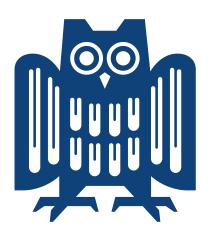# Comprehension of degraded speech: Exploring the role of attention and speed of processing in top-down prediction

Pratik Bhandari

Department of

Saarland University

A thesis submitted for the degree of

Doctor of Philosophy

xxxx 2022

Dedicated to . . .

Acknowledgements

Here I acknowledge lots of people including my GP, neurologists and counselors.

Pratik Bhandari
Universität des Saarlandes,
Saarbrücken
10 March 2021

Abstract

Listening in an adverse environment poses a challenge – it is difficult to understand what is being said when there is background noise, or when the speaker's speech signal is distorted. Nevertheless, listeners show remarkable success in understanding the distorted speech by utilizing context information to form predictions about upcoming linguistic events. The extent to which such top-down predictions are useful is still a matter of debate. Additionally, the role of other factors such as attention and rate of flow of information in degraded speech comprehension are understudied. In this thesis, I present a broader overview of the role of semantic predictions on degraded speech comprehension across lifespan, and how it interplays with attention, and rate of flow of information. In the first experiment (Chapter 4.1), I show that listeners can flexibly pay attention to a portion of speech stream; and attending to the sentence context is necessary to utilize the context and form top-down predictions. I show in the second experiment (Chapter 4.2) that as listeners utilize the context information, they form semantic predictions about upcoming linguistic events in a graded manner when the speech is moderately degraded. Semantic predictions are not restricted to only most highly likely sentence endings. I also argue for a novel metric to measure language comprehension, and show that sensory adaptation to degraded speech is disrupted by change in higher-level semantic features of speech. Perception, processing and comprehension of degraded speech is difficult and effortful. In the third experiment (Chapter 4.3), I show that when the rate of flow of information is changed by increasing the speed of speech, the facilitatory effect of predictability is observed even at a mildly degraded speech. That is, earlier (in Chapter 4.2), mild degradation was easier to process; but increase in speed rendered the speech processing difficult such that predictability had a facilitatory effect. To examine the general age differences in the facilitatory effect of predictability, I conducted a fourth experiment (Chapter 4.4) where younger adults (age range 18-30) and older adults (age range = ...) were recruited. Highest age difference in the use of sentence context for language comprehension was observed at the moderate level of speech degradation. This supported the hypothesis that sensory decline with aging tips older adults to rely on context information more than younger adults do. I show the neural markers of these age differences in the fifth experiment (Chapter 4.5) ... ... Taken together, this thesis and the

results herein support the views that sentence context and semantic prediction facilitate comprehension of moderately degraded speech in a graded manner, and reliance on context increases with age.

Contents

Contents

Contents

Contents

List of Figures

List of Tables

List of Abbreviations

HP, MP, LP . High-, Medium-, or Low-predictability

YA, OA . . . . Younger, or Older adults

ch . . . . channels

EN . . . . English

1

Introduction

One of the features that distinguishes us humans from other species is our ability to communicate using verbal language. We speak. We listen. We understand. This seemingly straight forward path of communication (i.e., speak-listen-understand) goes through a plenty of hindrances. One of them being adverse listening condition. We are hardly in a noise-free environment. There can be noise in the background during a conversation; poor signal transmission via a Zoom call can render a speaker's speech distorted; a listener can have hearing loss, or he might be wearing a hearing prostesis like cochlear implant and hearing aid. Despite any or all of these, the peripheral auditory processing, and the central language processing systems function such that most of the times we understand what is being said. How does the cognitive system achieve this feat of understanding speech that is distorted by so many internal and external factors? One of the 'tactics' that human language comprehenders rely upon is the use of preexisting linguistic knowledge, information about the world, and importantly, the context information to form predictions about linguistic events that are yet to be encountered.

Let's take the following sentence, for example:

(1) The day was breezy so the boy went outside to fly . . .

1. Introduction

Most of the listeners would expect the final word to be *a kite* in this sentence. The words preceding the ellipses provide a context. A reader, or a listener can utilize their knowledge about what a *boy* would ideally do *outside* when a day is *breezy.* This would lead the reader/listener to predict that the sentence continuation is most likely *a kite* and not any other word like a rocket (DeLong et al. 2005; cf. Nieuwland, Barr, et al. 2020). This particular sentence was used by DeLong et al. (2005) in a reading study which also applies to the auditory domain (Altmann and Kamide 2007).

While listening to a degraded speech, listeners can use the context information and form predictions. The nature of prediction in comprehension of speech that is degraded by noise vocoding is not a settled issue: Is prediction probabilistic, or is it an all-or-nothing phenomenon? However, one question that one might raise prior to the nature of prediction is, why study degraded speech comprehension in the first place when we do not come across it on our day-to-day conversation. There are two facets to the answer of this question. First, we cannot fully understand how speech comprehension system in humans operate unless we study how it responds to novel sounds. If predictive processing is the fundamental operating mechanism of human language comprehension as it is claimed to be, then it is imperative we examine the claim in different listening conditions, with novel auditory stimuli. After all, different paradigms have shown limits in predictive language processing accounts. Secondly, a X% of population has hearing difficulty, and y% receives cochlear implant. Degraded speech mimics the auditory input that cochlear implantees receive. To better understand language comprehension in this section of population, and to explore the difficulties that they face compared to normal hearing population, the use of degraded speech is important.

Therefore, we study both top-down predictive processes as well as bottom-up auditory processes and their interactions in comprehension of degraded speech. While doing so, we address the following challenges:

(i) **Replication crisis**

Almost all the disciplines of cognitive science – anthropology, computer

science, linguistics, neuroscience, and psychology – are suffering the so-called replication crisis (e.g., Sanderson and Roberts 2008; Aarts et al. 2015; Ebersole et al. 2016; Cockburn et al. 2020). The results of an experiment do not hold up consistently when conducted again by another group of researcher. For example, in a multi-lab collaborative study, Nieuwland, Barr, et al. (2020) did not find the same N400 effect at article (a/an) that DeLong et al. (2005) reported 15 years earlier. This thesis attempts to test if predictability effect in degraded speech comprehension is replicable.

(ii) **Limitations of predictive processing**

A number of authors claim that predictive processing is the fundamental nature of human cognition, and thus, by definition, also of language processing (Clark 2013; Lupyan and Clark 2015; Friston, Parr, et al. 2020; Friston, Sajid, et al. 2020). At the same time, many studies have shown limitations in predictive language processing, for example, due to memory, aging, experimental paradigm, literacy, etc. (Federmeier, Kutas, et al. 2010; Mishra et al. 2012; Huettig and Mani 2015; Huettig and Guerra 2019). This thesis examines the factors like, auditory attention and speed of information flow, that can be prerequisites and limitations of semantic prediction in understanding degraded speech.

(iii) **Debate on the nature of prediction**

There are at least two school of thoughts which argue that prediction is either a all-or-none phenomenon, or probabilistic in nature. In degraded speech comprehension, only one study has empirically investigated the theoretical postulation that prediction is restricted only for highly predictable sentence endings (Strauß et al. 2013). This thesis attempts to undertake the challenge of teasing apart the nature of prediction in degraded speech comprehension.

(iv) **Multitudes of speech distortion**

Speech signal is distorted by multiple factors in real life. Our conversational environment is noisy, or the signal is degraded during transmission. On the

top of that, people might speak fast or slow varying difficulty to process the information. Listeners can adapt to such degraded speech signal as shown by earlier studies; or trying to focus on the ever-changing lexical task might not let them adapt at all even when only one form of distortion is applied to the speech signal. When loaded with more than one form of distortion, it is not clear how the context will help the listeners' comprehension. And this thesis undertakes this challenge of understanding how multiple form of bottom-up processing difficulty interacts with the aid that top-down predictive processing provides, if at all.

The research goals of the thesis meet with these challenges. The studies addressing the challenges outlined above will contribute to do XYZ. In the next two sections, we present the research goals of this thesis, and the general overview of the all the chapters that follow.

1.1   Research goals

The interaction between top-down and bottom-up processes has been extensively discussed and examined in the literature of reading studies and clear speech comprehension. In this thesis, our aim to investigate such interaction, and the role of other top down process (e.g., attention) in comprehension of degraded speech. The goals of the research in correspondence to the central challenges of this thesis are outlined below.

Goal 1: The first goal of this thesis is to replicate the standard facilitatory effect of predictability given our experimental materials and method of speech degradation. As we have mentioned earlier, it is a problem in the field that well known effects are not replicable. If we can replicate the facilitatory effect of predictability, then it will move the research towards garnering more evidence of the *effect*. It will also provide a reliable foundation to test if other factors (e.g., speed of information processing) have any influence on the facilitatory effect.

Goal 2: The second goal of this thesis is to examine the nature of predictability effect. In the current literature of predictive processing, a debate exists about the nature of predictability effect: whether it is an all-or-none phenomenon or if it is probabilistic in nature. With carefully designed experiments, this thesis aims to tease apart this distinction between these two proposed nature of prediction in language comprehension.

Goal 3: The third goal of this thesis is to examine the role of top-down factors other than semantic prediction in degraded speech comprehension. Specifically, the role of attention to speech stream is of interest in this thesis, as we know from non-speech auditory perception literature that attention can modulate predictability effects. This thesis will aim to establish if attention to different parts of degraded speech stream aids, or hampers facilitatory effects of top-down predictions.

Goal 4: The fourth goal of this thesis is to examine if variation in information flow affects facilitatory effect of predictability. The flow of information is not constant in any conversation as speakers speak fast or slow. We will therefore test if an increase or decrease in speech rate impedes overall comprehension of degraded speech, and if it impedes or further aids the facilitatory effect of predictability.

Goal 5: The fifth goal of this thesis is to examine if listeners adapt to degraded speech. Studies that report adaptation and learning effect often overlook the effect of context that comes along with the perceptual property of the speech. It is not reasonable to argue that adaptation is independent of any lexical property of the speech that contains context information. Therefore, in this thesis we take into account the variability of context information across sentences even if they are at the same level of degradation. We will test if adaptation to bottom-up perceptual property of speech is influenced by top-down semantic property of speech.

Goal 6: The sixth goal of this thesis is to establish and consistently use a sensitive metric for the measurement of language comprehension. In the study of speech perception and language comprehension, different researchers use different measurement metric. This becomes a problem specially when the effect of context in comprehension is being discussed. In this thesis, we advocate for the use of a metric

that takes into account whether participants (in)correctly utilize the context evoking word in a sentence to measure language comprehension or word recognition accuracy.

1.2 Research contributions

The research reported in this thesis contributes to the studies of speech perception, language processing, predictive processing, language science, audiology, psycholinguistics, psychology, and broadly to cognitive science. The studies in this thesis helps us understand the processing and comprehension of novel speech that humans do not generally encounter day-to-day. Nevertheless, this opens a window towards understanding how human auditory and language processing mechanism work to help us understand speech in adverse listening condition. In an applied setup, this informs us about the speech perception in cochlear implantees. WRITE UP MORE HERE.

- **A metric of language comprehension**

  We propose and successfully use a metric of language comprehension that reflect listeners' use of context information. This metric does not merely measure how many words are correctly identified. Rather, it considers the fact that in the study of effect of predictability, how well a context is recognized should also be taken into account, and thus it measures word recognition accuracy in the sentences in which context is correctly recognized in the first place. The use of such a metric improves interpretation of contextual facilitation across studies, which lacks in the extant literature. Goal 6 corresponds to this research contribution which is brought about by consistent use of this metric in Chapters 6, 7 and 8.

- **Nature of prediction**

  First and foremost, we replicate previous findings that predictability indeed facilitates comprehension of degraded speech at a moderate level of degradation. This is of great value given the discussion on replication crisis in different domains of cognitive sciences. Our evidence shows that such facilitatory effect of prediction is graded in nature; it is not an all-or-none phenomenon. In

the existing debate between all-or-nothing vs graded prediction, our findings clearly show that prediction in degraded speech comprehension is graded in nature. Goals 1 and 2 correspond to this research contribution which is brought about by the experiment described in Chapter 6.

- **Limitation in predictive language processing**

  We show that predictive processing all by itself cannot explain how listeners understand degraded speech. Although top-down predictions facilitate comprehension, we show that attention to context is the prerequisite for such a contextual facilitation. Only when listeners attend to the context information and form its meaning representation, then top-down predictions can facilitate comprehension of degraded speech. Without proper attentional allocation, contextual facilitation cannot be observed. Goal 3 corresponds to this research contribution which is brought about by the experiment described in Chapter 5.

- **Variability in information flow and its effect on top-down prediction**

  We show that different rates of information flow – increase or decrease in rate of speech – have different effects in language comprehension that is brought about primarily by the effect in ease or difficulty in processing sentences with less predictable endings. Although the increase in the speed of information flow makes it difficult to encode the context information resulting in impeded contextual facilitation, the decrease in the speed does not necessarily enhance encoding the context information while listening to moderately degraded speech. This is one of the very few studies highlighting the role of speed of flow of information in contextual facilitation when the speech is modertely degraded. Goal 4 corresponds to this research contribution which is brought about by the experiment described in Chapter 7.

- **Adaptation to degraded speech**

  We show that listeners do not adapt to degraded speech which is opposite to some of the recent findings of speech perception experiments from other labs.

When listeners are working in a linguistic task in which the lexical cues vary on every trial, their cognitive resources are strained by lexical-semantic cues rather than acoustic-phonetic cues. Thus, they don't show any adaptation effect; every trial is effectively a *novel* trial for them. Goal 5 corresponds to this research contribution which is brought about by the experiment described mainly in Chapter 6, and continued in Chapter 7.

1.3 Overview of the thesis

The central theme of this thesis is the study of the interaction between top-down predictions and bottom-up auditory processes in degraded speeech comprehension. Although there are studies showing reliance on prediction at a moderate level of speech degradation, this thesis replicated such findings focusing on theoretical question of the nature of such predictive processing. The studies in this thesis examined the factors – attention and rate of information flow – that modulate semantic prediction, and also showed that listeners do not adapt to degraded speech contrary to some previously reported findings.

Chapter 2 provides a background on the rest of the chapters. It provides an overview of degraded speech comprehension, predictive language processing, and existing metrices of measuring word recognition and language comprehension.

Chapter 3 describes the stimuli used in all the experiments in this thesis. It provides an overview of cloze rating, speech processing, and online data collection method.

Chapter 4 provides an overview of the statistical tests employed for data analyses. Binomial logistic mixed effects modeling was performed on the data from all the experiments. This chapter provides a background on this statistical procedure, and how it is conduced on R.

Chapter 6

Chapter 7

Chapter 8

1.4   Dissimination of research findings

Some of the findings reported in this thesis are presented and published, in part, elsewhere as a means of dissemination of scientific findings. The list of presentations and publications that report on parts of the research described in this thesis is outlined below.

- Bhandari, P., Demberg, V., & Kray, J. (*in preparation*) Slow speech doesn't help while fast speech impedes comprhension of moderately degraded low predictability sentences.

- Bhandari, P., Demberg, V., & Kray, J. (*under review*) Predictability effects in degraded speech comprehension are reduced as a function of attention.

- Bhandari, P., Demberg, V., & Kray, J. (2021). Semantic predictability facilitates comprehension of degraded speech in a graded manner. *Frontiers in Psychology*

- Bhandari, P., Demberg, V., & Kray, J. (2021). Predictability facilitates comprehension but not adaptation to degraded speech in a graded manner. *Conference of the Society for the Neurobiology of Language*, Online (Presentation)

- Bhandari, P., Demberg, V., & Kray, J. (2021). Older adults make equal use of predictability in comprehension of degraded speech as younger adults do. *6th Annual Conference Aging and Cognition*, Online (Poster)

- Bhandari, P., Demberg, V., & Kray, J. (2021). Predictability facilitates comprehension of degraded speech in a graded manner. *Annual Meeting of Cognitive Neuroscience Society*, Virtual (Poster)

2
Background

In the previous chapter we outlined the research questions and goals of the studies in this dissertation. In so doing, we presented our motivation and relevant theoretical questions. To further elaborate on these motives, this chapter provides some background on the current understanding on the theoretical aspects of the research questions. We introduce speech degradation, and predictive language processing in the context of speech degradation. In addition, we also provide a background on the effects of speech degradation, and the measurement of speech perception and language comprehension. Understanding these fundamental concepts of top-down and bottom-up processes are essential for the chapters that follow. These concepts are briefly reiterated in the following chapters wherever relevant.

2.1   Speech distortion and degradation

In an ideal situation, spoken perception is a seamless process: a speaker produces an utterance, the speech signal transmits via some medium like air, and a listener perceives the signal as speech waves which enter her ears and start a cascade of mechanical-neural processes of hearing. However, speech perception is not so simple. There are primarily three sources distortion. Speech can be distorted by variability in speakers' production, like, accented speech, or soft and rapid

speech. Distortion can arise from listener-related factors like, hearing loss, or auditory processing disorder. It can also be a result of noise that appears during the transmission, like ambient noise, or poor medium of transmission (e.g., distortion in the telephone line). All these sources of distortion make listening conditions adverse. In laboratory setup, the effects of speech distortion, and the mechanism of listening in adverse listening condition are studied using artificial distortion of speech. Different forms of distorted speech manipulate different properties of speech signal. For example, speeding or slowing a speech signal (i.e. speech compression or expansion) manipulates its temporal property. Similarly, noise vocoding manipulates its spectral property. Noise vocoding removes the spectral detail of the speech signal only leaving its temporal and periodicity cues (see Section X.X.X). This method of speech degradation was initially developed as a means to reduce the information in speech signal to be transmitted through the telephone line (Dudley 1939; "The vocoder" 1940) — [Re-read this thoroughly]. Shannon and colleagues later used the same technique as an analogue to cochlear implant (Shannon, Zeng, et al. 1995; Loizou et al. 1999; Shannon, Fu, et al. 2004) – number of channels used in a cochlear implant are similar to the number of noise vocoding channels in terms of their speech output and intelligibility [... cite probably Wagner et al. ...]. Therefore, in addition to being a method of speech distortion to parametrically vary and control the quality of speech signal graded manner, noise vocoding also is a method of distortion that helps us understand the speech perception and language comprehension in cochlear implantees.

One of the main factors that determine the intelligibility of degraded speech is the number of noise vocoding channels.[1] The higher the number of noise vocoding channels, the more is the frequency specific information available in the degraded speech, consequently, higher is the intelligibility compared to the speech that is degraded with lesser number of noise vocoding channels. For example, listeners rate

---

[1]Throughout this thesis, speech distortion by noise vocoding is referred to as speech degradation, or spectral degradation of speech.

8 channels noise vocoded speech to be more intelligible and less effortful compared to 2 channels noise vocoded speech.

2.2 Prediction and comprehension of degraded speech

Listeners do not just rely on the quality of speech signal to understand degraded speech. That is, it is not just the number of noise vocoding channels that determines how well a person understand degraded speech. In addition to the bottom-up signal, listeners rely also on the context information. Before discussing on the role of top-down predictive processes on the comprehension of degraded speech, we first review the role of predictions in language comprehension in general.

2.2.1 Predictive processing and language comprehension

Research from various domains of cognitive (neuro)science, like emotion, vision, odor, and proprioception, has shown that perception and cognition primarily operates by predicting upcoming events (Stadler et al. 2012; Clark 2013; Seth 2013; Marques et al. 2018). Human language comprehension too has been claimed to be predictive in nature from as early as mid-twentieth century (e.g., Miller et al. 1951; McCullough 1958; Morton 1964). which in recent days has received overwhelming support from studies in psycholinguistics and cognitive neuroscience of language (DeLong et al. 2005; Lupyan and Clark 2015; Pickering and Gambi 2018; cf. Huettig and Mani 2015). Empirical evidence from a number of studies suggests that readers and listeners predict upcoming words in a sentence when the words are predictable from the preceding context (for reviews Staub 2015; Kuperberg and Jaeger 2016; Nieuwland 2019). For instance, such words are read faster and are skipped compared to the words that are less predictable from the context (Ehrlich and Rayner 1981; Frisson et al. 2005; Staub 2011). Applying the visual world paradigm, several studies have demonstrated that individuals show anticipatory eye movements towards the picture of the word that is predictable from the sentence context (Altmann and

Kamide 1999; Kamide et al. 2003; Ankener et al. 2018). The sentence-final word in a highly constraining sentence (e.g., *"She dribbles a ball."*) elicits a smaller N400 amplitude – a negative going EEG component that peaks around 400 ms post-stimulus and is considered as a neural marker of context-based semantic unexpectedness (Kutas and Federmeier 2011) – than that in a less constraining sentence (e.g., *"She buys a ball."*; Kutas and Hillyard (1984); Federmeier, Wlotko, et al. (2007)). Similarly, event-related words (e.g., *"luggage"*) elicited reduced N400 compared to event-unrelated words (e.g., *"vegetables"*) which were not predictable from the context (e.g., in an event of *"travel"*; Metusalem et al. (2012)). In sum, as the sentence context builds up, listeners make predictions about upcoming words in the sentence, and these in turn facilitate language comprehension. That is, individuals use the context available to them to generate predictions which aids understanding written and spoken language.

Limits of predictive language processing: It is important to note, and acknowledge that the ubiquity and universality of predictive language processing has not gone unquestioned (Huettig and Mani 2015). Apart from the debate on the nature of prediction, which we will come to later in this chapter, there are compelling evidence that question if prediction is necessary, even present at all in language comprehension. For example, Mishra et al. (2012) showed that literacy is a key factor that limits listeners' prediction about upcoming word. In a visual word paradigm study, they found that individuals with low literacy score showed less anticipatory eye movements compared to the individuals with high literacy score. They bolstered their finding in a neuroimaging study claiming that learning to read fundamentally changes the neural circuitry. It is therefore plausible that such structural change in the brain is manifested in linguistic behavior too. Similarly, cognitive aging has been shown to be a limiting factor in generating predictions. Smaller N400 amplitude and latency in older adults compared to younger adults have been shown as an evidence of inability of older adults in predictive processing in language processing. Among older adults, those with lower working working memory scores are shown to be further disadvantaged when it comes to the use of context information (Federmeier,

Mclennan, et al. 2002; Federmeier, Kutas, et al. 2010). A different line of argument that critiques the entire predictive processing domain comes from observations by Huettig and Guerra (2019). They claim that most of the studies in support of predictions are *artificial* which encourage participants to predict in a laboratory setup which does not apply to a real life interactions. They analyzed XYZ number of eye tracking studies and found that XYZ did not even mention the preview time; ZYC presented speech at a slow rate. Despite these critiques, there are plenty of experiments that have been replicated in pre-registerd multi-lab replication studies. Computational modeling of language processing also support the claim that language processing is predictive in nature. We are, nonetheless, aware of the critiques. And, in this thesis, we will study other factors (e.g., auditory temporal attention) that can interact with, and potentially limit top-down semantic predictions.

2.2.2 Predictive processing and comprehension of degraded speech

We have discussed above that individuals make predictions about not-yet-encountered linguistic unit based on available context information as the sentence unfolds: Top-down predictive and bottom-up perceptual processes interact dynamically in language comprehension. When the bottom-up perceptual input is less reliable, for example, in adverse listening conditions, it has been shown that listeners rely more on top-down predictions by narrowing down the predictions to smaller sets of semantic categories or words (e.g., Strauß et al. 2013; see also Corps and Rabagliati 2020). Obleser and colleagues (Obleser, Wise, et al. 2007; Obleser and Kotz 2010; Obleser and Kotz 2011), for instance, used sentences of two levels of semantic predictability (high and low) and systematically degraded speech signal by passing it through various numbers of noise vocoding channels ranging from 1 to 32 in a series of behavioral and neuroimaging studies (see also Hunter and Pisoni 2018). They found that semantic predictability facilitated language comprehension only at moderate levels of speech degradation. That is, participants relied more on the sentence context when the speech signal was degraded but it was *intelligible*

enough than when it was not degraded, or when it was highly degraded. At such moderate levels of speech degradation, accuracy of word recognition was found to be higher for words in high predictability sentences than the words in low predictability sentences (Obleser and Kotz 2010). For the extremes, i.e., when the speech signal was highly degraded (making the speech almost completely unintelligible) or when it was the least degraded (rendering the speech intelligible), the word recognition accuracy was similar across both levels of sentence predictability, meaning that predictability did not facilitate language comprehension. Sheldon et al. (2008a) 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 constraining sentences, a higher number of channels (i.e., more bottom-up information) was required in less constraining sentences to achieve the same level of accuracy. They also concluded that when speech is degraded, word recognition is facilitated by predictability and sentence context. Taken together, these studies conclude that at moderate levels of degradation, participants rely more on the top-down predictions generated by the sentence context and less on the bottom-up perceptual processing of unclear, less reliable and degraded speech signal (Obleser 2014).

Nature of prediction: One of the many debates in the literature of predictive language processing pertains this question: Is prediction probabilistic or is it an all-or-nothing phenomenon? This is a longstanding debate in sentence processing, and *clear* speech comprehension.

Whether the nature of prediction is similar in adverse listening condition has not come under theoretical questions among psycholinguists and cognitive scientists (cf. Strauß et al. 2013; Corps and Rabagliati 2020).

3

General methods

3.1 Experimental materials

As a part of a study in the research project A4 of SFB1102, sentences of different levels of predictability were created. Digital recordings of the sentences were degraded by noise vocoding and used in all experiments reported in this thesis. Speech was also distorted by its compression and expansion. Below we briefly describe how the sentences of different levels of predictability were obtained, and what methodology was used to create distorted versions of the speech.

3.2 Stimulus sentences

With an aim to create sentences of three levels of predictability (low, medium, and high), a triplet of 120 sentences – total of 360 sentences – were created from 120 nouns. Out of 120 nouns, 6 were repeated. All sentences were in present tense consisting of pronoun, verb, determiner, and object. These sentences were in Subject-Verb-Object form (e.g., Er fängt den Ball. EN: He catches the ball.). Some of these sentences were taken from Obleser and Kotz (2010). For each sentence, cloze probability ratings were collected from a group of young adults(n = 60; age range = 18 - 30 years). Mean cloze probabilities were 0.022 (SD = 0.027; range

= 0.00 − 0.09) for low-predictability sentences, 0.274 (SD = 0.134; range = 0.1 − 0.55) for medium-predictability sentences, and 0.752 (SD = 0.123; range = 0.56 − 1.00) for high-predictability sentences. The distribution of cloze probability across low-, medium-, and high-predictability sentences is shown in Figure XXX, and the cloze probability for individual sentence is shown in Table XX.

3.3   Speech processing

All 360 sentence were spoken by a female native speaker of German at a normal rate. The recordings were digitized at 44.1kHz with 32-bit linear encoding. Spoken sentences used in Chapter 5, 6, 7, and 8 were degraded by noise vocoding. In addition to degradation by noise vocoding, the sentences were distorted by compression and expansion of speech signal in Chapter 7.

3.3.1   Noise-vocoding

Noise vocoding is used to parametrically vary and control the quality of speech signal in a graded manner. It distorts a speech signal by dividing it into specific frequency bands corresponding to the number of vocoder channels. The frequency bands are analogous to the electrodes of cochlear implant (Shannon, Zeng, et al. 1995; Loizou et al. 1999; Shannon, Fu, et al. 2004). The amplitude envelope, i.e., the fluctuations of amplitude, within each frequency band is extracted and the spectral information within it is replaced by noise. This makes the vocoded speech difficult to understand although temporal characteristics and periodicity of perceptual cues are preserved (Rosen et al. 1999).

The spectral degradation conditions of 1, 4, 6, and 8 channels were achieved for each of the 360 recorded sentences using a customized script originally written by Darwin (n.d.) in Praat software. The speech signal was divided into 1, 4, 6, and 8 frequency bands between 70 and 9,000Hz. The boundary frequencies were approximately logarithmically spaced following cochlear-frequency position functions (Greenwood 1990; Erb 2014). The amplitude envelope of each band was

extracted and applied to band-pass filtered white noise in the same frequency ranges; the upper and lower bounds for band extraction are specified in Table X. Each of the modulated noise was then combined to produce degraded speech. Scaling was performed to equate the root-mean-square value of the original undistorted speech and the final degraded speech. This resulted into four levels of degradation: 1, 4, 6, and 8 channels noise vocoded speech.

Spectrograms of clear speech and noise-vocoded speech (1, 4, 6 and 8 channels) for the word *Aufgabe* are shown in Figure X.X. It shows that with a decrease in the number of noise vocoding channels, the information in speech signal reduces and becomes similar to noise.

3.3.2 Speech compression and expansion

As early as the mid-twentient century, investigators have reported that intelligibility does not drop significanlty when speech is speeded up to 2 times the normal speech rate (e.g., Garvey 1953). Speech rate was increased by chopping physical tapes. Digital algorithms like pitch-synchronous overlap-add technique (PSOLA, Charpentier and Stella 1986; Moulines and Charpentier 1990) developed in the 1980s and later (overlap-add technique based on waveform similarity, Verhelst and Roelands 1993) now allow us to speed up and slow down the speech rate in a controlled fashion.

In Chapter 7, we used Praat software that utilizes uniform time-compression algorithm (PSOLA) to create slow and fast speech with the compression factor of 1.35 and 0.65 respectively. PSOLA analyzes the pitch of an auditory signal, in the time domain of its digital waveform, to set pitch marks and then segments the signal into successive analysis windows centered around those pitch marks. To create synthesized speech (i.e., fast or slow speech), a new set of pitch marks are calculated and the analysis windows are rearranged. Depending on the time-compression factor, some analysis windows are deleted, and the remaining windows are concatenated by superimposing and averaging the neighboring analysis windows. The distortion

of phonemic properties of speech signals are minimal when accelerating and slowing down within the range of factor 2 or below (Moulines and Charpentier 1990).

In Chapter 7, we created fast and slow versions of 120 high-predictability sentences and 120 low-predictability sentences, but did not use medium-predictability sentences. These 480 recordings are then passed through 4 channels noise vocoding to use as experimental materials. Spectrograms of clear speech, fast speech, slow speech, and noise vocoded fast and slow speech for the word *Aufgabe* are shown in Figure X.X.

3.4 Data collection

The first generation of online experiments on human cognition began in the mid 1990s (for reviews, Musch and Reips 2000) with the advent of the internet (Berners-Lee et al. 1992). Welch and Krantz (1996) is considered to be the first online experiment that was conducted in 1995 as a part of tutorials in auditory perception. In their survey of researchers, Musch and Reips (2000) found that until 2000, there were already at least 2 psycholinguistics experiments conducted online; one of which studied the effect of context in shallow vs. deep encoding of words. Despite the difficulty in conducting online experiments, and skepticism of journals towards publishing results of online experiments, Musch and Reips (2000) expressed optimism: "At the moment, the number of Web experiments is still small, but a rapid growth can be predicted on the basis of the present results. We would not be surprised if within the next few years, a fair proportion of psychological experiments will be conducted on the Web." And by 2021, there is a significant growth of online experiments as technical and technological barriers are greatly reduced. There are many software and online platforms which psychologists and psycholinguists can use with minimal knowledge of computer programming to design, host and run their experiments, and get these data in a fairly structured format (Peirce et al. 2019; Anwyl-Irvine, Massonnié, et al. 2020; *Prolific Academic* n.d.; see also, Anwyl-Irvine, Dalmaijer, et al. 2021; Eyal et al. 2021). Online experiments have demonstrated advantages over laboratory experiments (Gadiraju et al. 2017; Johnson

et al. 2021). For example, a large pool of participants are available online which is usually not possible in laboratory experiments. Similarly, the participants in online experiments are more diverse than those in laboratory experiments. Taking these advantages into consideration, psychologists and psycholinguists have conducted online experiments for almost 3 decades now. Scientists who only conducted laboratory experiments, or who conducted online experiments only occasionally were forced to conduct online experiments almost exclusively due to the restrictions imposed by covid-19 lockdown (Gagné and Franzen 2021; Reips 2021). Since Welch and Krantz (1996)'s auditory perception experiment, a number of experiments have been conducted online in auditory domain (Leensen and Dreschler 2013; Woods et al. 2017; van Os et al. 2021; Seow and Hauser 2022) replicating laboratory findings (e.g., Cooke and Garcia Lecumberri 2021). The experiments reported in this thesis were also conducted online.

Initially, our experiments were designed to be conducted both in laboratory and online. As the laboratory was shut down due to covid-19 pandemic, we moved the laboratory experiments online as well. We recruited participants online via Prolific Academic (*Prolific Academic* n.d.). Participants were redirected to the experiments that were designed and hosted in Lingoturk(Pusse et al. 2016). Lingoturk is a local hosting platform that manages crowdsourcing experiments – it runs the experiments and stores the data. We report the details of each experiment in Chapters 5, 6, and 7.

4

General statistical approach

4.1 Linear regression

In this thesis, we use binomial mixed effects logistic regression models with crossed random effects (Baayen et al. 2008). These models are, simply speaking, extensions of logistic regression models. A logistic regression models a dependent variable (or an *outcome*, or a *response* variable) as a function of one or more independent predictor variables (or *factors*, or *explanatory* variables). That is, an outcome y is modeled as a function of explanatory variables $x_1, x_2, x_3..., x_n$, and an error term ε.

$$y = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + ... + \beta_n \cdot x_n + \varepsilon \qquad (4.1)$$

The intercept α, and the regression coefficients β_1, β_2, and β_3 for each explanatory variable are estimated to achieve the model that best fits the data. Analysis of Variance (ANOVA) is a special case of logistic regression (Chatterjee and Hadi 2012; Vasishth et al. 2022) that is one of the most common statistical tools in psychology and psycholinguistics. These linear regressions as shown above (Equation 4.1) and ANOVAS however, are not well suited for categorical data like response to multiple choice questions or yes/no questions, confidence ratings, etc. For example, in all the experiments in the current thesis, the response variables are response accuracy,

given binary correct/incorrect responses. Output of linear regression model ranges from $+\infty$ to $-\infty$ while accuracy (or probability) ranges from, 0 to 1. Additionally, simple regression models do not take into account the variability across individual participants and items. These problems in language sciences have been pointed out since as early as 1960s (Coleman 1964; Clark 1973). They are addressed to some extent by binomial logistic regression, and for our purpose by incorporating mixed effects model to binomial logistic regression (Baayen et al. 2008).

Below we briefly introduce binomial logistic regression and mixed effects model. Then we show a simple example of how binomial logistic mixed effects model is used in the experiments in this thesis.

4.2 Binomial logistic regression

The response variable in the experiments in this thesis are binary. Participants' written response to what they hear are coded as either correct or incorrect. A binomial logistic regression model is best suited for such a categorical data (Jaeger 2008). We use the term logistic regression model and binomial logistic regression model interchangeably henceforth.

As the name suggests, the output variable in a logistic regression model is in logit scale. The model therefore predicts logits of an outcome variable. Logits are log with base e, i.e. ln.

Probability ranges from 0 to 1 only while *odds* range from 0 to $+\infty$. Fitting a linear regression model with probability, or odds would assume the range to be between 0 and 1, or between 0 and $+\infty$ respectively. This restricts the range, and is incorrect for a linear model. Therefore, in a binomial logistic regression model, log-odds are used which range from $-\infty$ to $+\infty$.

A simple binomial logistic regression model is shown in Equation 4.2:

$$\ln(\frac{p}{1-p}) = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + ... + \beta_n \cdot x_n + \varepsilon \qquad (4.2)$$

This is equivalent to,

$$p = \frac{exp(\alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + ... + \beta_n \cdot x_n + \varepsilon)}{1 + exp(\alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + ... + \beta_n \cdot x_n + \varepsilon)} \qquad (4.3)$$

$$= \frac{exp(\ln(\frac{p}{1-p}))}{1 + exp(\ln(\frac{p}{1-p}))} \qquad (4.4)$$

where,

$$\ln(\frac{p}{1-p}) = logit(p) \qquad (4.5)$$

Log-odds of correct response obtained from Equation 4.2 can be transformed to probability of correct response. Equations 4.4, and 4.5 provide the relationship between probability, logit (or log-odds), and odds ($\frac{p}{1-p}$).

Some of the assumptions made for binomial logistic regression models are violated in our data. One of them being non-independence of observations, i.e., all data points are independent from one another. This assumption is violated in unbalanced design, and at times even for balanced design. Same participant responds to multiple trials of same experimental condition within an experiment. Although the design itself is balanced, after removal of outliers and/or trials which are not appropriate for comprehension measures (see section XXX for details), number of trials in analyses are unequal for each participant, item, and experimental condition. This introduces a bias in the model (Jaeger 2008).

Another intrinsic property or feature of logistic regression is that it assumes a common mean for each predictors. It has been shown that this is in fact not true: the effect of a predictor can vary depending on different random variables like participants, or items. To account for these variances, mixed effects models are used. In recent days, such statistical models are frequently used and advocated for by psycholingustis and statisticians (Th. Gries 2015; Meteyard and Davies 2020).

4.3 Mixed effects modeling

To overcome the limitations of logistic models, like violation of assumption of non-dependence of observations, and to account for the variability in the subject and/or item related parameter, mixed effects models are used. Mixed effects models contain 1) both linear and logistic regressions, and 2) *fixed effects* and *random effects*, hence the name *mixed effects*. Fixed effects term, e.g., levels of degradation assumes that all levels of degradation used in the experiment are independent from one another and they share a common residual variance. The random effects term with only varying intercept, e.g., subject as intercept, assumes that if there are 100 subjects then the mean accuracy of those 100 subjects are only a subset of possible global accuracies drawn from a set of population mean. When a slope, e.g., levels of predictability, is included to the random effects structure in addition to the varying intercept (e.g., subjects), then the model assumes that the effect of predictability on response accuracy varies across subjects.

4.4 Binomial logistic mixed effects modeling

A binomial logistic mixed effects model with varying intercepts and slopes for items and subjects is shown in Equation 4.6 below.

$$
\begin{aligned}
\ln(\frac{p}{1-p}) = \alpha + u_\alpha + w_\alpha + (\beta_1 + u_{\beta_1} + w_{\beta_1}) \cdot x_1 + \\
(\beta_2 + u_{\beta_2} + w_{\beta_2}) \cdot x_2 + ... + (\beta_n + u_{\beta_n} + w_{\beta_n}) \cdot x_n
\end{aligned}
\tag{4.6}
$$

where,

- α is the Intercept.
- Fixed effects: $\beta_1, \beta_2, ..., \beta_n$ are the coefficients (or effects) of $x_1, x_2, ..., x_n$.
- $\boldsymbol{u} = \langle u_\alpha, u_{\beta_1}, u_{\beta_2}, ..., u_{\beta_n} \rangle$: Varying intercept and slopes for random effect term like, *subject.*
- $\boldsymbol{w} = \langle w_\alpha, w_{\beta_1}, w_{\beta_2}, ..., w_{\beta_n} \rangle$: Varying intercept and slopes for random effect term like, *item.*

In this thesis, statistically, we examine the effect of predictability, speech degradation and speech rate (see section X.X, X.X, X.X) on response accuracy. And hence we use these variables in the fixed effects term. Subjects and items are used as random intercepts with by-subject and by-item slopes. The details of the models fitted to data from each experiment are given in Chapter X, X, X and X.

We therefore use binomial logistic mixed effects model as our primary statistical analysis tool in all the experiments reported in this thesis. We primarily follow the recommendations of Baayen et al. (2008), Barr et al. (2013), and Bates, Kliegl, et al. (2015).

4.5   Running the model in R

Data preprocessing and analyses were performed in R-Studio (Version 3.6.1; R Core Team, 2019; . . .). Accuracy was analyzed with Generalized Linear Mixed Models (GLMMs) with lmerTest (Kuznetsova et al. 2017) and lme4 (Bates, Mächler, et al. 2015) packages. Binary responses (correct responses coded as 1 and incorrect responses coded as 0) for all participants were fit with a binomial logistic mixed-effects model.

On the data in from each experiment, we fitted models with maximal random effects structure that included random intercepts for each participant and item (Barr et al. 2013). By-participant and by-item slopes included in the model are discussed in the Analysis sections of Chapter X, X, X. Model selection was based on Akaike Information Criterion (AIC) (Grueber et al. 2011; Richards et al. 2011) unless otherwise stated. Random effects not supported by the data that explained zero variance according to singular value decomposition were excluded to prevent overparameterization. This gave a more parsimonious model (Bates, Kliegl, et al. 2015) which was then extended separately with: i) item-related correlation parameters, ii) participant-related correlation parameter, and iii) both item- and participant-related correlation parameters. The best fitting model

among the parsimonious and extended models was then selected as the optimal model for our data.

Everyone knows what attention is.

— William James (James 1890)

5

Experiment 1: Predictability effect of degraded speech are reduced as a function of attention

In adverse listening conditions, when the bottom-up perceptual input is degraded, listeners tend to rely upon the context information, and form top-down semantic predictions. This provides contextual facilitation in understanding the degraded speech. Importantly, it is moderated by top-down attentional allocation to the context. The aim of this study was to examine the role of attention for understanding linguistic information in an adverse listening condition, i.e., when the speech was degraded. 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 have 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. As described in the previous chapter, these sentences varied in the degree of predictability (low, medium, high) as well as in the degree of speech degradation (4, 6, 8 noise-vocoding channels). Results indicated better word recognition for highly predictable sentences for moderate but 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.

5.1 Introduction

Understanding speech is highly automatized and seemingly easy when conditions are optimal. But conditions are often far from optimal in our day-to-day communication. As simple as a weak internet connection can create an adverse listening condition, for example, in an online meeting. Although the speech signal gets distorted or degraded at times, we do not frequently fail to understand such degraded speech. The use of context information helps us – the listeners – overcome the difficulty. Listeners successfully understand the speech by utilizing context information that contains information in a given situation about a topic of conversation, semantic and syntactic information of a sentence structure, world knowledge, visual information, etc.(Kaiser and Trueswell 2004; Knoeferle et al. 2005; Altmann and Kamide 2007; Xiang and Kuperberg 2015; for reviews, Stilp 2020). To utilize the context information, however, 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 (Wild et al. 2012; Eckert et al. 2016; Peelle 2018).

In this chapter we examine how attention modulates the predictability effects brought about by contextual information or cues at different levels of spectral degradation of speech. We address the existing unclarity in the literature regarding how listeners distribute their attentional resources in adverse listening conditions: 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.

We report a study that was run with an aim to investigate how the allocation of attentional resources induced by different task instructions influence 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 have heard, while in Experiment 2, they were instructed to repeat the whole sentence, and by this 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.

5.2   Background

As we have discussed earlier in Chapter 1 and Chapter 2, it is generally agreed upon that human language processing is predictive in nature, such that comprehenders generate expectations about upcoming linguistic material based on context information (for reviews, see Staub 2015; Kuperberg and Jaeger 2016; Pickering and Gambi 2018; Nieuwland 2019). When the bottom-up speech signal is less reliable and less informative in an adverse listening condition, listeners tend to rely more on the context information to support language comprehension (Sheldon et al. 2008a; Obleser and Kotz 2010; Amichetti et al. 2018). However, it is not just the quality of speech signal that determines and influences the reliance and use of predictive processing; attention to auditory input is important as well. Auditory attention allows a listener to focus on the speech signal of interest (for reviews, see Fritz et al. 2007; cf. 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

5. Experiment 1: Predictability effect of degraded speech are reduced as a function of attention

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 and Röder 2010; see also, Sanders and Astheimer 2008) which is the first negative going ERP[1] component peaking around 100 ms post-stimulus considered as a marker of auditory selective attention (**Thorton2007**; Näätänen and Picton 1987). 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, Xia, 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 accented or de-accented critical word. The N1 amplitude was larger for words with such attention probe than for words without a probe. These findings support the view that attention can be flexibly directed either by instructions towards a specific signal or by linguistic probes (Li, Zhang, Li, et al. 2017; see also, Brunellière et al. 2019). Thus, listeners are able to select a part or segment of stream of auditory stimuli to pay attention to.

The findings on the interplay of attention and prediction mentioned above come from studies most of which 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 speech stream plays in the contextual facilitation of degraded speech comprehension, although separate lines of research show that listeners attend to most informative portion of speech stream (e.g., Astheimer and Sanders 2011), and semantic predictability facilitates comprehension of degraded speech (e.g., Obleser and Kotz 2010). In two experiments, we therefore examined

---

[1]ERP = Event Related Potential

whether context-based semantic predictions are automatic during effortful listening to degraded speech, when participants are instructed to report only the final word of the sentence, or the entire sentence. We varied the task instructions to the listeners from Experiment 1 to Experiment 2 which required them to differentially attend to the target word, or to the target word including the context. We *hypothesized* that when listeners pay attention only to the contextually predicted target word, they do not form top-down predictions, i.e., 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 the facilitatory effect of target word predictability will be observed.

5.3 Experiment 1

This experiment was designed such that processing the context was not strictly necessary for the task. Listeners were asked to report the noun of the sentence that they heard which was in the final position of the sentence. This instruction did not require listeners to pay attention to the context which preceded the target word.

5.4 Methods

5.4.1 Participants

We recruited 50 participants online via Prolific Academic. One participant whose response accuracy was less than 50% across all experimental conditions was removed. Among the remaining 49 participants ($\bar{x} \pm \text{SD} = 23.31 \pm 3.53$ years; age range = 18 - 30 years), 27 were male and 22 were female. All participants were native speakers of German residing in Germany, and they did not have any speech-language disorder, hearing loss, or neurological disorder (all self-reported). All participants received 6.20 Euro 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 an informed consent in accordance with the declaration of Helsinki.

5.4.2   Stimuli

Materials used in the experiment were created by the method described in Chapter 3. Thy consisted of 360 German sentences spoken by a female native German speaker in an unaccented normal rate of speech. 120 nouns were used to create three categories of sentences differing in the cloze probability of the target words (nouns) which appeared as the final word of the sentence. Thus, we compared high, medium and low predictability sentences. The mean cloze probabilities of target words for low, medium and high predictability sentences were 0.022 ± 0.027 ($\bar{x} \pm$ SD; range $= 0.00$ - 0.09), 0.274 ± 0.134 ($\bar{x} \pm$ SD; range $= 0.1$ - 0.55), and 0.752 ± 0.123 ($\bar{x} \pm$ SD; range $= 0.56$ - 1.00) respectively. Speech degradation was achieved by noise vocoding through 1, 4, 6, and 8 channels.

Each participant was presented with 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 channels, ten 6 channels, and ten 8 channels 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 same degradation and predictability condition appeared consecutively.

5.4.3   Procedure

Participants were asked to use headphones or earphones. A sample of noise vocoded speech not used in the practice trial and the main experiment was provided so that the participants could adjust the loudness to a 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 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 inter-trial interval of 1000 ms.

5.5   Analyses

We preprocessed and analysed data in R-Studio (Version 3.6.3; R Core Team, 2020). At 1 channel, there were only 5 correct responses, one each from 5 participants among 49. Therefore, the 1 channel speech degradation condition was excluded from the analyses.

Response accuracy was analyzed with Generalized Linear Mixed Models (GLMMs) with lme4 (Bates, Mächler, et al. 2015) and lmerTest (Kuznetsova et al. 2017) packages following the procedure described in Chapter 4. Binary responses (correct/incorrect) for all participants were fit with a binomial logistic mixed-effects model(Jaeger 2006; Jaeger 2008). Noise condition (categorical; 4, 6, and 8 channels noise vocoding), target word predictability (categorical; high, medium, and low), global channel context (categorical; predictable channel context and unpredictable channel context), 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. Non-significant higher-order interactions were excluded from the fixed-effects structure and from the random-effects structure in a stepwise manner. Random effects not supported by the data that explained zero variance were excluded and a more parsimonious model was obtained (Bates, Kliegl, et al. 2015). Such a model was then extended separately with i) item-related correlation parameters, ii) participant-related correlation parameters, and iii) both item- and participant-related correlation parameters when applicable. Aiming for model parsimony, the best fitting model among the parsimonious and extended

models was then selected as the optimal model for our data. Model selection was based on Akaike Information Criterion (AIC) unless otherwise stated (Burnham and Anderson 2002; Grueber et al. 2011; Richards et al. 2011).

We applied treatment contrast for number of channels (8 channels as a baseline) and sliding difference contrast for target word predictability (low predictability vs. medium predictability, and low predictability vs. high predictability sentences). We report the results from the optimal model, and are shown in Table X.X.X.

5.6   Results and Discussion

Mean response accuracies for all experimental conditions are shown in Table X.X.X and Figure X.X.X. It shows that accuracy increases with an increase in the number of noise vocoding channels, i.e., with the decrease in speech degradation. However, accuracy does not increase with an increase in target word predictability. The results of statistical analyses confirmed these observations.

We found that there was a significant main effect of number of channels, indicating that response accuracy in the 8 channels noise vocoded speech was higher than in both 4 channels (β = -3.49, SE = 0.23, z (4246) = -15.30, $p < .001$) and 6 channels noise vocoded speech (β = -.69, SE = .22, z (4320) = -3.12, p = .002). That is, when the number of channels increased to 8, listeners made more correct responses (see Figure X.X.X). However, there was no significant main effect of target word predictability (β = -.07, SE = .17, z (4246) = -.42, p = .68, and β = -.003, SE = .16, z (4246) = -.02, p = .98), and no significant interaction between number of noise vocoding channels and target word predictability (all $ps > .05$).

The results of Experiment 1 indicated a decrease in response accuracy with an increase in speech degradation from 8 channels to 6 channels noise vocoding condition, and from 8 channels to 4 channels noise vocoding condition. However, response accuracy did not increase with an increase in target word predictability, and the interaction between number of noise vocoding channels and target word predictability was also absent, in contrast to previous findings (Obleser, Wise, et al.

2007; Obleser and Kotz 2011; see also Hunter and Pisoni 2018). These results suggest that the task instruction, which asked participants to only report the final word, indeed lead to neglecting the context, and therefore the facilitatory effect of prediction was not observed. However, to further test the hypothesis – as mentioned in the beginning of this chapter – that predictability effect is dependent on attentional effect, we conducted second experiment. In the second experiment, we changed the task instruction to draw participants attention on the entire sentence such that they would decode the whole sentence including the context.

5.7   Experiment 2

Following up on Experiment 1, we conducted Experiment 2 on a separate group of participants with a different task instruction. This experiment was intended to test the hypothesis that facilitatory effect of top-down predictions is observed only when listeners attention is unrestricted such that context information is also included within the attentional focus of a listener.

5.8   Methods

5.8.1   Participants and Materials

We recruited 48 participants ($\bar{x} \pm$ SD = 24.44 $\pm$ 3.5 years; age range = 18 - 31 years; 32 males) online via Prolific Academic. Same procedure as Experiment 1 was followed. We used the same materials that were used in Experiment 1.

5.8.2   Procedure

We followed the same procedure as in Experiment 1 with one difference: Instead of only the final word of a sentence, participants were asked to report the entire sentence by typing in what they heard.

5.9 Analyses

We followed the same data analyses procedure as in Experiment 1. The 1 channel noise vocoding condition was excluded from the analysis. We only considered the final words of the sentences (i.e., the target words) to be either correct or incorrect; accuracy of other preceding words were not considered in the analyses. Like Experiment 1, the results from the optimal model are reported.

5.10 Results and Discussion

Mean response accuracy for different conditions are shown in Table X.X.X. and are displayed in Figure X.X.X. It shows that the accuracy increased with an increase in both the number of noise vocoding channels, and the target word predictability. These observations are confirmed by the results of statistical analyses (Table X.X.X): We again found a main effect of number of noise vocoding channels such that response accuracy at 8 channels was higher than both 4 channels (β = -3.49, SE = .23, z (4320) = -15.29, p < .001), and 6 channels noise vocoding (β = -0.61, SE = .20, z (4320) = -3.07, p = .002).

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 (β = 1.25, SE = .28, z (4320) = 4.50, p < .001). We also found a statistically significant interaction between speech degradation and target word predictability (β = -.95, SE = .30, z (4320) = -3.14, p = .002). Subsequent subgroup analyses of each channel condition showed that the interaction was driven by the difference in response accuracy between high predictability sentences and low predictability sentences at 8 channels (β = 1.42, SE = .62, z (1440) = 2.30, p = .02), and 6 channels noise vocoding conditions (β = 1.14, SE = .34, z (1440) = 3.31, p < .001).

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 predictability and

speech degradation, which was not observed in Experiment 1, showed that semantic predictability facilitated the comprehension of degraded speech already at moderate degradation levels (like, 6 and 8 noise vocoding channels). In line with the findings from Experiment 1, response accuracy was better with a higher number of channels.

To test whether the difference between experimental manipulations is statistically significant, we combined the data from both the experiments in a single analysis. We ran another binomial linear mixed-effects model on response accuracy and followed the same procedure as Experiment 1 and Experiment 2 to obtain the optimal model. The model summary is shown in Table X.X.X. The model revealed that the critical interaction between experimental manipulation and target word predictability was indeed statistically significant (β = -.45, SE = .18, z (8566) = -2.55, p = .011);, i.e., the effect of predictability was larger in the group that was asked to type in the whole sentence. 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 for making use of the context information under degraded speech.

5.11   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 degraded.

In contrast to the first experiment, the results of the second experiment show an interaction between target word predictability and degraded speech. This is generally in line with 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 channels or 8 channels noise vocoded speech, Obleser, Wise, et al. 2007; Obleser and Kotz 2010). The important new

5. Experiment 1: Predictability effect of degraded speech are reduced as a function of attention

finding that our study adds to the present literature is that this predictability effect may be weakened or even lost, when listeners are instructed to report only the final word of the sentence that they heard, like in Experiment 1. The lack of predictability effect and 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 for the target word (noun) in our stimuli. Hence, this small change in task instructions from Experiment 1 to Experiment 2 sheds light on the role of top-down regulation of attention on using context for language comprehension in adverse listening conditions. In adverse listening conditions, language comprehension is generally effortful so that focusing attention to only a part of the speech signal seems much 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, Xia, 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.

From most theoretical accounts of language processing that align with predictive language processing, one would expect that listeners automatically form top-down predictions about upcoming linguistic stimuli based on prior context (McClelland and Elman 1986; Kuperberg and Jaeger 2016; Norris et al. 2016; Pickering and Gambi 2018; Friston, Parr, et al. 2020). Also, when speech is degraded, top-down predictions render a benefit in word recognition and language comprehension (e.g., Sheldon et al. 2008b; Sheldon et al. 2008a; Corps and Rabagliati 2020). 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

5. Experiment 1: Predictability effect of degraded speech are reduced as a function of attention

when speech is degraded. To this point, our findings shed light on the growing body of literature that indicate limitations of predictive language processing accounts (Mishra et al. 2012; Huettig and Mani 2015; Nieuwland, Politzer-Ahles, et al. 2018; Huettig and Guerra 2019).

In conclusion, this study provides a novel insight into the modulatory role of attention regulation in the interaction between top-down predictive and bottom-up auditory processes. We show that task instructions affect distribution of attention to the degraded 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. The findings of this study indicate limitations to predictive processing accounts of language comprehension.

6

Experiment 2: Semantic predictability facilitates comprehension of degraded speech in a graded manner

In the previous chapter we have shown that when listeners attend to the context, they gain facilitatory effect of predictability while listening to a degraded speech. However, there were a few lacunae in the study, which warrant further examination to bolster as well as extend this conclusion. There was an implicit assumption that all the correct responses were borne out of correct identification of the context evoking words (i.e., the verbs). We also did not consider if listeners adapted to the degraded speech and how such an adaptation could modulate contextual facilitation. And most importantly, we only showed that there is a difference between high and low predictability sentences, the granularity of predictability could not be tested. The aim of the study described in this chapter is to address these limitations and examine if listeners formed narrowed expectations or whether predictions are generated across a wide range of probable sentence endings in a graded manner. Here, we consider the accuracy of context identification (not just the accuracy of target word identification) as well as the possible learning effect due to adaptation to degraded speech. The results showed that in contrast to the *narrowed expectations* view postulated for *predictive processing in degraded speech comprehension*, listeners

probabilistically preactivate upcoming words from a wide range of semantic space, not limiting only to highly probable sentence endings. We also did not find any learning effects – We speculate that when there is a trial-by-trial variation in semantic feature (i.e., sentence predictability), listeners do not adapt to low-level perceptual property (i.e., speech quality).

6.1   Introduction

In the literature of speech perception and sentence processing, studies have argued that prediction is either probabilistic and graded, or it is all-or-none. Very few studies have investigated such theoretical questions within the domain of degraded speech comprehension (e.g. Strauß et al. 2013; see also Corps and Rabagliati 2020; van Os et al. 2021). As we have elaborated in Chapter 2, Strauß et al. (2013) postulated that listeners cannot preactivate less predictable sentence endings in an adverse listening condition: They proposed that the facilitatory effect of predictability is limited to only highly predictable sentence endings at a moderate level of spectral degradation of speech. Although the contextual facilitation is supported by the results of Obleser and colleagues, their study was conducted before Strauß et al. (2013), and were not designed to tell if prediction was graded.

In this chapter, our main aim is to replicate the previous findings of facilitatory effect of predictability, and extend them further by testing if listeners form *narrowed expectations* while listening to moderately degraded speech. In line with Strauß et al. (2013)'s argument, listeners' can form predictions that are restricted to only highly probable sentence endings. On the opposite, listeners' can garnet expectations about an upcoming word based on how likely the word is to appear in the context, and hence form a probabilistic prediction. We also test the possible learning effect due to potential perceptual adaptation and its effect on contextual facilitation if at all. We set a metric of measurement of language comprehension that considers whether or not listeners correctly identified the context information.

6.2 Background

Predictability effects in degraded speech perception: So far in this thesis, we have shown that predictability plays a facilitatory role in comprehension of degraded speech when listeners attend to the sentence context (Chapter 5). In their studies, Obleser and colleagues (Obleser, Wise, et al. 2007; Obleser and Kotz 2010; Obleser and Kotz 2011), used sentences of two levels of semantic predictability (high and low) and systematically degraded the speech signal by passing it through various numbers of noise vocoding channels ranging from 1 to 32 in a series of behavioral and neuroimaging studies. They observed contextual facilitation at moderate levels of speech degradation (at 4 channels, or 8 channels noise vocoding). That is, participants relied on the sentence context when the speech signal was degraded but intelligible enough. Accuracy of word recognition was found to be higher for highly predictable target words than for less predictable target words at such moderate levels of speech degradation (Obleser and Kotz 2010). Even at different cognitive load, Hunter and Pisoni (2018) observed a similar predictability effect. However, the granularity and the nature of prediction remains yet to be tested. None of the studies mentioned above were designed (or their data analyzed) to test if the the facilitatory effect of prediction is graded and probabilistic, or if it is narrowed and all-or-none.

Adaptation to degraded speech: With a repeated exposure to degraded speech, listeners' comprehension improves over time (Samuel and Kraljic 2009; Guediche et al. 2014) which we have comprehensively discussed in Chapter 2. Studies like those of Davis et al. (2005) and Erb et al. (2013) have shown that when participants are presented with a single noise condition, i.e., a single level of spectral degradation of speech, their response accuracy increases over the course of experiment. In these experiments, only one speech degradation level was presented in one block. So there was no uncertainty about the next-trial speech degradation from the participants' perspective. Importantly, there was no variation in predictability of target word. Similarly, the experiments which varied target word predictability and demonstrated the presence of contextual facilitation (e.g., Obleser, Wise, et al.

2007; Sheldon et al. 2008b; Sheldon et al. 2008a; Obleser and Kotz 2010; Strauß et al. 2013; Hunter and Pisoni 2018), did not test for adaptation to degraded speech. But we cannot ignore the facts that listeners learning effect might be absent with a change in the characteristics of auditory signal throughout the experiments (Mattys et al. 2012); and comprehension is impaired by any such trial-by-trial variability in the characteristics of distorted speech (Sommers et al. 1994; see also, Dahan and Magnuson 2006).

Measurement of language comprehension: How we measure language comprehension has not been to be guided by specific theoretical motive to a large extent in the existing literature. Studies that reported contextual facilitation in degraded speech comprehension have used proportion of correctly reported *final* words only (e.g., Sheldon et al. 2008b). Obleser??? measured language comprehension as accuracy of identification of sentence final word recognition.WHAT WAS THE EXACT MEASUREMENT?? Other studies like, Erb et al. (2013) and Hakonen et al. (2017) used *report scores* (Peelle 2013) that measure proportion of correctly recognized words per sentence. Such discrepancy makes cross-study comparison difficult, and more importantly, none of these measures take into account if listeners have correctly identified the context [cf. Amichetti et al. (2018); and prediction hinders study cited in Winn's paper.].

Stemming from the results of Chapter 5, and from the motivation driven by the open questions outlined above, the goals of the study in this chapter were threefold: The first goal was to replicate the previous findings of facilitatory effect of predictability, and test if listeners form *narrowed expectations.* Obleser and colleagues showed predictability effects (or contextual facilitation) only at a moderate level of speech degradation by using just two levels of sentence predictability (high and low). Our use of three levels of target-word predictability, in contrast to only two levels in Obleser and colleagues', will let us test the narrowed expectations view when we take into account the accuracy of context. If the listeners form a narrowed prediction only for high-cloze target words, then the facilitatory effect of semantic prediction will be observed only at these highly predictable sentence

6. Experiment 2: Semantic predictability facilitates comprehension of degraded speech in a graded manner

endings. Listeners' response to medium-cloze target words and low-cloze target words would be expected to be quite similar as these two will fall out of the range of narrowed prediction. However, if the listeners' predictions are not restricted to highly predictable target words, then they form predictions across a wide range of semantic context proportional to the probability of occurrence of the target word. In addition to highly predictable sentence endings, listeners will also form predictions for less predictable sentence endings. Such predictions, however, will depend on the probability of occurrence of the target words. In other words, listeners form predictions also for less expected sentence endings; and the semantic space of prediction depends on the probability of occurrence of those sentence endings. The addition of sentences with medium-cloze target words in the present study thus allows us to differentiate whether listeners form all-or-none prediction restricted to high-cloze target words, or a probabilistic prediction for words across a wide range of cloze probability.

There is a variation in the sentences we use, i.e., they are high, medium and low predictability sentences, and they are degraded at different levels of spectral degradation. So our second goal was to investigate the role of uncertainty about next-trial speech features on perceptual adaptation by varying the *global channel context* on the comprehension of degraded speech. To study this, we presented sentences of different levels of predictability blocked by each channel conditions (predictable channel context), and pseudo-randomized across all channels (unpredictable channel context). Based on previous findings, we expected that in the unpredictable channel context (i.e., when sentences are presented in a random order of spectral degradation) participants' word recognition performance will be worse than in the predictable channel context [i.e., when the sentences are blocked by noise-vocoding; Sommers et al. (1994); Garrido et al. (2011); Vaden et al. (2013)]. To further examine perceptual adaptation, we also considered the effect of trial number in the analyses of data.

6.3   Methods

6.3.1   Participants

We recruited two groups of participants via Prolific Academic and assigned them to one of the two groups: *unpredictable channel context* (n=48; $\bar{x} \pm$ SD = 24.44 $\pm$ 3.5 years; age range = 18-31 years; 16 females) and *predictable channel context* (n=50; $\pm$ SD = 23.6 $\pm$ 3.2 years; age range = 18-30 years; 14 females). All participants were native speakers of German residing in Germany. Exclusion criteria for participating in this study were self-reported hearing disorder, speech-language disorder, or any neurological disorder. All participants received monetary compensation for their participation. The study was approved by Deutsche Gesellschaft für Sprachwissenschaft (DGfS) Ethics Committee, and the participants provided consent in accordance with the Declaration of Helsinki.

6.3.2   Materials

We used the same stimuli described in Section X.X.X in Chapter X.X. The stimuli were digital recordings of 360 German sentences spoken by a female native speaker of German in a normal rate of speech. All sentences were in present tense consisting of pronoun, verb, determiner, and object (noun) in the Subject-Verb-Object form. We used 120 nouns to create three categorizes of sentences – high predictability sentences (HP sentencs), medium predictability sentences (MP sentences) and low predictability sentences (LP sentences) – that differed in cloze probability of sentence final target words. (See Appendix A for examples.) Their mean cloze probabilities were 0.022 $\pm$ 0.027 ($\bar{x} \pm$ SD; range = 0.00 - 0.09) for LP sentences, 0.274 $\pm$ 0.134 ($\bar{x} \pm$ SD; range = 0.1 - 0.55) for MP sentences, and 0.752 $\pm$ 0.123 ($\bar{x} \pm$ SD; range = 0.56 - 1.00) for HP sentences. The distribution of cloze probability across LP, MP and HP sentences are shown in Figure X.X.

In the unpredictable channel context, each participant was presented with 120 unique sentences: 40 HP, 40 MP and 40 LP sentences. Channel condition was also balanced across each sentence type, i.e., in each of HP, MP, and LP sentences,

ten sentences passed through each noise vocoding channels – 1, 4, 6, and 8 – were presented. This resulted into 12 experimental lists. The sentences in each list were pseudo-randomized, that is, not more than 3 sentences of same noise condition (i.e., same noise vocoding channel), or same predictability condition appeared consecutively. This randomization confirmed the uncertainty of next-trial speech quality (or degradation) in the global context of the experiment.

The same set of stimuli and experimental lists were used in the predictable channel context. Each participant was presented with 120 unique sentences blocked by channel conditions, i.e., blocked by noise vocoding channels. There were four blocks of stimuli. Thirty sentences were presented in each of the four blocks. In the first block, all sentences were 8 channels noise vocoded, followed by blocks of 6 channels, 4 channels, and 1 channel noise vocoded speech consecutively (Sheldon et al. 2008b). Within each block, 10 HP, 10 MP and 10 LP sentences were presented. All the sentences were pseudo-randomized so that not more than three sentences of the same predictability condition appeared consecutively in each block. This ascertained there was a certainty of next-trial speech quality (within each block) and an uncertianty of next-trial sentence predictability across all four blocks.

6.3.3   Procedure

Participants were asked to use headphones or earphones. A prompt to adjust loudness was displayed at the beginning of the experiment: A noise vocoded sound not used in the main experiment was presented, and participants were asked to adjust the loudness at their level of comfort. One spoken sentence was presented in each trial. Eight practice trials were presented before presenting 120 experimental trials. They were asked to enter what they had heard (i.e., to type in the entire sentence) *via* keyboard. Guessing was encouraged. The response was not timed. The experiment was about 40 minutes long.

6.4 Analyses

In the sentences used in our experiment, verbs evoke predictability of the sentence-final noun. Therefore, the effect of predictability (evoked by the verb) on language comprehension can be rightfully measured if we consider only those trials in which participants identify the verbs correctly. Verb-correct trials were considered as the sentence in which participants realized the context independent of whether they correctly understood the sentence final target noun. Morphological inflections and typos were considered as correct. We first filtered out those trials in which verbs were not identified correctly, i.e., trials with incorrect verbs. Therefore, we excluded 2469 out of 5760 trials in unpredictable channel context and 2374 out of 6000 trials in predictable channel context from the analyses. The 1 channel noise vocoding condition was dropped from the analyses as there were no correct responses in any of the trials in this condition.

We preprocessed and analysed data in R-Studio (Version 3.6.1; R Core Team, 2019) following the procedure described in Chapter 4.4.4. Response accuracy was analyzed with Generalized Linear Mixed Models (GLMMs) with lme4 (Bates, Mächler, et al. 2015) and lmerTest (Kuznetsova et al. 2017) packages. Binary responses (correct/incorrect) for all participants in both groups (predictable channel context and unpredictable channel context) were fit with a binomial logistic mixed-effects model(Jaeger 2006; Jaeger 2008). Noise condition (categorical; 4, 6, and 8 channels noise vocoding), target word predictability (categorical; HP, MP, LP), global channel context (categorical; predictable channel context and unpredictable channel context), and the interaction of noise condition 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 noise condition, target word predictability and their interaction. To find the optimal model for the data, non-significant higher-order interactions were excluded from the fixed-effects

structure (and from the random-effects structure) in a stepwise manner. Model selection was based on Akaike Information Criterion (AIC) (Grueber et al. 2011; Richards et al. 2011) unless otherwise stated. Random effects not supported by the data that explained zero variance according to singular value decomposition were excluded to prevent overparameterization. This gave a more parsimonious model (Bates, Kliegl, et al. 2015). Such a model was then extended separately with: i) item-related correlation parameters, ii) participant-related correlation parameter, and iii) both item- and participant-related correlation parameters. The best fitting model among the parsimonious and extended models, based on AIC, was then selected as the optimal model for our data.

We applied treatment contrast for noise condition (8 channels as a baseline; factor levels: 8 channels, 4 channels, 6 channels) and sliding difference contrast for target word predictability (factor levels: MP, LP, HP) and channel context (factor levels: unpredictable, predictable). The results from the optimal model are shown in Table X.X.X, and are reported below in the Results section.

6.5   Results

In this experiment, we tested i) whether predictability facilitates language comprehension only at a moderate level of spectral degradation, and ii) whether adaptation to degraded speech influences language comprehension. We observed that the mean response accuracy increased with an increase in number of noise vocoding channels from 4 to 6 to 8, and with an increase in target word predictability from low to medium to high (see Figure X.X). This trend is consistent across both the channel contexts; Figure X.X and Figure Y.Y show this trend for predictable channel context (i.e., blocked design) and unpredictable channel context (i.e., randomized design) respectively. Mean accuracies across all conditions are given in Table X and Y, and Figure X.X.

These observations are confirmed by the results of statistical analyses. We found a significant main effect of channel condition indicating that the response accuracy

6. Experiment 2: Semantic predictability facilitates comprehension of degraded speech in a graded manner

was higher in the 8 channels than in the 4 channels (β = -2.87, SE = 0.22, z (6917) = -13.1, p <.001) and 6 channels (β = -0.66, SE = 0.19, z (6917) = -3.42, p < .001). There was a significant main effect of target word predictability suggesting that response accuracy was lower at low predictability sentences than both high predictability sentences (β = 2.18, SE = 0.3, z (6917) = 7.2, p < .001) and medium predictability sentences (β = -0.52, SE = 0.27, z (6917) = -1.97, p = .049). We also found a significant interaction between channel condition and target word predictability (β = -0.71, SE = 0.29, z (6917) = -2.44, p = .015).

We performed a subsequent subgroup analyses on each noise channel condition. They revealed that the interaction was driven by the effect of predictability at 4 channels: The accuracy at high predictability sentences was higher than medium predictability sentences (β = 1.14, SE = 0.37, z (1608) = 3.1, p < .001), which in turn was also higher than low predictability sentences (β = 1, SE = 0.24, z (1608) = 4.2, p < .001). There was no significant difference in response accuracy between low predictability and high predictability sentences at both 6 channels (β = 0.33, SE = 0.32, z (2590) = 1.04, p = .3) and 8 channels (β = -0.014, SE = 0. 32, z (2719) = -0.04, p = .97). However, response accuracy was higher in high predictability than in medium predictability sentences at both 6 channels (β = 1.83, SE = 0.65, z (2590) = 2.83, p < .005) and 8 channels (β = 1.54, SE = 0.61, z (2719) = 2.54, p = .011).

We also found a significant main effect of global channel context which showed that the response accuracy was higher in predictable channel context than in unpredictable channel context (β = -0.27, SE = 0.14, z (6917) = -2.02, p = .04).

Further, to test the effect of practice on adaptation to degraded speech, we added trial number as a fixed effect in the maximal model. Note that there were 30 trials in each block in the predictable channel context (i.e., blocked design). For comparability, we divided unpredictable channel context (i.e., randomized design) into four blocks. Then following the same procedure as above, we obtained an optimal model. We did not find a significant main effect of trial number indicating that the response accuracy did not change throughout the experiment (β = -0.0004, SE = 0.01, z (6917) = -0.05, p = 0.97). It remained constant within each block

in the predictable channel context (β = -0.02, SE = 0.01, z (3291) = -1.43, p = 0.15) as well as in the unpredictable channel context (β = 0.01 SE = 0.01, z (3291) = 1.05, p = 0.29).

6.6 Discussion

The present study had three goals: i) to examine if previously reported facilitatory effect of semantic predictability is restricted to only highly predictable sentence endings; ii) to assess the role of perceptual adaptation on the facilitation of language comprehension by sentence predictability; and iii) to use and establish a sensitive metric to measure language comprehension that takes into account whether listeners benefited from the semantic context of the sentence they have listened to.

Results of our study showed the expected interaction between predictability and degraded speech, that is, language comprehension was better for high-cloze than for low-cloze target words when the speech signal was moderately degraded by noise-vocoding through 4 channels, while the effect of predictability was absent when speech was not intelligible (noise-vocoding through 1 channel). These results are fully in line with Obleser and Kotz (2010); we partly included identical sentences from their study in the present study (see Appendix A). Importantly, in contrast to their study, we had also created sentences with medium-cloze target words (which were intermediate between high-cloze and low-cloze target words) and found that the effect of predictability was also significant when comparing sentences with medium-cloze target words against sentences with low-cloze target words at 4 channels noise-vocoding condition. Recognition of a target word was dependent on its level of predictability (measured by cloze probability), and correct recognition was not just limited to high-cloze target words. These significant differences in response accuracy between medium-cloze and low-cloze target words, and between medium-cloze and high-cloze target words at noise-vocoding through 4 channels show that the sentence-final word recognition is facilitated by semantic predictability in a graded manner. This is in line with the findings from the ERP literature where it has been observed

6. Experiment 2: Semantic predictability facilitates comprehension of degraded speech in a graded manner

that semantic predictability, in terms of cloze probability of target word of a sentence, modulates semantic processing, indexed by N400, in a graded manner (DeLong et al. 2005; Wlotko and Federmeier 2012; Nieuwland, Politzer-Ahles, et al. 2018).

The interpretation of the observed graded effect of semantic predictability at the moderate level of spectral degradation (i.e., at noise-vocoding through 4 channels) provides a novel insight into how listeners form prediction when the bottom-up input is compromised. That is, in an adverse listening condition, listeners rely more on top-down semantic prediction than on bottom-up acoustic-phonetic cues. However, such a reliance on top-down prediction is not an all-or-none phenomenon; instead, listeners form a probabilistic prediction of the target word. The effect of target word predictability on comprehension is not sharply focused solely on high-cloze target words like a 'searchlight'. But rather it is spread across a wide range including low-cloze and medium-cloze target words. As the cloze probability of the target words decreases from high to low, the focus of the searchlight becomes less precise.

In conclusion, this study provides novel insights into predictive language processing when bottom-up signal quality is compromised and uncertain: We show that while processing moderately degraded speech, listeners form top-down predictions across a wide range of semantic space that is not restricted within highly predictable sentence endings. In contrast to the narrowed expectation view, comprehension of words ranging from low- to high-cloze probability, including medium-cloze probability, is facilitated in a graded manner while listening to a moderately degraded speech. We also found better speech comprehension when individuals were likely to have adapted to the noise condition in the blocked design compared to the randomized design. We did not find learning effects at the trial-to-trial level of perceptual adaption – it may be that the adaptation was hampered by variation in higher-level semantic features (i.e., target word predictability). We also argue that for the examination of semantic predictability effects during language comprehension, the analyses of response accuracy should be based on the trials in which context evoking words are correctly identified in the first place to make sure that listeners make use of the contextual cues instead of analyzing general word recognition scores.

<div style="text-align: right; font-size: 3em; font-weight: bold; color: gray;">7</div>

Experiment 3: Comprehension of degraded speech is modulated by the rate of speech

7.1 Introduction

Language comprehension depends on the features and quality of speech signal which is hampered in adverse listening conditions when speech is distorted, for example, due to change in rate of speech, and spectral degradation of speech. Studies have shown that individuals can benefit from sentence context to compensate for distorted speech, at least when the level of speech degradation is at an intermediate level (e.g., Obleser and Kotz 2010; Bhandari et al. 2021), and at varying speech rate (Aydelott and Bates 2004; cf. Goy et al. 2013). The goal of the present study is to examine if sentence context can provide benefit when speech that is degraded at an intermediate level is presented at different rates (from slow, normal, and fast); speech in day-to-day conversation is distorted and presents itself at varying rate (e.g., Krause and Braida 2004). 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.

7.1.1 Comprehension of degraded speech

There are a number of studies showing that speech intelligibility and language comprehension is hampered when the bottom-up input is less intelligible due to spectral degradation of the speech signal (Shannon, Zeng, et al. 1995; Davis et al. 2005). These studies have used noise vocoding as a methods of speech degradation. 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 lesser the number of vocoder channels, the lesser is the intelligibility. More attentional resources are required to process and comprehend such degraded speech as compared to clean speech (e.g., Wild et al. 2012; Eckert et al. 2016).

Listeners rely more on top-down predictions when the speech signal is less intelligible due to spectral degradation. 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 and 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, Wise, et al. 2007; Obleser and Kotz 2010; Obleser and Kotz 2011) found that at moderate levels of speech degradation, target words (the sentence final words) were better recognized when it was predictable from the sentence context than when it was 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 (Obleser and Kotz 2010; Bhandari et al. 2021).

7. Experiment 3: Comprehension of degraded speech is modulated by the rate of speech

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.

7.1.2 Comprehension of fast and slow speech

A change in speech rate manipulates the speech signal without producing any spectral degradation (Charpentier and Stella 1986; Moulines and 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 and Teece 2021; see also, Simantiraki and Cooke 2020), and its intelligibility and comprehension are reducec (Fairbanks and Kodman Jr. 1957; Peelle and Wingfield 2005; Schlueter et al. 2014). The comprehension deficit in fast speech has been linked to speed of processing (Gordon-Salant and Fitzgibbons 1995; Tun 1998; see also, Rönnberg et al. 2013) given the limited time available to decode and understand the information in the fast speech. It has also been suggested that the decoding and identification of incoming information in the fast speech puts a high demand on available cognitive resources (e.g., Rodero 2016) such that processing rapidly flowing information exhausts the cognitive resource required for language processing (Gordon-Salant and Fitzgibbons 2004; Janse 2009). Hence, intelligibility and comprehension of fast speech is reduced compared to normal speech. In contrast, the central auditory-language comprehension system is shown to be flexible to process slow speech without reducing intelligibility (Lerner et al. 2014; see also, Vagharchakian et al. 2012). However, some earlier studies have casted doubt on the processing advantage of slow speech (e.g., Nejime and Moore 1998; Kemper and Harden 1999; see also, Liu and Zeng 2006; Love et al. 2009). In sum, compression and expansion of speech have differential effects on speech intelligibility and language comprehension. Intelligibility and comprehension of fast speech is reduced while that of slow speech is generally increased compared to normal speech.

7. Experiment 3: Comprehension of degraded speech is modulated by the rate of speech

There are also a few studies so far that examined the role of predictability for understanding language when speech is fast or slow. 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 constrained 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 higher for slow speech. 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 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 can then continue to allocate resources in the same mode under faster presentation rate.

In sum, there is already some evidence from studies applying various paradigms that the predictability of the sentence context interacts with the speech rate (Sharit et al. 2003; Aydelott and Bates 2004; Wlotko and Federmeier 2015; Winn and Teece 2021). Benefits of using context information is not only limited under varying speech rates, but also when it is degraded (Sheldon et al. 2008b; Sheldon et al. 2008a; Obleser and Kotz 2010; Bhandari et al. 2021). In the present study we will examine whether context information is still useful when the speech rate is fast (or slow) and degraded at the same time.

7.1.3 Comprehension of degraded speech at varying rates of presentation

Real world listening is often a combination of more than one form of degradation and distortion of speech (Gordon-Salant and Fitzgibbons 1995; see also, Cooke, King, et al. 2014). For example, speakers' speech rate is variable, and there is a concurrent distortion due to background noise, reverberation, loss of spectral details in transmission, etc. Some of the earliest studies examined the 'devastating' effects of 'combinations of speech wave distortions', but the combination of multiple

7. Experiment 3: Comprehension of degraded speech is modulated by the rate of speech

distortions was not controlled for (Martin et al. 1956; Harris 1960). Gordon-Salant and Fitzgibbons (1995) systematically tested the effect of multiple, combined speech distortions. They found that when fast speech was presented in a background noise, or with a reverberation then the word recognition accuracy in these combined distortions was lower than when the speech was presented in any single distortion condition. Adams and Moore (2009) and Adams, Gordon-Hickey, et al. (2012) also showed that the addition of background noise had a detrimental effect on fast speech comprehension. With an increase in speech rate, listeners required higher signal-to-noise ratio (SNR) to achieve 50% word recognition accuracy across a wide range of speech rate (90 to 250 words per minute, in a step of 9 words per minute) and background noise (+15 dB SNR, +5 dB SNR, and 0 dB SNR). These studies showed that a combination of multiple distortions is worse than a single distortion. It is noteworthy that neither of the distortions mentioned above (e.g., reverberation, background noise) interfere with fine structure of a speech signal which spectral degradation does.

So far, the evidence for the detrimental effects of combination of spectral degradation and speech compression is sparse. Most of the evidence on this comes from cochlear implant users (e.g., Li, Zhang, Kang, et al. 2011; Su et al. 2016) as they are the group of listeners whose auditory input is spectrally degraded. Iwasaki et al. (2002) found that a change in speech rate from slow to fast reduced word recognition accuracy in cochlear implantees. Similarly, 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). Meng et al. (2019) used both clean speech and spectrally degraded speech from MSP (Mandarin speech perception) and MHINT (Mandarin hearing in noise test) corpora in normal-hearing listeners and cochlear implant users. They found that an increase in speech rate had much severe effect at spectrally degraded speech (4 channels sine-wave vocoded) than at clean speech. To achieve the same level of accuracy, listeners required degraded speech to be much slower than the normal speech. All of these studies indicated that when speech rate is increased, intelligibility and comprehension of

degraded speech is reduced. However, these studies did not examine the utility of context information when speech is distorted by more than one means, although the role of context and predictability in any one of these multiple distortions has been extensively discussed in the literature (Aydelott and Bates 2004; Sheldon et al. 2008b; Sheldon et al. 2008a; Obleser and Kotz 2010; Goy et al. 2013; Corps and Rabagliati 2020; Bhandari et al. 2021; Clark et al. 2021; van Os et al. 2021).

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. Whether semantic predictability is beneficial even for degraded speech presented at different rates (fast and slow) has not been investigated yet.

7.1.4 Study aims

We systematically examined if contextual facilitation at a moderate level of degradation varies with a change in speech rate. The aim was to investigate if the increase (or decrease) in speech rate decreases (or increases) 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 and 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 (or 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 (or 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 and Bates 2004; Wlotko and Federmeier 2015), 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 3A). 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 and Teece 2021), we expected contextual facilitation to be increased compared to normal speech (Experiment 3B). 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 and Bates 2004; Goy et al. 2013).

7.2   Experiment 3A

7.3   Methods

7.3.1   Participants

We recruited one group of participant (n=101; $\bar{x} \pm \text{SD} = 23.14 \pm \text{XX}$ years; age range = 18-31 years; 66 females, 1 preferred not to say) online via Prolific Academic. All participants were native speakers of German residing in Germany. Exclusion criteria for participating in this study were self-reported hearing disorder, speech-language disorder, or any neurological disorder. All participants received monetary compensation for their participation. The German Society for Language Science ethics committee approved the study and participants provided an informed consent in accordance with the declaration of Helsinki.

7.3.2   Materials

We used the stimuli created by the method described in Section X.X.X in Chapter X.X which consisted of 360 German sentences spoken by a female native German speaker in an unaccented normal rate of speech. Two categories of sentences that differed in the cloze probability of the target words (nouns) appearing the final

word of the sentence were created from 120 nouns. Thus, we compared high and low predictability sentences (abbreviated as HP and LP henceforth) which were sentences with low and high cloze target words respectively. This gave 240 sentences that consisted of pronoun, verb, determiner, and object (noun). The mean cloze probabilities of target words for low and high predictability sentences were 0.022 ± 0.027 ($\bar{x} \pm$ SD; range $= 0.00$ - 0.09) and 0.752 ± 0.123 ($\bar{x} \pm$ SD; range $= 0.56$ - 1.00) respectively.

These 240 sentences were passed compressed by a factor of 0.65 using PSOLA built-in in Praat to create fast and slow speech respectively. Speech degradation of all normal, slow and fast speech was achieved by noise vocoding through 4 channels.

Each participant was presented with 120 unique sentences: 60 HP and 60 LP sentences. Speech rate was also balanced across each predictability level. The participants received 30 sentences with normal speed and 30 with fast speed in each of the predictability conditions resulting into 4 experimental lists. The sentences in each list were pseudo-randomized, that is, not more than 3 sentences of same speed, or same predictability condition appeared consecutively.

7.3.3   Procedure

Participants were asked to use headphones or earphones. A sample of noise vocoded speech not used in the practice trial and the main experiment was provided so that the participants could adjust the loudness to a preferred level of comfort at the beginning of the experiment. The participants were instructed to listen to the sentences and to type in the entire sentence 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 understand, and in these cases, they were encouraged to guess what they might have heard. They were not informed about the speed of speech being slow/fast or normal. 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 inter-trial interval of 1000 ms.

7.4 Analyses

We have already conceded in the previous chapter X.X.X that "the effect of predictability (evoked by the verb) on language comprehension can be rightfully measured if we consider only those trials in which participants identify the verbs correctly." Therefore, we discarded the trials in which verbs were identified incorrectly – XXXX out of XXXX trials.

We preprocessed and analysed data in R-Studio (Version 4.1.1; R Core Team, 2021) following the procedure described in Chapter 4.4.4.

Response accuracy was analyzed with Generalized Linear Mixed Models (GLMMs) with lme4 (Bates, Mächler, et al. 2015) and lmerTest (Kuznetsova et al. 2017) packages. Binary responses (correct/incorrect) for all participants were fit with a binomial logistic mixed-effects model(Jaeger 2006; Jaeger 2008). Target word predictability (categorical; low and high), speech rate, or speed (categorical; xxx and xxx), and the interaction of predictability and speed 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, speed and their interaction.

We applied treatment contrast for target word predictability (low predictability as a baseline; factor levels: low predictability, high predictability) and speed (normal speed as a baseline; factor levels: normal speed, fast speed). The results from the maximal model are shown in Table X.X.X, and are reported below in the Results section.

7.5 Results and Discussion

Mean response accuracies across all conditions are presented in Table 2. We found a significant main effect of target word predictability ($\beta = 2.42$, SE $= .28$, $z = 8.55$, $p < .001$) and a significant main effect of speech rate ($\beta = $ -0.98, SE $= .24$, $z = $

7. Experiment 3: Comprehension of degraded speech is modulated by the rate of speech

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 ($\beta = 1.06$, 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. The results shown in Table 4 and Table 5 revealed that there was no significant main effect of speech rate at high predictability condition ($\beta = .02$, SE $= .34$, $z = .05$, $p = .96$). At low predictability condition, in contrast, we found a significant main effect of speech rate ($\beta = -.99$, 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 shown in Table 6 and Table 7 revealed that there was significant main effect of predictability in both normal speech ($\beta = 1.98$, SE $= 0.28$, $z = 7.05$, $p < .001$) and fast speech conditions ($\beta = 2.67$, SE $= .37$, $z = 7.14$, $p < .001$), but the effect appeared to be higher for fast speech ($\beta = 2.67$) than for normal speech ($\beta = 1.98$). 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 and 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.

7.6 Experiment 3B

7.7 Methods

7.7.1 Participants and Materials

We recruited 101 participants ($\bar{x} \pm \text{SD} = 23.49 \pm 3.26$ years; age range = 18-30 years; 60 females, 1 preferred not to say) online via Prolific Academic. Same procedure as Experiment 3A was followed.

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.

We followed the same steps as Experiment 3A to balance speech rate and predictability conditions and to pseudo-randomize these experimental conditions across all 4 lists.

7.7.2 Procedure

Same procedure as Experiment 3A was followed. We asked participants to report the entire sentence typing in what they heard. And guessing was encouraged.

7.8 Analyses

We followed the same data analyses procedure as in Experiment 3A. Only the trials with verb-correct responses were considered in the analyses of accuracy of sentence-final target-words (i.e., nouns); 5495 out of 12120 trials were removed before the final analyses.

7. Experiment 3: Comprehension of degraded speech is modulated by the rate of speech

Target word predictability (categorical; low and high), speech rate, or speed (categorical; normal and slow), and the interaction of predictability and speed were included in the fixed effects. Treatment contrast was applied to both target word predictability (low predictability as a baseline; factor levels: low predictability, high predictability) and speed (normal speed as a baseline; factor levels: normal speed, slow speed). The results from the maximal model are shown in Table X.X.X, and are reported below in the Results section.

7.9   Results and Discussion

Mean response accuracies for all experimental conditions are shown in Table 8. 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 9.

We conducted separate planned analyses of each predictability level; the procedure was identical to the analyses in Experiment 1. The results presented in Table 10 and Table 11 confirmed the findings of the main analysis reported above – there was no significant main effect of speech rate at both high predictability condition (β = .39, SE = .22, z = 1.79, p = .073), as well as at low predictability condition (β = -.11, SE = .16, z = -.71, p = .48). Hence, as opposed to the findings of Experiment 1, response accuracy was not affected by change in speech rate at both high and low predictability sentences.

7. Experiment 3: Comprehension of degraded speech is modulated by the rate of speech

As in Experiment 1, we conducted separate planned analyses of each speech rate. The results as shown in Table 12 and Table 13 revealed that there was a significant main effect of target word predictability at both slow speech rate ($\beta =$ 2.55, SE = .31, $z = 8.14$, $p < .001$) and normal speech rate ($\beta = 2.08$, SE = .27, $z = 7.68$, $p < .001$); the effect appeared to be higher for slow speech ($\beta = 2.55$) than for normal speech ($\beta = 2.08$). This, however, was a result of smaller accuracy at low predictability condition at slow speech rather than due to a higher accuracy at high predictability condition. This can also be seen in Table 8 and Figure 2.

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.

7.10 General Discussion

The main goals of the present study were to examine if semantic predictability facilitates comprehension of moderately degraded speech at different speech rates given that the ease of processing the 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 the 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 like Obleser and Kotz (2010) and Bhandari et al. (2021) in which participants were presented only

7. Experiment 3: Comprehension of degraded speech is modulated by the rate of speech

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 a little to no effect in comprehension of high predictability sentences at fast speech rate under moderate degradation level. 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 form predictions about upcoming target word. This difficulty is heightened when target words are not easily predictable from the context (Aydelott and Bates 2004). As a result, language comprehension in low predictability condition is impaired more than that in high predictability condition. This detrimental effect on low predictability condition also results in the reduced contextual facilitation in fast speech compared to normal speech observed 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 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 this ease of processing does not benefit intelligibility and comprehension of sentences with highly predictable target words any more than the sentences with less predictable target words. 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 listeners have more time to process the incoming information [CITE]. Additionally, there are conflicting empirical evidence whether increase or decrease of speech rate provides benefit in intelligibility, comprehension, and contextual facilitation. Given these state of affairs, our findings add new theoretical insights with empirical support into the 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 eases the processing of both high and low predictability to a similar extent. 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.

We note a limitation in our study. We tested only with one expansion factor of 1.35, and one compression factor of 0.65. Although less likely, it can be speculated that when the speech is expanded to a greater degree by including other expansion factors, an increase in facilitatory effect could be observed [cf.] which was not the case in the current study.

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.

7.11   Conclusions

8

Experiment 4: Older adults rely more on sentence context than on the auditory signal in comprehension of moderately degraded speech

9
Discussion and conclusion

9.1   Summary of the main findings

9.2   Theoretical and practical implications

9.3   Conclusion

A conclusion is the place where you got tired of thining.

— Martin H. Fischer (**Darwin1859**)

10
Ethics and funding

Appendices

A
The First Appendix

This first appendix includes an R chunk that was hidden in the document (using `echo = FALSE`) to help with readibility:

In 02-rmd-basics-code.Rmd

And here's another one from the same chapter, i.e. Chapter ??:

B

The Second Appendix, for Fun

Bibliography

Aarts, Alexander A. et al. (2015). "Estimating the reproducibility of psychological science". In: *Science* 349.6251. DOI: `10.1126/science.aac4716`.

Adams, Elizabeth M., Susan Gordon-Hickey, et al. (2012). "Effect of rate-alteration on speech perception in noise in older adults with normal hearing and hearing impairment". In: *American Journal of Audiology* 21.1, pp. 22–32. DOI: `10.1044/1059-0889(2011/10-0023)`.

Adams, Elizabeth M. and Robert E. Moore (2009). "Effects of speech rate, background noise, and simulated hearing loss on speech rate judgment and speech intelligibility in young listeners". In: *Journal of the American Academy of Audiology* 20.1, pp. 28–39. DOI: `10.3766/jaaa.20.1.3`.

Altmann, Gerry T.M and Yuki Kamide (Dec. 1999). "Incremental interpretation at verbs: restricting the domain of subsequent reference". In: *Cognition* 73.3, pp. 247–264. DOI: `10.1016/s0010-0277(99)00059-1`.

Altmann, Gerry T.M. and Yuki Kamide (2007). "The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing". In: *Journal of Memory and Language* 57.4, pp. 502–518. DOI: `10.1016/j.jml.2006.12.004`.

Amichetti, Nicole M. et al. (Jan. 2018). "Linguistic Context Versus Semantic Competition in Word Recognition by Younger and Older Adults With Cochlear Implants". In: *Ear & Hearing* 39.1, pp. 101–109. DOI: `10.1097/aud.0000000000000469`.

Ankener, Christine S., Mirjana Sekicki, and Maria Staudte (2018). "The influence of visual uncertainty on word surprisal and processing effort". In: *Frontiers in Psychology* 9, p. 2387. DOI: `10.3389/fpsyg.2018.02387`.

Anwyl-Irvine, Alexander, Edwin S. Dalmaijer, et al. (2021). "Realistic precision and accuracy of online experiment platforms, web browsers, and devices". In: *Behavior Research Methods* 53.4, pp. 1407–1425. DOI: `10.3758/s13428-020-01501-5`.

Anwyl-Irvine, Alexander L., Jessica Massonnié, et al. (2020). "Gorilla in our midst: An online behavioral experiment builder". In: *Behavior Research Methods* 52.1, pp. 388–407. DOI: `10.3758/s13428-019-01237-x`.

Astheimer, Lori B. and Lisa D. Sanders (2011). "Predictability affects early perceptual processing of word onsets in continuous speech". In: *Neuropsychologia* 49.12, pp. 3512–3516. DOI: `10.1016/j.neuropsychologia.2011.08.014`.

Aydelott, Jennifer and Elizabeth Bates (2004). "Effects of acoustic distortion and semantic context on lexical access". In: *Language and Cognitive Processes* 19.1, pp. 29–56. DOI: `10.1080/01690960344000099`.

Baayen, R. H., D. J. Davidson, and D. M. Bates (2008). "Mixed-effects modeling with crossed random effects for subjects and items". In: *Journal of Memory and Language* 59.4, pp. 390–412. DOI: `10.1016/j.jml.2007.12.005`.

Bibliography

Barr, Dale J. et al. (Apr. 2013). "Random effects structure for confirmatory hypothesis testing: Keep it maximal". In: *Journal of Memory and Language* 68.3, pp. 255–278. DOI: 10.1016/j.jml.2012.11.001.

Bates, Douglas, Reinhold Kliegl, et al. (2015). "Parsimonious Mixed Models". In: *arXiv*. arXiv: 1506.04967.

Bates, Douglas, Martin Mächler, et al. (2015). "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1. DOI: 10.18637/jss.v067.i01.

Berners-Lee, Tim et al. (1992). "World-wide web: The information universe". In: *Internet Research* 2.1, pp. 52–58. DOI: 10.1108/eb047254.

Bhandari, Pratik, Vera Demberg, and Jutta Kray (2021). "Semantic predictability facilitates comprehension of degraded speech in a graded manner". In: *Frontiers in Psychology* 3769. DOI: 10.3389/fpsyg.2021.714485.

Brunellière, Angèle, Cyril Auran, and Laurence Delrue (2019). "Does the prosodic emphasis of sentential context cause deeper lexical-semantic processing?" In: *Language, Cognition and Neuroscience* 34.1, pp. 29–42. DOI: 10.1080/23273798.2018.1499945.

Burnham, Kenneth P. and David R. Anderson (2002). "Basic Use of the Information-Theoretic Approach". In: *Model Selection and Multimodel Inference.* 2nd. Chap. 3, pp. 98–148. DOI: 10.1007/978-0-387-22456-5_3.

Charpentier, F. J. and M. G. Stella (1986). "Diphone Synthesis Using an Overlap-Add Technique for Speech Waveforms Concatenation." In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 2015–2018. DOI: 10.1109/icassp.1986.1168657.

Chatterjee, Samprit and Ali S. Hadi (2012). "Multiple linear regression". In: *Regression analysis by example.* John Wiley & Sons. Chap. 3, pp. 57–91.

Cherry, E. Colin (1953). "Some experiments on the recognition of speech, with one and with two ears". In: *Journal of the Acoustical Society of America* 25.5, pp. 975–979. DOI: 10.1121/1.1907229.

Clark, Andy (2013). "Whatever next? Predictive brains, situated agents, and the future of cognitive science". In: *Behavioral and Brain Sciences* 36.3, pp. 181–204. DOI: 10.1017/s0140525x12000477.

Clark, Catherine, Sara Guediche, and Marie Lallier (2021). "Compensatory cross-modal effects of sentence context on visual word recognition in adults". In: *Reading and Writing* 34.8, pp. 2011–2029. DOI: 10.1007/s11145-021-10132-x.

Clark, Herbert H (1973). "The language-as-a-fixed-effect fallacy: A critique of language statistics in psychological research". In: *Journal of Verbal Learning and Verbal Behavior* 12.4, pp. 335–359. DOI: 10.1016/S0022-5371(73)80014-3.

Cockburn, Andy et al. (2020). "Threats of a replication crisis in empirical computer science". In: *Communications of the ACM* 63.8, pp. 70–79. DOI: 10.1145/3360311.

Cole, Alissa (2020). "The effects of prediction and speech rate on lexical processing". Masters thesis. University of Maryland, p. 35.

Coleman, E. B. (1964). "Generalizing to a language population". In: *Psychological Reports* 14.1, pp. 219–226. DOI: 10.2466/pr0.1964.14.1.219.

Cooke, Martin and ML Garcia Lecumberri (2021). "Estimating the performance gap between lab and remote speech perception experiment". In: *Acoustical Society of America Journal* 149.4, A111–A111.

Bibliography

Cooke, Martin, Simon King, et al. (2014). "The listening talker: A review of human and algorithmic context-induced modifications of speech". In: *Computer Speech and Language* 28.2, pp. 543–571. DOI: `10.1016/j.csl.2013.08.003`.

Corps, Ruth E. and Hugh Rabagliati (Aug. 2020). "How top-down processing enhances comprehension of noise-vocoded speech: Predictions about meaning are more important than predictions about form". In: *Journal of Memory and Language* 113, p. 104114. DOI: `10.1016/j.jml.2020.104114`.

Dahan, Delphine and James S. Magnuson (2006). "Spoken Word Recognition". In: Elsevier, pp. 249–283. DOI: `10.1016/b978-012369374-7/50009-2`.

Darwin, Chris (n.d.). *Praat scripts for producing Shannon AM speech.* `http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/`. Accessed: 2021-03-08.

Davis, Matthew H. et al. (2005). "Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences." In: *Journal of Experimental Psychology: General* 134.2, pp. 222–241. DOI: `10.1037/0096-3445.134.2.222`.

DeLong, Katherine A, Thomas P Urbach, and Marta Kutas (July 10, 2005). "Probabilistic word pre-activation during language comprehension inferred from electrical brain activity". In: *Nature Neuroscience* 8.8, pp. 1117–1121. DOI: `10.1038/nn1504`.

Dincer D'Alessandro, Hilal et al. (2018). "The role of speech rate for Italian-speaking cochlear implant users: insights for everyday speech perception". In: *International Journal of Audiology* 57.11, pp. 851–857. DOI: `10.1080/14992027.2018.1498139`.

Dudley, Homer (1939). "The vocoder". In: *Bell Laboratories Record* 18.4, pp. 122–126.

Ebersole, Charles R. et al. (2016). "Many Labs 3: Evaluating participant pool quality across the academic semester via replication". In: *Journal of Experimental Social Psychology* 67, pp. 68–82. DOI: `10.1016/j.jesp.2015.10.012`.

Eckert, Mark A., Susan Teubner-Rhodes, and Kenneth I. Vaden (July 2016). "Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions". In: *Ear & Hearing* 37.1, 101S–110S. DOI: `10.1097/aud.0000000000000300`.

Ehrlich, Susan F. and Keith Rayner (1981). "Contextual effects on word perception and eye movements during reading". In: *Journal of Verbal Learning and Verbal Behavior* 20.6, pp. 641–655. DOI: `10.1016/S0022-5371(81)90220-6`.

Erb, J. et al. (June 26, 2013). "The Brain Dynamics of Rapid Perceptual Adaptation to Adverse Listening Conditions". In: *Journal of Neuroscience* 33.26, pp. 10688–10697. DOI: `10.1523/jneurosci.4596-12.2013`.

Erb, Julia (2014). "The neural dynamics of perceptual adaptation to degraded speech". Doctoral dissertation. Universität Leipzig, p. 211.

Eyal, Peer et al. (2021). "Data quality of platforms and panels for online behavioral research". In: *Behavior Research Methods*, pp. 1–20. DOI: `10.3758/s13428-021-01694-3`.

Fairbanks, Grant and Frank Kodman Jr. (1957). "Word intelligibility as a function of time compression". In: *The Journal of the Acoustical Society of America* 29.5, pp. 636–641.

Federmeier, Kara, Devon Mclennan, et al. (2002). "The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study". In: *Psychophysiology* 39.02, pp. 133–146.

Bibliography

Federmeier, Kara D., Marta Kutas, and Rina Schul (2010). "Age-related and individual differences in the use of prediction during language comprehension". In: *Brain and Language* 115.3, pp. 149–161. DOI: 10.1016/j.bandl.2010.07.006.

Federmeier, Kara D., Edward W. Wlotko, et al. (May 2007). "Multiple effects of sentential constraint on word processing". In: *Brain Research* 1146, pp. 75–84. DOI: 10.1016/j.brainres.2006.06.101.

Frisson, Steven, Keith Rayner, and Martin J. Pickering (2005). "Effects of contextual predictability and transitional probability on eye movements during reading". In: *Journal of Experimental Psychology: Learning Memory and Cognition* 31.5, pp. 862–877. DOI: 10.1037/0278-7393.31.5.862.

Friston, Karl J., Thomas Parr, et al. (2020). "Generative models, linguistic communication and active inference". In: *Neuroscience & Biobehavioral Reviews* 118, pp. 42–64. DOI: 10.1016/j.neubiorev.2020.07.005.

Friston, Karl J., Noor Sajid, et al. (2020). "Active listening". In: *Hearing Research* xxxx, p. 107998. DOI: 10.1016/j.heares.2020.107998.

Fritz, Jonathan B. et al. (2007). "Auditory attention - focusing the searchlight on sound". In: *Current Opinion in Neurobiology* 17.4, pp. 437–455. DOI: 10.1016/j.conb.2007.07.011.

Gadiraju, Ujwal et al. (2017). "Crowdsourcing versus the laboratory: Towards human-centered experiments using the crowd". In: *Evaluation in the crowd. Crowdsourcing and human-centered experiments*. Ed. by Daniel Archambault, Helen Purchase, and Tobias Hoßfeld. Springer, Cham, pp. 6–26. DOI: 10.1007/978-3-319-66435-4_2.

Gagné, Nathan and Léon Franzen (2021). "How to run behavioural experiments online: Best practice suggestions for cognitive psychology and neuroscience". In.

Garrido, Marta I, Raymond J Dolan, and Maneesh Sahani (Jan. 2011). "Surprise Leads to Noisier Perceptual Decisions". In: *i-Perception* 2.2, pp. 112–120. DOI: 10.1068/i0411.

Garvey, William David (1953). "The intelligibility of speeded speech". In: *Journal of Experimental Psychology* 45.2, pp. 102–108. DOI: 10.1037/h0054381.

Gordon-Salant, Sandra and Peter J. Fitzgibbons (1995). "Recognition of multiply degraded speech by young and elderly listeners". In: *Journal of Speech and Hearing Research* 38.5, pp. 1150–1156.

Gordon-Salant, Sandra and Peter J. Fitzgibbons (2004). "Effects of stimulus and noise rate variability on speech perception by younger and older adults". In: *The Journal of the Acoustical Society of America* 115.4, pp. 1808–1817. DOI: 10.1121/1.1645249.

Goy, Huiwen et al. (2013). "The Effects of Semantic Context and the Type and Amount of Acoustic Distortion on Lexical Decision by Younger and Older Adults". In: *Journal of Speech, Language, and Hearing Research* 56.6, pp. 1715–1732. DOI: 10.1044/1092-4388(2013/12-0053).

Greenwood, Donald D. (1990). "A cochlear frequency-position function for several species—29 years later". In: *Journal of the Acoustical Society of America* 87.6, pp. 2592–2605. DOI: 10.1121/1.399052.

Grueber, C. E. et al. (Jan. 27, 2011). "Multimodel inference in ecology and evolution: challenges and solutions". In: *Journal of Evolutionary Biology* 24.4, pp. 699–711. DOI: 10.1111/j.1420-9101.2010.02210.x.

Bibliography

Guediche, Sara et al. (2014). "Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research". In: *Frontiers in Systems Neuroscience* 7.Jan, pp. 1–16. DOI: 10.3389/fnsys.2013.00126.

Hafter, Ervin R., Anastasios Sarampalis, and Psyche Loui (2007). "Auditory Attention and Filters". In: *Auditory Perception of Sound Sources*. Vol. 29, pp. 115–142. DOI: 10.1007/978-0-387-71305-2_5.

Hakonen, Maria et al. (Aug. 4, 2017). "Predictive processing increases intelligibility of acoustically distorted speech: Behavioral and neural correlates". In: *Brain and Behavior* 7.9, e00789. DOI: 10.1002/brb3.789.

Harris, Donald J. (1960). "Combinations of distortion in speech". In: *Archives of Otolaryngology* 72.2, pp. 227–232.

Huettig, Falk and Ernesto Guerra (2019). "Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing". In: *Brain Research* 1706.June 2017, pp. 196–208. DOI: 10.1016/j.brainres.2018.11.013.

Huettig, Falk and Nivedita Mani (Sept. 7, 2015). "Is prediction necessary to understand language? Probably not". In: *Language, Cognition and Neuroscience* 31.1, pp. 19–31. DOI: 10.1080/23273798.2015.1072223.

Hunter, Cynthia R. and David B. Pisoni (2018). "Extrinsic cognitive load impairs spoken word recognition in high-and low-predictability sentences". In: *Ear and Hearing* 39.2, pp. 378–389. DOI: 10.1097/AUD.0000000000000493.

Iwasaki, Satoshi et al. (2002). "Contribution of speech rate to speech perception in multichannel cochlear implant users". In: *Annals of Otology, Rhinology and Laryngology* 111.8, pp. 718–721. DOI: 10.1177/000348940211100811.

Jaeger, T. Florian (2006). "Redundancy and syntactic reduction in spontaneous speech". Doctoral dissertation. Standford University, p. 231.

Jaeger, T. Florian (2008). "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models". In: *Journal of Memory and Language* 59.4, pp. 434–446. DOI: 10.1016/j.jml.2007.11.007.

James, William (1890). *The Principles of Psychology*. New York: Henry Holt.

Janse, Esther (2009). "Processing of fast speech by elderly listeners". In: *The Journal of the Acoustical Society of America* 125.4, pp. 2361–2373. DOI: 10.1121/1.3082117.

Johnson, Brian P. et al. (2021). "Crowdsourcing in cognitive and systems neuroscience". In: *The Neuroscientist*, p. 10738584211017018. DOI: 10.1177/10738584211017018.

Kaiser, E and J Trueswell (Dec. 2004). "The role of discourse context in the processing of a flexible word-order language". In: *Cognition* 94.2, pp. 113–147. DOI: 10.1016/j.cognition.2004.01.002.

Kamide, Yuki, Gerry T.M. Altmann, and Sarah L Haywood (July 2003). "The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements". In: *Journal of Memory and Language* 49.1, pp. 133–156. DOI: 10.1016/s0749-596x(03)00023-8.

Kemper, Susan and Tamara Harden (1999). "Experimentally disentangling what's beneficial about elderspeak from what's not". In: *Psychology and Aging* 14.4, pp. 656–670. DOI: 10.1037/0882-7974.14.4.656.

Knoeferle, Pia et al. (Feb. 2005). "The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events". In: *Cognition* 95.1, pp. 95–127. DOI: 10.1016/j.cognition.2004.03.002.

Bibliography

Koch, Xaver and Esther Janse (2016). "Speech rate effects on the processing of conversational speech across the adult life span". In: *The Journal of the Acoustical Society of America* 139.4, pp. 1618–1636. DOI: 10.1121/1.4944032.

Kok, Peter et al. (2012). "Attention reverses the effect of prediction in silencing sensory signals". In: *Cerebral Cortex* 22.9, pp. 2197–2206. DOI: 10.1093/cercor/bhr310.

Krause, Jean C. and Louis D. Braida (2004). "Acoustic properties of naturally produced clear speech at normal speaking rates". In: *The Journal of the Acoustical Society of America* 115.1, pp. 362–378. DOI: 10.1121/1.1635842.

Kuperberg, Gina R. and T. Florian Jaeger (Nov. 2016). "What do we mean by prediction in language comprehension?" In: *Language, Cognition and Neuroscience* 31.1, pp. 32–59. DOI: 10.1080/23273798.2015.1102299.

Kutas, Marta and Kara D. Federmeier (Jan. 10, 2011). "Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP)". In: *Annual Review of Psychology* 62.1, pp. 621–647. DOI: 10.1146/annurev.psych.093008.131123.

Kutas, Marta and Steven A. Hillyard (Jan. 1984). "Brain potentials during reading reflect word expectancy and semantic association". In: *Nature* 307.5947, pp. 161–163. DOI: 10.1038/307161a0.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen (2017). "lmerTest Package: Tests in Linear Mixed Effects Models". In: *Journal of Statistical Software* 82.13. DOI: 10.18637/jss.v082.i13.

Lange, Kathrin (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". In: *Frontiers in Human Neuroscience* 7, p. 263. DOI: 10.3389/fnhum.2013.00263.

Lange, Kathrin and Brigitte Röder (2010). "Temporal orienting in audition, touch, and across modalities". In: *Attention and time*, pp. 393–405.

Leensen, Monique C.J. and Wouter A. Dreschler (2013). "Speech-in-noise screening tests by internet, Part 3: Test sensitivity for uncontrolled parameters in domestic usage". In: *International Journal of Audiology* 52.10, pp. 658–669. DOI: 10.3109/14992027.2013.803610.

Lerner, Y. et al. (2014). "Temporal scaling of neural responses to compressed and dilated natural speech". In: *Journal of Neurophysiology* 111.12, pp. 2433–2444. DOI: 10.1152/jn.00497.2013.

Li, Junfeng, Risheng Xia, et al. (2014). "Investigation of objective measures for intelligibility prediction of noise-reduced speech for Chinese, Japanese, and English". In: *The Journal of the Acoustical Society of America* 136.6, pp. 3301–3312. DOI: 10.1121/1.4901079.

Li, Xiaoqing, Yuping Zhang, Lin Li, et al. (2017). "Attention is shaped by semantic level of event-structure during speech comprehension: An electroencephalogram study". In: *Cognitive Neurodynamics* 11.5, pp. 467–481. DOI: 10.1007/s11571-017-9442-4.

Li, Yongxin, Guoping Zhang, Hou-yong Kang, et al. (2011). "Effects of speaking style on speech intelligibility for Mandarin-speaking cochlear implant users". In: *The Journal of the Acoustical Society of America* 129.6, EL242–EL247. DOI: 10.1121/1.3582148.

Liu, Sheng and Fan-Gang Zeng (2006). "Temporal properties in clear speech perception". In: *The Journal of the Acoustical Society of America* 120.1, pp. 424–432. DOI: 10.1121/1.2208427.

Bibliography

Loizou, Philipos C., Michael Dorman, and Zhemin Tu (Oct. 1999). "On the number of channels needed to understand speech". In: *The Journal of the Acoustical Society of America* 106.4, pp. 2097–2103. DOI: 10.1121/1.427954.

Love, Tracy, Matthew Walenski, and David Swinney (2009). "Slowed speech input has a differential impact on on-line and off-line processing in children's comprehension of pronouns". In: *Journal of Psycholinguistic Research* 38.3, pp. 285–304. DOI: 10.1007/s10936-009-9103-9.

Lupyan, Gary and Andy Clark (Aug. 2015). "Words and the World". In: *Current Directions in Psychological Science* 24.4, pp. 279–284. DOI: 10.1177/0963721415570732.

Marques, Tiago et al. (Apr. 16, 2018). "The functional organization of cortical feedback inputs to primary visual cortex". In: *Nature Neuroscience* 21.5, pp. 757–764. DOI: 10.1038/s41593-018-0135-z.

Martin, D. W., R. L. Murphy, and Albert Meyer (1956). "Articulation reduction by combined distortions of speech waves". In: *The Journal of the Acoustical Society of America* 28.4, pp. 597–601. DOI: 10.1121/1.1918111.

Mattys, Sven L. et al. (2012). "Speech recognition in adverse conditions: A review". In: *Language and Cognitive Processes* 27.7-8, pp. 953–978. DOI: 10.1080/01690965.2012.705006.

McClelland, James L. and Jeffrey L. Elman (1986). "The TRACE model of speech perception". In: *Cognitive Psychology* 18.1, pp. 1–86. DOI: 10.1016/0010-0285(86)90015-0.

McCullough, Constance M. (1958). "Context aids reading". In: *The Reading Teacher* 11.4, pp. 225–229.

Meng, Qinglin et al. (2019). "Time-compression thresholds for Mandarin sentences in normal-hearing and cochlear implant listeners". In: *Hearing Research* 374, pp. 58–68. DOI: 10.1016/j.heares.2019.01.011.

Meteyard, Lotte and Rob Davies (2020). "Best practice guidance for linear mixed-effects models in psychological science". In: *Journal of Memory and Language* 112, p. 104092. DOI: 10.31234/osf.io/h3duq.

Metusalem, Ross et al. (May 2012). "Generalized event knowledge activation during online sentence comprehension". In: *Journal of Memory and Language* 66.4, pp. 545–567. DOI: 10.1016/j.jml.2012.01.001.

Miller, George A., George A. Heise, and William Lichten (1951). "The intelligibility of speech as a function of the context of the test materials". In: *Journal of Experimental Psychology* 41.5, pp. 329–335.

Mishra, Ramesh K et al. (2012). "Spoken language-mediated anticipatory eye- movements are modulated by reading ability - Evidence from Indian low and high literates". In: *Journal of Eye Movement Research* 5.1, pp. 1–10. DOI: 10.16910/jemr.5.1.3.

Morton, John (1964). "The effects of context on the visual duration threshold for words". In: *British Journal of Psychology* 55.2, pp. 165–180. DOI: 10.1111/j.2044-8295.1964.tb02716.x.

Moulines, Eric and Francis Charpentier (1990). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones". In: *Speech Communication* 9.1990, pp. 453–467.

Müller, Jana Annina et al. (2019). "Effect of speech rate on neural tracking of speech". In: *Frontiers in Psychology* 10. DOI: 10.3389/fpsyg.2019.00449.

Musch, Jochen and Ulf-Dietrich Reips (2000). "A brief history of web experimenting". In: *Psychological Experiments on the Internet*, pp. 61–87. DOI: 10.1016/b978-012099980-4/50004-6.

Näätänen, Risto and Terence Picton (1987). "The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure". In: *Psychophysiology* 24.4, pp. 375–425. DOI: 10.1111/j.1469-8986.1987.tb00311.x.

Nejime, Yoshito and Brian C. J. Moore (1998). "Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss". In: *The Journal of the Acoustical Society of America* 103.1, pp. 572–576. DOI: 10.1121/1.421123.

Nieuwland, Mante S, Stephen Politzer-Ahles, et al. (2018). "Large-scale replication study reveals a limit on probabilistic prediction in language comprehension". In: *eLife* 7, pp. 1–24. DOI: 10.7554/elife.33468.

Nieuwland, Mante S. (2019). "Do 'early' brain responses reveal word form prediction during language comprehension? A critical review". In: *Neuroscience and Biobehavioral Reviews* 96, pp. 367–400. DOI: 10.1016/j.neubiorev.2018.11.019.

Nieuwland, Mante S., Dale J. Barr, et al. (2020). "Dissociable effects of prediction and integration during language comprehension: Evidence from a largescale study using brain potentials". In: *Philosophical Transactions of the Royal Society B: Biological Sciences*. DOI: 10.1098/rstb.2018.0522.

Norris, Dennis, James M. McQueen, and Anne Cutler (2016). "Prediction, Bayesian inference and feedback in speech recognition". In: *Language, Cognition and Neuroscience* 31.1, pp. 4–18. DOI: 10.1080/23273798.2015.1081703.

Obleser, Jonas (Dec. 2014). "Putting the Listening Brain in Context". In: *Language and Linguistics Compass* 8.12, pp. 646–658. DOI: 10.1111/lnc3.12098.

Obleser, Jonas and S. A. Kotz (2010). "Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network". In: *Cerebral Cortex* 20.3, pp. 633–640. DOI: 10.1093/cercor/bhp128.

Obleser, Jonas and Sonja A. Kotz (2011). "Multiple brain signatures of integration in the comprehension of degraded speech". In: *NeuroImage* 55.2, pp. 713–723. DOI: 10.1016/j.neuroimage.2010.12.020.

Obleser, Jonas, R. J. S. Wise, et al. (Feb. 28, 2007). "Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions". In: *Journal of Neuroscience* 27.9, pp. 2283–2289. DOI: 10.1523/jneurosci.4663-06.2007.

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

Peelle, Jonathan E (2013). "Cortical responses to degraded speech are modulated by linguistic predictions". In: *Proceedings of Meetings on Acoustics ICA2013*. Vol. 19. 1. Acoustical Society of America, p. 060108.

Peelle, Jonathan E. (2018). "Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior". In: *Ear and Hearing* 39.2, pp. 204–214. DOI: 10.1097/AUD.0000000000000494.

Peelle, Jonathan E. and Arthur Wingfield (2005). "Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech". In:

Journal of Experimental Psychology: Human Perception and Performance 31.6, pp. 1315–1330. DOI: 10.1037/0096-1523.31.6.1315.

Peirce, Jonathan et al. (2019). "PsychoPy2: Experiments in behavior made easy". In: *Behavior Research Methods* 51.1, pp. 195–203. DOI: 10.3758/s13428-018-01193-y.

Pickering, Martin J. and Chiara Gambi (Oct. 2018). "Predicting while comprehending language: A theory and review." In: *Psychological Bulletin* 144.10, pp. 1002–1044. DOI: 10.1037/bul0000158.

Prolific Academic (n.d.). https://www.prolific.co. Accessed: 2020-02-02.

Pusse, Florian, Asad Sayeed, and Vera Demberg (2016). "LingoTurk: Managing crowdsourced tasks for psycholinguistics". In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pp. 57–61. DOI: 10.18653/v1/n16-3012.

Reips, Ulf-Dietrich (2021). "Web-based research in psychology". In: *Zeitschrift für Psychologie.*

Richards, Shane A., Mark J. Whittingham, and Philip A. Stephens (Aug. 2011). "Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework". In: *Behavioral Ecology and Sociobiology* 65.1, pp. 77–89. DOI: 10.1007/s00265-010-1035-8.

Rodero, Emma (2016). "Influence of speech rate and information density on recognition: The moderate dynamic mechanism". In: *Media Psychology* 19.2, pp. 224–242. DOI: 10.1080/15213269.2014.1002942.

Rönnberg, Jerker et al. (2013). "The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances". In: *Frontiers in Systems Neuroscience* 7.31, pp. 1–17. DOI: 10.3389/fnsys.2013.00031.

Rosen, Stuart, Andrew Faulkner, and Lucy Wilkinson (Dec. 1999). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants". In: *The Journal of the Acoustical Society of America* 106.6, pp. 3629–3636. DOI: 10.1121/1.428215.

Samuel, Arthur G. and Tanya Kraljic (Aug. 2009). "Perceptual learning for speech". In: *Attention, Perception, & Psychophysics* 71.6, pp. 1207–1218. DOI: 10.3758/app.71.6.1207.

Sanders, Lisa D. and Lori B. Astheimer (2008). "Temporally selective attention modulatesn early perceptual processing: Event-related potential evidence". In: *Perception and Psychophysics* 70.4, pp. 732–742. DOI: 10.3758/PP.70.4.732.

Sanderson, Stephen K. and Wesley W. Roberts (2008). "The evolutionary forms of the religious life: A cross-cultural, quantitative analysis". In: *American Anthropologist* 110.4, pp. 454–466. DOI: 10.1111/j.1548-1433.2008.00078.x.

Schlueter, Anne et al. (2014). "Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms". In: *The Journal of the Acoustical Society of America* 135.3, pp. 1541–1555. DOI: 10.1121/1.4863654.

Seow, Tricia X.F. and Tobias U. Hauser (2022). "Reliability of web-based affective auditory stimulus presentation". In: *Behavior Research Methods* 54.1, pp. 378–392. DOI: 10.3758/s13428-021-01643-0.

Seth, Anil K. (Nov. 2013). "Interoceptive inference, emotion, and the embodied self". In: *Trends in Cognitive Sciences* 17.11, pp. 565–573. DOI: 10.1016/j.tics.2013.09.007.

Bibliography

Shannon, R. V., F.-G. Zeng, et al. (Oct. 13, 1995). "Speech Recognition with Primarily Temporal Cues". In: *Science* 270.5234, pp. 303–304. DOI: 10.1126/science.270.5234.303.

Shannon, Robert, Qian-Jie Fu, and John Galvin Iii (Apr. 1, 2004). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation". In: *Acta Oto-Laryngologica* 124.0, pp. 50–54. DOI: 10.1080/03655230410017562.

Sharit, Joseph et al. (2003). "Effects of age, speech rate, and environmental support in using telephone voice menu systems". In: *Human Factors* 45.2, pp. 234–251. DOI: 10.1518/hfes.45.2.234.27245.

Sheldon, Signy, M. Kathleen Pichora-Fuller, and Bruce A. Schneider (Oct. 2008a). "Effect of age, presentation method, and learning on identification of noise-vocoded words". In: *The Journal of the Acoustical Society of America* 123.1, pp. 476–488. DOI: 10.1121/1.2805676.

Sheldon, Signy, M. Kathleen Pichora-Fuller, and Bruce A. Schneider (Feb. 2008b). "Priming and sentence context support listening to noise-vocoded speech by younger and older adults". In: *The Journal of the Acoustical Society of America* 123.1, pp. 489–499. DOI: 10.1121/1.2783762.

Simantiraki, Olympia and Martin Cooke (2020). "Exploring listeners' speech rate preferences". In: *INTERSPEECH*, pp. 1346–1350. DOI: 10.21437/Interspeech.2020-1832.

Sommers, Mitchell S., Lynne C. Nygaard, and David B. Pisoni (Sept. 1994). "Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude". In: *The Journal of the Acoustical Society of America* 96.3, pp. 1314–1324. DOI: 10.1121/1.411453.

Stadler, Waltraud et al. (2012). "Repetitive TMS Suggests a Role of the Human Dorsal Premotor Cortex in Action Prediction". In: *Frontiers in Human Neuroscience* 6. DOI: 10.3389/fnhum.2012.00020.

Staub, Adrian (2011). "The effect of lexical predictability on distributions of eye fixation durations". In: *Psychonomic Bulletin and Review* 18.2, pp. 371–376. DOI: 10.3758/s13423-010-0046-9.

Staub, Adrian (Aug. 2015). "The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation". In: *Language and Linguistics Compass* 9.8, pp. 311–327. DOI: 10.1111/lnc3.12151.

Stilp, Christian (2020). "Acoustic context effects in speech perception". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 11.1, pp. 1–18. DOI: 10.1002/wcs.1517.

Strauß, Antje, Sonja A. Kotz, and Jonas Obleser (Aug. 2013). "Narrowed Expectancies under Degraded Speech: Revisiting the N400". In: *Journal of Cognitive Neuroscience* 25.8, pp. 1383–1395. DOI: 10.1162/jocn_a_00389.

Su, Qiaotong et al. (2016). "Effects of within-talker variability on speech intelligibility in Mandarin-speaking adult and pediatric cochlear implant patients". In: *Trends in Hearing* 20, pp. 1–16. DOI: 10.1177/2331216516654022.

Th. Gries, Stefan (2015). "The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models". In: *Corpora* 10.1, pp. 95–125. DOI: 10.3366/cor.2015.0068.

"The vocoder" (1940). In: *Nature* 145.3665, p. 157. DOI: 10.1038/145157a0.

Bibliography

Tóth, Brigitta et al. (2020). "The effects of speech processing units on auditory stream segregation and selective attention in a multi-talker (cocktail party) situation". In: *Cortex* 130, pp. 387–400. DOI: 10.1016/j.cortex.2020.06.007.

Tun, Patricia A. (1998). "Fast noisy speech: Age differences in processing rapid speech with background noise". In: *Psychology and Aging* 13.3, pp. 424–434. DOI: 10.1037/0882-7974.13.3.424.

Vaden, Kenneth I. et al. (2013). "The cingulo-opercular network provides word-recognition benefit". In: *Journal of Neuroscience* 33.48, pp. 18979–18986. DOI: 10.1523/JNEUROSCI.1417-13.2013.

Vagharchakian, Laurianne et al. (2012). "A temporal bottleneck in the language comprehension network". In: *Journal of Neuroscience* 32.26, pp. 9089–9102. DOI: 10.1523/JNEUROSCI.5685-11.2012.

Van Os, Marjolein, Jutta Kray, and Vera Demberg (2021). "Recognition of minipairs in (un)predictive sentence contexts in two types of noise". In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 43.43, pp. 2943–2949.

Vasishth, Shravan et al. (2022). "Hypothetical repeated sampling and the t-test". In: *Linear mixed models in linguistics and psychology: A comprehensive introduction (DRAFT)*. Chap. 2.

Verhelst, Werner and Marc Roelands (1993). "An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech". In: *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 2. IEEE, pp. 554–557. DOI: 10.1109/icassp.1993.319366.

Welch, Norma and John H. Krantz (1996). "The World-Wide Web as a medium for psychoacoustical demonstrations and experiments: Experience and results". In: *Behavior Research Methods, Instruments, and Computers* 28.2, pp. 192–196. DOI: 10.3758/bf03204764.

Wild, Conor J. et al. (2012). "Effortful listening: The processing of degraded speech depends critically on attention". In: *Journal of Neuroscience* 32.40, pp. 14010–14021. DOI: 10.1523/JNEUROSCI.1528-12.2012.

Winn, Matthew B. and Katherine H. Teece (2021). "Slower speaking rate reduces listening effort among listeners with cochlear implants". In: *Ear and Hearing* 42.3, p. 584. DOI: 10.1097/aud.0000000000000958.

Wlotko, Edward W. and Kara D. Federmeier (Apr. 3, 2012). "Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension". In: *Psychophysiology* 49.6, pp. 770–785. DOI: 10.1111/j.1469-8986.2012.01366.x.

Wlotko, Edward W. and Kara D. Federmeier (2015). "Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading". In: *Cortex* 68, pp. 20–32. DOI: 10.1016/j.cortex.2015.03.014.

Woods, Kevin J.P. et al. (2017). "Headphone screening to facilitate web-based auditory experiments". In: *Attention, Perception, and Psychophysics* 79.7, pp. 2064–2072. DOI: 10.3758/s13414-017-1361-2.

Xiang, Ming and Gina Kuperberg (2015). "Reversing expectations during discourse comprehension". In: *Language, Cognition and Neuroscience* 30.6, pp. 648–672. DOI: 10.1080/23273798.2014.995679.