lines mark a slope of zero. *$p \leq .05$.
SNR = signal-to-noise ratio; n.s. = not significant.

(Figure 7), when nuisance variables (Nativeness, Sex, and Age) were accounted for (the pattern of results was similar for nonparametric analyses and for analyses limited to data from native English speakers; see Supplemental Materials). In other words, motivation and absorption increased, and effort decreased over stories for individuals listening to masked stories compared with individuals listening to clear stories.

These results suggest that the effort of listening to speech in 12-talker babble is reduced and story absorption increased when the listener has had time to get used to the noise. In addition, the results depicted in Figure 6 are consistent with those of Experiment 2b, showing that absorption, enjoyment, and comprehension do not seem to be adversely affected by a moderate level of interfering multitalker babble, even when individuals find listening more effortful in babble compared with clear stories (second and third story presentation in Figure 6).

General Discussion

In the experiments presented here, we used engaging, spoken stories to investigate how a moderate level of an interfering babble masker, resulting in an estimated intelligibility level of >94% (native English speakers) and >87% (nonnative English speakers), affects positive (absorption, enjoyment) and negative (effort) listening

experiences. We also examined how thematic knowledge and prolonged listening changes these experiences. We show that story absorption and enjoyment are only minimally affected (if at all) by decreasing speech-to-masker levels, whereas listening effort clearly increases. We further demonstrate that thematic knowledge about a story presented in multitalker babble helps to increase absorption and enjoyment and to reduce listening effort. Finally, we were surprised to discover that absorption and enjoyment increase and listening effort decreases over time as individuals listen to several successive stories in multitalker babble, over a ~40-min period.

NAS Is Sensitive to Spoken Stories

The results of Experiment 1 (Figure 1) show that motivation, absorption, and enjoyment are higher for stories from *The Moth* podcast compared with a section from the *Sleep With Me* podcast. *The Moth* stories have broad appeal and are meant to engage the listener. The *Sleep With Me* podcast, in contrast, aims to help listeners fall asleep, that is, the goal is to disengage listeners. That we find higher rating scores for *The Moth* stories compared with the *Sleep* story for almost every participant indicates that the NAS developed for written narratives (Kuijpers et al., 2014), and that we modified for spoken stories, is valid. This work provides the

justification for using the scale in our other experiments. Moreover, our work may also be interesting to scholars in the fields of literature and media studies who traditionally do not focus on spoken narratives, but may begin to, given the increasing number of spoken narratives available as podcasts.

Masking Increases Listening Effort but Influences Story Absorption and Enjoyment Much Less

In Experiment 2, we acoustically masked a spoken story and speech snippets derived from it by adding 12-talker babble at SNRs of +12 dB and +4 dB to the speech stimuli. These SNRs are typical of those encountered in real-world settings such as schools, restaurants, transport, and public spaces (Olsen, 1998; Smeds et al., 2015).

A 12-talker babble masker at +12 dB SNR and +4 dB SNR reduced intelligibility for speech snippets presented in random order by about 0.6% and 3.7% (native English speakers) and 1% and 9.3% (nonnative English speakers), respectively, relative to clear speech. Speech intelligibility was generally reduced for nonnative compared with native English speakers (Figure 2). Intelligibility scores for speech snippets may not fully reflect the intelligibility of the same sentences during story listening because of the extended context the story provided; the intelligibility of stories degraded the same way is probably even higher. Nevertheless, ratings for comprehension during story listening decreased and effort increased with decreasing SNRs (Figure 3). Story comprehension was also rated lower by nonnative compared with native English speakers. These results for story listening mirror intelligibility results for speech snippets. Our results were expected given similar observations in previous speech intelligibility and listening effort studies using isolated sentences (Duncan & Aarts, 2006; Krueger et al., 2017a, 2017b; Obleser & Kotz, 2010; Obleser et al., 2007; van Wijngaarden et al., 2002).

Story absorption was unaffected by masking, although story enjoyment was lower for the story in 12-talker babble compared with the story presented clearly. However, neither absorption nor enjoyment differed between stories presented at +12 dB SNR and +4 dB SNR, or between native and nonnative English speakers. In contrast, intelligibility of speech snippets (Figure 2) and rated effort during story listening were higher for +4 dB SNR compared with +12 dB SNR conditions (Figure 3). Experiment 4 further demonstrates that effort is higher for masked stories, whereas motivation, absorption, and enjoyment do not differ (Figure 6, i.e., for the second and third story presentation). The current results are consistent with cognitive control and neuroeconomic accounts positing that a person is willing to exert cognitive resources to engage in a task

and, as a result, may experience effort, if the task (here, listening to an engaging story) is rewarding (Eckert et al., 2016; Shenhav et al., 2017; Westbrook & Braver, 2015). Our results may thus indicate that under the masking conditions used here, in which intelligibility was verified to be high, listening to speech is still rewarding, despite measurable effort.

Absorption, enjoyment, as well as effort varied highly across individuals, particularly for stories presented in multitalker babble (Figure 3). Correlations between absorption and effort and between enjoyment and effort (Figure 4) indicate that 8% and 17% of this variance, respectively, is shared among these dimensions. Higher absorption and enjoyment were associated with reduced listening effort. Although less listening effort may enable individuals to find stories more absorbing and enjoyable, the source of the shared variance is unclear. Our observation introduces the intriguing possibility that individuals who enjoy hearing stories told by friends, family, or colleagues may experience less effort and, in turn, are more willing to engage in listening situations such as bars, cafes, and restaurants despite acoustic demands. Alternatively, a person may also report less effort despite experiencing it when listening is absorbing and enjoyable. The low proportion of shared variance also implies that listening effort, absorption, and enjoyment are independent enough that they should all be considered (cf. Matthen, 2016) to understand why people engage in listening and why they may not.

Thematic Knowledge Increases Story Absorption and Enjoyment and Reduces Listening Effort

The current study demonstrates that thematic knowledge is associated with increased absorption in a story, as well as increased enjoyment and comprehension, and reduced listening effort. Engagement is thought to require the integration of story information with world and thematic knowledge (Busselle & Bilandzic, 2008, 2009; Gerrig & Mumper, 2017), and expertise with the Harry Potter universe may have fostered engagement with the 6-min audio summary of the movies used here. The summary poses challenges to creating and updating a mental model for listeners who are unfamiliar with the Harry Potter series because of its fast pace: It necessarily skips details which a listener with thematic knowledge may be able to fill in. That listeners with higher Harry Potter scores find the summary more absorbing and enjoyable is consistent with previous work using written narratives, indicating that thematic knowledge can increase engagement (Green, 2004).

Research using brief sentences has shown that speech intelligibility under acoustic challenges increases with semantic context (Cohen & Faulkner, 1983; Dubno

et al., 2000; Miller et al., 1951; Obleser & Kotz, 2010; Obleser et al., 2007; Pichora-Fuller et al., 1995; Signoret et al., 2011). Our Harry Potter score assessed knowledge that provides a thematic thread *between* sentences, not simply context within a sentence. The current data thus demonstrate that broad thematic knowledge spanning discrete sentences can increase speech comprehension (see also Holmes et al., 2018).

We also demonstrate that thematic knowledge can reduce listening effort. Individuals likely draw substantially on autobiographical, thematic, and world knowledge when engaging with stories in everyday listening situations (Gerrig & Mumper, 2017), particularly in social contexts with friends, families, or colleagues. Isolated-sentence measures of speech intelligibility and listening effort, which do not draw on the same sources, may not fully capture what listeners experience in real life. The work here highlights the possibilities afforded through use of engaging, naturalistic materials and assessing positive and aversive listening experiences to better understand who is at risk of disengagement and social withdrawal.

Story Motivation and Absorption Increase and Listening Effort Decreases Over Time

The duration of spoken stories in real life may vary substantially depending on whether these are short anecdotes or gossip, or extended event descriptions. Experiments 2 and 3 assessed listening experiences for relatively short stories with durations of 5 to 6 min. Experiment 4 was designed to investigate whether listening experiences change when individuals listen to degraded speech over a longer period, here about 40 min (although with short 2–3 min breaks after each of the four stories). We observed that motivation and absorption increased and listening effort decreased over time as participants listened to four successive stories in multitalker babble.

We had originally anticipated that story listening in babble over an extended period would result in decreased absorption and enjoyment and increased effort. This hypothesis was based on the idea that listeners may become fatigued over time due to the continuous cognitive investment required to understand degraded speech (Hess & Ennis, 2014; McGarrigle et al., 2014; Phillips, 2016; Reitan & Wolfson, 2000, 2004). Instead, our data suggest that listeners adapt to speech in multitalker babble. The observed increase in motivation and absorption and decrease in effort are consistent with reports of speech intelligibility improvement when individuals listen to degraded speech for an extended period (Davis et al., 2005; Eisner et al., 2010; Erb et al., 2013; Erb & Obleser, 2013; Huyck & Johnsrude, 2012; Samuel & Kraljic, 2009). These previous studies suggest listeners

undergo perceptual learning when exposed to at least some kinds of degraded speech materials. Notably, the observed reduction of effort and increase in absorption cannot be due to familiarity with the speaker's voice because each story was spoken by a different narrator. Noise-vocoding was used to degrade speech in previous work, however, and it is not clear whether listeners can learn to hear speech in the presence of masking multi-talker babble—our data suggest that perhaps they can.

There may be a few reasons why the results of Experiment 4 did not reveal an increase in effort or a decrease in absorption and enjoyment as individuals listened to multiple spoken stories masked by babble. First, stories from The Moth podcast have broad appeal and aim to engage. Our student participant pool may have realized over the course of the experiment that the stimulus materials are interesting, which, in turn, may have motivated them to listen. Second, participants were young, normal-hearing individuals, and the 40 min of story listening in multitalker babble at +4 dB SNR (with 2–3 min breaks) may not have been demanding enough to result in disengagement and fatigue. We expect that such effects may become more apparent for older people and people with hearing impairment (Hess & Ennis, 2014). Lastly, the 12-talker babble used in the current study provided relatively predictable masking of the story because the amplitude envelope of the babble noise was relatively flat—such consistency in the masker may not have been as distracting as a more variable, or more intelligible, masker, and thus less detrimental to story engagement (Busselle & Bilandzic, 2009; Kuijpers et al., 2014).

Conclusions

In the current study, naturalistic, spoken stories, masked with multitalker babble at a level that still afforded high intelligibility (with effort), were used to investigate a variety of concurrent listening experiences. We investigated how moderate masking affects both positive (absorption, enjoyment) as well as negative (effort) listening experiences. Our results show that although listening effort certainly increases with acoustic challenges, at the same time, individuals continue to find a story absorbing and enjoyable. This pattern of results highlights the unique experiences with naturalistic stories that may not be observed with the isolated sentence materials that are typically used to test speech recognition. We also demonstrate that thematic knowledge makes story listening more enjoyable and absorbing and less effortful. Finally, we show that effort experienced by individuals who listen to several stories in babble noise decreases over time. These results indicate that under masking conditions, in which intelligibility is high, listening to speech is still rewarding, despite effort.

This work provides an important step toward understanding listening challenges and benefits in real-world listening situations and opens new avenues to better understand why some people disengage from listening whereas others persist despite challenges.

ORCID iD
Björn Herrmann 🔟 https://orcid.org/0000-0001-6362-3043

Supplemental material
Supplemental material for this article is available online.

References
Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 19, 1061–1170. https://doi.org/10.1037/0278-7393.19.5.1061

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear & Hearing*, 38, e39–e48. DOI: 10.1097/AUD.0000000000000361

Bamberg, M. (2010). Who am I? Narration and its contribution to self and identity. *Theory & Psychology*, 21, 1–22. https://doi.org/10.1177/0959354309355852

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bilandzic, H., & Busselle, R. W. (2017). Beyond metaphors and traditions: Exploring the conceptual boundaries of narrative engagement. In: F. Hakemulder, M. M. Kuijpers, E. S. Tan, K. Bálint, & M. M. Doicaru (Eds.), *Narrative absorption* (pp. 11–27). John Benjamins Publishing Company.

Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement. *Communication Theory*, 18, 255–280. https://doi.org/10.1111/j.1468-2885.2008.00322.x

Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology*, 12, 321–347. https://doi.org/10.1080/15213260903287259

Cohen, G., & Faulkner, D. (1983). Word recognition: Age differences in contextual facilitation effects. *British Journal of Psychology*, 74, 239–251. https://doi.org/10.1111/j.2044-8295.1983.tb01860.x

Cohen, J. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication & Society*, 4, 245–264. https://doi.org/10.1207/S15327825MCS0403_01

Cohen, J., & Tal-Or, N. (2017). Antecedents of identification: Character, text, and audiences. In: F. Hakemulder, M. M. Kuijpers, E. S. Tan, K. Bálint, & M. M. Doicaru (Eds.), *Narrative absorption* (pp. 133–153). John Benjamins Publishing Company.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience*, 23, 3423–3431. https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134, 222–241. https://doi.org/10.1037/0096-3445.134.2.222

DeSantis, L., & Ugarriza, D. N. (2000). The concept of theme as used in qualitative nursing research. *Western Journal of Nursing Research*, 22, 351–372. https://doi.org/10.1177/019394590002200308

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, Article 112. 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, 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, 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, 235–247. https://doi.org/10.1177/0165025413479475

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

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *The Journal of Neuroscience*, 30, 7179–7186. doi: 10.1523/JNEUROSCI.4040-09.2010

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *The Journal of Neuroscience*, *33*, 10688–10697. https://doi.org/10.1523/JNEUROSCI.4596-12.2013

Erb, J., & Obleser, J. (2013). Upregulation of cognitive control networks in older adults' speech comprehension. *Frontiers in Systems Neuroscience*, *7*, Article 116. doi: 10.3389/fnsys.2013.00116

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, *15*, 870–878. doi: 10.1006/nimg.2001.1037

Gerrig, R. J., & Mumper, M. L. (2017). How readers' lives affect narrative experiences. In: M. Burke & E. T. Troscianko (Eds.), *Cognitive literary science: Dialogues between literature and cognition* (pp. 239–257). Oxford University Press.

Graesser, A. C., Olde, B., & Klettke, B. (2002). How does the mind construct and represent stories? In: M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 229–262). Lawrence Erlbaum Associates Publishers.

Green, M. C. (2004). Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse Processes*, *38*, 247–266. https://doi.org/10.1207/s15326950dp3802_5

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory*, *14*, 311–327. https://doi.org/10.1111/j.1468-2885.2004.tb00317.x

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

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. doi: 10.3109/14992027.2015.1117663

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

Hess, T. M., & Ennis, G. E. (2014). Assessment of adult age differences in task engagement: The utility of systolic blood pressure. *Motivation and Emotion*, *38*, 844–854. doi: 10.1007/s11031-014-9433-2

Hjørland, B. (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content ... and relevance. *Journal of the American Society for Information Science and Technology*, *52*, 774–778. https://doi.org/10.1002/asi.1131

Hodgson, M., Steininger, G., & Razavi, Z. (2007). Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments. *The Journal of the Acoustical Society of America*, *121*, 2023–2033. https://doi.org/10.1121/1.2535571

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*, 483–492. doi: 10.1080/14992027.2018.1432901

Huyck, J. J., & Johnsrude, I. S. (2012). Rapid perceptual learning of noise-vocoded speech requires attention. *Journal of the Acoustical Society of America*, *131*, EL236–EL242. https://doi.org/10.1121/1.3685511

Johnsrude, I. S., & Rodd, J. M. (2016). Factors that increase processing demands when listening to speech. In: G. Hickok & S.L. Small (Eds.), *Neurobiology of language* (pp. 491–502). Elsevier Academic Press.

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

Krueger, M., Schulte, M., Brand, T., & Holube, I. (2017a). Development of an adaptive scaling method for subjective listening effort. *The Journal of the Acoustical Society of America*, *141*, 4680–4693. https://doi.org/10.1121/1.4986938

Krueger, M., Schulte, M., Zokoll, M. A., Wagener, K. C., Meis, M., Brand, T., & Holube, I. (2017b). Relation between listening effort and speech intelligibility in noise. *American Journal of Audiology*, *26*, 378–392. doi: 10.1044/2017_AJA-16-0136

Kuijpers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. M. (2014). Exploring absorbing reading experiences. *Scientific Study of Literature*, *4*, 89–122. https://doi.org/10.1075/ssol.4.1.05kui

Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. *Ear & Hearing*, *37*, 77S–84S. doi: 10.1097/AUD.0000000000000304

Lunner, T., & Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *Journal of the American Academy of Audiology*, *18*, 604–617. doi: 10.3766/jaaa.18.7.7

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, *3*, 173–192. https://doi.org/10.1111/j.1745-6924.2008.00073.x

Matthen, M. (2016). Effort and displeasure in people who are hard of hearing. *Ear & Hearing*, *37*(Suppl 1), 28S–34S. doi: 10.1097/AUD.0000000000000292

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*, 953–978. https://doi.org/10.1080/01690965.2012.705006

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, *53*, 433–440. doi: 10.3109/14992027.2014.890296

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*, 329–335. https://doi.org/10.1037/h0062491

Nabi, R. L., & Krcmar, M. (2004). Conceptualizing media enjoyment as attitude: Implications for mass media effects research. *Communication Theory*, *14*, 288–310. https://doi.org/10.1111/j.1468-2885.2004.tb00316.x

Oatley, K. (1999). Meetings of minds: Dialogue, sympathy, and identification, in reading fiction. *Poetics*, *26*, 439–454. https://doi.org/10.1016/S0304-422X(99)00011-X

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*, 633–640. doi: 10.1093/cercor/bhp128

Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience*, *27*, 2283–2289. https://doi.org/10.1523/JNEUROSCI.4663-06.2007

Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. *American Journal of Audiology*, *7*, 21–25. doi: 10.1044/1059-0889 (1998/012)

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

Phillips, N. A. (2016). The implications of cognitive aging for listening and the framework for understanding effortful listening (FUEL). *Ear & Hearing*, *37*, 44S–51S. doi: 10.1097/AUD.0000000000000309

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 & Hearing*, *37*(Suppl 1), 5S–27S. doi: 10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*, 593–608. https://doi.org/10.1121/1.412282

Pickett, J. M., & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*, *6*, 151–164. https://doi.org/10.1177/002383096300600304

Pollack, I., & Pickett, J. M. (1963). The intelligibility of excerpts from conversation. *Language and Speech*, *6*, 165–171. https://doi.org/10.1177/002383096300600305

Pollack, I., & Pickett, J. M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior*, *3*, 79–84. https://doi.org/10.1016/S0022-5371(64)80062-1

Poulsen, A. T., Kamronn, S., Dmochowski, J. P., Parra, L. C., & Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific Reports*, *7*, 43916. doi: 10.1038/srep43916

Reitan, R. M., & Wolfson, D. (2000). Conation: A neglected aspect of neuropsychological functioning. *Archives of Clinical Neuropsychology*, *15*, 443–453. https://doi.org/10.1016/S0887-6177(99)00043-8

Reitan, R. M., & Wolfson, D. (2004). The differential effect of conation on intelligence test scores among brain-damaged and control subjects. *Archives of Clinical Neuropsychology*, *19*, 29–35. https://doi.org/10.1093/arclin/19.1.29

Rusnock, C. F., & Bush, P. M. (2012). An evaluation of restaurant noise levels and contributing factors. *Journal of Occupational and Environmental Hygiene*, *9*, 108–113. https://doi.org/10.1080/15459624.2012.683716

Ryan, M.-L. (2007). Toward a definition of narrative. In: D. Herman (Ed.), *The Cambridge companion to narrative* (pp. 22–35). Cambridge University Press.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*, 1207–1218. https://doi.org/10.3758/APP.71.6.1207

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Signoret, C., Johnsrude, I. S., Classon, E., & Rudner, M. (2011). Combined effects of form- and meaning-based predictability on perceived clarity of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *44*, 277–285. doi: 10.1037/xhp0000442

Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology*, *26*, 183–196. doi: 10.3766/jaaa.26.2.7

Strauss, D. J., & Francis, A. L. (2017). Toward a taxonomic model of attention in effortful listening. *Cognitive, Affective & Behavioral Neuroscience*, *17*, 809–825. https://doi.org/10.3758/s13415-017-0513-0

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, *28*, 455–462. doi: 10.1044/jshr.2803.455

van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America*, *111*, 1906–1916. https://doi.org/10.1121/1.1456928

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Frontiers in Psychology*, *7*, Article 345. https://doi.org/10.3389/fpsyg.2016.00345

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective & Behavioral Neuroscience*, *15*, 395–415. https://doi.org/10.3758/s13415-015-0334-y

Westgate, E. C., & Wilson, T. D. (2018). Boring thoughts and bored minds: The MAC model of boredom and cognitive engagement. *Psychological Review*, *125*, 689–713. https://doi.org/10.1037/rev0000097

Wright, P. C., McCarthy, J., & Meekison, L. (2003). Making sense of experience. In: M. A. Blythe, A. F. Monk, K. Overbeeke, & P. C. Wright (Eds.), *Funology: Human-computer interaction series* (pp. 43–53). Kluwer Academic Publishers.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear & Hearing*, *31*, 480–490. doi: 10.1097/AUD.0b013e3181d4f251

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, *23*, 1028–1034. doi: 10.3758/s13423-015-0864-x

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*, 292–297. https://doi.org/10.1111/j.1467-9280.1995.tb00513.x