

ARTICLE

Predictability effects in degraded speech comprehension are reduced as a function of attention

Pratik Bhandari^{1,2*} , Vera Demberg^{2,3} and Jutta Kray¹

¹Department of Psychology, Saarland University, Saarbrücken, Germany; ²Department of Language Science and Technology, Saarland University, Saarbrücken, Germany; ³Department of Computer Science, Saarland University, Saarbrücken, Germany

*Corresponding author. Email: pratikb@coli.uni-saarland.de

(Received 21 January 2022; Revised 30 June 2022; Accepted 04 July 2022)

Abstract

The aim of this study was to examine the role of attention in understanding linguistic information even in a noisy environment. To assess the role of attention, we varied task instructions in two experiments in which participants were instructed to listen to short sentences and thereafter to type in the last word they heard or to type in the whole sentence. We were interested in how these task instructions influence the interplay between top-down prediction and bottom-up perceptual processes during language comprehension. Therefore, we created sentences that varied in the degree of predictability (low, medium, and high) as well as in the degree of speech degradation (four, six, and eight noise-vocoding channels). Results indicated better word recognition for highly predictable sentences for moderate, though not for high, levels of speech degradation, but *only* when attention was directed to the whole sentence. This underlines the important role of attention in language comprehension.

Keywords: temporal attention; auditory attention; speech perception; bottom-up processing; top-down prediction; semantic prediction; perceptual adaptation; noise-vocoded speech

1. Introduction

Spoken language comprehension seems like an easy, automatized process. But intelligibility and comprehension of speech can be rendered difficult in our daily conversations due to adverse listening conditions like background noise and distortion of the speech signal (e.g., Chen & Loizou, 2011; Fontan et al., 2015). For example, the voice of a person talking on the other end of a telephone connection can sound robotic and difficult to understand when the signal quality or transmission is poor. Perception and comprehension of speech in such an adverse condition is effortful (Pals et al., 2013; Strauss & Francis, 2017; Winn et al., 2015). To deal with perceptual difficulties, listeners rely on top-down prediction based on the context that has been

understood so far (Obleser & Kotz, 2010; Pichora-Fuller, 2008; Sheldon et al., 2008b). The context can contain information about a topic of the conversation, syntactic information about the structure of the sentence, world knowledge, visual information, and so forth (Altmann & Kamide, 2007; Brothers et al., 2020; Kaiser & Trueswell, 2004; Knoeferle et al., 2005; Xiang & Kuperberg, 2015; for reviews, see Ryskin & Fang, 2021; Stilp, 2020).

To utilize context information, listeners must attend to it and build up a meaning representation of what has been said. Listeners attend to the context information in clear speech with minimal effort, but processing and comprehending degraded speech is more effortful and requires more attentional resources (Eckert et al., 2016; Peelle, 2018; Wild et al., 2012). However, it is less clear how listeners distribute attentional resources: On the one hand, listeners can attend throughout the whole stream of speech and may thereby profit from the context information to predict sentence endings. On the other hand, listeners can focus their attention on linguistic material at a particular time point in the speech stream and, as a result, miss critical parts of the sentence context. If the goal is to understand a specific word in an utterance, there is a trade-off between allocating attentional resources to the perception of that word vs. allocating resources also to the understanding of the linguistic context and generating predictions.

The aim of this study was to investigate how the allocation of attentional resources induced by different task instructions influences language comprehension and, in particular, the use of context information under adverse listening conditions. To examine the role of attention on predictive processing under degraded speech, we conducted two experiments in which we manipulated task instructions. In Experiment 1, participants were instructed to only repeat the final word of the sentence they heard, while in Experiment 2, they were instructed to repeat the whole sentence, thus drawing attention to the entire sentence including the context. In both experiments, we varied the degree of predictability of sentence endings as well as the degree of speech degradation. In the following, we first summarize the findings of studies that have investigated predictive language processing in the comprehension of degraded speech, and then results on the role of attention and task instruction in speech perception.

1.1. Predictive processing and language comprehension under degraded speech

It is broadly agreed that human comprehenders generate expectations about upcoming linguistic material based on context information (for reviews, see Kuperberg & Jaeger, 2016; Nieuwland, 2019; Pickering & Gambi, 2018; Staub, 2015). These expectations are formed while a sentence unfolds. The claims about the predictive nature of language comprehension are based on a variety of behavioral and electrophysiological experimental measures including eye-tracking and electroencephalography (EEG). For instance, in the well-known visual world paradigm, listeners fixate on a picture of an object (e.g., a cake) that is predictable based on the prior sentence context (e.g., ‘The boy will eat the ...’) even before hearing the final target word (e.g., Altmann & Kamide, 1999, 2007; Ankenier et al., 2018). Moreover, highly predictable words are read faster and are skipped more often compared to less predictable words (Frisson et al., 2005; Rayner et al., 2011).

In EEG studies, the N400, a negative-going EEG component that usually peaks around 400 ms poststimulus, is considered as a neural marker of semantic

unexpectedness (Kutas & Federmeier, 2011). For instance, in the highly predictable sentence context ‘The day was breezy so the boy went outside to fly ...’, DeLong et al. (2005) found that the amplitude of the N400 component for the expected continuation ‘a kite’ was much smaller than for the unexpected continuation ‘an airplane’. Although these studies demonstrated that as the sentence context builds up, listeners form predictions about upcoming words in the sentence, the universality and ubiquity of predictive language processing have been questioned (see Huettig & Mani, 2016). Also, the use of context for top-down prediction can be limited by factors like literacy (Mishra et al., 2012), age, and working memory (Federmeier et al., 2002, 2010), as well as by the experimental setup (Huettig & Guerra, 2019). While these language comprehension studies investigating predictive processing have used clean speech and sentence reading, the present study focuses on examining how attention influences the use of context to form top-down predictions under adverse listening conditions.

There is already some evidence that when the bottom-up speech signal is less reliable due to degradation, listeners tend to rely more on the context information to support language comprehension (Amichetti et al., 2018; Obleser & Kotz, 2010; Sheldon et al., 2008a). For example, Sheldon et al. (2008a, Figure 2) estimated that for both younger and older adults, the number of noise-vocoding channels required to achieve 50% accuracy varied as a function of sentence context. Compared to highly predictable sentences, a greater number of channels (i.e., more bottom-up information) was required in less predictable sentences to achieve the same level of accuracy. Therefore, they concluded that when speech is degraded, predictable sentence context facilitates word recognition. Obleser et al. (2007) found that at a moderate level of spectral degradation, listeners’ word recognition accuracy was higher for highly predictable sentence contexts than for less predictable ones. However, while listening to the least degraded speech, there was no such beneficial effect of sentence context (see also Obleser & Kotz, 2010). Hence, especially when the bottom-up speech signal is less reliable due to moderate degradation, information available from the sentence context is used to enhance language comprehension, suggesting that there is a dynamic interaction between top-down predictive and bottom-up sensory processes in language comprehension (Bhandari et al., 2021).

1.2. Attention and predictive language processing

It is not only the quality of speech signal that influences the reliance on and use of predictive processing; attention to auditory input is also important. Auditory attention allows a listener to focus on the speech signal of interest (for reviews, see Fritz et al., 2007; Lange, 2013). For instance, it has been shown that a listener can attend to and derive information from one stream of sound among many competing streams as demonstrated in the well-known *cocktail party effect* (Cherry, 1953; Hafter et al., 2007). When a participant is instructed to attend to only one of the two or more competing speech streams in a diotic or dichotic presentation, response accuracy to the attended speech stream is higher than to the unattended speech (e.g., Tóth et al., 2020). Similarly, when a listener is presented with a stream of tones (e.g., musical notes varying in pitch, pure tones of different harmonics) but attends to any one of the tones appearing at a specified time point, this is reflected in a larger amplitude of N1 (e.g., Lange & Röder, 2010; see also Sanders & Astheimer, 2008) which is the first

negative-going ERP component, peaking around 100 ms poststimulus, considered as a marker of auditory selective attention (Näätänen & Picton, 1987; Thornton et al., 2007). Hence, listeners can draw attention to and process one among multiple competing speech streams.

So far, most previous studies investigated listeners' attention within a single speech stream by using acoustic cues like accentuation and prosodic emphasis. For example, Li et al. (2014) examined whether the comprehension of critical words in a sentence context was influenced by a linguistic attention probe such as 'ba' presented together with an accented or deaccented critical word. The N1 amplitude was larger for words with such an attention probe than for words without a probe. These findings support the view that attention can be flexibly directed either by instructions toward a specific signal or by linguistic probes (Li et al., 2017; see also Brunellière et al., 2019). Thus, listeners are able to select a part or segment of a stream of auditory stimuli to pay attention to.

The findings on the interplay of attention and prediction mentioned above come from studies which, for the most part, used a stream of *clean speech* or *multiple streams of clean speech* in their experiments. They cannot tell us about the attention–prediction interplay in *degraded speech comprehension*. Specifically, we do not know what role attention to a segment of a speech stream plays in the contextual facilitation of degraded speech comprehension, although separate lines of research show that listeners attend to the most informative portion of the speech stream (e.g., Astheimer & Sanders, 2011), and semantic predictability facilitates comprehension of degraded speech (e.g., Obleser & Kotz, 2010).

1.3. The present study

We examined whether context-based semantic predictions are automatic during effortful listening to degraded speech, when participants are instructed to report either the final word of the sentence or the entire sentence. We manipulated semantic predictions and speech degradation by orthogonally varying cloze probability of target words and number of channels for the noise-vocoding of speech in a factorial design. Noise-vocoded speech is difficult to understand, as the frequency-specific information of a specific bandwidth is replaced with white noise while temporal cues are preserved (e.g., Corps & Rabagliati, 2020; Davis et al., 2005; Shannon et al., 1995).

In two experiments, we varied the task instructions to the listeners, which required them to differentially attend to the target word. In Experiment 1, listeners were asked to report the noun which was in the final position of the sentence that they heard. This instruction did not require listeners to pay attention to the context. Hence, processing the context was not strictly necessary for the task. In Experiment 2, listeners were asked to report the entire sentence by typing in everything they heard. Thus, the listeners' attention in Experiment 2 was not focused on any specific part of the sentence. We hypothesized that when listeners pay attention only to the contextually predicted target word, as they might choose to do in Experiment 1, they do not form top-down predictions, that is, there should not be a facilitatory effect of target word predictability. In contrast, when listeners attend to the whole sentence, they do form expectations, such that a facilitatory effect of target word predictability will be observed.

2. Experiment 1

2.1. Method

2.1.1. Participants

We recruited 50 participants online via Prolific Academic (Prolific, 2014). One participant whose response accuracy was less than 50% across all experimental conditions was removed. Among the remaining 49 participants (M age \pm SD = 23.31 \pm 3.53 years; age range = 18–30 years), 27 were male and 22 were female. All participants were native speakers of German and did not have any speech-language disorder, hearing loss, or neurological disorder (all self-reported). All participants received 6.20 euros as monetary compensation for their participation. The experiment was approximately 40 minutes long. The German Society for Language Science ethics committee approved the study and participants provided informed consent in accordance with the Declaration of Helsinki.

2.1.2. Materials

We used the same materials from our previous study (Bhandari et al., 2021). They consist of 360 German sentences spoken by a female native German speaker, unaccented, at a normal rate of speech. The sentences were recorded and digitized at 44.1 kHz with 32-bit linear encoding. All sentences consisted of pronoun, verb, determiner, and object (noun) (e.g., stimuli sentences with their English translations see Supplementary Material). We used 120 nouns to create three types of sentences differing in the cloze probability of the target words (nouns) which mostly appeared as the final word of the sentence. We thereby compared sentences with low, medium, and high cloze target words.

The cloze probability ratings for each of these sentences were measured in a norming study with a separate group of participants (n = 60; age range = 18–30 years). Mean cloze probabilities for sentences with low cloze target words (low predictability sentences), medium cloze target words (medium predictability sentences) and high cloze target words (high predictability sentences) were 0.022 ± 0.027 ($M \pm SD$; range = 0.00–0.09), 0.274 ± 0.134 ($M \pm SD$; range = 0.10–0.55), and 0.752 ± 0.123 ($M \pm SD$; range = 0.56–1.00), respectively.

The speech signal was divided into 1, 4, 6, and 8 frequency bands between 70 and 9,000 Hz to create four different levels of speech degradation for each of the 360 recorded sentences. Frequency boundaries were approximately logarithmically spaced, determined by cochlear-frequency position functions (Erb, 2014; Greenwood, 1990). A customized Praat script originally written by Darwin (2005) was used to create noise-vocoded speech. Boundary frequencies for each noise-vocoding condition are given in Table 1.

2.1.3. Procedure

Participants were asked to use headphones or earphones. A sample of vocoded speech not used in the practice trial or the main experiment was provided so that the participants could adjust the volume to their preferred level of comfort at the beginning of the experiment. The participants were instructed to listen to the sentences and to type in the target word (noun) by using the keyboard. The time for typing in the response was not limited. They were also informed at the beginning of the experiment that some of the sentences would be ‘noisy’ and not easy to

Table 1. Boundary frequencies (in Hz) for 1-, 4-, 6-, and 8-channel noise-vocoding conditions

Number of channels		Boundary frequencies								
1	70	9,000								
4	70	423	1,304	3,504	9,000					
6	70	268	633	1,304	2,539	4,813	9,000			
8	70	207	423	764	1,304	2,156	3,504	5,634	9,000	

understand, and in these cases, they were encouraged to guess what they might have heard. Eight practice trials with different levels of speech degradation were given to familiarize the participants with the task before presenting all 120 experimental trials with an intertrial interval of 1,000 ms.

Each participant had to listen to 40 high predictability, 40 medium predictability, and 40 low predictability sentences. Levels of speech degradation were also balanced across each predictability level, so that for each of the three predictability conditions (high, medium, and low predictability), ten 1-channel, ten 4-channel, ten 6-channel, and ten 8-channel noise-vocoded sentences were presented, resulting in 12 experimental lists. The sentences in each list were pseudo-randomized so that no more than three sentences of the same degradation and predictability condition appeared consecutively.

2.2. Analyses

We performed data preprocessing and analyses in RStudio (R version 3.6.3; R Core Team, 2020). At 1-channel, there were only five correct responses, one each from 5 participants out of 49. Therefore, the 1-channel speech degradation condition was excluded from the analyses.

Accuracy was analyzed using Generalized Linear Mixed Models (GLMMs) with lmerTest (Kuznetsova et al., 2017) and lme4 (Bates et al., 2015) packages. Binary responses (categorical: correct and incorrect) for all participants were fit with a binomial linear mixed-effects model (Jaeger, 2006, 2008). Correct responses were coded as 1 and incorrect responses were coded as 0. Number of channels (categorical: 4-channel, 6-channel, and 8-channel noise-vocoding), target word predictability (categorical: high predictability sentences, medium predictability sentences, low predictability sentences), and the interaction of number of channels and target word predictability were included in the fixed effects.

We first fitted a model with maximal random effects structure that included random intercepts for each participant and item (Barr et al., 2013). Both by-participant and by-item random slopes were included for number of channels, target word predictability, and their interaction, which was supported by the experiment design. Based on the previous findings on perceptual adaptation (e.g., Cooke et al., 2022; Davis et al., 2005; Erb et al., 2013; but see also Bhandari et al., 2021), we further added trial number (centered) in the fixed effect structure to control for whether the listeners adapted to the degraded speech. We report the results of the model that includes trial number as fixed effects.¹

¹lmer(response ~ 1 + channels × predictability + trial number + (1 + channels × predictability || participant) + (1 + channels × predictability || item) ...

We applied treatment contrast for number of channels (8-channel as a baseline) and sliding difference contrast for target word predictability (low predictability vs. medium predictability, and low predictability vs. high predictability sentences). The code and data are available in the following publicly accessible repository: <https://osf.io/t6unj/>.

2.3. Results and discussion

Mean response accuracy for all experimental conditions is shown in Table 2 and Fig. 1. We found that accuracy increased with an increase in the number of noise-vocoding channels, that is, with a decrease in speech degradation. However, accuracy did not increase with an increase in target word predictability. The results of statistical analysis confirmed these observations (see Table 3).

Table 2. Response accuracy (mean and standard error of the mean) across all levels of speech degradation and target word predictability in Experiment 1

Number of channels	Target word predictability	Mean	Standard error
4	High	62.65	2.24
	Medium	63.43	2.03
	Low	63.99	1.83
6	High	95.60	0.94
	Medium	95.54	1.05
	Low	95.16	1.10
8	High	98.16	0.84
	Medium	96.75	1.04
	Low	97.91	0.97

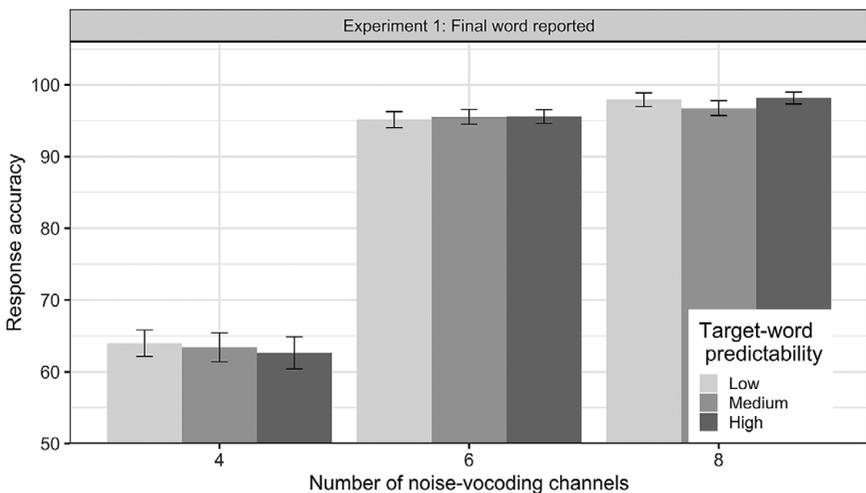
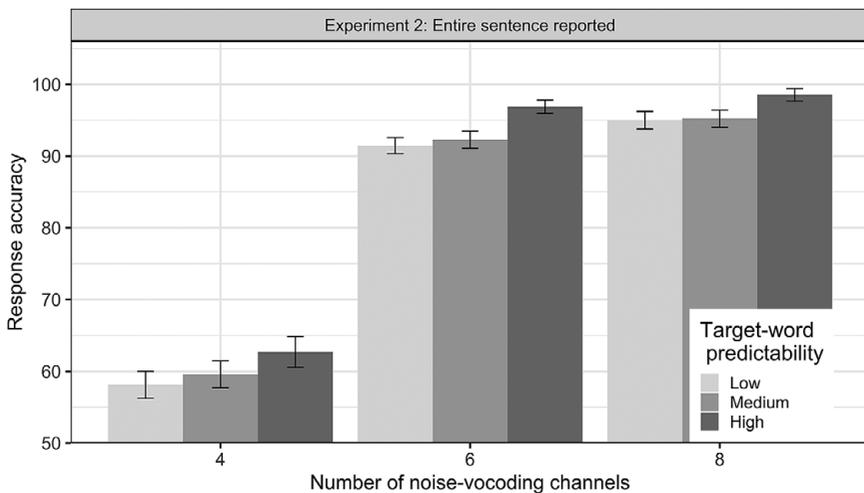


Fig. 1. Mean response accuracy across all conditions in Experiment 1. Accuracy increased only with an increase in the number of noise-vocoding channels. There is no change in accuracy with an increase or decrease in target-word predictability. Error bars represent standard error of the means.

Table 3. Estimated effects of the model accounting for the correct word recognition in Experiment 1

Fixed effects	Estimate of β	Standard error	z	p
(Intercept)	4.17	0.25	16.73	<0.001
Noise condition (four channels)	-3.50	0.22	-16.19	<0.001
Noise condition (six channels)	-0.70	0.21	-3.29	<0.001
Target word predictability (Low-Medium)	0.30	0.36	0.84	0.40
Target word predictability (High-Low)	0.50	0.43	1.16	0.25
Noise condition (four channels) \times Target word predictability (Low-Medium)	-0.22	0.39	-0.57	0.57
Noise condition (six channels) \times Target word predictability (Low-Medium)	-0.34	0.44	-0.76	0.44
Noise condition (four channels) \times Target word predictability (High-Low)	-0.54	0.45	-1.18	0.24
Noise condition (six channels) \times Target word predictability (High-Low)	0.04	0.50	0.09	0.03
Trial number	0.001	0.002	0.48	0.63

**Fig. 2.** Mean response accuracy across all conditions in Experiment 2. Accuracy increased with an increase in number of noise-vocoding channels and target-word predictability. Error bars represent standard error of the means.

There was a significant main effect of number of channels, indicating that response accuracy for the 8-channel vocoded speech was higher than for both 4-channel ($\beta = -3.50$, $SE = 0.22$, $z(4,410) = -16.19$, $p < 0.001$) and 6-channel vocoded speech ($\beta = -0.70$, $SE = 0.21$, $z(4,410) = -3.29$, $p = 0.001$), that is, when the number of channels increased to 8, listeners gave more correct responses (see Fig. 2). There was, however, no significant main effect of target word predictability ($\beta = 0.30$, $SE = 0.36$, $z(4,410) = .84$, $p = 0.40$, and $\beta = 0.50$, $SE = 0.43$, $z(4,410) = 1.16$, $p = 0.25$), and no interaction between number of channels and target word predictability (all $p > 0.05$). There was also no significant main effect of trial number ($\beta = 0.001$,

SE = 0.002, $z(4,410) = .48$, $p = 0.63$) suggesting that the listeners' performance did not improve over time.

These results indicated a decrease in response accuracy with an increase in speech degradation from the 8-channel to the 6-channel noise-vocoding condition, and from the 8-channel to the 4-channel noise-vocoding condition. However, response accuracy did not increase with an increase in target word predictability, and the interaction between number of channels and target word predictability was also absent, in contrast to previous findings (Obleser & Kotz, 2010; Obleser et al., 2007; see also Hunter & Pisoni, 2018). These results suggest that the task instruction, which asked participants to report only the final word, indeed led to neglecting the context. Although participants were able to neglect the context, there was still uncertainty about the speech quality of the next trial; hence, they could not adapt to the different levels of degraded speech.

To confirm that the predictability effect (or contextual facilitation) is replicable and dependent on attentional focus, we conducted a second experiment in which we changed the task instruction to draw participants' attention to decoding the whole sentence.

3. Experiment 2

3.1. Method

3.1.1. Participants and materials

We recruited 48 participants (M age \pm $SD = 24.44 \pm 3.55$ years; age range = 18–31 years; 32 males) online via Prolific Academic. The same procedure was followed as in Experiment 1, and the same stimuli were used.

3.1.2. Procedure

Participants were presented with sentences at a comfortable volume level. They were asked to use headphones or earphones, and a prompt was presented before the experiment began to adjust the volume to their level of comfort. Eight practice trials were presented, followed by 120 experimental trials. The participants were instructed to report the entire sentence by typing in what they heard. We did not limit the response time.

3.2. Analysis

We followed the same data analysis procedure as in Experiment 1. The 1-channel speech degradation condition was excluded from the analysis. We did not consider whether listeners reported other words in a sentence correctly; only the final words of the sentences (target words) were considered as either correct or incorrect responses. As in Experiment 1, we report the results from the maximal model supported by the design.²

²glmer (response ~ 1 + channels × predictability + trial number + (1 + channels × predictability || participant) + (1 + channels × predictability || item) ...

3.3. Results and discussion

Mean response accuracy for different conditions is shown in Table 4 and Fig. 2. We found that accuracy increased when the number of noise-vocoding channels increased, as well as when the target word predictability increased. The results of statistical analysis confirmed these observations (Table 5): We again found a main effect of number of channels, such that response accuracy at 8-channel was higher than for both 4-channel ($\beta = -3.51$, $SE = 0.24$, $z(4,320) = -14.64$, $p < 0.001$), and 6-channel noise-vocoding ($\beta = -0.65$, $SE = 0.22$, $z(4,320) = -2.93$, $p = 0.003$). Similar to Experiment 1, the main effect of trial number was not significant ($\beta = 0.002$, $SE = 0.002$, $z(4,320) = 1.11$, $p = 0.27$) indicating that the response accuracy did not increase over the course of the experiment.

In contrast to Experiment 1, there was also a main effect of target word predictability: Response accuracy in high predictability sentences was significantly higher than in low predictability sentences ($\beta = 1.42$, $SE = 0.47$, $z(4,320) = 3.02$, $p = 0.003$). We also found a statistically significant interaction between speech degradation and target word predictability ($\beta = -1.14$, $SE = 0.50$, $z(4,320) = -2.30$, $p = 0.02$). Subsequent subgroup analyses of each channel condition showed that the interaction was driven by the difference in response accuracy between high predictability

Table 4. Response accuracy (mean and standard error of the mean) across all levels of speech degradation and target word predictability in Experiment 2

Number of channels	Target word predictability	Mean	Standard error
4	High	62.71	2.14
	Medium	59.58	1.88
	Low	58.13	1.88
6	High	96.88	0.93
	Medium	92.29	1.21
	Low	91.46	1.12
8	High	98.54	0.86
	Medium	95.21	1.19
	Low	95.00	1.23

Table 5. Estimated effects of the model accounting for the correct word recognition in Experiment 2

Fixed effects	Estimate of β	Standard error	z	p
(Intercept)	4.09	0.24	16.79	<0.001
Noise condition (four channels)	-3.51	0.24	-14.64	<0.001
Noise condition (six channels)	-0.65	0.22	-2.93	0.003
Target word predictability (Low-Medium)	-0.08	0.34	-0.23	0.82
Target word predictability (High-Low)	1.42	0.47	3.02	0.003
Noise condition (four channels) \times Target word predictability (Low-Medium)	0.02	0.38	0.05	0.96
Noise condition (six channels) \times Target word predictability (Low-Medium)	-0.13	0.43	-0.31	0.76
Noise condition (four channels) \times Target word predictability (High-Low)	-1.14	0.50	-2.30	0.02
Noise condition (six channels) \times Target word predictability (High-Low)	-0.23	0.57	-0.41	0.68
Trial number	0.002	0.002	1.11	0.27

sentences and low predictability sentences in the 8-channel ($\beta = 1.42$, $SE = 0.62$, $z(1,440) = 2.30$, $p = 0.02$) and 6-channel noise-vocoding conditions ($\beta = 1.14$, $SE = 0.34$, $z(1,440) = 3.31$, $p < 0.001$); at 4 channel, the difference in response accuracy between high and low predictability sentences was not significant ($\beta = 0.28$, $SE = 0.18$, $z(1,440) = 1.59$, $p = 0.11$).

In contrast to Experiment 1, these results indicate an effect of target word predictability; that is, response accuracy was higher when the target word predictability was high as compared to low. Also, the interaction between target word predictability and speech degradation, which was not observed in Experiment 1, showed that semantic predictability facilitated the comprehension of degraded speech already at moderate levels (like 6- or 8-channel). In line with the findings from Experiment 1, response accuracy was better with a higher number of channels.

We combined the data from both experiments in a single analysis to test whether participants' response accuracy changes across the experiments, that is, to test whether the difference between experimental manipulations is statistically significant. We ran a binomial linear mixed-effects model on response accuracy and followed the same procedure as in Experiments 1 and 2. A full random effects structure supported by the study design was modeled.³ The model summary is shown in Table 6. The model revealed that there was no significant main effect of experimental group ($\beta = 0.04$, $SE = 0.26$, $z(8,730) = .15$, $p = 0.88$) indicating that the overall response accuracy did not change with the change in instructions from Experiments 1 and 2. However, the critical interaction between experimental group and target

Table 6. Estimated effects of the best-fitting model accounting for the correct word recognition in both experiments

Fixed effects	Estimate of β	Standard error	z	p
(Intercept)	4.19	0.20	20.72	<0.001
Noise condition (four channels)	-3.56	0.20	-18.19	<0.001
Noise condition (six channels)	-0.59	0.18	-3.28	0.001
Target word predictability (Low-Medium)	0.13	0.26	0.50	0.62
Target word predictability (High-Low)	0.98	0.34	2.93	0.003
Experimental group	0.04	0.26	0.15	0.88
Noise condition (four channels) \times Target word predictability (Low-Medium)	-0.12	0.29	-0.40	0.69
Noise condition (six channels) \times Target word predictability (Low-Medium)	-0.30	0.34	-0.87	0.38
Noise condition (four channels) \times Target word predictability (High-Low)	-0.84	0.35	-2.42	0.02
Noise condition (six channels) \times Target word predictability (High-Low)	-0.11	0.38	-0.29	0.77
Noise condition (four channels) \times Experimental group	-0.10	0.25	-0.41	0.68
Noise condition (six channels) \times Experimental group	0.1	0.28	-0.36	0.72
Target word predictability (High-Low) \times Experimental group	0.465	0.20	2.34	0.02
Trial number	0.001	0.001	0.93	0.35

³ glmer (response ~ 1 + channels \times predictability + channels \times experiment + predictability \times experiment + trial number + (1 + channels \times predictability || participant) + (1 + channels \times predictability + channels \times experiment + predictability \times experiment || item) ...

word predictability was statistically significant ($\beta = 0.46$, $SE = 0.20$, $z(8,730) = 2.34$, $p = 0.02$), that is, the effect of predictability was larger in the group that was asked to type in the whole sentence (Experiment 2) than in the group that was asked to type only the sentence-final target word (Experiment 1). Together, these findings suggest that the change in task instruction, which draws attention either to the entire sentence or only to the final word, is critical to whether the context information is used under degraded speech. But degraded speech comprehension is not reduced by binding listeners' attention allocation to one part of the speech stream.

4. General discussion

The main goals of the present study were to investigate whether online semantic predictions are formed in comprehension of degraded speech when task instructions encourage attention to the processing of the context information, or only to the critical target word. The results of two experiments revealed that attentional processes clearly modulate the use of context information for predicting sentence endings when the speech signal is moderately degraded.

In contrast to the first experiment, the results of our second experiment show an interaction between target word predictability and degraded speech. This is generally in line with the few existing studies that found a facilitatory effect of predictability at different levels of speech degradation when the participants were instructed to pay attention to the entire sentence (e.g., at 4-channel, or at 8-channel; Bhandari et al., 2021; Obleser & Kotz, 2010; Obleser et al., 2007). The important new finding that our study adds to the present literature is that this effect may be weakened or lost when listeners are instructed to report only the final word of the sentence that they heard (Experiment 1). The lack of predictability effect (or contextual facilitation) can most likely be attributed to listeners not successfully decoding the meaning of the verb of the sentence, as the verb is the primary predictive cue in our stimuli for the target word (noun). Hence, this small change in task instructions from Experiment 1 to Experiment 2 sheds light on the role of top-down regulation of attention in using context for language comprehension in adverse listening conditions. In an adverse listening condition, language comprehension is generally effortful, so that focusing attention on only a part of the speech signal seems beneficial in order to enhance stimulus decoding. However, the results of this study also show that this comes at the cost of neglecting the context information that could be beneficial for language comprehension. Our findings hence demonstrate that there is a trade-off between the use of context for generating top-down predictions vs. focusing all attention on a target word. Specifically, the engagement in the use of context and generation of top-down predictions may change as a function of attention (see also Li et al., 2014). This claim is also corroborated by the significant change in predictability effects (or contextual facilitation) from Experiment 1 to Experiment 2, in the combined dataset. Findings from the irrelevant-speech paradigm also support our conclusion. It has been shown that the predictability of unattended speech has no effect on the main experimental task (e.g., memorization of auditorily presented digits). Wöstmann and Obleser (2016) did not find predictability effects when the participants ignored the degraded speech (see also Ellermeier et al., 2015). An alternative explanation of 'participants neglecting the context' could be that the participants did not listen to the context at all, or they heard but did not process the context. However,

irrelevant-speech paradigm studies show that listeners cannot avoid listening to the speech presented to them; to-be-ignored speech has been shown to interfere with the main experimental task (e.g., LeCompte, 1995). It is not implausible that the listeners listened to the context but did not do a deep processing. This is not incompatible with our first explanation, as in either case, attention to the final word leaves the listeners with limited resources to process and form a representation of the context information.

At this point, we note the differences in response accuracies across different levels of speech degradation, and contextual facilitation therein. At 8-channel condition, the speech was least degraded, and listeners recognized more words than in the 4- or 6-channel conditions, which is in line with prior studies that have found an increase in intelligibility and word recognition with an increase in number of channels (e.g., Davis et al., 2005; Obleser et al., 2011). Speech signal passed through 4-channel noise-vocoding was most degraded. Therefore, in the second experiment, at 4-channel, attending to the entire sentence did not confer contextual facilitation because decoding the context itself was difficult. Listeners could not utilize the context differentially across high and low predictability sentences to generate semantic predictions. At 6-channel – a moderate level of degradation – listeners could attend to, identify, and decode the context; hence we observed the significant difference in response accuracy between high and low predictability sentences. We observed a similar contextual facilitation at 8-channel as well. This is in line with previous findings (e.g., Obleser et al., 2007; but see also Obleser & Kotz, 2010) which show that predictability effects can be observed at a moderate degradation level of 8-channel or less. To summarize, our results indicate that there was a very strong difference in intelligibility between 4- and 6-channel conditions, but that the difference in intelligibility between 6- and 8-channel conditions was minor. Note, however, that even for 8-channel, low predictability sentences were not always understood correctly.

Considering theoretical accounts of predictive language processing (Friston et al., 2020; Kuperberg & Jaeger, 2016; McClelland & Elman, 1986; Norris et al., 2016; Pickering & Gambi, 2018), one would expect that listeners automatically form top-down predictions about upcoming linguistic stimuli based on prior context. Also, when speech is degraded, top-down predictions render a benefit in word recognition and language comprehension (e.g., Corps & Rabagliati, 2020; Sheldon et al., 2008a, 2008b). Results of our study revealed new theoretical insights by showing that this is not always the case. Top-down predictions are dependent on attentional processes (see also Kok et al., 2012), directed by task instructions; thus they are not *always* automatic, and predictability does not *always* facilitate language comprehension of degraded speech. To this point, our findings shed light on the growing body of literature indicating limitations of predictive language processing accounts (Huettig & Guerra, 2019; Huettig & Mani, 2016; Mishra et al., 2012; Nieuwland et al., 2018).

Results from both experiments show that the effect of trial number is not significant. In contrast to previous studies (e.g., Davis et al., 2005; Erb et al., 2013) we did not observe adaptation to noise-vocoded speech. In those studies, there was certainty about the speech quality of the next trial, as the participants were presented with only one level of spectral degradation (only 4-channel or only 6-channel noise-vocoding), and crucially with no specific regard to semantic predictability. On the contrary, in our study, listeners were always uncertain about the speech quality of the next trial as well as its semantic predictability. Because of this changing context, the perceptual system of the participants may not retune itself (cf. Goldstone, 1998;

Mattys et al., 2012). This is also in line with our prior finding that listeners do not adapt to degraded speech when there is a trial-by-trial variation in perceptual and semantic features (Bhandari et al., 2021).

We also should note the limitations of the current study. In our experiments, we have used short Subject–Verb–Object sentences in which the verb is predictive of the noun, and we have given participants the somewhat unnatural task of reporting the last word of a sentence. In more naturalistic sentence comprehension, participants would normally aim to understand the full utterance, and would most likely not have restricted goals such as first and foremost decoding a word in a specific position of the sentence. Instead, the speaker would usually indicate important words or concepts via pitch contours, stress, or intonation patterns, which would then direct the attention of a listener. Furthermore, the sentences uttered in most day-to-day conversations are longer, and context information builds up more gradually – information from several words is usually jointly predictive of upcoming linguistic units. Similarly, the design of our experiments limits our ability to discern whether participants generated predictions online while processing the speech, or while typing in the words after listening to the degraded speech.

To conclude, we show that task instructions affect distribution of attention to the noisy speech signal. This, in turn, means that when insufficient attention is given to the context, top-down predictions cannot be generated, and the facilitatory effect of predictability is substantially reduced.

Supplementary Materials. To view supplementary materials for this article, please visit <http://doi.org/10.1017/langcog.2022.16>.

Data Availability Statement. The code and data mentioned above are available in the public repository of Open Science Framework – <https://osf.io/t6unj/>.

Conflict of Interest. We conducted this research with no relationship, financial or otherwise, that could be a potential conflict of interest.

References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57(4), 502–518. <https://doi.org/10.1016/j.jml.2006.12.004>
- Amichetti, N. M., Atagi, E., Kong, Y. Y., & Wingfield, A. (2018). Linguistic context versus semantic competition in word recognition by younger and older adults with cochlear implants. *Ear and Hearing*, 39(1), 101–109. <https://doi.org/10.1097/AUD.0000000000000469>
- Ankenier, C. S., Sekicki, M., & Staudte, M. (2018). The influence of visual uncertainty on word surprisal and processing effort. *Frontiers in Psychology*, 9, 1–17. <https://doi.org/10.3389/fpsyg.2018.02387>
- Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, 49(12), 3512–3516. <https://doi.org/10.1016/j.neuropsychologia.2011.08.014>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

- Bhandari, P., Demberg, V., & Kray, J. (2021). Semantic predictability facilitates comprehension of degraded speech in a graded manner. *Frontiers in Psychology*, 12, 714485.
- Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the extra mile: Effects of discourse context on two late positivities during language comprehension. *Neurobiology of Language*, 1, 135–160.
- Brunellière, A., Auran, C., & Delrue, L. (2019). Does the prosodic emphasis of sentential context cause deeper lexical-semantic processing? *Language, Cognition and Neuroscience*, 34(1), 29–42.
- Chen, F., & Loizou, P. C. (2011). Predicting the intelligibility of vocoded speech. *Ear and Hearing*, 32(3), 331–338. <https://doi.org/10.1038/mp.2011.182>.doi
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25(5), 975–979. <https://doi.org/10.1121/1.1907229>
- Cooke, M., Scharenborg, O., & Meyer, B. T. (2022). The time course of adaptation to distorted speech. *The Journal of the Acoustical Society of America*, 151(4), 2636–2646.
- Corps, R. E., & Rabagliati, H. (2020). How top-down processing enhances comprehension of noise-vocoded speech: Predictions about meaning are more important than predictions about form. *Journal of Memory and Language*, 113, 104114. <https://doi.org/10.1016/j.jml.2020.104114>
- Darwin, C. (2005). Praat scripts for producing Shannon AM speech [Computer software]. Retrieved March 8, 2021, from http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/.
- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., 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(2), 222–241. <https://doi.org/10.1037/0096-3445.134.2.222>
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121. <https://doi.org/10.1038/nn1504>
- 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(Suppl 1), 101S–110S. <https://doi.org/10.1097/AUD.0000000000000300>.Is
- Ellermeier, W., Kattner, F., Ueda, K., Doumoto, K., & Nakajima, Y. (2015). Memory disruption by irrelevant noise-vocoded speech: Effects of native language and the number of frequency bands. *The Journal of the Acoustical Society of America*, 138(3), 1561–1569.
- Erb, J. (2014). *The neural dynamics of perceptual adaptation to degraded speech*. Doctoral Dissertation, Max Planck Institute for Human Cognitive and Brain Sciences.
- Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *Journal of Neuroscience*, 33(26), 10688–10697. <https://doi.org/10.1523/JNEUROSCI.4596-12.2013>
- Federmeier, K., McLennan, D., de Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39(02), 133–146.
- Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149–161. <https://doi.org/10.1016/j.bandl.2010.07.006>
- Fontan, L., Tardieu, J., Gaillard, P., Woisard, V., & Ruiz, R. (2015). Relationship between speech intelligibility and speech comprehension in babble noise. *Journal of Speech, Language, and Hearing Research*, 58, 977–986. <https://doi.org/10.1044/2015>
- Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(5), 862–877. <https://doi.org/10.1037/0278-7393.31.5.862>
- Friston, K. J., Parr, T., Yufik, Y., Sajid, N., Price, C. J., Holmes, E., & Square, Q. (2020). Generative models, linguistic communication and active inference. *Neuroscience & Biobehavioral Reviews*, 118, 42–64. <https://doi.org/10.1016/j.neubiorev.2020.07.005>
- Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention - Focusing the searchlight on sound. *Current Opinion in Neurobiology*, 17(4), 437–455. <https://doi.org/10.1016/j.conb.2007.07.011>
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585–612. <https://doi.org/10.1146/annurev.psych.49.1.585>

- Greenwood, D. D. (1990). A cochlear frequency-position function for several species – 29 years later. *The Journal of the Acoustical Society of America*, 87(6), 2592–2605. <https://doi.org/10.1121/1.399052>
- Hafer, E. R., Sarampalis, A., & Loui, P. (2007). Auditory attention and filters. *Auditory Perception of Sound Sources*, 29, 115–142. https://doi.org/10.1007/978-0-387-71305-2_5
- Huetting, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, 1706, 196–208. <https://doi.org/10.1016/j.brainres.2018.11.013>
- Huetting, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, 31(1), 19–31. <https://doi.org/10.1080/23273798.2015.1072223>
- Hunter, C. R., & Pisoni, D. B. (2018). Extrinsic cognitive load impairs spoken word recognition in high- and low-predictability sentences. *Ear and Hearing*, 39(2), 378–389. <https://doi.org/10.1097/AUD.0000000000000493>
- Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. Stanford University.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94(2), 113–147. <https://doi.org/10.1016/j.cognition.2004.01.002>
- Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95(1), 95–127. <https://doi.org/10.1016/j.cognition.2004.03.002>
- Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, 22(9), 2197–2206. <https://doi.org/10.1093/cercor/bhr310>
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59. <https://doi.org/10.1080/23273798.2015.1102299>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed-effects models. *Journal of Statistical Software*, 82, 1–26. <http://doi.org/10.18637/jss.v082.i13>
- Lange, K. (2013). The ups and downs of temporal orienting: A review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Frontiers in Human Neuroscience*, 7, 1–14. <https://doi.org/10.3389/fnhum.2013.00263>
- Lange, K., & Röder, B. (2010). Temporal orienting in audition, touch, and across modalities. In *Attention and Time* (pp. 393–405). Oxford University Press.
- LeCompte, D. C. (1995). An irrelevant speech effect with repeated and continuous background speech. *Psychonomic Bulletin & Review*, 2(3), 391–397.
- Li, X., Lu, Y., & Zhao, H. (2014). How and when predictability interacts with accentuation in temporally selective attention during speech comprehension. *Neuropsychologia*, 64, 71–84. <https://doi.org/10.1016/j.neuropsychologia.2014.09.020>
- Li, X., Zhang, Y., Li, L., Zhao, H., & Du, X. (2017). Attention is shaped by semantic level of event-structure during speech comprehension: An electroencephalogram study. *Cognitive Neurodynamics*, 11(5), 467–481. <https://doi.org/10.1007/s11571-017-9442-4>
- 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(7–8), 953–978. <https://doi.org/10.1080/01690965.2012.705006>
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- Mishra, R. K., Singh, N., Pandey, A., & Huetting, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability – Evidence from Indian low and high literates. *Journal of Eye Movement Research*, 5(1), 1–10. <https://doi.org/10.16910/jemr.5.1.3>
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425. <https://doi.org/10.1111/j.1469-8986.1987.tb00311.x>

- Nieuwland, M. S. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience and Biobehavioral Reviews*, 96, 367–400. <https://doi.org/10.1016/j.neubiorev.2018.11.019>
- Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., von Grebmer Zu Wolfsturn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., ... Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, 7, 1–24. <https://doi.org/10.7554/elife.33468>
- Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, 31(1), 4–18. <https://doi.org/10.1080/23273798.2015.1081703>
- Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20(3), 633–640. <https://doi.org/10.1093/cercor/bhp128>
- Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in comprehension of degraded speech. *Neuroimage*, 55(2), 713–723. <https://doi.org/10.1016/j.neuroimage.2010.12.020>
- Obleser, J., Wise, R. J. S., Alex Dresner, M., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, 27(9), 2283–2289. <https://doi.org/10.1523/JNEUROSCI.4663-06.2007>
- Pals, C., Sarampalis, A., & Başkent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research*, 56(4), 1075–1084. [https://doi.org/10.1044/1092-4388\(2012\)12-0074](https://doi.org/10.1044/1092-4388(2012)12-0074)
- 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>
- Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology*, 47(Suppl 2), S72–S82. <https://doi.org/10.1080/14992020802307404>
- Pickering, M., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144(10), 1002. <https://doi.org/10.1037/bul0000158>
- Prolific. (2014). *Prolific academic*. <https://www.prolific.co>.
- Rayner, K., Slattery, T. J., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 514–528. <https://doi.org/10.1037/a0020990>
- Ryskin, R., & Fang, X. (2021). The many timescales of context in language processing. In *Psychology of learning and motivation - Advances in research and theory* (1st ed.). Elsevier. <https://doi.org/10.1016/bs.plm.2021.08.001>
- R Core Team (2020). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Available at: <https://www.R-project.org>
- Sanders, L. D., & Astheimer, L. B. (2008). Temporally selective attention modulates early perceptual processing: Event-related potential evidence. *Perception and Psychophysics*, 70(4), 732–742. <https://doi.org/10.3758/PP.70.4.732>
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–304. <https://doi.org/10.1126/science.270.5234.303>
- Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008a). Effect of age, presentation method, and learning on identification of noise-vocoded words. *The Journal of the Acoustical Society of America*, 123(1), 476–488. <https://doi.org/10.1121/1.2805676>
- Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008b). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, 123(1), 489–499. <https://doi.org/10.1121/1.2783762>
- Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, 9(8), 311–327. <https://doi.org/10.1111/lnc3.12151>
- Stilp, C. (2020). Acoustic context effects in speech perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 11(1), 1–18. <https://doi.org/10.1002/wcs.1517>
- Strauss, D. J., & Francis, A. L. (2017). Toward a taxonomic model of attention in effortful listening. *Cognitive, Affective and Behavioral Neuroscience*, 17(4), 809–825. <https://doi.org/10.3758/s13415-017-0513-0>

- Tóth, B., Honbolygó, F., Szalárdy, O., Orosz, G., Farkas, D., & Winkler, I. (2020). The effects of speech processing units on auditory stream segregation and selective attention in a multi-talker (cocktail party) situation. *Cortex*, 130, 387–400. <https://doi.org/10.1016/j.cortex.2020.06.007>
- Thorton, A.R.D., Harmer, M., & Lavoie, B.A. (2007). Selective attention increases the temporal precision of the auditory N100 event-related potential. *Hearing Research*, 230(1-2), 73–79. <https://doi.org/10.1016/j.heares.2007.04.004>
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience*, 32(40), 14010–14021. <https://doi.org/10.1523/JNEUROSCI.1528-12.2012>
- Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear and Hearing*, 36(4), e153–e165. <https://doi.org/10.1097/AUD.0000000000000145>
- Wöstmann, M., & Obleser, J. (2016). Acoustic detail but not predictability of task-irrelevant speech disrupts working memory. *Frontiers in Human Neuroscience*, 10, 538.
- Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30(6), 648–672. <https://doi.org/10.1080/23273798.2014.995679>

Cite this article: Bhandari, P., Demberg, V. & Kray, J. (2022). Predictability effects in degraded speech comprehension are reduced as a function of attention *Language and Cognition* 14: 534–551. <https://doi.org/10.1017/langcog.2022.16>