Speaking fast and slow: How speech rate affects contextual facilitation in degraded speech comprehension

Pratik Bhandari[1,2], Vera Demberg[2,3], Jutta Kray[1]

[1]Department of Psychology, Saarland University

[2]Department of Language Science and Technology, Saarland University

[3]Department of Computer Science, Saarland University

Financial disclosures/conflicts of interest: This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102. The authors have no conflict of interest, financial, or otherwise, to declare.

Data, experimental materials, and scripts used for data analyses are openly available at the open science framework (osf) repository [https://osf.io/b2czt/?view_only=6e99c094ac1146daad7bf0bef3a030cc].

All correspondence should be addressed to:
Pratik Bhandari, Campus Building A1.3, Room 2.11, Saarland University, 66123, Saarbrücken, Germany. E-mail: pratikb@coli.uni-saarland.de

ABSTRACT

Purpose: The aim of the study was to examine the effect of speed of information flow, i.e., speech rate, on the facilitatory effect of predictability when the speech is moderately degraded.

Method: We created sentences with two levels of predictability (high and low) that were spectrally degraded at 4 channels noise vocoding. In Experiment 1 (n=101) we compared normal speech and fast speech (compressed by a factor of 0.65), and in Experiment 2 (n=101) we compared normal speech and slow speech (expanded by a factor of 1.35). Participants listened to high and low predictability sentences presented at different speech rates and reported the entire sentence that they heard.

Results: Our results showed that contextual facilitation is reduced in fast speech as it selectively impedes comprehension of low predictability sentences. In contrast, slow speech does not significantly improve contextual facilitation; instead, slow sentences are overall harder to understand than the sentences presented at normal speech rate.

Conclusion: Facilitatory effect of predictability observed at a moderate level of degradation for normal speech rate is not amplified by providing more time for processing by slow speech rate. With an increase in speech rate, however, the facilitatory effect is significantly reduced.

INTRODUCTION

When speech is degraded, its intelligibility and comprehension are reduced. For example, degradation by noise vocoding strips off the speech from its spectral properties rendering it difficult to understand (Shannon et al., 1995; Davis et al., 2005). Studies have undisputedly shown that semantic predictability facilitates comprehension of moderately degraded speech (Obleser et al., 2010; Bhandari et al., 2021). That is, listeners utilize context information and form predictions about upcoming linguistic units which in turn facilitates the comprehension of the degraded speech. Such degraded speech is intrinsically effortful to process (Eckert et al., 2016; Wild et al., 2012), and limits the cognitive resources available to encode the context information and form predictions (cf. Huettig & Janse, 2016). Additionally, studies on clear speech have suggested that the demand on cognitive resources to process speech is moderated by the rate of flow of incoming information, i.e., by increase or decrease in speech rate (Gordon-Salant & Fitzgibbons, 1995). More processing time is available for slow speech, and less so for fast speech. We know from the studies on clear speech that although contextual facilitation is reduced in fast speech, the evidence of enhanced contextual facilitation in slow speech is mixed (Aydelott & Bates, 2004; Goy et al., 2013; Koch & Janse, 2016). Specifically for degraded speech, there is no clear evidence how the facilitatory effect at a moderate level of degradation is affected by different speech rates (Iwasaki, 2002; Meng et al., 2019; see also Winn & Teece, 2021). The aim of the present study was therefore to investigate the interplay between the rate of information flow (fast and slow speech rate) and the facilitatory effect of semantic predictability (i.e., contextual facilitation) at a moderate level of spectral degradation. In the following, we first summarize the impact of speech degradation in language comprehension, and its interaction with sentence context, and then the influence of speech rate on language comprehension, and its interaction with sentence context.

Comprehension of degraded speech

There are a number of studies showing that language comprehension is hampered when the bottom-up input is less intelligible due to spectral degradation of the speech signal (e.g., Shannon et al., 1995; Davis et al., 2005). These studies have used noise vocoding as a method of spectral degradation of speech. Here the speech signal is first divided into a specific number of frequency bands that corresponds to the number of vocoder channels. The amplitude envelope within each frequency band is extracted, and the spectral information within it is replaced by noise. The resulting vocoded speech contains temporal cues of the original speech, but it is difficult to understand – the smaller the number of vocoder channels, the more reduced is the intelligibility. More effort and cognitive resources are required to process and comprehend such degraded speech as compared to clean speech (e.g., Eckert et al., 2016; Wild et al., 2012).

When the speech signal is less intelligible due to spectral degradation, listeners rely more on top-down predictions (e.g., Sheldon et al., 2008). Hence, they use the context information of the sentence to narrow down their predictions to a smaller set of semantic categories or words (Strauß et al., 2013; see also, Corps & Rabagliati, 2020). However, it is important that the context itself is ‘intelligible enough’ which is the case when the speech is only moderately degraded. For example, Obleser and colleagues (Obleser et al., 2007; Obleser & Kotz, 2010) found that at moderate levels of speech degradation, target words (the sentence final words) were better recognized when they were predictable from the sentence context than when they were unpredictable. When the speech signal is clear or only very mildly degraded, there is typically no effect of predictability on comprehension, as even unpredictable words can be understood well in this condition (intelligibility is at ceiling). In contrast, when the speech signal is extremely degraded (for instance, at 1 channel noise vocoding), no facilitation from the context can be observed as the context itself cannot be understood and hence it cannot help with comprehension (Bhandari et al., 2021; Obleser & Kotz, 2010).

Taken together, these studies show that semantic predictability facilitates comprehension of degraded speech at a moderate level of spectral degradation, for example, at 4 channels noise vocoding, when the speech is intelligible enough for the listeners to understand and form meaning representation of the context to generate predictions about upcoming word in a sentence.

Comprehension of fast and slow speech

A change in speech rate (by uniform compression or expansion) manipulates the speech signal but does not by itself produce any spectral degradation (Charpentier & Stella, 1986; Moulines & Charpentier, 1990; Schlueter et al., 2014). Understanding fast speech is more effortful compared to normal and slow speech (e.g., Müller et al., 2019; Winn & Teece, 2021; see also, Simantiraki & Cooke, 2020), and its intelligibility and comprehension are reduced (Fairbanks & Kodman Jr., 1957; Peele & Wingfield, 2005; Schlueter et al., 2014). With an increased speech rate, processing demand increases as less time is available to process the incoming information (Gordon-Salant & Fitzgibbons, 1995; Rodero, 2016; see also, Rönnberg et al., 2013). This exhausts the cognitive resources required for language processing (Gordon-Salant & Fitzgibbons, 2004; Janse, 2009). Since cognitive resources are required also for generating predictions (Pickering & Gambi, 2018), it can be expected that in fast speech, the effect of predictability is reduced as the resources are reduced to encode and process the context information. In contrast, the central auditory-language comprehension system is shown to be flexible to process slow speech without reducing intelligibility (Lerner et al., 2014). So, it can be expected that slow speech does not limit cognitive resources, and therefore processing context information to generate predictions in slow speech is not different than that in normal speech. Alternatively, slow speech provides more time to buffer the auditory information at the lower level of information processing hierarchy (Ghitza & Greenberg, 2009; Vagharchakian et al., 2012), and consequently provides more time for central language comprehension system to use the context information and form predictions. Studies from Visual World Paradigm also support this claim that slow speech provides more time for speech processing, and for semantic prediction (e.g., Fernandez et al., 2020; Huettig & Janse, 2016). However, some earlier studies have casted doubt on the processing advantage of slow speech arguing that slow speech is perceived as overly artificial, and it demands high working memory (e.g., Kemper & Harden, 1999; Nejime & Moore, 1998; see also, Liu & Zeng, 2006; Love et al., 2009). Therefore, given these competing accounts, it is unclear whether the effect of predictability increases in slow speech compared to normal speech. Nonetheless, it is clear that compression and expansion of speech have different effects on speech intelligibility and language comprehension. Fast speech reduces intelligibility and comprehension, but the evidence for the effect of slow speech on language comprehension is mixed.

A few studies have directly examined the role of predictability in understanding language when speech is fast or slow by using clear speech. For instance, Aydelott and Bates (2004) used a priming paradigm to examine the effects of contextual cues, which were target words embedded in sentences, and compared fast speech to normal speech. Target words were either congruent to the sentence context (100% cloze probability, i.e., in a constraining sentence context), incongruent (0% cloze probability, i.e., in an implausible sentence), or neutral (cloze probability not mentioned). Results indicated no reduction in facilitatory effect of contextual cues (congruent versus neutral target words) at fast speech compared to normal speech. In contrast, they found a reduced inhibitory effect (incongruent versus neutral target words). They argued that the constraining sentence context was easy to process – fast speech did not interfere with the earlier stage of activation of words that matched the context (i.e., in congruent trials). In contrast, the inhibition effect was reduced because there was less time to build up the representation of words in implausible sentence contexts so that less inhibition of the incongruent target word was needed. However, in a replication study of Aydelott and Bates (2004), Goy et al. (2013) found that the facilitatory effect was reduced in fast speech compared to normal speech. They argued that the fast speech slowed down the activation of potential target words that matched the context, which effectively reduced the contextual facilitation. In a recent study, Winn and Teece (2021) did not observe an increase in contextual facilitation for slow speech compared to normal speech, although the intelligibility was slightly higher for slow speech among cochlear implantees. In another experiment, Koch and Janse (2016) presented participants with a question-answer sequence of varying length across a wide range of normal and fast speech from clear speech of Spoken Dutch Corpus (Oostdijk, 2000). They did not find any effect of predictability on word recognition. However, target word predictability and target word position in the sentences were not systematically controlled for in their study.

The effects of predictability at varying rates of presentation have been also investigated with self-paced reading studies. For example, Wlotko and Federmeier (2015) presented participants with context evoking sentences followed by sentences containing a target word that was either *expected* (mean cloze probability of 74%) or *unexpected* (either same or different semantic category, both with cloze probability of approximately 0%). They found that the facilitation effect (as reflected in the N400 amplitude) was reduced at the sentences that were presented fast compared to the ones that were presented slow. They suggested that at fast rate of presentation, predictive preactivation of words was not common: There was not enough time to activate proper representation for processing of upcoming word. In the same study, however, when the fast presentation was followed by slow presentation in separate blocks, semantic facilitation effect was not reduced. That is, increase in the flow of information did not *always* impair the ability to predict (see also, Cole, 2020). They argued that once the brain is engaged in predictive comprehension mode, for example, first in the slow presentation rate, it then continues to allocate resources in the same mode under faster presentation rate.

To summarize, there is already some evidence from studies applying various paradigms that the predictability of the sentence context interacts with the speech rate (Aydelott & Bates, 2004; Sharit et al., 2003; Winn & Teece, 2021; Wlotko & Federmeier, 2015). Predictability effect is generally reduced for fast speech, while the findings are in conflict in the case of slow speech.

Speech rate and contextual facilitation of moderately degraded speech

Predictions about upcoming linguistic units are generated as a listener forms the meaning representation of the context information from the auditory input as it unfolds in time. At a moderate level of spectral degradation, understanding the context and forming predictions facilitate comprehension. The effect of predictability observed at the moderate degradation level no longer exists if the listener does not understand the context. Therefore, it is important that the speed at which the speech unfolds remains within the listener's limit to buffer and process the auditory information (Vagharchakian et al., 2012) so that one can form the representation of the context and have sufficient time to generate predictions.

There are a plenty of studies on the role of speech rate on the intelligibility and comprehension of degraded speech, but these do not consider predictability effects. For example, Meng et al. (2019) found that an increase in speech rate had a much more severe effect on spectrally degraded speech (4 channels sine-wave vocoded) than on clear speech. To achieve the same level of accuracy, listeners required degraded speech to be much slower than normal speech. Among cochlear implantees whose speech input is spectrally degraded (Shannon et al., 2004), Iwasaki et al. (2002) found that a change in speech rate from slow to fast reduced word recognition accuracy. Their speech perception was impaired with increased rate of speech, and it was improved when the speech rate was decreased (e.g., Dincer D'Alessandro et al., 2018). In Winn et al. (2021)'s study, there was no significant difference in facilitatory effect of semantic predictability between slow speech and normal speech rates. This was attributed to listeners' "repair" strategy at normal speech rate such that they made sensible guesses about the words that fit the given context. There have not been similar studies to study the effect of speech rate on predictive processing of moderately degraded speech, specifically among normal hearing listeners.

Taken together, the utility of semantic predictability in comprehension of degraded speech is fairly established whereas its effects in comprehension of both fast speech as well as slow speech is inconsistent. Similarly, prediction itself is time and resource consuming mechanism (Pickering & Gambi, 2018) which is affected by a comprehender's processing speed (e.g., Huettig & Janse, 2016). However, the role of speed of incoming information (i.e., speech rate) on a listener's ability to form predictions, and hence its interplay with the facilitatory effect of semantic predictability at a moderately degraded speech remains under-studied and unexplained.

Study aims

We systematically examined whether contextual facilitation at a moderate level of degradation varies with a change in speech rate. The aim was to investigate whether the increase (and decrease) in speech rate reduces (and amplifies) the facilitatory effect of semantic predictability that has been observed for moderately degraded speech at a normal speech rate. Semantic predictability was manipulated by varying the cloze probability of target words, and moderate degradation was achieved by noise vocoding of speech through 4 channels; Obleser & Kotz (2010) and Bhandari et al. (2021) have reported 4 channels noise vocoding to be the moderate degradation level at which contextual facilitation is observed. Speech rate was manipulated by compression (and expansion) of the moderately degraded speech, by uniform pitch synchronous overlap-add technique that acts upon the temporal envelope of the speech signal, to make it fast (and slow).

To achieve the goal, we conducted two experiments in which listeners were required to listen to the sentences and type in the entire sentence they hear. Sentence comprehension (word recognition accuracy) for high and low predictability sentences were assessed in fast speech (Experiment 1), and slow speech (Experiment 2). Because the processing demand increases, and a limited time is to be available to process the context and form predictions (e.g., Aydelott & Bates, 2004; Wlotko & Federmeier, 2015; see also, Pickering & Gambi, 2018), we expected that the contextual facilitation (i.e., the difference between high and low predictability sentences) will be reduced for fast speech compared to normal speech (Experiment 1). However, for slow speech due to abundance of time to process the degraded speech and the context, and reduction in effortful processing (e.g., Winn & Teece, 2021), we expected contextual facilitation to be increased compared to normal speech (Experiment 2). We expected that both increase and decrease in contextual facilitation will be primarily driven by the ease of processing high predictability sentences as compared to low predictability sentences (Aydelott & Bates, 2004; Goy et al., 2013).

EXPERIMENT 1

METHOD

Participants

We recruited 101 native speakers of German (*mean age* ± *SD* = 23.14 ± 3.31 years, age range = 18-30 years, 66 female, 1 preferred not to say) using the crowd-sourcing platform Prolific Academic. None of the participants reported any hearing loss, language-related disorder, cognitive impairment, or any neurological disorder. All participants received 6.20 Euro as monetary compensation for their participation in the experiment which was approximately 40 minutes long. The German Society for Language Science ethics committee approved the study and participants provided an informed consent in accordance with the declaration of Helsinki.

Materials

A subset of 240 German sentences were taken from Bhandari et al. (2021).[1] These sentences were read by a female native German speaker in an unaccented speech at a normal rate. The sentences were recorded and digitized at 44.1 kHz with 32-bit linear encoding. The sentence structure was Subject-Verb-Object (SVO sentences) in which verb was predictive of the noun (for examples, see Table 1). 120 nouns were used to create two types of sentences – high and low predictability – that differed in the cloze probability of the sentence-final target words (i.e., the nouns). The cloze probabilities of highly predictable and less predictable target words were 0.752 ± 0.123 (M ± SD; range = 0.56 - 1.00) and 0.022 ± 0.027 (M ± SD; range = 0.00 - 0.09) respectively, which were measured in a norming study among a separate group of participants (n = 60; age range = 18-30 years).

All the 240 recordings were compressed by a factor of 0.65 to create fast speech by using the Praat software that applies the pitch-synchronous overlap-add (PSOLA) approach (Charpentier & Stella, 1986; Moulines & Charpentier, 1990). Then we passed the recordings of all 240 normal speech and 240 fast speech sentences through 4 channels noise vocoding, using a customized Praat script originally written by Darwin (2005), to create moderately degraded speech. Frequency boundaries for 4 channels noise vocoding (70 Hz, 423 Hz, 1304 Hz, 3504 Hz, 9000 Hz) were approximately logarithmically spaced and determined by cochlear-frequency position functions (Erb, 2014; Greenwood, 1990).

Procedure

[1] Stimuli, and codes used in analyses are available in the public repository at [https://osf.io/b2czt/?view_only=6e99c094ac1146daad7bf0bef3a030cc].

Participants were asked to use headphones or earphones. A sample of noise vocoded speech not used in the main experiment and the practice trials was provided at the beginning of the experiment. Participants were asked to adjust the loudness to a preferred level of comfort. They were instructed to listen to the sentences and type in the keyboard what they heard, i.e., the entire sentence. Participants were informed at the beginning of the experiment that some sentences that they hear would be 'noisy' and not easy to understand, and in such cases, guessing was encouraged. We did not impose time limit on the participants' response. To familiarize the participants with the task, 8 practice trials at different levels of speech degradation were presented at the beginning of the experiment. After the practice trials, the participants were presented 120 experimental trials with an inter-trial interval of 1000 ms.

Four experimental lists were constructed such that each participant was presented with 60 high predictability and 60 low predictability sentences. Speech rate was also balanced across each predictability condition in each list. For each predictability condition, 30 sentences with fast speech and 30 with normal speech were presented. Sentences were pseudo-randomized so that no more than three sentences of the same predictability level, or same speech rate appeared consecutively.

Statistical analyses

Data pre-processing and analyses were performed in RStudio (R version 4.1.1; R Core Team, 2021). We analysed accuracy using Generalized Linear Mixed Models (GLMMs) with R packages lmerTest (Kuznetsova et al., 2017) and lme4 (Bates et al., 2015). We first filtered out the trials in which participants incorrectly identified the context (i.e., verb). In the remaining trials, binary responses (correct and incorrect) on target word (i.e., noun) recognition 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. Target word predictability (categorical; high, low), speech rate (categorical; normal speech and fast speech), and the interaction of target word predictability and speech rate were included in the fixed effects. We 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 target word predictability, speech rate, and their interaction. We applied treatment contrast for both predictability and speech rate, mapping low predictability and normal speech to the intercept.

RESULTS AND DISCUSSION

Mean response accuracies across all conditions are presented in Table 2. We found a significant main effect of target word predictability (β = 2.42, SE = .28, z = 8.55, p < .001) and a significant main effect of speech rate (β = -0.98, SE = .24, z = 4.16, p < .001). These suggested that participants' response accuracy was higher for the high predictability sentences than for the low predictability sentences, and for normal speech than for fast speech. We also found a significant interaction between target word predictability and speech rate (β = 1.06, SE = .42, z = 2.50, p = .01). As it can be seen in Figure 1, the effect of target word predictability was reduced at fast speech. These results are shown in Table 3.

Separate planned analyses of each predictability level were performed following the same procedure described above in the Analysis section. There was no significant main effect of speech rate at high predictability condition (β = .02, SE = .34, z = .05, p = .96). At low predictability condition, in contrast, we found a significant main effect of speech rate (β = -.99, SE = .27, z = -3.72, p < .001). Hence, response accuracy decreased at fast speech only for the low predictability condition.

Separate planned analyses of each speech rate revealed that there was significant main effect of predictability in both normal speech (β = 1.98, SE = 0.28, z = 7.05, p < .001) and fast speech conditions (β = 2.67, SE = .37, z = 7.14, p < .001), but the effect appeared to be higher for fast speech (β = 2.67) than for normal speech (β = 1.98). This, however, is a result of significant reduction in accuracy at low predictability condition at fast speech rather than due to an increase in accuracy at high predictability condition. This can also be seen in Table 2 and Figure 1.

These results indicated an increase in response accuracy with an increase in target word predictability only at a normal speech rate. Fast speech rate significantly affected accuracy at low predictability condition such that the contextual facilitation was essentially reduced. These findings align with previous studies conducted with clean speech that found fast speech to reduce contextual facilitation (e.g., Aydelott & Bates, 2004). The results of the first experiment showed ease of processing high predictability sentences compared to low predictability sentences at moderately degraded fast speech. We conducted another experiment to examine if slowing down the speech rate eases processing of both low and high predictability sentences and increases the contextual facilitation in comprehension of moderately degraded speech.

EXPERIMENT 2

METHOD

Participants and Materials

We recruited 101 participants (*mean age* ± *SD* = 23.49 ± 3.26 years, age range = 18-30 years, 60 female, 1 preferred not to say) online via Prolific Academic following the same procedure and criteria as in Experiment 1. We used the same sentences that were used in Experiment 1. However, the auditory recordings were expanded by a factor of 1.35 to create slow speech. All other procedure to create stimuli were identical to Experiment 1 resulting in two sets (one set of normal and one set of slow speech) of 240 sentences (120 high and 120 low predictability sentences) which were passed through 4 channels noise vocoding.

Procedure

The procedure was also identical to Experiment 1. Four experimental lists were constructed such that each participant was presented with 60 high predictability and 60 low predictability sentences. Speech rate was also balanced across each predictability condition in each list. For each predictability condition, 30 sentences with slow speech and 30 with normal speech were presented. Sentences were pseudo-randomized so that no more than three sentences of the same predictability level, or same speech rate appeared consecutively.

Statistical analyses

Data analyses procedure was the same as Experiment 1. We applied treatment contrast for both predictability and speech rate, mapping low predictability and normal speech to the intercept.

RESULTS AND DISCUSSION

Mean response accuracies for all experimental conditions are shown in Table 4. There was a significant main effect of target word predictability indicating that participants' response accuracy was higher for the high predictability condition than for the low predictability condition (β = 2.58, SE = .30, z = 8.65, p < .001). In contrast to Experiment 1, we did not find a significant main effect of speech rate (β = -.08, SE = .15, z = .57, p = .568), nor there was a significant interaction between speech rate and target word predictability (β = .44, SE = .27, z = 1.65, p = .099). These suggested that there was no change in participants' response accuracy with a reduction in speech rate, nor did the contextual facilitation significantly increase or decrease with slowing down of the speech rate. It can be seen in Figure 2 that the effect of target word predictability did not change with speech rate. These results are shown in Table 5.

In contrast to Experiment 1, the findings of Experiment 2 did not indicate differential effect of speech rates in the comprehension of high and low predictability sentences. While the results of Experiment 1 showed that speeding up the speech rate significantly reduced the accuracy of low predictability sentences, such a reduction was not observed in Experiment 2 when the speech rate was slowed down. Although listeners' response accuracy was reduced at both fast and slow speech rates as compared to normal speech rate, their ability to utilize context information was only impaired by fast speech.

GENERAL DISCUSSION

The main goal of the present study was to examine whether speech rate modulates the facilitatory effect of semantic predictability in comprehension of moderately degraded speech given that the ease of processing sentences varies with their predictability and speed. The results of the two experiments revealed that fast speech selectively impedes the comprehension of low predictability sentences, while slow speech has no effect on contextual facilitation at a moderate level of degradation.

In both experiments, our results showed a significant main effect of predictability at normal speech rate, i.e., we observed a facilitatory effect of semantic predictability at normal speech rate under moderate degradation level of 4 channels noise vocoding. This replicates the findings from earlier studies (Obleser & Kotz, 2010; Bhandari et al., 2021) in which participants were presented only with normal speech rate, and contextual facilitation was observed at 4 channels noise vocoding. At this moderate degradation level, listeners were able to decode the context and form its meaning representation. Consequently, they generated predictions about the upcoming target word in a sentence even in low predictability condition depending on the contextual constrain of the sentences (Bhandari et al., 2021; see also, Strauß et al., 2013).

The expected interaction between speech rate and target word predictability in Experiment 1 showed that comprehension of degraded speech was significantly impaired for low predictability sentences at fast speech rate. In contrast, there was little to no effect in comprehension of high predictability sentences at fast speech rate. Listening to degraded speech itself requires more attentional resources compared to clean speech (Wild et al., 2012). When presented as a fast speech, spectral degradation imposes additional cognitive demands; and less time is available to process the auditory signal. In such a rapidly unfolding event, it is difficult to decode the context information and form its meaning representation from the degraded speech to generate predictions about upcoming target word. This difficulty is amplified when target words are not easily predictable from the context (Aydelott & Bates, 2004). As a result, language comprehension in the low predictability condition is impaired more than that in the high predictability condition in Experiment 1.

In contrast to Experiment 1, we did not find the expected interaction between speech rate and target word predictability in Experiment 2, i.e., decrease in speech rate did not differentially affect the comprehension of high or low predictability sentences although the comprehension tended to decrease at slow speech rate compared to normal speech rate. As opposed to Experiment 1, we did not observe a significant change in contextual facilitation at a slow speech rate at a moderate degradation level in Experiment 2. Slowing down the speech provides listeners more time to process the information, including the context that is important to generate predictions. However, our findings show that the added time does not benefit intelligibility and comprehension of sentences with highly predictable target words any more than normally available time with normal speech rate does. This claim is in line with the results of Winn and Teece (2021) who reported that intelligibility does not increase when the speech is slowed down even though there is a difference in the intelligibility between high and low predictability sentences at both normal and slow speech rates.

Accounts from speech perception and predictive language processing point to a common expectation: contextual facilitation is enhanced when comprehenders have more time to process the presented information (e.g., DeLong et al, 2021; Huettig & Guerra, 2019; Kuperberg & Jaeger, 2016). However, there is conflicting empirical evidence on whether an increase or a decrease of speech rate provides benefit in intelligibility, comprehension, and contextual facilitation. Our findings show this interplay among spectral degradation, speech rate, and semantic prediction. Although reducing the speech rate provides time to process the information (including the context) in the degraded speech, this does not necessarily ease the processing of both high or low predictability sentences differentially. And thus, no increased facilitatory effect is observed at slow speech rate. In contrast, increasing the speech rate adds more cognitive load on the top of the effort required to process degraded speech. This results in difficulty in processing and understanding the rapidly unfolding sentences among which this difficulty is further increased when the target words are not easily predictable.

Of note, the generalisation of our results is also limited. First, we tested only with one expansion factor of 1.35, and one compression factor of 0.65. It can be speculated that when the speech is expanded to different levels by including other expansion factors, an increase in facilitatory effect could be observed. There could be an optimal trade-off between slowing down of the speech with more time to process (Fernandez et al., 2020) and the speech still remaining intelligible. Second, we only tested younger adults. We did not examine the effect of cognitive aging in contextual facilitation of comprehension of fast and slow speech. Older adults have delayed processing speed such that slow speech generally improves their speech intelligibility and language comprehension. And semantic context benefits older adults more than younger adults in adverse listening conditions. Therefore, the effect of slow speech could be different in older than what we found in younger adults in adverse listening condition.

To conclude, we show that processing speed and constrains in attentional and cognitive resources are key factors that influence contextual facilitation of moderately degraded speech. When enough time is available to process information, i.e., at slow speech, contextual facilitation does neither increase nor decrease. However, at a time crunch of information processing, i.e., at fast speech, contextual facilitation is reduced such that the fast speech is detrimental to understanding words that are not easily predictable from the context.

ACKNOWLEDGMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102.

REFERENCES

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*, *19*(1), 29–56.

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., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1).

Bhandari, P., Demberg, V., & Kray, J. (2021). Semantic predictability facilitates comprehension of degraded speech in a graded manner. *Frontiers in Psychology*, *3769*. https://doi.org/10.3389/fpsyg.2021.714485

Charpentier, F. J., & Stella, M. G. (1986). Diphone synthesis using an overlap-add technique for speech waveforms concatenation. *ICASSP'86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2015–2018.

Cole, A. (2020). *The effects of prediction and speech rate on lexical processing* (pp. 1–35) [Bachelor's thesis, University of Maryland].

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.

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.

Darwin, C. (2005). Praat scripts for producing Shannon AM speech [*Computer software*]. http:// http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/ (Last viewed 08 March 2021)

DeLong, K. A., Chan, W. H., & Kutas, M. (2021). Testing limits: ERP evidence for word form preactivation during speeded sentence reading. *Psychophysiology*, *58*(2), 1–7.

Dincer D'Alessandro, H., Boyle, P. J., Ballantyne, D., De Vincentiis, M., & Mancini, P. (2018). The role of speech rate for Italian-speaking cochlear implant users: Insights for everyday speech perception. *International Journal of Audiology*, *57*(11), 851–857.

Eckert, M. A., Teubner-Rhodes, S., & Vaden Jr, 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.

Erb, J. (2014). *The neural dynamics of perceptual adaptation to degraded speech* (p. 211) [Doctoral dissertation]. Universität Leipzig.

Fairbanks, G., & Kodman Jr., F. (1957). Word intelligibility as a function of time compression. *The Journal of the Acoustical Society of America*, *29*(5), 636–641.

Fernandez, L. B., Engelhardt, P. E., Patarroyo, A. G., & Allen, S. E. M. (2020). Effects of speech rate on anticipatory eye movements in the Visual World Paradigm: Evidence from aging, native, and non-native language processing. *Quarterly Journal of Experimental Psychology*, *73*(12), 2348–2361.

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*(1–2), 113–126.

Gordon-Salant, S., & Fitzgibbons, P. J. (1995). Recognition of multiply degraded speech by young and elderly listeners. *Journal of Speech and Hearing Research*, *38*(5), 1150–1156.

Gordon-Salant, S., & Fitzgibbons, P. J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *The Journal of the Acoustical Society of America*, *115*(4), 1808–1817.

Goy, H., Pelletier, M., Coletta, M., & Pichora-Fuller, M. K. (2013). The effects of semantic context and the type and amount of acoustic distortion on lexical decision by younger and older adults. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1715–1732.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *Journal of the Acoustical Society of America*, *87*(6), 2592–2605.

Huettig, 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.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31.

Iwasaki, S., Ocho, S., Nagura, M., & Hoshino, T. (2002). Contribution of speech rate to speech perception in multichannel cochlear implant users. *Annals of Otology, Rhinology and Laryngology*, *111*(8), 718–721.

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech* [Doctoral Dissertation]. Standford 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.

Janse, E. (2009). Processing of fast speech by elderly listeners. *The Journal of the Acoustical Society of America*, *125*(4), 2361–2373.

Kemper, S., & Harden, T. (1999). Experimentally disentangling what's beneficial about elderspeak from what's not. *Psychology and Aging*, *14*(4), 656–670.

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America*, *139*(4), 1618–1636.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13).

Lerner, Y., Honey, C. J., Katkov, M., & Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *Journal of Neurophysiology*, *111*(12), 2433–2444.

Liu, S., & Zeng, F.-G. (2006). Temporal properties in clear speech perception. *The Journal of the Acoustical Society of America*, *120*(1), 424–432.

Love, T., Walenski, M., & Swinney, D. (2009). Slowed speech input has a differential impact on on-line and off-line processing in children's comprehension of pronouns. *Journal of Psycholinguistic Research*, *38*(3), 285–304.

Meng, Q., Wang, X., Cai, Y., Kong, F., Buck, A. N., Yu, G., Zheng, N., & Schnupp, J. W. H. (2019). Time-compression thresholds for Mandarin sentences in normal-hearing and cochlear implant listeners. *Hearing Research*, *374*, 58–68.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(1990), 453–467.

Müller, J. A., Wendt, D., Kollmeier, B., Debener, S., & Brand, T. (2019). Effect of speech rate on neural tracking of speech. *Frontiers in Psychology*, *10*(MAR).

Nejime, Y., & Moore, B. C. J. (1998). Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss. *The Journal of the Acoustical Society of America*, *103*(1), 572–576.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, *20*(3), 633–640.

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.

Oostdijk, N. (2000). The spoken Dutch corpus: Overview and first evaluation. *2nd International Conference on Language Resources and Evaluation, LREC 2000*, *January 2000*.

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1315–1330.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological bulletin*, *144*(10), 1002.

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R- project.org

Rodero, E. (2016). Influence of speech rate and information density on recognition: The moderate dynamic mechanism. *Media Psychology*, *19*(2), 224–242.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(JUNE), 1–17.

Schlueter, A., Lemke, U., Kollmeier, B., & Holube, I. (2014). Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms. *The Journal of the Acoustical Society of America*, *135*(3), 1541–1555.

Shannon, R. V., Fu, Q. J., & Galvin 3rd, J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta otolaryngologica. Supplementum*, (552), 50-54.

Shannon, R. v., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303–304.

Sharit, J., Czaja, S. J., Nair, S., & Lee, C. C. (2003). Effects of age, speech rate, and environmental support in using telephone voice menu systems. *Human Factors*, *45*(2), 234–251.

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). 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.

Simantiraki, O., & Cooke, M. (2020). Exploring listeners' speech rate preferences. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 1346–1350.

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: Revisiting the N400. *Journal of Cognitive Neuroscience*, *25*(8), 1383–1395.

van Os, M., Kray, J., & Demberg, V. (2021). Recognition of minimal pairs in (un)predictive sentence contexts in two types of noise. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43(43), 2943–2949.

Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., & Dehaene, S. (2012). A temporal bottleneck in the language comprehension network. *Journal of Neuroscience*, *32*(26), 9089–9102.

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.

Winn, M. B., & Teece, K. H. (2021). Slower speaking rate reduces listening effort among listeners with cochlear implants. *Ear and Hearing*, *42*(3), 584.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20–32.

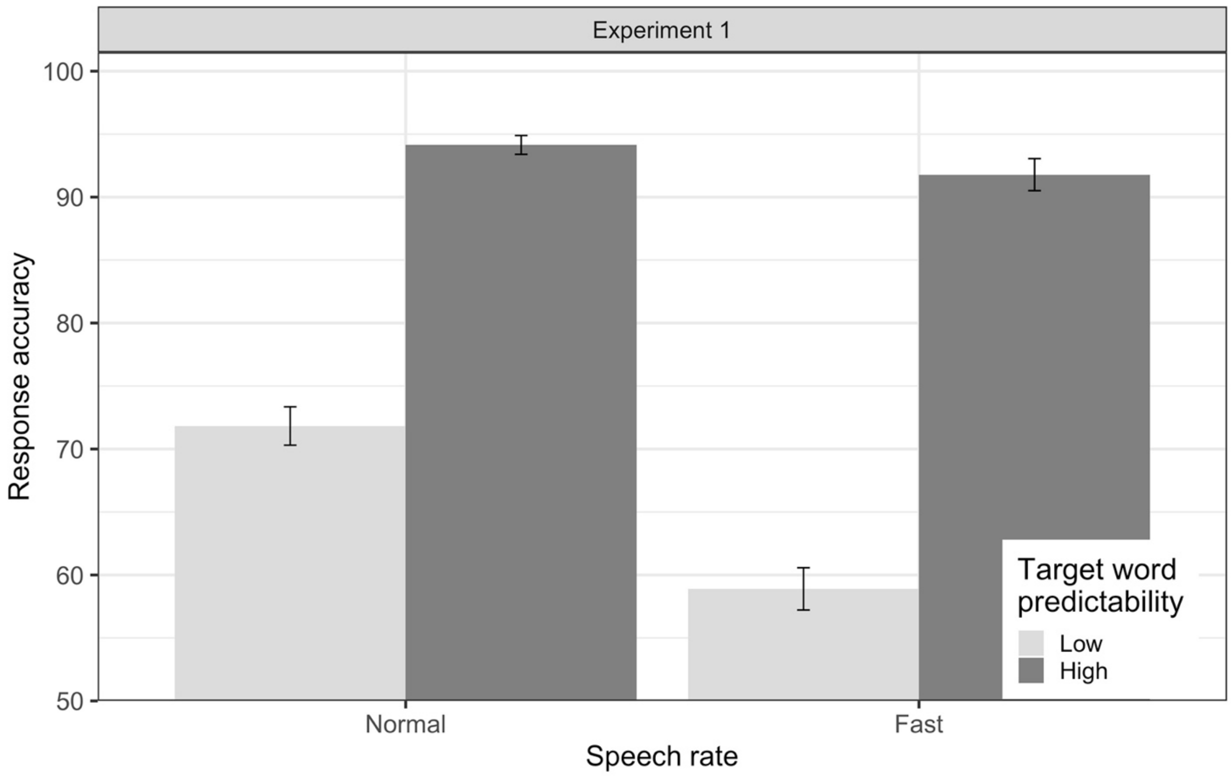

Figure 1: Mean response accuracy across all conditions in Experiment 1. There was no decrease in contextual facilitation from normal to fast speech rate. Error bars represent standard error of the means.

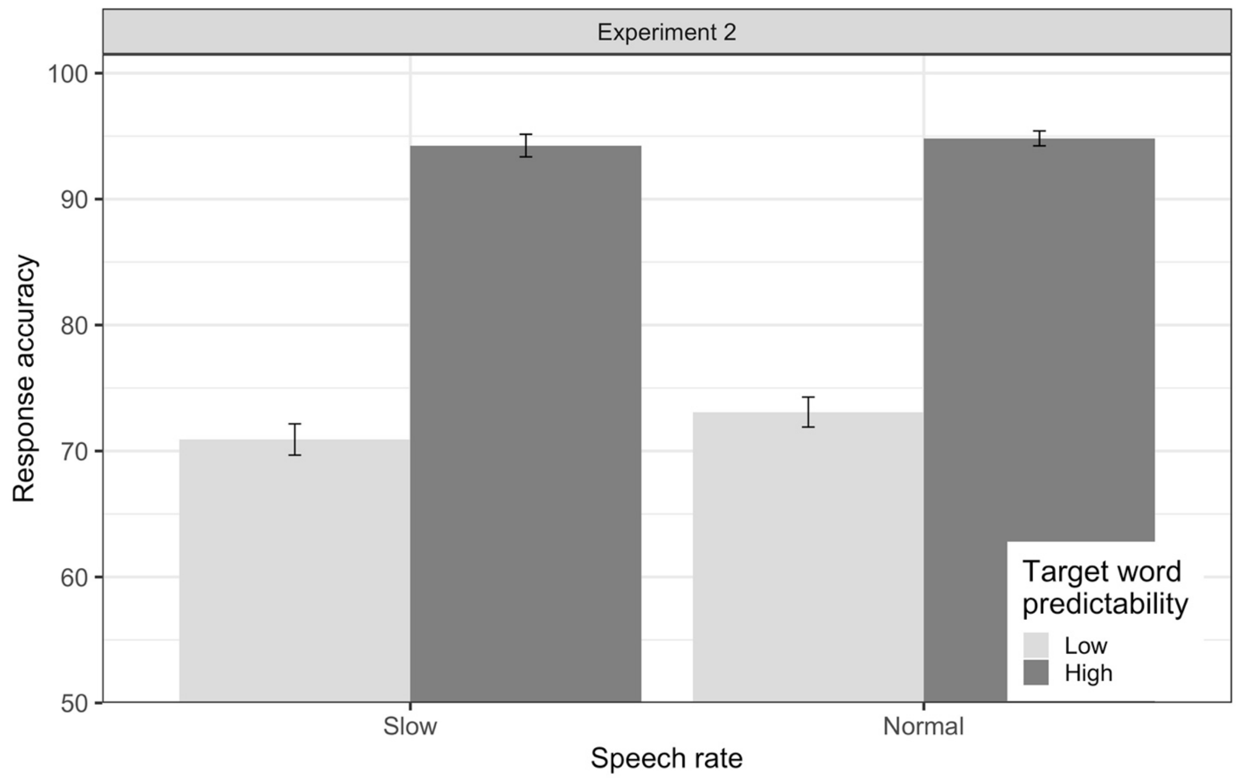

***Figure 2*: Mean response accuracy across all conditions in Experiment 2. There was no increase in contextual facilitation from slow to normal speech rate. Error bars represent standard error of the means.**

Table 1: Example sentences.

Target word predictability	Sentences	English translation
Low	Sie *findet* die *Baelle*.	She finds the balls.
High	Sie *jongliert* die *Baelle*.	She juggles the balls.

Table 2: Response accuracy (Mean and Standard error of the mean) across all rates of speech and levels of target word predictability in Experiment 1.

Speed	Target word predictability	Mean	Standard error
Fast	High	91.78	1.23
	Low	58.93	1.54
Normal	High	94.13	1.01
	Low	71.82	1.37

3

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

Fixed effects	Estimate of β	Standard error	*z*	*p*
(Intercept)	1.34	.24	5.58	<.001
Speech rate (Fast)	-.98	.24	-4.16	<.001
Target word predictability (High)	2.42	.28	8.55	<.001
Speech rate × Target word predictability	1.06	.42	2.50	.012

Model:

glmer (response~ 1 + Rate + Predictability + Rate:Predictability +

(1+ Rate + Predictability + Rate:Predictability || subject) +

(1+ Rate + Predictability + Rate:Predictability || item) …

Table 4: Response accuracy (Mean and Standard error of the mean) across all rates of speech and levels of target word predictability in Experiment 2.

Speed	Target word predictability	Mean	Standard error
Slow	High	94.25	.89
	Low	70.92	1.09
Normal	High	94.82	0.70
	Low	73.09	1.02

6

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

Fixed effects	Estimate of β	Standard error	*z*	*p*
(Intercept)	1.41	.23	6.20	<.001
Speech rate (Slow)	-.08	.14	-.57	.568
Target word predictability (High)	2.58	.30	8.65	<.001
Speech rate × Target word predictability	.44	.27	1.65	.099

Model:

glmer (response~ 1 + Rate + Predictability + Rate:Predictability +

(1+ Rate + Predictability + Rate:Predictability || subject) +

(1+ Rate + Predictability + Rate:Predictability || item) …