Degraded speech comprehension: The influence of predictability, attention and speech rate

Pratik Bhandari

Department of Psychology

Saarland University

A thesis submitted for the degree of

Doctor of Philosophy

January 2023

-

Acknowledgements

Oh this is hard! PhD and the journey until here. How could I have done it without the wonderful people in my life?

Jutta and Vera have been amazing mentors. The German word "doktormutter" describes what they have been to me, academically, better than the English word "supervisor" does. Words don't do justice: It is rare to find a supervisor who understands your

The interdisciplinary discussions in Vera's lab have pushed me beyond my comfort zone to learn about different dimensions of human and machine language. It has been an honor and a pleasure to be her student, to have worked with her. I couldn't have learnt

Pratik Bhandari
Universität des Saarlandes,
Saarbrücken
10 March 2021

Abstract

Some abstract here

Contents

List of Figures

List of Tables

List of Abbreviations

HP . . . . High predictability

MP . . . . Medium predictability

LP . . . . Low predictability

ch . . . . channels

DE . . . . German

EN . . . . English

MEG . . . . . Magnetoencephalography

ERP . . . . . . Event Related Potential

SPIN . . . . . Speech In Noise

SNR . . . . . . Signal to Noise Ratio

PSOLA . . . . Pitch synchronous overlap add technique

SD . . . . Standard deviation

1
Introduction

One of the features that distinguishes us, humans, from other species is our ability to communicate using verbal language (Hauser et al., 2002; Lieberman, 2013; Pinker & Jackendoff, 2005). We speak. We listen. We understand. This seemingly straightforward path of communication goes through plenty of hindrances. One of them is an adverse listening condition caused by background noise and speech distortion (e.g., Chen & Loizou, 2011; Fontan et al., 2015). Human comprehenders rely on top-down predictive and bottom-up auditory processes to understand spoken language. Language comprehension in adverse listening conditions is aptly described by the noisy channel model of communication (Gibson et al., 2013, 2019; Levy, 2008; C. E. Shannon, 1948) schematically represented in Figure 1.1 below.

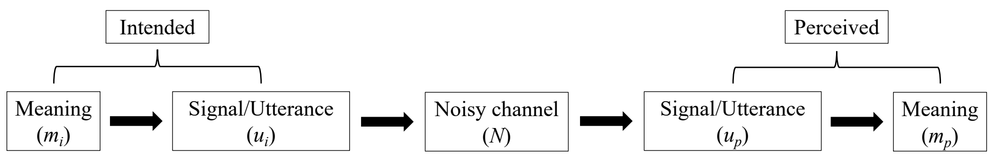

Figure 1.1: Schematic representation of the noisy channel model of communication

The speaker produces an utterance u_i with a meaning m_i that she intends to send. The utterance is encoded into a signal and sent through a channel of transmission. During transmission, some external noise disrupts the signal. The receiver (e.g., a listener) perceives the signal as u_p and decodes it to recover the meaning as m_p. The human language comprehension system is assumed to be engaged in optimal Bayesian decoding that uses all the sources of information (e.g., prior semantic knowledge, context information, world knowledge, etc.) and infers the intended meaning from the perceived utterance that it receives from a noisy channel of communication (Gibson et al., 2013; Levy, 2008; cf. Markman & Otto, 2011).

For successful communication to occur, the message recovered by the listener must be approximately equal to the message intended to be sent by the speaker. Let's take an example. X sees a spherical object flying towards Y. So, she intends to warn him about it: a "spherical object which is played by two teams of 11 players each in a big playground" is about to hit Y. To convey this message, X utters BALL. Due to external noise, X's (i.e., the speaker's) utterance is distorted, so Y (i.e., the listener) perceives the utterance as HALL. The listener then interprets that the speaker's message is intended to point him to the "building where lectures take place in their university" they were trying to find. (In this case of unsuccessful communication, or due to the listener wrongly identifying the speaker's intended message, Y gets hit by a ball.)

We assume that the goal of a listener is to identify the message m_i that is most likely from the perceived utterance u_p, taking into account the external noise (N) and the prior likelihood of the speaker uttering u_i. This can be expressed formally as:

$$\hat{m_p} = \underset{m_p}{\operatorname{argmax}} P(m_p, u_p, N, u_i, m_i) \tag{1.1}$$

This sequence of events from the intended message m_i to the perceived message m_p can be graphically represented in a Bayesian network (Bruineberg et al., 2021; Darwiche, 2010; Pearl, 1985) in Figure 1.2 (cf. Figure 1.1).

Figure 1.2 models the dependencies among the events, which shows that the external noise and the speaker's utterance are *independent*; however, the listener's

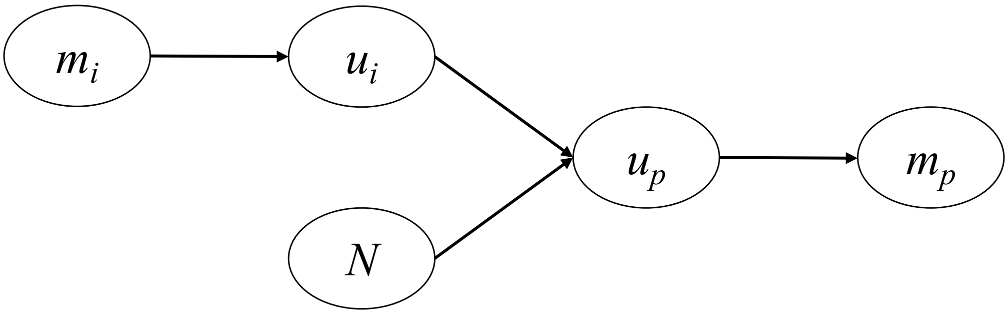

Figure 1.2: Bayesian network representation of the noisy channel model of communication

perception of the uttered message is also dependent on the noise. The communication in the noisy channel, represented as a Bayesian network, can now be expressed as:

$$\hat{m_p} = \underset{m_p}{\operatorname{argmax}} P(m_p|u_p) * P(u_p|u_i, N) * P(u_i|m_i) * P(m_i) \tag{1.2}$$

Equation (1.2) can be interpreted easily from its corresponding representation in Figure 1.2.

It shows:

- $P(m_p|u_p)$: the probability of inferring a meaning m_p (e.g., a building where lectures take place) from a perceived utterance u_p (e.g., hall)
- $P(u_p|u_i, N)$: bottom-up auditory information, i.e., the probability of the listener hearing a particular utterance u_p (e.g., hall) given that the speaker has uttered an utterance u_i (e.g., ball) in the noisy channel N (e.g., background noise, signal distortion, etc.)
- $P(u_i|m_i)P(m_i)$: prior information (e.g., top-down semantic knowledge, information about the speaker, etc.), i.e., the probability of a speaker uttering u_i with an intended message m_i with the probability that the intended message is m_i

The channel of transmission can become noisy due to factors like background noise present in a conversation, a poor signal transmission of a telephone call that distorts the speaker's speech, hearing loss of a listener, hearing aid or cochlear implant worn by a listener, and so on. To understand speech in such a noisy channel of communication, a listener puts different weights on the distorted bottom-up auditory input $P(u_p|u_i, N)$ vs her prior information $P(u_i|m_i)P(m_i)$ (e.g., knowledge about the speaker, world, context information, etc.).

Studies in clean speech and reading comprehension have demonstrated that listeners and readers use prior knowledge and context information to form semantic predictions about the linguistic events yet to be encountered.

Let's take the following sentence, for example:

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

Most readers would expect the final word to be *kite* in this sentence (DeLong et al., 2005; cf. Nieuwland et al., 2020). Here, the words up to the final word of the sentence provide a context: A reader can utilize their knowledge about what a *boy* would ideally do *outside* on a *breezy* day. It leads the reader to predict that the sentence continuation is most likely *kite* and not an improbable word like *rocket*. Similar results are observed in the auditory domain as well. Listeners use context information from what they have heard and form predictions about an upcoming word (e.g., Altmann & Kamide, 2007; Ankener, 2019). That is, human language comprehension is predictive in nature, such that listeners engage in predictive language processing (Section 2.2.1, Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). In a noisy channel, listeners' engagement in predictive processing is influenced by the noise in the signal (Obleser et al., 2007; Sheldon et al., 2008a).

Based on the theoretical accounts of the noisy channel model of communication and the predictive language processing (Christiansen & Chater, 2015; Ferreira & Lowder, 2016; Friston, Parr, et al., 2020; Hale, 2001; Levy, 2008; McClelland & Elman, 1986; Norris et al., 2016; Pickering & Gambi, 2018), this thesis investigates the interaction between top-down predictive and bottom-up auditory processes. We examine how top-down predictive processes facilitate language comprehension in a noisy channel created by acoustic degradation of speech. We examine the nature of predictability effects in a noisy channel. We investigate what the optimal level of noise in the signal is, i.e., the optimal level of speech degradation, for the effect of top-down predictive processes to be most efficient or facilitatory for language comprehension. By manipulating different factors of top-down as well as bottom-up processes (e.g., speech rates, attention allocation to different parts of the speech stream), we examine their role in aiding (or interfering) the comprehension of degraded speech. While doing so, we address the following research goals.

1.1 Research goals

(1) **Replication of the predictability effect in a noisy channel**

Almost all the disciplines of cognitive science — anthropology, computer science, linguistics, neuroscience, and psychology — are suffering the so-called replication crisis (Aarts et al., 2015; Cockburn et al., 2020; Ebersole et al., 2016; Minocher et al., 2021; Sanderson & Roberts, 2008). The results of an experiment do not hold up consistently when another group of researchers conduct it again. For example, in a multi-lab collaborative study, Nieuwland et al. (2020) did not find the same N400 effect in English articles (a/an) that DeLong et al. (2005) had reported 15 years earlier. The first goal of this thesis is to test if we can replicate the facilitatory effect of semantic predictability in language comprehension in a noisy channel (e.g., Obleser et al., 2007; Sheldon et al., 2008a). Replication of the predictability effect in comprehension of degraded speech will help gather evidence in favour of this *effect of interest*. It will also provide a reliable foundation to test if (and how) other factors (e.g., speed of information processing) influence and interact with the facilitatory effect of predictability.

(2) **Nature of prediction**

There are at least two schools of thought which argue that prediction is either all-or-nothing or probabilistic (Coltheart, 2004; Kuperberg & Jaeger, 2016; Luke & Christianson, 2016). These debates generally centre around reading comprehension and clean speech comprehension. The discussion about the nature of prediction in a noisy channel like degraded speech is sparse. Specifically, in degraded speech comprehension, only one study has empirically investigated the theoretical postulation that prediction is restricted only to highly predictable sentence endings (StrauSS et al., 2013). Therefore, the second goal of this thesis is to examine the nature of the predictability effect. With carefully designed experiments and materials, this thesis aims to test the distinction between all-or-nothing and probabilistic predictions in degraded speech comprehension.

(3) **Boundary conditions of predictive language processing**

Several authors claim that predictive processing is the fundamental nature of human cognition and, thus, by definition, also of language processing (A. Clark, 2013; Friston, Parr, et al., 2020; Friston, Sajid, et al., 2020; Kuperberg, 2021; Lupyan & Clark, 2015). At the same time, an increasing number of studies are showing boundary conditions and prerequisite conditions for predictive language processing (K. D. Federmeier et al., 2010; Huettig & Guerra, 2019; Huettig & Mani, 2016; Mishra et al., 2012). For example, attention can reverse the effect of predictability in non-speech auditory perception (cf. Kok et al., 2012). In a noisy channel (i.e., degraded speech), attention to a part or parts of a speech stream can modulate, or that can limit the predictability effect as different parts of the speech stream contain different linguistic units; each linguist unit (e.g., each word in a sentence) carries its own meaning that serves the entire message (e.g., the entire sentence). Therefore, the third goal of this thesis is to examine the role of auditory attention that can act as a prerequisite for semantic predictions or limit the automaticity of predictive processing in degraded speech comprehension.

This thesis aims to test whether attention to different parts of degraded speech stream aids or hampers facilitatory effects of top-down predictions.

(4) **Adaptation to degraded speech**

Despite the difficulty in understanding speech in a noisy channel, listeners rapidly adapt to degraded speech (Rosen et al., 1999). When the properties of speech vary in the dimension of both acoustic-phonetic cues as well as lexical-semantic cues, adaptation can be difficult. The fourth goal of this thesis is to examine if listeners adapt to degraded speech when both degradation level and predictability of speech are varied. We test if an adaptation to the bottom-up perceptual property of speech is influenced by its top-down semantic property.

(5) **Speed of information processing**

Unlike the visual scene that opens in the spatial dimension, speech signal flows in the temporal dimension. This challenges the listeners to process information at different speeds and timescales; more time is available to process the information in slow speech, while less time is available for fast speech (Lerner et al., 2014). Listeners build up the meaning representation as they process the speech to predict upcoming linguistic units. The fifth goal of this thesis is to examine whether a change in information flow, i.e., speech rate, affects the facilitatory effect of predictability. We test if an increase or decrease in speech rate impedes the intelligibility of speech over a noisy channel and whether it impedes or further aids the predictability effect in the noisy channel.

(6) **Metric for measurement of language comprehension**

Different researchers have used different measurement metrics in the study of speech perception and language comprehension (Amichetti et al., 2018; Obleser et al., 2007; Peelle, 2013; Sheldon et al., 2008a). The measurement is inconsistent across studies which becomes a problem, especially when the effect of context in comprehension is under discussion: cross-study comparison does not give a clear picture of the predictability effect in this case. Therefore, the sixth goal of this thesis is to establish and consistently use a sensitive metric for the measurement of language comprehension that takes into account whether participants (in)correctly use the context-evoking word in a sentence.

Studies addressing the research goals outlined above will primarily contribute to elaborating and developing the existing theories of predictive language processing and furthering the understanding of spoken language comprehension in a noisy channel, especially degraded speech comprehension. Below we present the contributions of the research presented in this thesis.

1.2 Research contributions

The research reported in this thesis examines theoretical questions of predictive language processing and its boundary conditions when spoken language comprehension takes place through a noisy channel. It contributes to the studies of speech perception, language comprehension, predictive coding, language science, audiology, psycholinguistics, psychology, and, broadly, cognitive science. In an applied setup, this informs translational/clinical researchers about language comprehension in cochlear implantees.

- **Graded effect of predictability**

  We replicate the previous finding of the predictability effect showing that predictability facilitates comprehension of degraded speech at moderate levels of degradation (e.g., Obleser et al., 2007). Additionally, in the current debate of all-or-nothing vs graded prediction, our findings indicate that prediction across the noisy channel of degraded speech is graded in nature rather than being restricted to a narrow space of highly predictable sentence endings. Goals 1 and 2 correspond to this research contribution brought about by the experiments described in Chapters 5 and 6.

- **Attention in predictive language processing**

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

- **Absence of perceptual adaptation**
  We show that listeners do not adapt to degraded speech when lexical-semantic cues are taken into consideration. This is in contrast with the previous findings of speech perception experiments, some of which disregard the trial-by-trial variation in sentence context (e.g., Davis et al., 2005; Erb et al., 2013). When listeners are engaged 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 do not show any adaptation effect; every trial is effectively a novel trial for them. Goal 4 corresponds to this research contribution brought about by the experiments described mainly in Chapters 6 and 7.

- **Change in information flow and its effect on top-down prediction**
  We show that different rates of information flow — increase or decrease in the rate of speech — have different effects on language comprehension. Intelligibility of speech decreases with both increase and decrease of speech rate. However, the increase in speech rate is particularly detrimental to comprehension of degraded speech as it increases the difficulty in processing sentences with less predictable endings. This is one of the few studies highlighting the role of speed of flow of information in the contextual facilitation of degraded speech. Goal 5 corresponds to this research contribution brought about by the experiment described in Chapter 7.

- **A metric of language comprehension**

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

1.3 Overview of the thesis

The central theme of this thesis is the study of predictive processing in language comprehension across a noisy channel. On the grounds of predictive language processing and the noisy channel model of communication, we investigate how and to what extent listeners use context information while listening to degraded speech. We replicate and extend prior findings, which claim that predictability facilitates language comprehension at moderate levels of speech degradation. Furthermore, the boundary conditions of predictive processing are tested, examining the effect of different rates of information flow in the predictability effect. We test for the presence of perceptual adaptation and find evidence against the learning effect and adaptation to degraded speech.

Chapter 2 provides a background on the rest of the chapters. It provides an overview of degraded speech comprehension and predictive language processing. The current status of the debate on these topics is also presented.

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

Chapter 4 describes the statistical tests employed for data analyses. Binomial logistic mixed effects modelling is performed on the data from all the experiments. This chapter provides a background on this statistical procedure and how it is operated on the statistical software `R`.

Chapter 5 presents two experiments that address the first and the third research goal. These experiments are conducted to examine the predictability effect in degraded speech comprehension and the role of auditory attention. Participants in both experiments are presented with the speech degraded at different levels of degradation and sentences of different levels of predictability. Participants in Experiment 1A are asked to type in only the final word of a sentence; this did not bind their attention to the sentence context. In contrast, the participants in Experiment 1B are asked to type in the entire sentence that they heard, which required them to attend to the sentence context as well. We replicate the previously reported predictability effects in the noisy channel only when participants attended to the entire sentence, including the context. We show that top-down predictions cannot be generated at moderate levels of degradation when insufficient attention is given to context. We discuss the limitation in the existing theories of predictive language processing, which commit to the automaticity of prediction. We show the importance of *attention* in language comprehension. We end this chapter with the note that the measurement of language comprehension can be further refined and the nature of the predictability effect tested.

Chapter 6 addresses the first, the second, the fourth, and the sixth research goals. The predictability effect partially replicated in Chapter 5 is further examined in this chapter. We use a refined metric of measurement of language comprehension that takes into consideration whether listeners correctly identified the context. We observe predictability effects at a moderate level of speech degradation, thereby consistently replicating the facilitatory effect of predictability. We find the predictability effects to be graded in nature and discuss it in the light of existing theories of predictive processing. We also show that regardless of the certainty about the next-trial degradation level, listeners do not adapt to degraded speech when its lexical-semantic property varies every trial. At the end of this chapter, we note the intrinsic difficulty of processing degraded speech and open the question that the predictability effects could be further enhanced (or limited) with more (or less) time available to process the degraded speech.

Chapter 7 addresses the questions opened in Chapter 6. In two experiments, it addresses the fourth, the fifth, and the sixth research goals. We use the same metric of measurement of language comprehension as Chapter 6, which takes into account listeners' correct identification of the context. Listeners are presented with a moderately degraded speech at which the predictability effect is observed in Chapter 6. In Experiment 7A, the moderately degraded speech is presented at normal and fast speech rates. In Experiment 7B, the speech rates are normal and slow. For fast speech, however, both intelligibility and the predictability effect are reduced, driven by the difficulty in processing words that are less predictable from the context. Although more time is available to process the context of the degraded speech at a slow speech rate, there is no increase in the facilitatory effect of predictability with a reduced speech rate; instead, intelligibility is reduced in slow speech compared to normal speech. This chapter discusses on the limitations of predictive processing driven by the constraints in cognitive resources.

Chapter 8 summarizes the findings of all the studies. It presents the theoretical implications of this thesis and the future direction it points to. The general limitations of the studies are briefly discussed.

Chapter 9 presents ethical approval that was obtained to run the experiments on human subjects. The source that provided funding to conduct the research presented in this thesis is disclosed.

1.4 Dissemination of research findings

Some of the findings reported in this thesis are presented and published elsewhere to disseminate scientific findings to a broader audience. The presentations and publications that report on parts of the research described in this thesis are outlined below.

1. *Introduction*

- Bhandari, P., Demberg, V., & Kray, J. (2022). The effect of speech rate on contextual facilitation of degraded speech comprehension. *Architectures and Mechanisms for Language Processing*, 2022-09-07–2022-09-09.

- Bhandari, P., Demberg, V., & Kray, J. (2022). Predictability effects in degraded speech comprehension are reduced as function of attention. *Architectures and Mechanisms for Language Processing*, 2022-09-07–2022-09-09.

- Bhandari, P., Demberg, V., & Kray, J. (*under review*) Speaking fast and slow: How speech rate affects contextual facilitation in degraded speech comprehension.

- Bhandari, P., Demberg, V., & Kray, J. (2022). Predictability effects in degraded speech comprehension are reduced as a function of attention. *Language and Cognition*, 1-18. doi:10.1017/langcog.2022.16

- Bhandari, P., Demberg, V., & Kray, J. (2022). The effect of speech rate in comprehension of degraded speech. *International Max Planck Research School (IMPRS) Conference*, 2022-06-01–2022-06-03.

- Bhandari, P., Demberg, V., & Kray, J. (2022). Predictability effects are reduced as a function of attention. *Annual Convention of American Psychological Association*, 2022-05-25–2022-05-28.

- Bhandari, P., Demberg, V., & Kray, J. (2021). Semantic predictability facilitates comprehension of degraded speech in a graded manner. *Frontiers in Psychology*, 12:714485. doi: 10.3389/fpsyg.2021.714485

- 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*, 2021-10-05–2021-10-08.

1. Introduction

- Bhandari, P., Demberg, V., & Kray, J. (2021). Predictability facilitates comprehension of degraded speech in a graded manner. *Annual Meeting of Cognitive Neuroscience Society*, 2021-03-13–2021-03-16.

2
Background

In the previous chapter, we outlined the theoretical background and the research goals of the studies in this dissertation. We stated that the central theme of this thesis is to investigate the interaction between top-down predictive and bottom-up auditory processes in language comprehension. Grounding on the noisy channel model of communication and predictive language processing, the studies in this thesis manipulate the auditory processes $P(u_p|u_i, N)$, and prior information $P(u_i, m_i)$ in the form of semantic context available in a sentence. In this chapter, we provide background on the noisy channel created and used to introduce variations in bottom-up processing in the studies presented in this thesis. We also elaborate on the predictive language processing in the noisy channel and the evidence of its limits and nature. Understanding these fundamental concepts of top-down and bottom-up processes is essential for the chapters that follow; these concepts are briefly reiterated in the following chapters wherever relevant. Additionally, this chapter outlines the gaps in previous research that this thesis fills in.

2.1 Speech distortion and degradation

Most of the existing frameworks of spoken language comprehension are inspired by the experiments conducted with clean speech, the condition of "artificial normalcy"

(Mattys et al., 2012). However, spoken language communication generally occurs outside the artificial normalcy, alongside different sources of noise and disruption. Probabilistic models of language comprehension, like the noisy channel model of communication (Gibson et al., 2013, 2019; Levy, 2008; C. E. Shannon, 1948) in Figures 1.1 and 1.2 show that the speech signal uttered by the speaker gets disrupted and distorted due to the noise ($u_i \rightarrow u_p \leftarrow N$) . Distortion can occur at these three points or sources: encoding, transmission, and decoding (Mattys et al., 2012). Speech can be distorted while encoding the utterance u_i due to the variability in speakers' production, like accented or slow and fast speech. Distortion can arise while decoding the signal u_p due to listener-related factors, like hearing loss or auditory processing disorder. It can also result from an external noise that appears during the transmission, like ambient noise or poor transmission medium (e.g., distortion in the telephone line). These different sources of distortion make a listening condition adverse by affecting the time and frequency-related properties/cues of the speech signal, i.e., temporal envelope cues and spectral details of speech, respectively. The temporal envelope cues are the slow variations in the amplitude of the speech signal over time (Moon et al., 2014; Moon & Hong, 2014), while the spectral details are the frequency-specific information about the speech. The temporal envelope cues reflect the prosodic information of the speech and are used in lexical-semantic and syntactic processing (Greenberg, 1996; Schneider & Pichora-Fuller, 2001; Sheldon et al., 2008b). The spectral details provide information about the sound production reflecting the vocal tract's resonant properties, speech signal frequency range, energy distribution across frequency bands, etc. (Roberts et al., 2011; Shannon et al., 1995; Shannon et al., 2004).

In an experimental setup, a noisy channel can be created artificially by digital signal processing (see Section 3.1.2) to investigate the response of the speech perception system to distorted speech and to study language comprehension in an adverse listening condition. For example, signal compression or expansion acts upon the temporal property of the speech and makes it fast or slow (i.e., change its speed), and an optimal level of speech expansion/compression does not distort the spectral property of speech (see Section 3.1.2). In addition to speech compression and expansion in Chapter 7, throughout the studies in this thesis, we implement noise-vocoding to manipulate the spectral property of speech and create a noisy channel of communication.

Noise-vocoding removes the spectral details of the speech signal in a controlled manner, only leaving intact its temporal and periodicity cues (see Section 3.1.2). This method of speech degradation was initially developed as a means to reduce the information in speech signals to be transmitted through the telephone line (Clendeninn, 1940; Dudley, 1939). Shannon and colleagues later modified and used this technique as an analogue to cochlear implants such that the number of channels used in a cochlear implant is similar to the number of noise-vocoding channels in terms of their speech output and intelligibility (Loizou et al., 1999; Shannon et al., 1995; Shannon et al., 2004; cf. Orena & Colby, 2021). Therefore, in addition to being a method of speech distortion to parametrically vary and control the quality of speech signals in a graded manner, noise-vocoding is also a method of distortion that is used to understand the speech perception and language comprehension in cochlear implantees (e.g., Patro & Mendel, 2020; Shannon et al., 2004; M. Winn, 2016)

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 the frequency-specific information available in the degraded speech, and the higher the intelligibility compared to the speech that is degraded with a lesser number of noise-vocoding channels. For example, listeners are shown to rate 8-channel noise-vocoded speech to be more intelligible and less effortful than 2 channels noise-vocoded speech (e.g., Obleser & Kotz, 2011; Sohoglu et al., 2012). In our studies, we create a noisy channel with different degradation levels and intelligibility by noise-vocoding the speech signal through 1, 4, 6 and 8 channels. The details of the artificial distortions are described in Chapter 3.

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

2.2 Prediction and comprehension of degraded speech

In addition to the quality of speech signals, listeners rely on context information and form top-down predictions to understand speech in adverse listening conditions. Below, we first review the role of predictions in language comprehension in general, and then we discuss the role of top-down predictive processes in comprehension of degraded speech in particular.

2.2.1 Predictive language processing

Research from various domains of cognitive (neuro)science, like emotion, vision, odour, and proprioception (the sensation of one's body position and movement, Tuthill & Azim, 2018), has shown that perception and cognition can be described under the framework of predictive processing; they primarily operate by predicting upcoming events (A. Clark, 2013; Marques et al., 2018; Seth, 2013; Stadler et al., 2012; cf. Bowers & Davis, 2012; Jones & Love, 2011; Pierce & Ollason, 1987). Despite a long-standing scepticism and doubt about the usefulness of prediction in language processing (Forster, 1981; Jackendoff, 2002; Van Petten & Luka, 2012), human language comprehension too has been claimed to be predictive in nature from as early as the mid-twentieth century (e.g., McCullough, 1958; Miller et al., 1951; Morton, 1964) which in recent days has received overwhelming support from computational linguistics, psycholinguistics and cognitive neuroscience of language (e.g., DeLong et al., 2005; Demberg et al., 2013; Heyselaar et al., 2021; Lupyan & Clark, 2015; Pickering & Gambi, 2018). Empirical evidence from several studies suggests that readers and listeners predict upcoming words in a sentence when the words are predictable from the preceding context (Kuperberg & Jaeger, 2016; Nieuwland, 2019; for reviews, Staub, 2015). For instance, predictable words are skipped and read faster than the words that are less predictable from the context (Ehrlich & Rayner, 1981; Frisson et al., 2005; Staub, 2011). Applying the visual world paradigm, studies have demonstrated that individuals show anticipatory eye movements towards a picture of an object (e.g., *cake*) that is predictable from the preceding sentence context (e.g., *The boy will eat a...*) even before hearing the final target word (Altmann & Kamide, 1999; Ankener et al., 2018; Kamide et al., 2003). Similar results have been observed in a virtual world setup with naturalistic scenes (e.g., Heyselaar et al., 2021). The sentence-final word in a highly constraining sentence (e.g., *"She dribbles a ball."*) elicits a smaller N400 amplitude[2] than a less constraining sentence (e.g., *"She buys a ball."*, K. Federmeier et al., 2007; Kutas & Hillyard, 1984). 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 the 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 that aids understanding of written and spoken language.

But, what is prediction?

The history of *prediction* in language science is rocky (Husband & Bovolenta, 2020). People have been sceptical that language processing is predictive in nature. Different people mean different things when they use the word prediction. As Kuperberg & Jaeger (2016) put it, *prediction* has become a loaded term; it is used alongside other similar terms like *integration* (K. D. Federmeier, 2007), *anticipation*, *expectation* (Van Petten & Luka, 2012), *preparedness* (Ferreira & Chantavarin, 2018), etc.

This thesis uses the word *prediction* in the following minimal sense. As a sentence unfolds, listeners encounter the context information in the sentence and form its meaning representation, i.e., an internal representation of the context. Before they hear the next word, i.e., before they encounter new bottom-up information, they generate an expectation[3] about the new word based on the meaning

[2] N400 is a negative-going EEG component that peaks around 400 ms post-stimulus and is considered a neural marker of context-based semantic unexpectedness (Kutas & Federmeier, 2011).

[3] Henceforth, we use the word expectation and prediction interchangeably.

representation of the context. They could form a prediction about only the semantic feature of the next word, or they could predict the exact word (meaning prediction vs word-form prediction).

In reading studies and clean speech comprehension, there are opposing views. One view is that the comprehenders predictively preactivate the upcoming linguistic unit solely based on the top-down information (i.e., predictive *preactivation*). In contrast, the opposing view is that the comprehenders wait for the bottom-up information to activate the representation (e.g., phonological and semantic representation) of the new information and its neighbours[4], then use the top-down information to select the best representation. To clarify it further, let's take the example sentence *(1)* presented in Chapter 1 on page 4: *The day was breezy so the boy went outside to fly a*_____. Upon listening to this context, the listeners can form a high degree of belief that the next word will be 'kite'. Before even hearing it, listeners preactivate the representation of "kite" in their mental lexicon. Alternatively, they could wait until they hear the auditory input "kite", which activates "kite" and its phonological (and semantic) neighbours in the mental lexicon, then use the top-down information to select the most likely word that completes the sentence. Either way, top-down processes facilitate comprehension.

While listening in an adverse condition, it is unlikely that a listener follows the latter strategy of waiting for the bottom-up input to activate the representations and then selecting the most likely one based on the top-down information (Kuperberg & Jaeger, 2016). When speech is distorted, it is difficult to form the context representation in the first place (cf. cue-based retrieval, Kaufeld, 2021; Martin, 2016). Once a listener has formed a meaning representation of the context, she cannot afford to again wait for the bottom-up input to activate phonological and semantic representations of upcoming words; the uncertainty about the bottom-up information is persistent (see the phoneme restoration effect (Warren, 1970), the McGurk effect (McGurk & MacDonald, 1976), and the Ganong effect (Ganong,

[4]The Neighborhood Activation Model of Luce & Pisoni (1998) proposes that an auditory input of a word activates its neighbourhood words, which can be similar acoustically. The neighbourhood density is supposed to depend on the word frequency as well.

1980) in speech perception). Thus, once the listener has formed a representation of the context, she uses this top-down information to predictively preactivate what the upcoming word can be. Such predictive preactivation can take different forms: it can be a probabilistic (graded) or deterministic (all-or-nothing) prediction. These differences in the nature of prediction are discussed below.

2.2.2 Facilitatory effect of predictability

We have discussed above that individuals make predictions about not-yet-encountered linguistic units based on available context information as a 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 an adverse listening condition, it has been shown that listeners rely more on top-down processes by narrowing down the predictions to smaller sets of semantic categories or words (Corps & Rabagliati, 2020; StrauSS et al., 2013). Obleser and colleagues (Obleser et al., 2007; Obleser & Kotz, 2010), for instance, used sentences of two levels of semantic predictability (high and low) and systematically degraded speech signals by passing them through various numbers of noise-vocoding channels ranging from 1 to 32 in a series of behavioural and neuroimaging studies (see also Hunter & Pisoni, 2018). They found that semantic predictability facilitated language comprehension only at moderate levels of speech degradation. That is, participants relied more on sentence context when the speech signal was degraded though *intelligible enough* than when it was not degraded or highly degraded. At such moderate levels of speech degradation, word recognition accuracy was found to be higher for words in high predictability sentences than the words in low predictability sentences (Obleser & Kotz, 2010). For the extremes, i.e., when the speech signal was highly degraded (making the speech almost entirely 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. (2008b) 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. A higher number of channels (i.e., more bottom-up information) was required in less constraining sentences to achieve the same level of accuracy as highly constraining sentences. They also concluded that word recognition is facilitated by predictability and sentence context when the speech is degraded. Taken together, these studies conclude that at moderate levels of degradation, participants rely more on the top-down predictions generated by a sentence context and less on the bottom-up perceptual processing of an unclear, less reliable, and degraded speech signal (Obleser, 2014). However, these studies are agnostic about the nature of prediction, i.e., if it is probabilistic or deterministic.

Nature of prediction

A debate in the literature on predictive language processing pertains to this question: Is prediction probabilistic, or is it an all-or-nothing phenomenon? For instance, the garden path phenomenon was explained as a parser's irreversible prediction about the sentence structure; if its predicted parsing fails (or if it turns out to be incorrect), then the parser reanalyzes the sentence and reformulates another prediction (e.g., Ferreira & Clifton Jr, 1986; see also Demberg et al., 2013; Slattery et al., 2013). In recent days, the support for the probabilistic nature of prediction comes, for example, from ERP studies that show an inverse and graded relationship between the magnitude of the N400 effect evoked by a word and its predictability measured by cloze probability[5] (e.g., DeLong et al., 2005; K. D. Federmeier et al., 2007), or *surprisal*[6] (Frank et al., 2015; cf. Brothers et al., 2015).

These discussions come from reading studies and spoken language comprehension in clear speech. Although a few frameworks of language processing speculate that language comprehension in adverse listening conditions can be predictive (e.g.,

[5]Cloze probability of a word is the proportion of participants who provide that word as the next word of a sentence, in an offline norming task, given the preceding words of the sentence (Staub et al., 2015; Taylor, 1953). Its value ranges from 0 to 1.

[6]Surprisal is a measure of the change in probability mass (or simply put, the change in expectation) as predictions are proven wrong with an encounter of new words in a sentence, discourse, etc. (Hale, 2001; Smith & Levy, 2008).

Lowder & Ferreira, 2016; Ryskin et al., 2018), so far, only StrauSS et al. (2013) have investigated the nature of prediction in degraded speech comprehension. They proposed an “expectancy searchlight model”, which suggests that listeners form *narrowed expectations* from a restricted semantic space only when the sentence endings are highly predictable. They rule out the graded nature of predictability. However, their approach to predictability was confounded by verb-noun association (discussed in Chapter 6). In contrast to their study, we systematically vary the predictability of the target word and examine the graded vs probabilistic nature of prediction in degraded speech comprehension. We argue that the facilitatory effect of predictability is graded in nature; it is not an all-or-nothing phenomenon focused solely on highly predictable sentence endings.

2.2.3 Limits of predictive language processing

It is important to note and acknowledge that the ubiquity and universality of predictive language processing have not gone unquestioned (Huettig & Mani, 2016). Apart from the debate on the nature of prediction, which we will come to later in this chapter, there is compelling evidence that questions the necessity of prediction in language comprehension. For example, Mishra et al. (2012) showed that literacy is a critical factor that limits listeners’ predictions about an upcoming word. In a visual world paradigm study, they found that individuals with lower literacy showed less anticipatory eye movements than those with higher literacy. They bolstered their finding in a neuroimaging study claiming that learning to read fundamentally changes the neural circuitry (Hervais-Adelman et al., 2019). It is, therefore, plausible that such structural change in the brain manifests in linguistic behaviour. Similarly, Scholman et al. (2020) demonstrated that literacy is predictive of readers’ sensitivity to discourse signals available in the context for predictiong upcoming content. Cognitive ageing is also reported as a limiting factor in generating predictions. Smaller N400 amplitude and latency in older adults compared to younger adults have been shown as evidence of the inability of older adults to engage in predictive language processing (K. Federmeier et al., 2002). Furthermore, among older adults, those with lower working memory scores are shown to be further disadvantaged when using context information (K. D. Federmeier et al., 2010). Another line of argument that critiques predictive processing comes from the observations of Huettig & Guerra (2019). They analyzed participants' anticipatory eye movements in the visual world paradigm and showed that listeners predict the target word only in an *artificial* setup of long preview time coupled with slow speech (cf. Fernandez et al., 2020; Heyselaar et al., 2021).

In this thesis, we study additional top-down and bottom-up processes that can interact with and potentially limit the facilitatory effect of predictability. For example, current theories of predictive processing are poor in explaining the role of *attention* in semantic prediction (e.g., Christiansen & Chater, 2015; Ferreira & Lowder, 2016; Friston, Sajid, et al., 2020; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). For example, in their prediction-by-production account, Pickering & Gambi (2018) emphasize that listeners use their speech production mechanism in speech perception and comprehension to predict what their interlocutor will say next. Their framework attempts to paint a big picture of prediction — using the motor system — but it does not consider how a listener's strategy of attending to only a part of a speech stream in adverse listening conditions influences linguistic predictions. We argue that attention to context information is critical in forming semantic predictions, especially in degraded speech comprehension (cf. Kok et al., 2012). By manipulating listeners' attention allocation to parts of a speech stream and information content in the sentences, we show that attention to context information is a prerequisite for the listeners to generate predictions. We also investigate the effect of bottom-up processes, like speech rate, on top-down processes (i.e., predictability effect in degraded speech comprehension). The extant findings on the effects of speech rate on the facilitatory effect of predictability have been mixed both in clear and degraded speech comprehension (e.g., Aydelott & Bates, 2004; Goy et al., 2013; Iwasaki et al., 2002; M. B. Winn & Teece, 2021). We demonstrate a scope for current theories of predictive language processing to incorporate the instances of varying predictability effects at fast and slow speech rates and the effects of attention on degraded speech comprehension.

2.3 Adaptation to degraded speech

Listeners quickly adapt to novel speech with artificial acoustic distortions (Dupoux & Green, 1997). Repeated exposure to distorted speech, improves listeners' comprehension improves over time (Guediche et al., 2014; for reviews, see Samuel & Kraljic, 2009). When the noise condition, like speech degradation level, is constant throughout the experiment, listeners adapt to it, and the performance (e.g., word recognition) improves with as little as 20 minutes of exposure (Rosen et al., 1999). For example, Davis et al. (2005)Experiment 1 presented listeners with 6 channels noise-vocoded speech and found an increase in the proportion of correctly reported words over the course of the experiment. Similarly, Erb et al. (2013) presented participants with 4 channels noise-vocoded speech and reached a similar conclusion. In these experiments, only one speech degradation level (6- or 4-channel noise-vocoded speech) was presented in one block. There was no uncertainty about the next-trial speech degradation from the participants' perspective. Additionally, semantic feature (i.e., target word predictability) was not varied. When multiple types or levels of degraded speech signals are presented in a (pseudo-)randomized order within a block, a listener is uncertain about the signal quality of any upcoming trial. This can influence perceptual adaptation such that it might be totally absent with the change in the characteristics of the auditory signals throughout an experiment (Mattys et al., 2012). In addition, trial-by-trial variability in the characteristics of distorted speech can impair word recognition (Sommers et al., 1994; see also Dahan & Magnuson, 2006). Only a limited number of studies have looked at how the (un)certainty about next-trial speech quality and semantic features influence adaptation. For example, in a word-recognition task, Vaden et al. (2013) presented words at +3dB SNR and +10dB SNR in a pseudo-random order; the goal was to minimize the certainty about the noise level within the block. They proposed that an adaptive control system might be involved to optimize the task performance when the listeners are uncertain about an upcoming trial (Eckert et al., 2016; Vaden Jr et al., 2016; Vaden et al., 2015). However, we cannot make a firm conclusion about perceptual adaptation *per se* from their studies as they do not report the performance change throughout the experiment. Similarly, Obleser and colleagues (Hartwigsen et al., 2015; Obleser et al., 2007; Obleser & Kotz, 2010) also presented listeners with noise-vocoded sentences (ranging from 2 to 32 channels noise-vocoding) in a pseudo-randomized order but did not report the presence or absence of perceptual adaptation. On the one hand, repeated exposure is shown to lead to perceptual adaptation to degraded speech. On the other hand, uncertainty about speech quality is speculated to impair word recognition. We argue that a trial-by-trial variation in a higher-level semantic feature of speech hinders listeners' perceptual system from retuning itself to adapt to the lower-level auditory features of the degraded speech (cf. Nahum et al., 2008). In contrast to prior studies, we show that listeners do not adapt to degraded speech despite repeated exposure to the same degraded speech as long as its semantic predictability is uncertain.

2.4 Summary

In this chapter, we provided an overview of the concepts that will be repeated in the following chapters. We introduced the concept of speech distortion and degradation. Digital signal processing methods used in this process will be discussed in Chapter 3 (Section 3.1.2). Importantly, we provided an overview of how predictive language processing aids language comprehension, as well as its limitations. We discussed perceptual adaptation to degraded speech and the role of uncertainty about next-trial in adaptation. At each step, we presented the motivation behind the studies in this thesis and the gaps in the literature these studies fill in. In the next chapter, we will discuss the methods that are common in all the experiments (Chapters 5, 6, and 7) in developing materials and collecting data.

3
General methods

This chapter provides an overview of the experimental materials used in the experiments described in Chapters 5, 6, and 7. Sentences used as experimental material were common in all the experiments, and the signal processing method was also common. Here, we also present an overview of online data collection.

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. The 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.

3.1.1 Stimulus sentences

With an aim to create sentences of three levels of predictability (low, medium, and high), a triplet of 120 sentences — a 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

Figure 3.1: Distribution of cloze probability ratings of target words in low, medium and high predictability sentences

Subject-Verb-Object form (e.g., *Er fängt den Ball.* EN: He catches the ball.). Some of these sentences were taken from Obleser & 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 of low, medium and high probability sentences are shown in Table 3.1 and the distribution of cloze probability across low, medium, and high predictability sentences is shown in Figure 3.1. The cloze probabilities of the target words in each sentence are shown in Appendix A.

Table 3.1: Cloze probabilities of high, medium and low predictability sentences

	Cloze probability	
Predictability	Mean ± SD	Range
Low	0.022 ± 0.027	0.00 0.09
Medium	0.274 ± 0.134	0.1 0.55
High	0.752 ± 0.123	0.56 1.00

3.1.2 Speech processing

All 360 sentences 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 Chapters 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.

Noise-vocoding

Noise-vocoding is used to parametrically vary and control the speech quality 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 a cochlear implant (Loizou et al., 1999; Shannon et al., 1995; Shannon 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 noise-filtering makes the vocoded speech difficult to understand, although its 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 (2005) in Praat software (Boersma, 2001). 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 (Erb, 2014; Greenwood, 1990; Rosen et al., 1999). 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 3.2. Modulated noise bands were then combined to produce a degraded speech. Scaling was performed to equate the root-mean-square value of the original undistorted speech and the final degraded speech. This resulted in four levels of degradation: 1-, 4-, 6-, and 8-channel noise-vocoded speech.

Spectrograms of clear speech and noise-vocoded speech for the sentence *Er löest die Aufgabe* are shown in Figure 3.2. It shows that with a decrease in the number of noise-vocoding channels, the information in speech signal reduces and becomes noise-like.

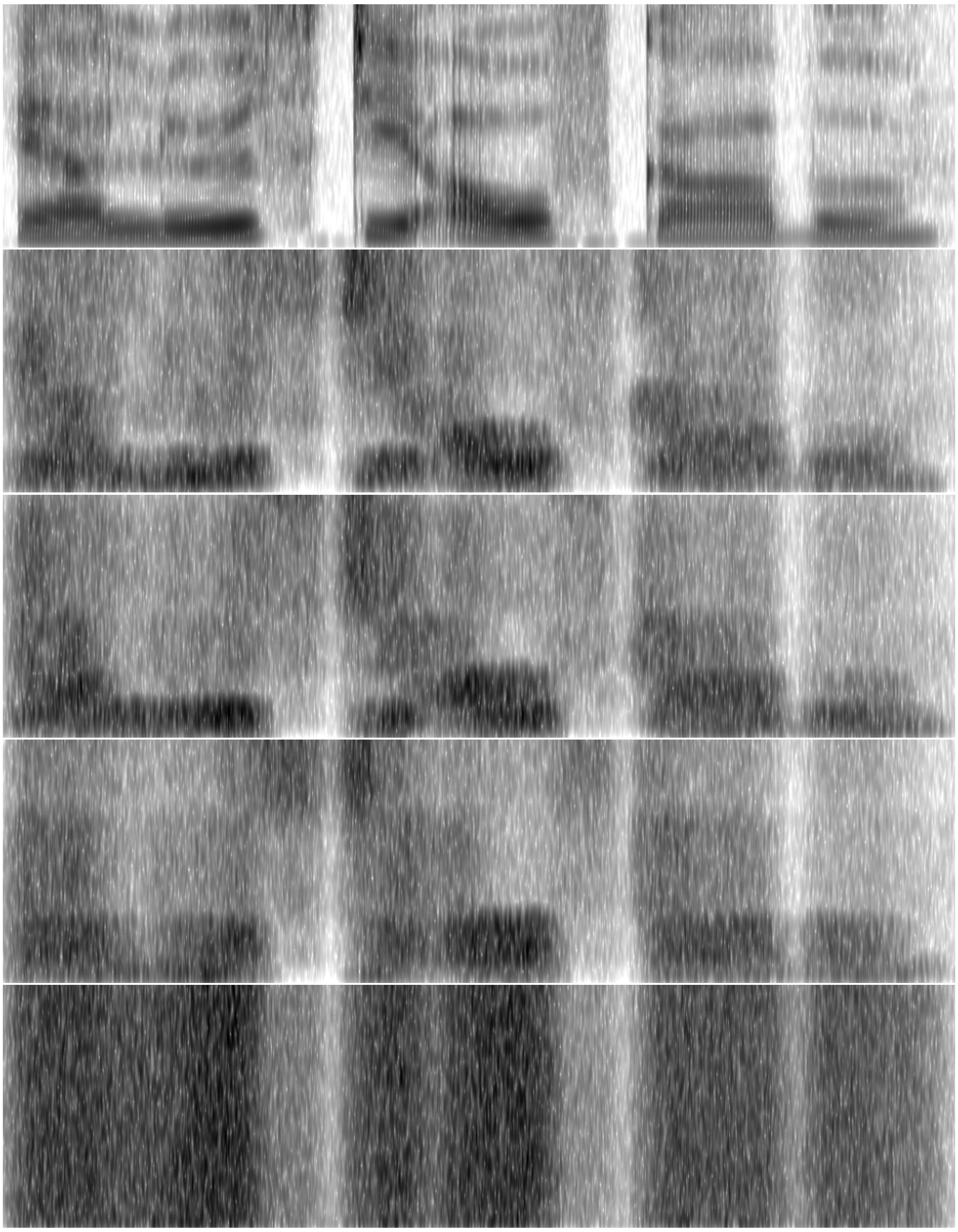

Figure 3.2: Spectrograms of clear speech, and degraded speech arranged with a decreasing number of noise-vocoding channels (8, 6, 4 and 1 band) for the sentence 'Er löest die Aufgabe.'

Table 3.2: Boundary frequencies (in Hz) for 1, 4, 6 and 8 channels noise-vocoding conditions

Number of channels	Boundary frequencies								
1	70	9000							
4	70	423	1304	3504	9000				
6	70	268	633	1304	2539	4813	9000		
8	70	207	423	764	1304	2156	3504	5634	9000

The primary motivation to degrade speech signals by noise-vocoding is twofold: On the practical side, noise-vocoding simulates the frequency selectivity with a cochlear implant or sensory-neural hearing loss. This provides insight into speech perception and language comprehension in special populations (older adults with hearing loss, patients with cochlear implants). On the experimental side, noise-vocoding preserves the temporal periodicity cues of the speech; we can investigate the importance of specific suprasegmental cues in speech perception. Noise-vocoding reduces the fine structure cues that carry the pitch-related suprasegmental information and allows the study of temporal amplitude envelope cues, which carry the suprasegmental information involved in lexical processing; noise-vocoding preserves these cues. Most importantly, it allows fine-grained control over speech intelligibility by varying the number of vocoder channels.

Speech compression and expansion

Temporal compression and expansion are used as a method to simulate fast and slow speech and to study the effect of acoustic degradation (which is the change in speech rate) and increase or decrease in information flow. As early as the mid-twentieth century, investigators have reported that intelligibility does not drop significantly 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 the pitch-synchronous overlap-add technique (PSOLA, Charpentier & Stella, 1986; Moulines & Charpentier, 1990) developed in the 1980s and later (e.g., Verhelst & 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 a uniform time-compression algorithm (PSOLA) to create slow and fast speech with the compression factor of 1.35 and 0.65, respectively. A schematic representation of waveforms of different speech rates — normal, slow and fast — is shown in Figure 3.3.

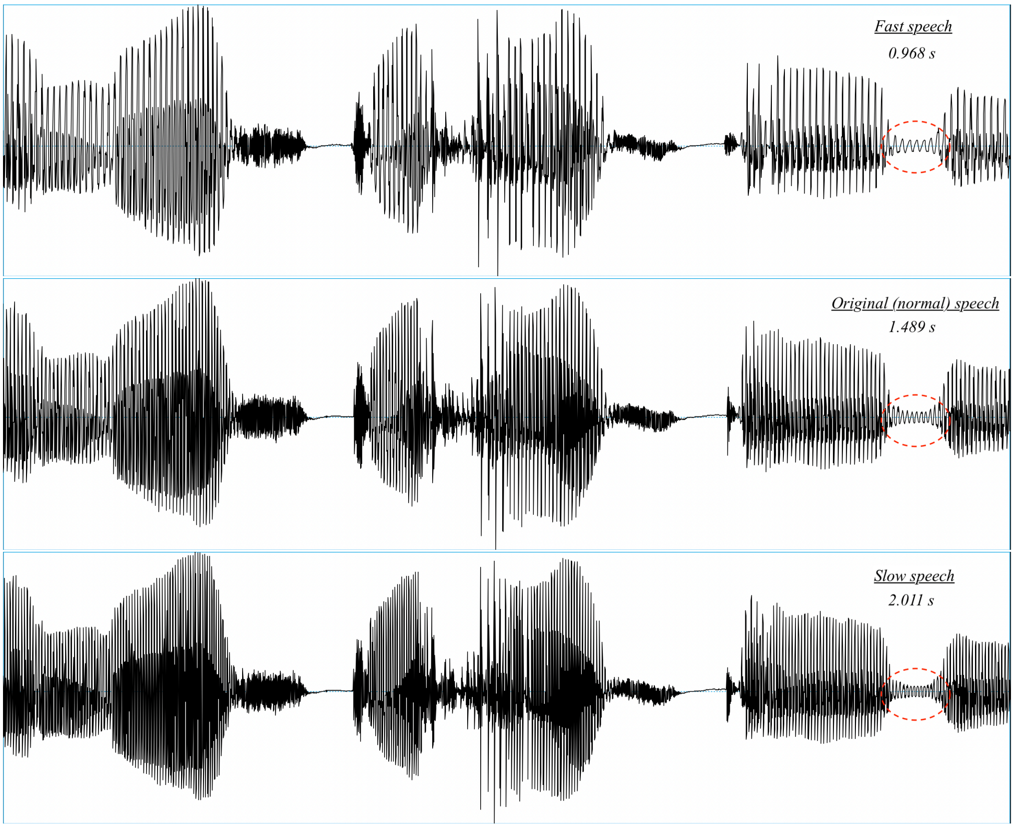

Figure 3.3: Schematic representation of waveforms of fast, normal, and slow speech rates for the sentence 'Er löest die Aufgabe' with the duration of each speech rates in second. Note the circled portion of the waveform, which examplifies that PSOLA eliminates and duplicates the parts of the original waveform to create fast and slow speech respectively.

PSOLA creates fast or slow speech in three steps: analysis, modification, and synthesis (Charpentier & Stella, 1986; Taleb, 2020). In the analysis step, it first sets pitch marks in an audio file and then creates segments of it (i.e., it segments the signal into successive analysis windows centred around those pitch marks). Then in the modification step, depending on the time-compression/expansion factor, it deletes (or duplicates) those segments and sets a new set of pitch marks. Finally, in the synthesis step, it adds the new segments back to the audio file (i.e., it rearranges the analysis window) and creates fast or slow speech as required. The distortion of phonemic properties of speech signals are minimal when accelerating and slowing down within the range of factor 2 or below (Moulines & Charpentier, 1990; cf. Longster, 2003).

We create fast and slow versions of 120 high and 120 low predictability sentences. These 480 recordings are then passed through 4 channels noise-vocoding to use as experimental materials. As discussed earlier, the main aim of manipulating this bottom-up process is to investigate the effect of change in the rate of information flow (i.e., change in the speech rate) on the top-down processes of contextual facilitation in degraded speech comprehension.

3.2 Data collection on the web

Traditionally, behavioural experiments with human participants are conducted in a laboratory setup. In recent years, there has been a surge of experiments that are conducted on the web (Reips, 2021). The first generation of online experiments to study human cognition began in the mid-1990s (for reviews, Musch & Reips, 2000) with the advent of the internet (Berners-Lee et al., 1992). Welch & Krantz (1996) was the first online experiment that was conducted in 1995 as a part of tutorials in auditory perception (Musch & Reips, 2000). In their survey of researchers, Musch & Reips (2000) discovered that until 2000, there were already at least two 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 scepticism of journals towards publishing results of online experiments, Musch & 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. (p. 85)

By 2022, there has been significant growth in 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 retrieve these data in a fairly structured format (A. L. Anwyl-Irvine et al., 2020; Peirce et al., 2019; Prolific, 2014; see also, A. Anwyl-Irvine 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 is available online, which is usually not possible in laboratory experiments. Similarly, the participants in online experiments are more diverse than in laboratory experiments. Considering these advantages, psychologists and psycholinguists have conducted online experiments for almost three decades. Scientists who only conducted laboratory experiments and occasional online experiments were forced to conduct their experiments almost exclusively on the web due to the restrictions imposed by covid-19 lockdown (Gagné & Franzen, 2021; Reips, 2021). Since Welch & Krantz (1996)'s auditory perception experiment, a number of experiments have been conducted online in the auditory domain (Leensen & Dreschler, 2013; Seow & Hauser, 2022; van Os et al., 2021; Woods et al., 2017) replicating laboratory findings (e.g., Cooke & Garcia Lecumberri, 2021). The experiments reported in this thesis were also conducted online.

Initially, our experiments were designed to be conducted both in the laboratory and online. As the laboratory was shut down due to covid-19 pandemic-related lockdown (M. Schmitt, personal communication, March 16, 2020), we moved the laboratory experiments to the online platform. We recruited participants online via Prolific Academic (Prolific, 2014). We used Prolific's filters to recruit only native speakers of German residing in Germany who reported not having any hearing loss, speech-language disorder, or cognitive impairment. 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 mixed effects models

As the name suggests, the linear mixed effects model (LME) is a linear regression model that consists of both fixed and random effects. It allows modelling the underlying structure of the data, which includes the standard fixed effects like the levels of speech degradation and the levels of target word predictability, as well as random effects like items and participants. These random effects are assumed to be random samples drawn from the general population. In this thesis, the dependent variable (an *outcome* or a *response* variable) is binary (correct vs incorrect response). So, we use binomial logistic mixed effects models with crossed random effects to model the data (Baayen et al., 2008).

A linear mixed effects model can be written as:

$$\begin{aligned} y = \alpha + u_\alpha + w_\alpha + (\beta_1 + u_{\beta_1} + w_{\beta_1}) \cdot x_1 + \\ (\beta_2 + u_{\beta_2} + w_{\beta_2}) \cdot x_2 + ... + \\ (\beta_n + u_{\beta_n} + w_{\beta_n}) \cdot x_n \end{aligned} \tag{4.1}$$

where,

- y is the dependent variable, like participant's response (correct vs. incorrect)
- 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 contrast to linear regression models, mixed effects models allow to simultaneously account for the effects of two random variables, like item and participants. The variance in the categorical dependent variable is also preserved, which would otherwise be eliminated by averaging in linear regression models. We discuss these issues and the motivation to use the mixed effects model in this thesis in more detail below in this chapter.

4.1.1 Linear regression and its limitations

In linear regression, a dependent variable (or an *outcome*) is modelled as a function of one or more independent predictor variables (*factors* or *explanatory* variables). That is, an outcome y is modelled as a function of explanatory variables $x_1, x_2, x_3..., x_n$, and an error term ε.

$$y = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + ... + \beta_n \cdot x_n + \varepsilon \tag{4.2}$$

Analysis of Variance (ANOVA), also a form of linear regression (Chatterjee & Hadi, 2012; Vasishth et al., 2022), compares the means and variances of two or more conditions. As expressed in Equation (4.2), regression models can only model fixed effects. Although ANOVA can account for one random effect at a time, it still averages out the variance in the second random effect. These problems of using ANOVA in language sciences have been pointed out as early as the 1960s (H. H. Clark, 1973; Coleman, 1964). We elaborate on them in the context of the data of our experiments as follows.

Modeling two random effects simultaneously, and Variability in the data

As mentioned above, a simple linear regression model, including ANOVA, does not model the effect of two random effects simultaneously, which a mixed effects model does. In the traditional ANOVA approach, researchers often run two separate regression models (Lorch & Myers, 1990) by averaging raw data across participants, and items. Averaging eliminates the variability in the data. Additionally, comparing the means of a categorical variable (correct vs incorrect responses) even when transformed into accuracy or proportion scale is hard to interpret sensibly compared to a continuous variable like reaction time (for discussion, see Bolker et al., 2009; Jaeger, 2008). The statistical remedy for these problems in analyzing the data of our experiments is to apply mixed effects models.

Non-independence of observations

Some of the assumptions made for regression models are violated in our data. One of them is the non-independence of observations, i.e., all data points are independent of one another. This assumption is violated in an unbalanced design, and sometimes even for a balanced design. The same participant responds to multiple trials of the same experimental condition within an experiment. Although the design itself is balanced, after the removal of outliers and trials which are not appropriate for comprehension measures (for details, see Section 6.2.3), the number of trials in the analysis is unequal for each participant, item, and experimental condition. This inequality or unbalance introduces a bias in the regression model (Jaeger, 2008). A mixed effects model is best suited for such unbalanced data.

Common mean for each predictor

An intrinsic property or feature of the linear regression model is that it assumes a common mean for each predictor. It has been shown that this is, in fact, not true in the actual data: the effect of a predictor can vary depending on random variables like participant or item. Mixed effects models take into account such inter-participant and inter-item variability present in the data. For example, in mixed effects models, the random effects term with only varying intercept, e.g., participant as an intercept, assumes that if there are 100 participants, then the mean accuracy of those 100 participants is only a subset of possible global accuracies drawn from a set of the population mean. When a slope, e.g., levels of predictability, is included in the random effects structure in addition to the varying intercept (e.g., participants), then the model assumes that the effect of predictability on response accuracy varies across participants. Such variance across participants (or across items) is present in the real data and can be modelled in a mixed effects model but not in a linear regression model.

Bounded output variable

Linear models assume an output variable to not be bounded within a narrow range and to be on a continuous scale. In our data, the output variable (correct vs incorrect response) has bounded outcomes on $[0, 1]$. To fit a linear model, it can be transformed into a proportion scale. Even though it is a continuous variable, the proportion scale (i.e., response accuracy) has a range (0,1). Additionally, the transformation of discrete variables brings a host of problems that we have already discussed above (e.g., loss of variability by averaging raw data). Binomial logistic mixed effects models, on the other hand, transform[1] the output variable into a *logit* scale, log with base e, i.e. ln, with a range $(-\infty, +\infty)$. Therefore, these mixed effects models do not violate the model assumptions regarding the range of the outcome variables.

Thus, Equation (4.1) can also be written as:

$$\begin{aligned} \ln(\frac{p}{1-p}) = \alpha + u_\alpha + w_\alpha + (\beta_1 + u_{\beta_1} + w_{\beta_1}) \cdot x_1 + \\ (\beta_2 + u_{\beta_2} + w_{\beta_2}) \cdot x_2 + \\ ... + (\beta_n + u_{\beta_n} + w_{\beta_n}) \cdot x_n \end{aligned} \tag{4.3}$$

[1]Such transformation is brought about by a generalized linear mixed effects model with a canonical logit link function (Malik et al., 2020).

This is equivalent to,

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

where,

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

Log-odds of correct response obtained from Equation (4.3) can also be transformed into the probability of correct response. Equations (4.4) and (4.5) provide the relationship between probability, logits (or log-odds), and odds ($\frac{p}{1-p}$).

We have presented the advantages of mixed effects models over linear (regression) models. We use the binomial logistic mixed effects model as the statistical analysis tool in all experiments reported in this thesis. Below we discuss how the model that best fits our data was selected.

4.2 Model selection and Running mixed effects models in R

The underlying structure of given data can be explained by different approximate statistical models. We intend to select a model that best fits our data. 'Best fit' can be objectively measured by Akaike Information Criterion, Bayesian Information Criterion, and Likelihood Ratio Test, among others (Akaike, 1973; Schwarz, 1978).

In this thesis, we first build a complex (or maximal) model by including all predictors (target word predictability, speech degradation level, speech rate), their interactions, and co-variates (e.g., trial number) in the fixed effects (cf. Bondell et al., 2010). The model is fitted with a maximal random effects structure that includes 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 Chapters 5, 6, and 7.

Model selection was based on the backward-selection heuristics on the fixed effects (cf. Matuschek et al., 2017). To find the best fitting model for the data, non-significant higher-order interactions were excluded from the fixed-effects structure in a stepwise manner. Similarly, random effects not supported by the data that explained zero variance according to singular value decomposition were excluded to prevent overparameterization (Bates, Kliegl, et al., 2015). This gave a more parsimonious model, which was then extended separately with: i) item-related correlation parameters, ii) participant-related correlation parameters, and iii) both item- and participant-related correlation parameters. Among the parsimonious model and extended models, the model with the smallest AIC was selected as the best fitting model for our data (Grueber et al., 2011; Richards et al., 2011).

Data preprocessing and analyses were performed in R-Studio (Version 3.6.1; R Core Team, 2019; Version 3.6.3; R Core Team 2020; Version 4.1.3; R Core Team, 2022). 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. Contrast coding of each factor and the model description are presented in the Analysis section of the chapters that follow.

4.3 Summary

In this chapter, we introduced the statistical tool used for data analysis in this thesis. We discussed the limitations of traditional linear regression-based models like ANOVA and outlined the motivations for using mixed effects models. To capture the variability of our data without averaging out across participants or items, and to account for the effect of two random effects — participant and item — simultaneously, we fit mixed effects models to our data.

Details of each dataset corresponding to each experiment are presented in Chapters 5, 6, and 7.

5

Predictability effects 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 context information and form top-down semantic predictions, which provides contextual facilitation in understanding the degraded speech. Importantly, it is also affected by the allocation of attention to the context in a top-down manner. The aim of this study was to examine the role of attention in 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 heard or to type in the whole sentence. We were interested in how these task instructions influence the interplay between top-down predictions and bottom-up perceptual processes during language comprehension. The sentences varied in the degree of predictability (low, medium, and high) as well as in the levels of speech degradation (1-, 4-, 6-, and 8-channel noise-vocoding). Results indicated better word recognition for highly predictable sentences for moderate, though not for high, levels of speech degradation, but only when attention was directed to the whole sentence.

5.1 Introduction

When there is noise in the signal, listeners overcome the difficulty of understanding speech by using context information. The 'context information' can be information in a given situation about a topic of conversation, semantic and syntactic information of a sentence structure, world knowledge, visual information, or even information about neighbouring phonemes (Altmann & Kamide, 2007; Kaiser & Trueswell, 2004; Knoeferle et al., 2005; Xiang & Kuperberg, 2015; for reviews, see Ryskin & Fang, 2021; Stilp, 2020). For example, in the *phoneme restoration effect* (Samuel, 1996; Warren, 1970), a phoneme of one or more words in a sentence is replaced with white noise or a coughing sound. Participants are unable to notice such 'noisy' words in a sentence, as they perceptually restore the missing sound in those words from the context information. To utilize the context information in a sentence, listeners must attend to it and build a meaning representation of what has been said.

Processing and comprehending degraded speech is more effortful and requires more attentional resources than clear speech (Eckert et al., 2016; Hunter & Pisoni, 2018; Peelle, 2018; Wild et al., 2012). In this chapter, we examine how attention modulates the predictability effects brought about by context information at different levels of spectral degradation of speech. We address the existing unclarity in the literature regarding the distribution of attentional resources in an adverse listening condition: 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 and allocating resources also to understanding the linguistic context and generating predictions.

The study reported in this chapter was conducted to investigate how the allocation of attentional resources induced by different task instructions influences language comprehension and, in particular, the use of context information in communication through a noisy channel, i.e., when the speech is degraded. To examine the role of attention on predictive processing under degraded speech, we ran two experiments in which we manipulated the task instructions. In Experiment 1, participants were instructed to repeat only the final word of the sentence they heard, while in Experiment 2, they were instructed to repeat the whole sentence, 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 Chapters 1 and 2, it is generally agreed upon that human language processing is predictive in nature, and comprehenders generate expectations about upcoming linguistic materials based on the context available to them (for reviews, see Kuperberg & Jaeger, 2016; Nieuwland, 2019; Pickering & Gambi, 2018; Staub, 2015). When the bottom-up speech signal is less informative in an adverse listening condition, listeners tend to rely more on top-down lexical-semantic cues from the context to support speech perception and language comprehension (Amichetti et al., 2018; Ganong, 1980; McGurk & MacDonald, 1976; Obleser & Kotz, 2010; Sheldon et al., 2008b; Warren, 1970). However, it is not just the quality of speech signal that determines and influences the reliance and use of predictive processing; attention to the auditory input is essential too. Auditory attention allows a listener to focus on the speech signal of interest (for reviews, see Fritz et al., 2007; Lange, 2013). For instance, a listener can attend to and derive information from one stream of sound among many competing streams, as demonstrated in the well-known *cocktail party effect* (Cherry, 1953; Hafter et al., 2007). When a participant is instructed to attend to only one of the two or more competing speech streams in a diotic or dichotic presentation, response accuracy to the attended speech stream is higher than to the unattended speech (e.g., Tóth et al., 2020). Similarly, when a listener is presented with a stream of tones (e.g., musical notes varying in pitch, pure tones of different harmonics) but attends to any one of the tones appearing at a specified time point, this is reflected in a larger amplitude of N1[1] (e.g., Lange & Röder, 2010; see also, Sanders & Astheimer, 2008). Hence, listeners can draw attention to and process one among multiple competing speech streams, as well as orient their attention in the temporal dimension within an unfolding sound stream.

So far, most previous studies have investigated listeners' attention within a single speech stream using acoustic cues like accentuation and prosodic emphasis. For example, J. Li et al. (2014) examined whether the comprehension of critical words in a sentence context was influenced by a linguistic attention probe such as "ba" presented together with an accented or a de-accented critical word. The N1 amplitude was larger for words with such an attention probe than for words without a probe. These findings support the view that attention can be flexibly directed either by instructions towards a specific signal or by linguistic probes (X. Li et al., 2017; see also Brunellière et al., 2019). Thus, listeners are able to select a part or segment of a stream of auditory stimuli to selectively allocate their 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 are not informative about the attention-prediction interplay in degraded speech comprehension. Specifically, we do not know what role attention to a segment of the speech stream plays in the contextual facilitation of degraded speech comprehension. The studies that report predictability effects in degraded speech comprehension do not systematically examine the role of attention (e.g., Amichetti et al., 2018; Obleser & Kotz, 2010; Sheldon et al., 2008b). Their conclusion that *semantic predictability facilitates comprehension of degraded speech* is based on listeners' attention to the entire sentence, which

[1] N1 or N100 is a negative-going EEG component that peaks around 100 ms post-stimulus. It is considered as a neural marker of auditory selective attention (Näätänen & Picton, 1987; Thornton et al., 2007).

is not compared to any other experimental condition manipulating attentional allocation. Therefore, in two experiments, we examined whether context-based semantic predictions are automatic during effortful listening to degraded speech when participants are instructed to report either 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 not binding the context or 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 replicating the prior behavioural findings (e.g., Obleser & Kotz, 2010).

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.

5.3.1 Methods

Participants

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

Materials

Materials used in the experiment were created by the method described in Chapter 3 (Section 3.1). That is, there were 120 unique sentences in each of these three categories: low predictability, medium predictability and high predictability. Mean cloze probabilities of the target words of low, medium and high predictability sentences were 0.022 ± 0.027 ($M \pm SD$; range $= 0.00 - 0.09$), 0.274 ± 0.134 ($M \pm SD$; range $= 0.1 - 0.55$), and 0.752 ± 0.123 ($M \pm SD$; range $= 0.56 - 1.00$) respectively. All 360 sentences were then noise-vocoded through 1, 4, 6, and 8 channels to create degraded speech.

Procedure

Participants were asked to use headphones or earphones. A sample of vocoded speech not used in the practice trial or the main experiment was provided so that the participants could adjust the volume to their preferred level of comfort at the beginning of the experiment. The participants were instructed to listen to the sentences and type in the target word (noun) using the keyboard. The time for typing in the response was not limited. They were informed at the beginning of the experiment that some of the sentences would be 'noisy' and not easy to understand. In these cases, guessing was encouraged. Eight practice trials with different levels of speech degradation were given to familiarize the participants with the task before presenting all 120 experimental trials with an intertrial interval of 1,000 ms.

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

5.3.2 Analyses

Out of 5880 trials from 49 participants, there were only five correct responses, one each from 5 participants at 1-channel. Therefore, the 1-channel speech degradation condition was excluded from the analyses.

Accuracy was analyzed using Generalized Linear Mixed Models (GLMMs) following the procedure described in Chapter 4 (Section 4.2) with lmerTest (Kuznetsova et al., 2017) and lme4 (Bates, Mächler, et al., 2015) packages. Binary responses (categorical: correct and incorrect) for all participants were fit with a binomial linear mixed-effects model. Correct responses were coded as 1, and incorrect responses were coded as 0. Number of channels (categorical: 4-channel, 6-channel, and 8-channel noise-vocoding), target word predictability (categorical: high predictability sentences, medium predictability sentences, low predictability sentences), and the interaction of number of channels and target word predictability were included in the fixed effects.

We fitted a model with a maximal random effects structure that included random intercepts for each participant and item (Barr et al., 2013). Both by-participant and by-item random slopes were included for number of channels, target word predictability, and their interaction, which was supported by the experiment design. Based on the previous findings on perceptual adaptation (e.g., Cooke et al., 2022; Davis et al., 2005; Erb et al., 2013), we further added trial number (centred) in the fixed effect structure to control for whether the listeners adapted to the degraded speech. We applied treatment contrast for number of channels (8-channel as a baseline) and sliding difference contrast for target word predictability (low predictability vs medium predictability and low predictability vs high predictability).

5.3.3 Results and discussion

Mean response accuracies (in percentage) for all experimental conditions aggregated across all participants and items are shown in Table 5.1 and Figure 5.1. It

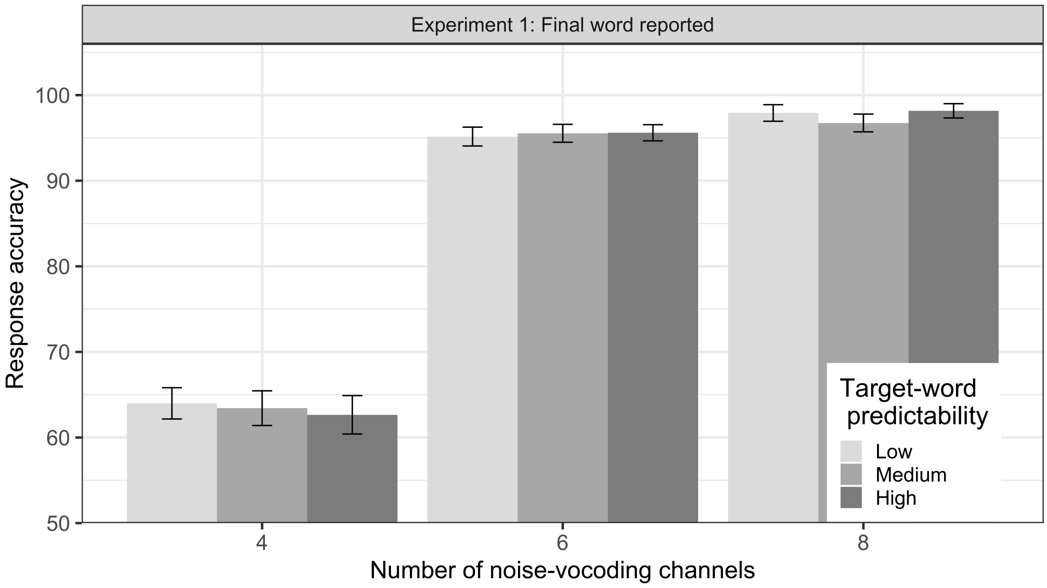

Figure 5.1: Mean response accuracy across all conditions in Experiment 1. Accuracy increased only with an increase in the number of noise-vocoding channels. There is no change in accuracy with an increase or decrease in target word predictability. Error bars represent the standard error of the means.

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. These observations aligned with the results of the statistical analyses (Table 5.2).

Table 5.1: Response accuracy (mean and standard error of the mean) across all levels of speech degradation and target word predictability in Experiment 1

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

There was a significant main effect of number of channels, indicating that response accuracy for the 8-channel vocoded speech was higher than for both 4-channel (β = -3.50, SE = .22, z(4,410) = -16.19, p < .001) and 6-channel vocoded speech (β = -.70, SE = .21, z(4,410) = -3.29, p = .001), that is, when the number of channels increased to 8, listeners gave more correct responses (see Figure 5.1). There was, however, no significant main effect of target word predictability (β = .30, SE = .36, z(4,410) = .84, p = .40, and β = .50,SE = .43, z(4,410) = 1.16, p = .25), and no interaction between number of channels and target word predictability (all ps > .05). There was also no significant main effect of trial number (β = .001, SE = .002, z(4,410) = .48, p = .63) suggesting that the listeners' performance did not improve over time. These results indicated a decrease in response accuracy with an increase in speech degradation from the 8-channel to the 6-channel noise-vocoding condition and from the 8-channel to the 4-channel noise-vocoding condition. However, response accuracy did not increase with an increase in target word predictability, and the interaction between number of channels and target word predictability was also absent, in contrast to previous findings (e.g., Obleser et al., 2007; Obleser & Kotz, 2011; see also Hunter & Pisoni, 2018). These results suggest that the task instruction, which asked participants to report only the final word, indeed led to neglecting the context. Although participants were able to neglect the context, there was still uncertainty about the speech quality of each subsequent trial; hence, they could not adapt to the different levels of degraded speech.

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

Table 5.2: Estimated effects of the model accounting for the correct word recognition in Experiment 1

Fixed effects	Estimate	Std. Error	z value	p value
Intercept	4.17	.25	16.73	<.001
Noise condition (4-channel)	-3.50	.22	-16.19	<.001
Noise condition (6-channel)	-.70	.21	-3.29	<.001
Target word predictability (Low-Medium)	.30	.36	.84	.40
Target word predictability (High-Low)	.50	.43	1.16	.25
Noise condition (4-channel) × Target word predictability (Low-Medium)	-.22	.39	-.57	.57
Noise condition (6-channel) × Target word predictability (Low-Medium)	-.34	.44	-.76	.44
Noise condition (4-channel) × Target word predictability (High-Low)	-.54	.45	-1.18	.24
Noise condition (6-channel) × Target word predictability (High-Low)	.04	.50	.09	.03
Trial number	.001	.002	.48	.63

5.4 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 the facilitatory effect of top-down predictions is observed only when the listeners' attention is unrestricted so that the context information is also included within the listener's attentional focus.

5.4.1 Methods

Participants and Materials

We recruited a new group of 48 participants (M age $\pm SD = 24.44 \pm 3.5$ years; age range = 18-31 years; 32 males) online via Prolific Academic. The same stimuli were used as in Experiment 1.

Procedure

Participants were presented with sentences at a comfortable volume level. They were asked to use headphones or earphones, and a prompt was presented before the experiment began to adjust the volume to their level of comfort. Eight practice trials were presented, followed by 120 experimental trials. In contrast to Experiment 1, the participants were instructed to report the entire sentence, instead of reporting only the sentence-final word, by typing in what they heard. We did not limit the response time.

5.4.2 Analyses

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

5.4.3 Results and discussion

Mean response accuracy for different conditions is shown in Table 5.3 and Figure 5.2. We found that accuracy increased when the number of noise-vocoding channels increased, as well as when the target word predictability increased. The results of the statistical analysis confirmed these observations (Table 5.4).

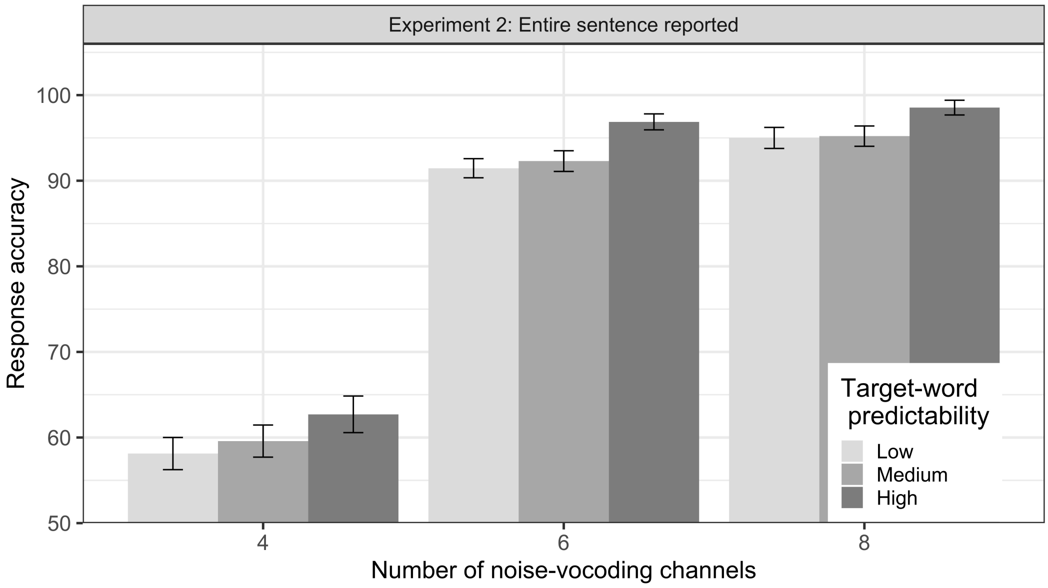

Figure 5.2: Mean response accuracy across all conditions in Experiment 2. Accuracy increased with an increase in number of noise-vocoding channels and target word predictability. Error bars represent the standard error of the means.

Table 5.3: Response accuracy (mean and standard error of the mean) across all levels of speech degradation and target word predictability in Experiment 2

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

We again found a main effect of number of channels, such that response accuracy at 8-channel was higher than for both 4-channel (β = -3.51, SE = .24, z(4,320) = -14.64, p < .001), and 6-channel noise-vocoding (β = -.65, SE = .22, z(4,320) = -2.93, p = .003). Similar to Experiment 1, the main effect of trial number was not significant (β = .002, SE = .002, z(4,320) = 1.11, p = .27) indicating that the response accuracy did not increase over the course of the experiment.

In contrast to Experiment 1, there was also a main effect of target word predictability: Response accuracy in high predictability sentences was significantly higher than in low predictability sentences (β = 1.42, SE = .47, z(4,320) = 3.02, p = .003). We also found a statistically significant interaction between speech degradation and target word predictability (β = -1.14, SE = .50, z(4,320) = -2.30, p = .02). 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 in the 8-channel (β = 1.42, SE = .62, z(1,440) = 2.30, p = 0.02) and 6-channel noise-vocoding conditions (β = 1.14, SE = 0.34, z(1,440) = 3.31, $p < .001$); at 4 channel, the difference in response accuracy between high and low predictability sentences was not significant (β = .28, SE = .18, z(1,440) = 1.59, p = .11).

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

We combined the data from both experiments in a single analysis to test whether participants' response accuracy changes across the experiments, that is, to test whether the difference between experimental manipulations is statistically significant. We ran a binomial linear mixed-effects model on response accuracy and followed the same procedure as in Experiments 1 and 2. A full random effects structure supported by the study design was modelled. The model summary is shown in Table 6. The model revealed no significant main effect of experimental group (β = .04, SE = .26, z(8,730) = .15, p = .88), indicating that the overall response accuracy did not change with the change in instructions from Experiment 1 to Experiment 2. However, the critical interaction between experimental group

Table 5.4: Estimated effects of the model accounting for the correct word recognition in Experiment 2

Fixed effects	Estimate	Std. Error	z value	p value
Intercept	4.09	.24	16.79	<.001
Noise condition (4-channel)	-3.51	.24	-14.64	<.001
Noise condition (6-channel)	-.65	.22	-2.93	.003
Target word predictability (Low-Medium)	-.08	.34	-.23	.82
Target word predictability (High-Low)	1.42	.47	3.02	.003
Noise condition (4-channel) × Target word predictability (Low-Medium)	.02	.38	.05	.96
Noise condition (6-channel) × Target word predictability (Low-Medium)	-.13	.43	-.31	.76
Noise condition (4-channel) × Target word predictability (High-Low)	-1.14	.50	-2.30	.02
Noise condition (6-channel) × Target word predictability (High-Low)	-.23	.57	-.41	.68
Trial number	.002	.002	1.11	.27

and target word predictability was statistically significant (β = .46, SE = .20, z(8,730) = 2.34, p = .02). That is, the effect of predictability was larger in the group that was asked to type in the whole sentence (Experiment 2) than in the group that was asked to type only the sentence-final target word (Experiment 1). Together, these findings suggest that the change in task instruction, which draws attention either to the entire sentence or only to the final word, is critical to whether the context information is used under degraded speech. Nonetheless, degraded speech comprehension is not reduced by binding listeners' attention allocation to one part of the speech stream. The model summary is shown in Table 5.5.

Table 5.5: Estimated effects of the best-fitting model accounting for the correct word recognition in both experiments

Fixed effects	Estimate	Std. Error	z value	p value
Intercept	4.19	.20	20.72	<.001
Noise condition (4-channel)	-3.56	.20	-18.19	<.001
Noise condition (6-channel)	-.59	.18	-3.28	.001
Target word predictability (Low-Medium)	.13	.26	.50	.62
Target word predictability (High-Low)	.98	.34	2.93	<.003
Experimental group	.04	.26	.15	.88
Noise condition (4-channel) × Target word predictability (Low-Medium)	-.12	.29	-.40	.69
Noise condition (6-channel) × Target word predictability (Low-Medium)	-.30	.34	-.87	.38
Noise condition (4-channel) × Target word predictability (High-Low)	-.84	.35	-2.42	.02
Noise condition (6-channel) × Target word predictability (High-Low)	-.11	.38	-.29	.77
Noise condition (4-channel) × Experimental group	-.10	.25	-.41	.68
Noise condition (6-channel) × Experimental group	-.10	.28	-.36	.72
Target word predictability (High-Low) × Experimental group	-.47	.20	2.34	.02
Trial number	.001	.001	.93	.35

5.5 Conclusion

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 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., Obleser et al., 2007). Obleser and colleagues reported that at the 8-channel noise-vocoded speech, key word recognition[2] was higher in high predictability sentences than in low predictability sentences. Listeners were required to attend to the entire sentence in those studies as well. Therefore, the findings of Experiment 2 replicate this facilitatory effect of predictability that was observed in Obleser and colleagues' behavioural experiments.

The important new finding that our study adds to the present literature is that this effect of semantic predictability (i.e., contextual facilitation of degraded speech comprehension) may be weakened or lost when listeners are instructed to report only the final word of the sentence that they heard (Experiment 1). The lack of predictability effect (or contextual facilitation) can most likely be attributed to listeners not successfully decoding the meaning of the verb of the sentence, as the verb is the primary predictive cue in our stimuli (e.g., Sie *jongliert* die Bälle) for the target word (noun: *Bälle*). Findings from auditory attention literature also support this explanation: When listeners' attention is focused on one feature of an auditory stimulus and the rest are not attended to, then they are not accessed (filter mechanism, Lange, 2013; change detection, Sanford et al., 2006; Sturt et al., 2004)

Hence, this small change in task instructions from Experiment 1 to Experiment 2 sheds light on the role of top-down attenion regulation in using context for language comprehension in adverse listening conditions. In a noisy channel created

[2]Notice the difference in measurement metrics. Obleser et al. (2007) calculated response accuracy as the number of correct keywords identified, while we calculated it as the correct identification of the sentence-final target word.

by degraded speech, language comprehension is generally effortful, so focusing attention on only a part of the speech signal seems beneficial in order to enhance stimulus decoding. However, the results of this study also show that this comes at the cost of neglecting the context information that could be beneficial for language comprehension. Our findings hence demonstrate that there is a trade-off between the use of context for generating top-down predictions vs focusing all attention on a target word. Specifically, the engagement in the use of context and generation of top-down predictions may change as a function of attention (see also J. Li et al., 2014). This claim is also corroborated by the significant change in predictability effects (or contextual facilitation) from Experiment 1 to Experiment 2 in the combined dataset.

Findings from the irrelevant-speech paradigm also support our conclusion. It has been shown that the predictability of unattended speech has no effect on the main experimental task (e.g., memorization of auditorily presented digits). Wöstmann & Obleser (2016) did not find predictability effects when the participants ignored the degraded speech (see also Ellermeier et al., 2015). An alternative explanation of 'participants neglecting the context' could be that the participants did not listen to the context at all, or they heard but did not process the context. However, irrelevant-speech paradigm studies show that listeners cannot avoid listening to the speech presented to them; to-be-ignored speech has been shown to interfere with the main experimental task (e.g., Lecompte, 1995). It is not implausible that the listeners listened to the context but did not do deep processing (cf. Ferreira & Lowder, 2016). This is not incompatible with our first explanation, as in either case, attention to the final word leaves the listeners with limited resources to process and form a representation of the context information.

Considering the theoretical accounts of predictive language processing (Friston, Parr, et al., 2020; Kuperberg & Jaeger, 2016; McClelland & Elman, 1986; Norris et al., 2016; Pickering & Gambi, 2018), one would expect that listeners automatically form top-down predictions about upcoming linguistic stimuli based on prior context.

Also, when speech is degraded, top-down predictions render a benefit in word recognition and language comprehension (e.g., Sheldon et al., 2008a, 2008b). Results of our study revealed new theoretical insights by showing that this is not always the case. Top-down predictions are dependent on attentional processes (see also Kok et al., 2012), directed by task instructions; thus, they are not *always* automatic, and predictability does not *always* facilitate comprehension of degraded speech. To this point, our findings shed light on the growing body of literature indicating limitations of predictive language processing accounts (Huettig & Guerra, 2019; Huettig & Mani, 2016; Mishra et al., 2012; Nieuwland et al., 2018).

Results from both experiments show that the effect of trial number is not significant. In contrast to previous studies (e.g., Davis et al., 2005; Erb et al., 2013), we did not observe adaptation to noise-vocoded speech. In those studies, there was certainty about the speech quality of the next trial, as the participants were presented with only one level of spectral degradation (only 4-channel or only 6-channel noise-vocoding) and crucially with no specific regard to semantic predictability. On the contrary, in our study, listeners were always uncertain about the speech quality of the next trial as well as its semantic predictability. Because of this changing context, the perceptual system of the participants may not retune itself (Goldstone, 1998; Mattys et al., 2012). However, there was no experimental condition in the current study in which participants were certain about the next-trial speech degradation. It cannot be discarded entirely that certainty about speech degradation would retune the perceptual system, and listeners would adapt to the degraded speech. This is one of the limitations of the current study.

One could object to the metric of calculating accuracy in Experiment 2, but it should be noted that for a valid comparison of the results between the two experiments, we can only consider the accuracy of the sentence-final target word in Experiment 2. Participants' response of the words preceding the sentence-final target word in Experiment 1 was not available; in fact, it was the whole point of the instruction given to the participants: Direct their attention only to the sentence-final target word but not to the words preceding it. Hence, we find a discrepancy between the result of prior studies (Obleser et al., 2007; Obleser & Kotz, 2010) and our study (Experiment 2) regarding the degradation level at which contextual facilitation is observed. Nonetheless, the conclusion from these studies and our study is consistent: as long as listeners attend to the sentence context, semantic context facilitates language comprehension, but, such a facilitatory effect is not observed when the degradation is at an extreme level, like 1-channel noise-vocoding.

In conclusion, this study provides a novel insight into the modulatory role of attention in the interaction between top-down predictive and bottom-up auditory processes. We show that task instructions affect the 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.

5.6 Summary

This chapter reported studies which replicated the previous finding that semantic predictability facilitates language comprehension when speech degradation is not at an extreme level. That is, when the channel of transmission N is noisy, listeners put less weight on the degraded auditory input $P(u_p|u_i, N)$ and more weight on the prior $P(u_i|m_i)$ derived from the context information that facilitates language comprehension. Importantly, we showed in this chapter that contextual facilitation (i.e., facilitatory effect of predictability) is observed only when the listeners attend to the entire sentence, including the context. In the next chapter, we further investigate this effect; specifically, we examine the granularity of the predictability effect at moderate levels of speech degradation. We also examine whether (un)certainty about next-trial speech degradation and predictability influences perceptual adaptation to degraded speech.

6

Semantic predictability facilitates comprehension of degraded speech in a graded manner

In the previous chapter we concluded that in a noisy channel, predictability facilitates comprehension of degraded speech only when listeners attended to the context. We also pointed out a few limitations of the study. In Experiment 2 of Chapter 5, there was an implicit assumption that all the noun-correct responses were borne out of a correct identification of the context evoking words (i.e., verbs). For a comparable analysis and results between Experiments 1 and 2, we only considered the accuracy of noun identification. In this chapter, we take into account listeners' identification of context (i.e., a verb that precedes the noun) while calculating the response accuracy. Importantly, we replicated the predictability effects in degraded speech comprehension in the previous chapter, showing a difference between comprehension of high and low predictability sentences. In this chapter, we extend it further and examine if the predictability is graded or all-or-nothing. We only showed that listeners do not adapt to degraded speech when there is a trial-by-trial variability in speech degradation. Here we report two experiments, investigating if listeners' adaptation to degraded speech is affected by such variability by comparing it against another condition in which speech degradation is kept constant. The results showed that in contrast to the "narrowed expectations" view postulated for the predictive processing of degraded speech, listeners probabilistically preactivate upcoming words from a wide range of semantic space, not limiting only to the highly probable sentence endings. We also did not find any learning effect on repeated exposure to degraded speech. We speculate that when there is a trial-by-trial variation in semantic feature (e.g., sentence predictability), listeners do not adapt to low-level perceptual property (e.g., speech quality) regardless of the certainty about the degradation level.

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 an all-or-nothing phenomenon (e.g., Coltheart, 2004; Kuperberg & Jaeger, 2016; Luke & Christianson, 2016). Very few studies have investigated such theoretical questions within the domain of adverse listening conditions, specifically in degraded speech comprehension (e.g., Corps & Rabagliati, 2020; StrauSS et al., 2013; cf. van Os et al., 2021). StrauSS et al. (2013) posited 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 many studies support the general idea of Strauß and colleagues that predictability facilitates comprehension of degraded speech (e.g., Hunter & Pisoni, 2018; Obleser & Kotz, 2010; Sheldon et al., 2008a), there have been no studies so far after Strauß and colleagues', to our knowledge, which examined the nature of predictability specifically in degraded speech comprehension.

In this chapter, our main aim is to attempt a replication of the previous findings of these predictability effects 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 generate expectations about an upcoming word based on how likely the word is to appear in the given context, and hence form a probabilistic prediction. We also test the presence of perceptual adaptation and its interplay with contextual facilitation. We set a metric of measurement of language comprehension that considers whether listeners correctly identified the context information.

6.2 Background

6.2.1 Predictability effects in degraded speech perception

We discussed in Chapter 2 (Section 2.2.2) that some studies (Obleser et al., 2007; Obleser & Kotz, 2010; Sheldon et al., 2008a; see also Hunter & Pisoni, 2018) have already shown the facilitatory effect of predictability in comprehension of degraded speech. For example, Obleser and colleagues compared high and low predictability sentences and observed contextual facilitation, in terms of the difference in response accuracy between high and low predictability sentences, at 8 channels and 4 channels noise-vocoded speech in their (2007) and (2010) studies respectively. However, these neuroimaging studies were not designed to test the nature of predictability effects.

In a modified experimental design, StrauSS et al. (2013) varied the target word predictability by manipulating its expectancy (i.e., how expected the target word is given the verb) and typicality (i.e., co-occurrence of target word and the preceding verb). They reported that at a moderate level of spectral degradation, N400 responses at highly predictable (strong-context and high-typical) words were smallest. In contrast, they found that the N400 effect (in terms of the amplitude of the N400 component) were largest at the strong-context, low typical words and the weak-context, low-typical words; the responses at the latter two were not statistically different from each other. The authors interpreted these findings as a facilitatory effect of sentence predictability which might be limited to only highly predictable sentence endings at a moderate level of spectral degradation. They proposed it as an *expectancy searchlight model*, and suggested that listeners form *narrowed expectations* from a restricted semantic space when the sentence endings are highly predictable. Their model posits that when the sentence endings are less predictable, listeners cannot preactivate those less predictable sentence endings in an adverse listening condition, namely, a severe spectral degradation. It is similar to the earlier observations made by Rayner et al. (2006; see Brothers & Kuperberg, 2021 for discussion), who found that reading times at low and medium predictability words were shorter than high predictability words, but it is contrary to the view that readers and listeners form a probabilistic prediction of upcoming word in a sentence. For example, Nieuwland et al. (2018)[1] showed in a large-scale replication study of DeLong et al. (2005) that the N400 amplitude at the sentence-final noun is directly proportional to its cloze probability across a range of high- and low-cloze words (see also Kochari & Flecken, 2019; Nicenboim et al., 2020). Heilbron et al. (2022) also showed that a probabilistic prediction model outperforms a constrained guessing model in predicting listeners' neural activities (MEG and EEG recordings), suggesting that linguistic prediction is probabilistic and it is not limited only to the highly predictable sentence endings, but it operates broadly in a wide range of probable sentence endings. However, when put in perspective with our research question, these studies were conducted in conditions without noise or degraded speech. And the ones that examined degraded speech comprehension used only two levels of semantic predictability (high and low). The granularity and the nature of prediction remain to be tested in degraded speech comprehension.

6.2.2 Adaptation to degraded speech

We discussed in Chpater 2 that studies have shown evidence for perceptual adaptation to artificially distorted speech. When exposed to noise-vocoded speech, listeners' word recognition accuracy is shown to increase over the course of experiment.

[1]However, Nieuwland et al. (2018) could not replicate DeLong et al. (2005)'s finding that comprehenders predict word-form. The N400 effect at the English articles *a/an* were not replicable.

Davis et al. (2005) and Erb et al. (2013) presented participants with 4- and 6-channel noise-vocoded speech, respectively in a single block. They found that the proportion of correctly reported words increased over the course of the experiment. It is important to note that in these experiments, listeners were not uncertain about the speech quality of any upcoming trial, i.e., the *global channel context* was certain or predictable. Additionally, there was no systematic variation of the semantic features of the words presented to the participants.

Listeners gradually map the representation of degraded lexicons around their "true" (or clear) schema/exemplars on repeated exposure (Goldstone, 1998; Nosofsky, 1986). With repeated exposure, the representation of degraded input comes closer to the exemplar, thereby, improving the performance. This feature-mapping mechanism proposes that the listeners map the whole feature of the sensory input. However, the higher level features of a speech (e.g., semantic property, like predictability) can also influence the acoustic realization of a degraded word, i.e., bottom-up processing of the degraded speech and perceptual adaptation (Gold & Watanabe, 2010; Goldstone, 1998; cf. Nahum et al., 2008). Listeners can assign weight to different dimensions of a speech stimuli (acoustic-phonetic and lexical-semantic). Performance improves over time when listeners give more attentional weight to the acoustic-phonetic dimension (cf. Haider & Frensch, 1996).

Thus, when multiple levels of degraded speech signals are presented in a (pseudo-)randomized order, then the listener is uncertain about the speech quality of any upcoming trial, i.e., the *global channel context* is certain or predictable. Such changes in the auditory features of the speech signal throughout the experiment is likely to render the perceptual adaptation totally absent (Mattys et al., 2012). Additionally, the trial-by-trial variability in the characteristics of speech signal can also impair word recognition (Sommers et al., 1994; see also Dahan & Magnuson, 2006). Very few studies have tried to study the influence of (un)certaintly about next-trial speech quality, and semantic feature in perceptual adaptation. For example, in a study by Vaden et al. (2013), participants were presented with words in a background noise at +3dB SNR and +10dB SNR in a pseudo-random order. They argued that an adaptive control system is involved to optimize task performance when there is an uncertainty about next trial. Similarly, Obleser and colleagues (Hartwigsen et al., 2015; Obleser et al., 2007; Obleser & Kotz, 2010) also presented listeners with multiple noise-vocoded speech (ranging from 2 to 32 channels noise-vocoding) in a pseudo-random order. However, none of these studies reported the change in listeners' performance over the course of the experiment. So, we cannont derive a conclusion from these studies regarding the effect of (un)certainty of the *global channel context* in perceptual adaptation and contextual facilitation.

As previously mentioned in Chapter 2 (Section 2.3), there are two conflicting findings in the literature on perceptual adaptation: On the one hand, repeated exposure to degraded speech leads to perceptual adaptation, and consequently improves word recognition over the course of experiment. On the other hand, uncertainty about next-trial speech quality is detrimental to word recognition.

6.2.3 Measurement of language comprehension

How we measure language comprehension has rarely been guided by any specific theoretical motive in the existing literature (cf. Amichetti et al., 2018). There is a discrepancy across studies in how language comprehension in degraded speech in quantified. Some studies that reported contextual facilitation in degraded speech comprehension used proportion of correctly reported *final* words only (e.g., Hunter & Pisoni, 2018; Obleser & Kotz, 2010; Sheldon et al., 2008a). Obleser et al. (2007) quantified language comprehension as the proportion of correctly identified key words in SPIN sentences. Erb et al. (2013) and Hakonen et al. (2017) used *report scores* (Peelle, 2013) that measure proportion of correctly recognized words per sentence as an index of language comprehension. Such inconsistencies make cross-study comparison difficult. None of the metrics outlined here take into account if listeners have correctly identified the context, which should be the most important factor to be considered in the first place. It is clear that if the context is misidentified, then the listeners are highly likely to misidentify the succeeding words (Marrufo-Pérez et al., 2019).

6.2.4 The present study

Stemming from the results of Chapter 5, and from the motivations outlined in the beginning of this chapter, the goals of the present study were threefold: The first goal was to attemnt to replicate the facilitatory effect of predictability examining the nature of predictability, i.e., to test if listeners form narrowed expectations. Obleser and colleagues (e.g., Obleser et al., 2007; Obleser & Kotz, 2010) have shown predictability effects (or contextual facilitation) to appear only at a moderate level of speech degradation by using only two levels of predictability (low and high). Our use of three levels of target word predictability (low, medium and high) will let us test the narrowed expectations view (StrauSS et al., 2013) by also taking into account the accuracy of context. If the listeners form narrowed predictions only for the target words with high cloze probability, then the facilitatory effect of semantic prediction will be observed only at these highly predictable sentence endings. Listeners' response to the target words with medium and low cloze probability would be quite similar, since these two fall out of the range of narrow prediction. However, if the listeners' predictions are not restricted to highly predictable target words, then they form predictions across a wide range of 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 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. In contrast to prior studies (e.g., Obleser & Kotz, 2010), inclusion of sentences with medium cloze target words thus allows us to differentiate whether listeners form all-or-nothing 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 low, medium and high 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 randomized across all levels of predictability, but i) blocked by each noise-vocoding channel, i.e. spectral degradation (*predictable channel context*) and ii) pseudo-randomized across all noise-vocoding channels (*unpredictable channel context*). Based on the 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, Garrido et al., 2011; Sommers et al., 1994; Vaden et al., 2013). To further examine perceptual adaptation, we also considered the effect of trial number in the analyses.

And third, we aimed at measuring language comprehension with a metric that reflects the participants' correct or incorrect identification of the context. If participants do not understand the context and we only measure the recognition of final word, this might not truly reflect the effect of contextual facilitation.

6.3 Methods

6.3.1 Participants

We recruited a group of participant via Prolific Academic (Prolific, 2014) and assigned them to the *predictable channel context* (n = 50; M age $\pm SD$ = 23.6$\pm$ 3.2 years; age range = 18-30 years; 14 females). Another group of 48 participants (n = 48; M age $\pm SD$ = 24.44$\pm$ 3.5 years; age range = 18-31 years; 16 females) from Experiment 2 of Chapter 5 were recruited and assigned to the *unpredictable channel context*. All participants were native speakers of German and reportedly did not have any speech-language disorder, hearing loss, or neurological disorder.

They received a monetary compensation of 6.20 Euro for their participation in the approximately 40 minutes long experiment.

6.3.2 Materials

We used the same stimuli described in Chapter 3 (Section 3.1). 120 sentences each for low predictability, medium predictability, and high predictability sentences that differed in the cloze probability of sentence final target words were used. Their mean cloze probabilities in the low, medium and high predictability sentences were 0.022 ± 0.027 ($M \pm SD$; range $= 0.00 - 0.09$), 0.274 ± 0.134 ($M \pm SD$; range $= 0.1 - 0.55$), and 0.752 ± 0.123 ($M \pm SD$; range $= 0.56 - 1.00$), respectively. All 360 sentences were then noise-vocoded through 1, 4, 6, and 8 channels to create degraded speech.

In the unpredictable channel context, each participant was presented with 120 unique sentences: 40 low predictability, 40 medium predictability, and 40 high predictability sentences. Channel condition (i.e., degradation level) was also balanced across each sentence type, i.e., in each of low, medium, and high predictability sentences, 10 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 three 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, one for each channel condition (i.e., degradation level). Thirty sentences were presented in each of the four blocks. In the first block, all sentences were 8 channels noise-vocoded, followed by the blocks of 6 channels, 4 channels, and 1 channel noise-vocoded speech consecutively (Sheldon et al., 2008a). Within each block, 10 low predictability, 10 medium predictability and 10 high predictability 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 confirmed there was a certainty of next-trial speech quality (within each block) and an uncertainty of next-trial sentence predictability across all four blocks.

6.3.3 Procedure

Participants were asked to use earphones or headphones. A sample of vocoded speech not used in the practice trial or the main experiment was presented to adjust the volume to their preferred level of comfort at the beginning of the experiment. The participants were instructed to listen and report the entire sentences by typing in the everything they heard using the keyboard. The time for typing in the response was not limited. They were informed at the beginning of the experiment that some of the sentences would be 'noisy' and not easy to understand. In these cases, guessing was encouraged. Eight practice trials with different levels of speech degradation were provided to the participants to familiarize them with the task before presenting all 120 experimental trials with an intertrial interval of 1,000 ms.

Each participant were presented with 40 high predictability, 40 medium predictability, and 40 low predictability sentences and the levels of speech degradation were balanced across each predictability level. For each of the three predictability conditions (high, medium, and low predictability), ten 1-channel, ten 4-channel, ten 6-channel, and ten 8-channel noise-vocoded sentences were presented, resulting in a total of 30 sentences at each degradation level. In the *unpredictable channel context*, the sentences were pseudo-randomized across degradation predictability conditions. No more than three sentences of the same degradation and predictability condition appeared consecutively. In the *predictable channel context*, sentences in were blocked by degradation level. There were four blocks corresponding to each degradation level. Within each block, predictability was randomized such that no more than three sentences of the same predictability condition appeared consecutively. The lists of both unpredictable and predictable channel contexts are presented in Appendix B and C.

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 identified the context independent of the succeeding words. 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 remaining trials in this condition.

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

Separately for each group (i.e., for predictable and unpredictable channel context), we first fitted a model with a maximal random effects structure that included random intercepts for each participant and item (Barr et al., 2013). Both by-participant and by-item random slopes were included for number of channels, target word predictability, and their interaction, which was supported by the experiment design. Following the procedure described in Chapter 4 (Section 4.2), we selected the optimal model, not necessarily the maximal model, that best fit our data. We applied treatment contrast for number of channels (8-channel as a baseline; factor levels: 8-channel, 4-channel, 6-channel) and sliding difference contrast for target word predictability (low vs medium predictability and low vs high predictability) and channel context (factor levels: unpredictable vs predictable).

6.5 Results and discussion

Mean response accuracy for different conditions in both channel contexts is shown in Tables 6.1, 6.2, and Figure 6.1. We found that accuracy increased when the number of noise-vocoding channels increased, and when the target word predictability increased. The results of the statistical analysis confirmed these observations (Table 6.3).

Table 6.1: Response accuracy (mean and standard error of the mean) across all levels of speech degradation and target word predictability in the predictable channel context

Number of channels	Target word predictability	Mean	Standard error
4	Low	71.59	2.74
	Medium	86.53	1.99
	High	93.53	1.42
6	Low	93.73	1.33
	Medium	96.21	1.08
	High	98.75	1.02
8	Low	97.84	0.80
	Medium	97.52	1.04
	High	99.38	0.59

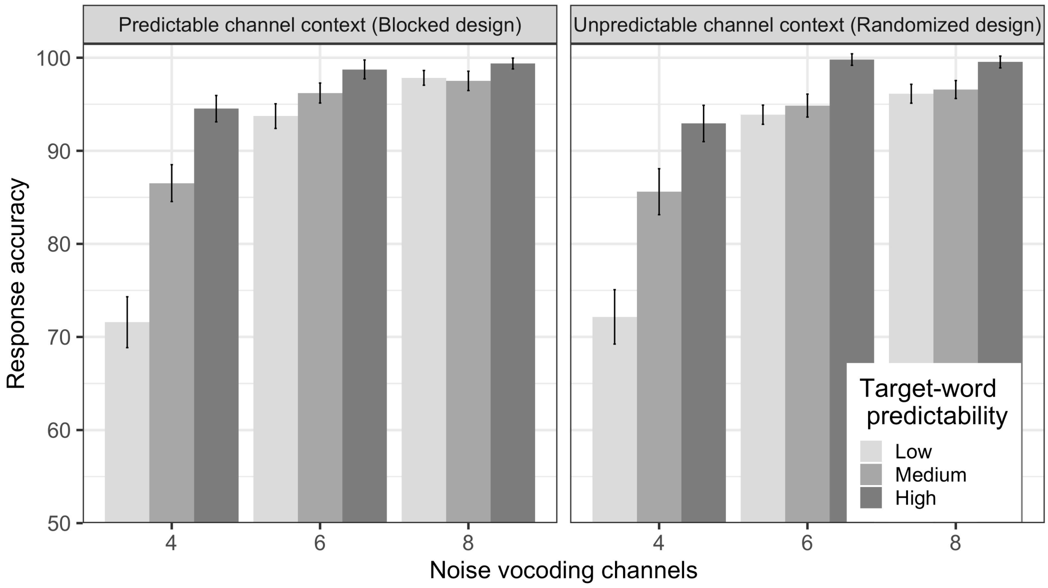

Figure 6.1: Mean response accuracy across all conditions in Experiment 2. Accuracy increased only with an increase in the number of noise-vocoding channels in both channel contexts. Only in the unpredictable globlal channel context, at the 4-channel noise-vocoding condition, we find a graded effect of prediction. Error bars represent the standard error of the means.

Table 6.2: Response accuracy (mean and standard error of the mean) across all levels of speech degradation and target word predictability in the unpredictable channel context

Number of channels	Target word predictability	Mean	Standard error
4	Low	72.16	2.93
	Medium	85.61	2.47
	High	92.94	1.96
6	Low	93.88	1.04
	Medium	94.86	1.24
	High	99.81	0.62
8	Low	96.14	1.02
	Medium	96.59	0.97
	High	99.55	0.64

We found a significant main effect of number of channels. The response accuracy at 8-channel was higher than for both 4-channel (β = -2.87, SE = .22, z(6917) = -13.1, p <.001), and 6-channel noise-vocoding (β = -.66, SE = .19, z(6917) = -3.42, p < .001). There was a significant main effect of target word predictability suggesting that the response accuracy in low predictability sentences was lower than in high predictability sentences (β = 2.18, SE = .3, z(6917) = 7.2, p < .001) and medium predictability sentences (β = -.52, SE = .27, z(6917) = -1.97, p = .049).

Table 6.3: Estimated effects of the model accounting for the correct word recognition

Fixed effects	Estimate	Std. Error	z value	p value
Intercept	5.09	.24	21.38	<.001
Noise condition (4-channel)	-2.87	.22	-13.10	<.001
Noise condition (6-channel)	-.66	.19	-3.42	.001
Target word predictability (Low-Medium)	-.52	.27	-1.97	.049
Target word predictability (High-Low)	2.18	.30	7.21	<.001
Noise condition (4-channel) × Target word predictability (Low-Medium)	-.71	.29	-2.44	.015
Global channel context × (Unpredictable - Predictable)	-..27	.14	-2.02	.043

We also found a significant interaction between number of channels and target word predictability (β = -.71, SE = .29, z(6917) = -2.44, p = .015). The interaction was driven by the effect of predictability at 4-channel condition: The accuracy in high predictability sentences was higher than in medium predictability sentences (β = 1.14, SE = .37, z(1608) = 3.1, p < .001) which in turn was higher than low predictability sentences (β = 1, SE = .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-channel (β = .33, SE = .32, z(2590) = 1.04, p = .3) and 8-channel (β = -.014, SE = .32, z(2719) = -.04, p = .97) conditions.

A subgroup analysis was also performed on each channel context. There was 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 = -.27, SE = .14, z(6917) = -2.02, p = .04).

To further address the question of perceptual adaptation, following the findings of Chapter 5, we also added trial number in the fixed effect. 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 in the analysis. We did not find a significant main effect of trial number indicating that the response accuracy did not change throughout the experiment (β = -.0004, SE = .01, $z(6917)$ = -.05, p = .97). It remained constant within each block in the predictable channel context (β = -.02, SE = .01, $z(3291)$ = -1.43, p = .15) as well as in the unpredictable channel context (β = .01 SE = .01, $z(3291)$ = 1.05, p = .29).

6.6 Conclusion

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.

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 by noise-vocoding through 1 channel. Listeners could not even identify the context at this severe degradation level. These results are fully in line with Obleser & Kotz (2010); we partly included identical sentences from their study in the present study (see Appendix A). Importantly, in contrast to their study, we 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 the sentences with low-cloze and high-cloze target words in 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 4-channel noise-vocoding condition show that the sentence-final word recognition is facilitated by semantic predictability in a graded manner, especially at a moderate level of speech degradation. This is in line with the findings from other experimental paradigms, including but not limited to the ERP literature where it has been observed that the 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; Nieuwland et al., 2018; Wlotko & Federmeier, 2012).

The interpretation of the observed graded effect of semantic predictability at the moderate level of spectral degradation 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-nothing phenomenon. Instead, listeners form a probabilistic prediction about the target word. The effect of target word predictability on comprehension is not sharply focused solely on high-cloze target words like a 'searchlight' as proposed by Strauß and colleagues. Rather, it is spread across a wide range including low- 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.

One could argue that the participants in our experiment "guessed" the verb after first correctly identifying the noun in a sentence. To rule out this possible explanation of our findings, we conducted an additional analysis compareing the *forward predictability effect* (from verb to noun) to the size of *backward predictability effect* (correct identification of the noun based on the final verb). If the observed effect is simply a guessing phenomenon, then we would expect that both forward and backward predictability effects are similar in size. If, on the other hand, understanding the verb really helps to shape predictions of the upcoming noun, and this helps intelligibility, then the forward prediction effect should be larger. The results of this complementary analysis (Table **??**) support the findings of the main analysis reported in the Results section. In the backward predictability analysis, there was no graded effect of predictability, and the backward effect of "guessing" the verb *jongliert* after recognizing the noun *Bälle*, if present at all, was smaller than the forward effect of predicting the noun after recognizing the verb in the sentence *Sie jongliert die Bälle*. This further corroborates our argument that the listeners, in fact, made use of the verb-evoked context to form predictions about the upcoming noun, not the other way around, in a graded manner when the speech was moderately degraded.

There was no learning effect or perceptual adaptation to degraded speech at the trial-by-trial level. We reason that the adaptation was hampered by a constant variation in the higher-level semantic feature (i.e., target word predictability).

The results of the analyses of trial number on the effect of channel context to capture trial-by-trial perceptual adaptation showed that the response accuracy did not increase over the course of experiment. This suggests that listeners' performance remained constant over the course of experiment regardless of certainty about the next-trial spectral degradation. One way by which perceptual adaptation occurs is when the perceptual system of a listener retunes itself to the sensory properties of the auditory signal, which can be facilitated by a feedback from higher-level lexical information (Goldstone, 1998; Mattys et al., 2012; cf. Davis et al., 2005). We reason that the trial-by-trial variability in the spectral resolution of the speech signal in the unpredictable channel context prevented perceptual adaptation. Although there was certainty about the quality of speech signal within a block in the predictable channel context, we did not observe trial-by-trial perceptual adaptation in this condition either. This is contrary to previous studies showing that listeners adapt to degraded speech when the global context of speech quality is predictable (Davis et al., 2005; Erb et al., 2013). However, the crucial difference between those studies and the study reported here is the manipulation of target word predictability. For example, Erb et al. (2013) presented sentences with only low predictability target words from the G-SPIN test. We, on the contrary, parametrically varied target word predictability from low to medium and high. Note that we presented target words in a randomized order in both channel contexts. This alone introduces trial-by-trial uncertainty in the predictable channel context and possibly hinders trial-by-trial perceptual adaptation. As Goldstone (1998), p. 588 notes – "one way in which perception becomes adapted to tasks and environments is by increasing the attention paid to perceptual dimensions and features that are important, and/or by decreasing attention to irrelevant [perceptual] dimensions and features" (see also Gold & Watanabe, 2010). A similar prediction is made by the Reverse Hierarchy Theory on auditory perception (Ahissar et al., 2009; Nahum et al., 2008). It posits that listeners first have access to the higher level features of a speech signal. If their task is to comprehend language, then they may not be able to access the lower level perceptual features. Consequently, they cannot adapt to the speech in an adverse listening condition. In our study, listeners paid more attention to semantic properties of the sentences (i.e., contextual cues and target word predictability) than to the perceptual properties (i.e., spectral resolution or speech quality) as the instruction was focused on "language comprehension" rather than "perception". We speculate this might have resulted in the absence of trial-by-trial perceptual adaptation to degraded speech, even when next-trial speech quality was predictable. Therefore, in both predictable and unpredictable channel contexts, perceptual learning of the degraded speech was hindered by trial-by-trial variation of either one (target word predictability) or both properties (target word predictability and spectral degradation level) of the speech stimuli.

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. The experimental paradigm in which sentence context is presented visually on a screen could likely circumvent the problem of context mis-identification, hence, one could simply use the word recognition score without considering the context recognition (e.g., van Os et al., 2021). However, such a paradigm introduces a confound: linguistic materials are processed differently in the brain when presented auditorily vs visually, and listeners do not necessarily identify visually presented context (e.g., visual half-field recognition superiority, McKeever & VanDeventer, 1977)

6.7 Summary

This chapter primarily investigated the nature of predictability effect. The experiment reported here provides a novel insight into predictive language processing when bottom-up signal quality is compromised and uncertain: We showed that while processing a 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 expectations* view, comprehension of words ranging from low- to high-cloze probability, including the medium-cloze probability, is facilitated in a graded manner. This contextual facilitation is observed while listening to a moderately degraded speech. Regardless of (un)certainty about the next-trial speech quality, we found that listeners do not adapt to degraded speech when semantic predictability varies constantly, i.e., higher level semantic features interfere with the lower-level perceptual properties. In the next chapter, we investigate the effect of changes in the lower-level bottom-up processing on top-down semantic predictions. Specifically, we examine how the change in speech rate moderates semantic predictions at the moderate level of speech degradation. The findings from the experiments reported in this chapter and the preceding chapter show that when listeners attend to the sentence context, predictability facilitates comprehension of moderately degraded speech. In the next chapter, we ask the question: Does an increase or decrease in the effortful perception of degraded speech along with an increase or decrease in the rate of information flow further increases or decreases the contextual facilitation?

7

Comprehension of degraded speech is modulated by the rate of speech

Abstract kind of text here.

7.1 Introduction

On the one hand, clean speech perception and reading studies have shown that contextual facilitation decreases with an increase in presentation rate (e.g., fast speech), with a mixed evidence on the enhancement of contextual facilitation with a decrease in presentation rate (e.g., slow speech). On the other hand, it has been shown that semantic predictability facilitates language comprehension at a moderate level of spectral degradation (e.g., at 4-channel noise-vocoded speech) while degraded speech perception is inherently effortful. Considering these two lines of research and their inconsistencies, the study reported in this chapter aimed to examine how a change in speech rate modulates contextual facilitation in language comprehension when the speech is moderately degraded by noise-vocoding through 4 channels. To this end, we conducted two experiments: In Experiment 1, we compared participants' word recognition in a sentence while they listened to the moderately degraded speech presented at a normal and fast speech rates (compressed by a factor of 0.65). In Experiment 2, we compared a separate group of participants' word recognition in a moderately degraded speech presented at a normal and slow speech rates (expanded by a factor of 1.35). The sentences varied in the degree of predictability of the sentence-final target word (high and low predictability). Results of this study demonstrate that fast speech limits the time for lexical processing. This time constraint interferes with the lexical processing of words disproportionately affecting the low predictability sentences on top of effortful listening of the moderately degraded speech In contrast, slow speech does not amplify the contextual facilitation — we found the lexical processing, context representation, and semantic predictions to be optimal at the normal speech rate when the speech is moderately degraded.

7.2 Background

When speech is degraded, its intelligibility and comprehension are reduced. For example, degradation by noise-vocoding reduces the spectral properties of speech rendering it difficult to understand (Davis et al., 2005; Shannon et al., 1995). Studies have shown that semantic predictability facilitates comprehension of moderately degraded speech (e.g., 4-channels noise-vocoded speech, Obleser & Kotz, 2010), which we have also replicated in the study presented in the previous chapter (Chapter 5). That is, listeners utilize context information and form predictions about upcoming linguistic units, which in turn facilitates the comprehension of the degraded speech. However, prediction is a time- and resource-consuming mechanism (Pickering & Gambi, 2018) such that an increase or decrease in speech rate can modulate a listener's ability to use available context information and generate linguistic predictions (cf. Cole, 2020; Ito et al., 2016). More processing time is available at slow presentation rates (slow speech) and less at fast presentation rates (fast speech). So, the contextual facilitation is reduced in fast speech. However, the evidence of enhanced contextual facilitation in slow speech is mixed (Aydelott & Bates, 2004; Goy et al., 2013; Koch & Janse, 2016). Specifically for degraded speech, there is no clear evidence on how the facilitatory effect at a moderate level of degradation is affected by different speech rates (Iwasaki et al., 2002; Meng et al., 2019; M. B. Winn & Teece, 2021).

Degraded speech is intrinsically effortful to listen to (Eckert et al., 2016; Wild et al., 2012). An increase (or decrease) in listening effort by a change in speech rate limits the cognitive resources available to encode the context information and form predictions (cf. Huettig & Mani, 2016). Therefore, the aim of the present study was to investigate the effect of a change in speech rate on listeners' ability to generate predictions while listening to a moderately degraded speech. One line of studies shows that at a moderate level of degradation (e.g., 4-channel noise-vocoding), semantic predictions facilitate language comprehension (e.g., Obleser et al., 2007). Another line of studies shows that an increase or decrease in speech rate modulates the predictability effect, i.e., contextual facilitation (e.g., Aydelott & Bates, 2004; Goy et al., 2013). The current study is driven by an interest to bring these two lines of research together to understand the effect of speech rate on the facilitatory effect of predictability in degraded speech comprehension. We wanted to investigate how contextual facilitation at a moderate level of spectral degradation is affected by the change in speech rate. We expected that the contextual facilitation in degraded speech comprehension would be reduced by an increase in speech rate, while a decrease in speech rate would increase the contextual facilitation.

7.2.1 Comprehension of fast and slow speech

A change in speech rate (by uniform time-compression or expansion) manipulates the speech signal but does not by itself produce a spectral degradation (Charpentier & Stella, 1986; Moulines & Charpentier, 1990; Schlueter et al., 2014; but see Longster, 2003). Understanding fast speech is more effortful compared to normal and slow speech (e.g., Müller et al., 2019; M. B. Winn & Teece, 2021; see also Simantiraki & Cooke, 2020), and its intelligibility and comprehension are reduced (Fairbanks & Kodman Jr., 1957; Peelle & Wingfield, 2005; Schlueter et al., 2014). With an increased speech rate, processing demand increases as less time is available to process the incoming information (Gordon-Salant & Fitzgibbons, 1995; Rodero, 2016; see also Rönnberg et al., 2013). Furthermore, some authors argue that the cognitive resources required for language processing are exhausted (Gordon-Salant & Fitzgibbons, 2004; Janse, 2009). Since cognitive resources are also required to encode and process the context information for generating predictions (Pickering & Gambi, 2018), it can be expected that the effect of predictability is reduced in fast speech. Studies comparing older and younger adults show that reduced intelligibility and comprehension in fast speech is associated with the limit of the central auditory processing system to process fast speech, identify the word, and activate its meaning (Wingfield et al., 1999; Wingfield et al., 2006; see also Connolly et al., 1990; Poldrack et al., 1998). Lerner et al. (2014) also show that the central auditory-language processing system is flexible and can rescales itself according to the speed of incoming information, i.e., the information processing speed in the auditory-language processing system can change in accordance with the change in speech rate, but there is an upper limit to the system's flexibility. Beyond a certain maximum speed of speech rate, the processing of fast speech becomes difficult.

In contrast, the central auditory-language comprehension system is shown to be flexible in processing slow speech without reducing its intelligibility to a certain lower limit of its rescaling capacity (Lerner et al., 2014). So, it can be expected that slow speech does not limit cognitive resources, and therefore processing context information to generate predictions in slow speech is not different from normal speech. Alternatively, slow speech provides more time to buffer the auditory information at the lower level of the information processing hierarchy (Ghitza & Greenberg, 2009; Vagharchakian et al., 2012) and consequently provides more time for the central auditory-language comprehension system to use the context information and form predictions. Studies from Visual World Paradigm also support this claim that slow speech provides more time for speech processing and semantic predictions (Fernandez et al., 2020; Huettig & Mani, 2016). However, some studies have cast doubt on the processing advantage of slow speech, arguing that slow speech is perceived as overly artificial and demands high working memory (e.g., Kemper & Harden, 1999; Nejime & Moore, 1998; see also Liu & Zeng, 2006; Love et al., 2009).

In both younger and older adults, Sommers et al. (2020) found that slow speech does not render additional benefit in a sentence comprehension task in noise, even when supported by a visual context. Therefore, given these competing accounts, it is unclear whether the effect of predictability increases in slow speech compared to normal speech. Nonetheless, it is clear that a change in speech rate has different effects on speech intelligibility and language comprehension: Fast speech reduces intelligibility and comprehension, but the evidence for the beneficial/neutral effect of slow speech on language comprehension is mixed.

A few studies have directly examined the role of fast and slow speech in listeners' use of and benefit from semantic context by using clean speech. For instance, Aydelott & 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 the facilitatory effect of contextual cues (congruent versus neutral target words) at fast speech compared to normal speech. In contrast, they found a reduced inhibitory effect (incongruent versus neutral target words). They argued that the constraining sentence context was easy to process — fast speech did not interfere with the earlier stage of activation of words that matched the context (i.e., in the 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 less inhibition of the incongruent target word was needed. However, in a replication study of Aydelott & 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, effectively reducing the contextual facilitation. In a recent study, M. B. Winn & Teece (2021) did not observe an increase in contextual facilitation for slow speech compared to normal speech, although the intelligibility was slightly higher for slow speech among cochlear implantees. In another experiment, Koch & Janse (2016) presented participants with a question-answer sequence of varying lengths across a wide range of normal and fast speech from the clean speech of Spoken Dutch Corpus (Oostdijk, 2000). They did not find any effect of predictability on word recognition. However, their study did not systematically control target word predictability and target word position in the sentences.

The effects of varying presentation rates on semantic predictability have also been investigated with self-paced reading studies. For example, Wlotko & 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 in the sentences that were presented fast compared to the ones that were presented slow. They suggested that at a fast presentation rate, predictive preactivation of words was not common: There was not enough time to activate proper representation to process upcoming words. In the same study, however, the semantic facilitation effect was not reduced when the slow presentation followed the fast presentation in separate blocks. That is, an increase in the flow of information did not always impair the ability to predict. They argued that once the brain is engaged in predictive comprehension mode, for example, first in the slow presentation rate, it then continues to allocate resources in the same mode under a faster presentation rate. Dambacher et al. (2012) also showed that the N400 effect was delayed and smaller at a fast presentation rate compared to slow presentation rates.

To summarize, there is already some evidence from studies applying various paradigms that the predictability of the sentence context interacts with the speech rate (Aydelott & Bates, 2004; Dambacher et al., 2012; Ito et al., 2016; Sharit et al., 2003; M. B. Winn & Teece, 2021; Wlotko & Federmeier, 2015). The predictability effect is generally reduced for fast speech, while the findings are inconsistent in the case of slow speech. Fast speech interferes with the lexical processing and activation of words that match the context, as limited time would be available to form expectations about an upcoming word (Dambacher et al., 2012; Goy et al., 2013). In contrast, slow speech can provide more time than a normal speech rate for listeners to form a rich context representation and generate a prediction about an upcoming word (cf. Huettig & Guerra, 2019).

7.2.2 Speech rate and contextual facilitation of moderately degraded speech

Predictions about upcoming linguistic units are generated as a listener forms a meaning representation of context information from a speech signal. Such linguistic predictions facilitate comprehension of degraded speech when the degradation is at a moderate level. However, the effect of predictability observed at the moderate degradation level no longer exists if the listener does not understand the context. Therefore, it is essential that the speech rate remains within the listener's limit to buffer and process the auditory information (Vagharchakian et al., 2012) so that the listener can form the representation of the context and have sufficient time to generate predictions.

Several studies have examined the role of speech rate on the intelligibility and comprehension of degraded speech but without considering predictability effects. For example, Meng et al. (2019) found that an increase in speech rate had a much more severe effect on spectrally degraded speech (4-channel sine-wave vocoded) than on clean speech. To achieve the same level of accuracy, listeners required degraded speech to be much slower than normal speech rate. Among cochlear implantees whose speech input is spectrally degraded (Shannon et al., 2004), Iwasaki et al. (2002) found that a change in speech rate from slow to fast reduced word recognition accuracy. Their speech perception was impaired with an increased speech rate, and it was improved when the speech rate was decreased (e.g., Dincer D'Alessandro et al., 2018). M. B. Winn & Teece (2021)'s study showed no significant difference in the facilitatory effect of semantic predictability between slow and normal speech rates. This was attributed to listeners' "repair" strategy at normal speech rate such that they made sensible guesses about the words that fit the given context. Similar to the studies conducted with clean speech, these studies also indicate that an increase in the speed of degraded speech is detrimental to its intelligibility, while its intelligibility increases with a decrease in speech rate.

Taken together, the utility of semantic predictability in comprehension of degraded speech is fairly established. However, the findings about the effect of speech rate on predictability effect in degraded speech comprehension are inconsistent. Similarly, prediction itself is a time- and resource-consuming mechanism (Pickering & Gambi, 2018) which is affected by a comprehender's processing speed (e.g., Huettig & Mani, 2016). However, the role of the speed of incoming information (i.e., speech rate of a degraded speech) on a listener's ability to form predictions, and hence its interplay with the facilitatory effect of semantic predictability at a moderately degraded speech, remains unclear. It can be speculated from the findings of the abovementioned studies that fast speech reduces the availability of time and resources to process the speech signal and generate predictions, in addition to the effortful listening of degraded speech. It can similarly be speculated that slow speech provides listeners more time than normal speech to process the words and form a representation of context information, in addition to reducing the effortful listening of degraded speech.

7.2.3 The present study

Semantic predictability has been shown to facilitate degraded speech comprehension when the degradation level is moderate at normal speech rate. We systematically examined whether contextual facilitation at a moderate level of degradation varies with a change in speech rate. That is, the aim of this study was to investigate whether an increase (and decrease) in speech rate reduces (and amplifies) the facilitatory effect of semantic predictability. Semantic predictability was manipulated by varying the cloze probability of target words in a sentence, and moderate degradation was achieved by noise-vocoding of speech through 4 channels. 4-channel noise-vocoding has been shown to be the moderate degradation level in the previous chapter, similar to the findings of Obleser & Kotz (2010). Speech rate was manipulated by time-compression (and expansion) of the moderately degraded speech to make it fast (and slow).

To achieve our aim, we conducted two experiments in which listeners were instructed to listen to the sentences and type in the entire sentence they heard. Sentence comprehension (word recognition accuracy) for high and low predictability sentences was assessed in fast speech (Experiment 1) and slow speech (Experiment 2). The processing demand increases, and a limited time is to be available to process the context and generate predictions with an increase in speech rate (e.g., Aydelott & Bates, 2004; Wlotko & Federmeier, 2015; see also Pickering & Gambi, 2018). Therefore, we expected that the contextual facilitation (i.e., the increase in word recognition accuracy in high predictability sentences compared to low predictability sentences) would be reduced for fast speech compared to normal speech (Experiment 1). However, for slow speech, due to an abundance of time to process the degraded speech and the context and reduced listening effort (e.g., M. B. Winn & Teece, 2021), we expected contextual facilitation to be increased compared to normal speech (Experiment 2). We expected that both increase and decrease in contextual facilitation would be primarily driven by the ease of processing high predictability sentences compared to low predictability sentences (Aydelott & Bates, 2004; Goy et al., 2013).

7.3 Experiment 1

7.3.1 Methods

Participants

We recruited one group of participant (n=101; (M age $\pm SD$ = 23.14 $\pm$ 3.31 years; age range = 18-31 years; 66 females, 1 preferred not to say) online via Prolific Academic. All participants were native speakers of German and did not have any speech-language disorder, hearing loss, or neurological disorder (all self-reported). All participants received 6.20 Euro as monetary compensation for their participation in the approximately 40 minutes long experiment.

Materials

We used a subset of materials created by the method described in Chapter 3 (Section 3.1) in this experiment. 120 sentences each for low predictability and high predictability sentences that differed in the cloze probability of sentence final target words were used. Mean cloze probabilities of the target words of low and high predictability sentences were 0.022 ± 0.027 ($M \pm SD$; range $= 0.00 - 0.09$) and 0.752 ± 0.123 ($M \pm SD$; range $= 0.56 - 1.00$) respectively. The audio recodings of all 240 sentences were compressed by a factor of 0.65 in Praat software to create fast speech (see Chapter 3 Section 3.1.2 for details). Then the recordings of speech signal at fast rate and normal rate were both passed through 4 channels of noise-vocoding to create moderately degraded speech stimuli of two types: fast speech and normal speech (see Chapter 3 Section 3.1.2 for details).

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.

Procedure

Participants were asked to use earphones or headphones. A sample of vocoded speech not used in the main experiment and the practice trial was provided so that the participants could adjust the volume to a preferred level of comfort at the beginning of the experiment. The participants were instructed to listen and report the entire sentences by typing in the everything they heard using the keyboard. The time for typing in the response was not limited. They were informed at the beginning of the experiment that some of the sentences would be 'noisy' and not easy to understand. In these cases, guessing was encouraged. To familiarize the participants with the task, eight practice trials with different levels of speech degradation were provided before presenting all 120 experimental trials with an intertrial interval of 1,000 ms.

Each participant had to listen to 60 high and 60 low predictability sentences. Speech rate was also balanced across each predictability condition. For each predictability condition, 30 sentences with fast speech and 30 with normal speech were presented. Sentences were pseudo-randomized so that no more than three sentences of the same predictability level or speech rate appeared consecutively. A total of four lists were constructed (see Appendix D).

7.3.2 Analyses

We have already shown in the previous chapters that predictability effects, i.e., contextual facilitation in language comprehension can be rightfully measured only if we consider the trials in which participants accurately identify the context. Verbs are predictive of the nouns in our stimuli (e.g., *Sie jongliert die Bälle*). Therefore, we discarded the trials in which verbs were identified incorrectly, which were XXXX out of XXXX trials.

Accuracy was analyzed using Generalized Linear Mixed Models (GLMMs) following the procedure described in Chapter 4 (Section 4.2) similar to the preceding chapters. Binary responses (categorical: correct and incorrect) for all participants were fit with a binomial linear mixed-effects model. Correct responses were coded as 1, and incorrect responses were coded as 0. Speech rate (categorical: fast speech and slow speech), target word predictability (categorical: high predictability sentences and low predictability sentences), and the interaction of speech rate (i.e., speed) and target word predictability were included in the fixed effects. We fitted a model with a 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 speech rate, target word predictability, and their interaction, which was supported by the experiment design. We applied treatment contrast for both predictability and speech rate, mapping low predictability and normal speech to the intercept.

7.3.3 Results and discussion

Table 7.1: Response accuracy (mean and standard error of the mean) across all levels of speech rate and target word predictability in Experiment 1

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

Mean response accuracies (in percentage) for all experimental conditions aggregated across all participants and items are shown in Table 7.1 and Figure 7.1. It shows that accuracy increases with an increase in target word predictability, but it decreases with an increase in speech rate. The results of the statistical analysis confirmed these observations (Table 7.2).

Table 7.2: Estimated effects of the model accounting for the correct word recognition in Experiment 1

Fixed effects	Estimate	Std. Error	z value	p value
Intercept	1.34	.24	5.58	<.001
Speech rate (Fast)	-.98	.24	-4.16	<.001
Target word predictability (High)	2.42	.28	8.55	<.001
Speech rate *times* Target word predictability	1.06	.42	2.50	.012

We found a significant main effect of target word predictability (β = 2.42, SE = .28, z = 8.55, p < .001) and a significant main effect of speech rate (β = -.98, SE = .24, z = 4.16, p < .001) suggesting that participants' response accuracy was

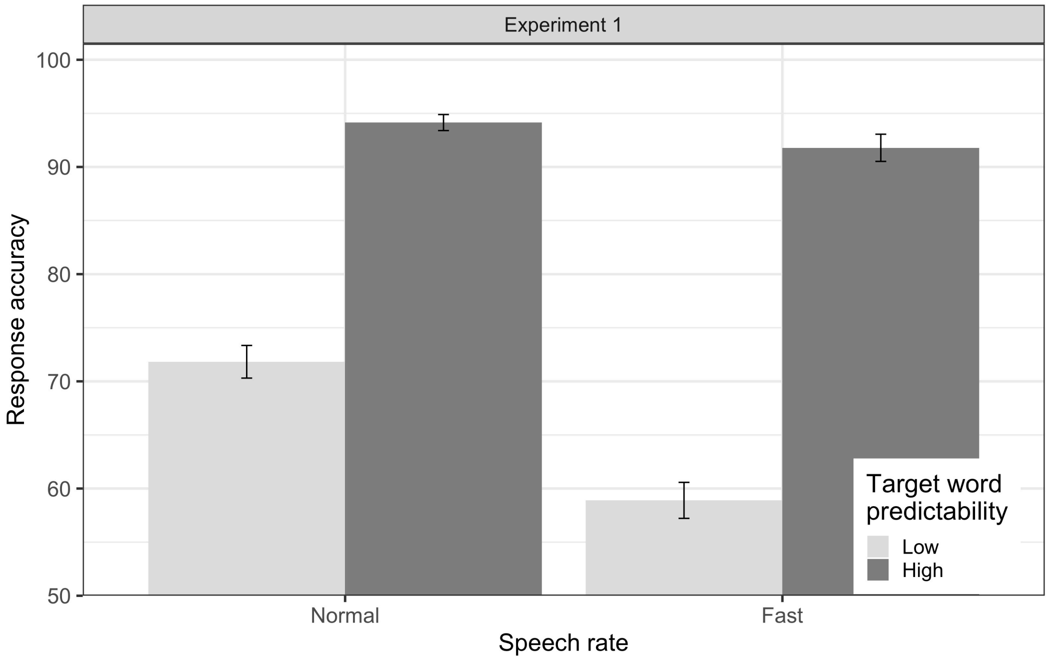

Figure 7.1: Mean response accuracy across all conditions in Experiment 1. Accuracy increased only with an increase in the target-word predictability and a decrease in speech rate. Error bars represent the standard error of the means.

higher for the high predictability sentences than for the low predictability sentences and for normal speech than for fast speech. We also found a significant interaction between target word predictability and speech rate (β = 1.06, SE = .42, z = 2.50, p = .01). These findings show that the effect of target word predictability, i.e., contextual facilitation was reduced at fast speech (see Figure 7.1).

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

Separate planned analyses of each speech rate revealed that there was significant main effect of predictability in both normal speech (β = 1.98, SE = .28, z = 7.05, p < .001) and fast speech (β = 2.67, SE = .37, z = 7.14, p < .001), but the effect appeared to be higher for fast speech (β = 2.67) than for normal speech (β = 1.98).

This resulted from a significant reduction in accuracy at the low predictability condition rather than due to an increase in accuracy at high predictability condition in the fast speech.

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 reduced the accuracy in the low predictability condition. It suggests that fast speech incurs a cost in processing the low predictability sentences and reduces the contextual facilitation in degraded speech comprehension. These findings also align with previous studies conducted with clean speech, which reported that fast speech reduces contextual facilitation (e.g., Aydelott & Bates, 2004).

We conducted another experiment to examine if slowing down the speech rate eases the processing of both low and high predictability sentences and increases the contextual facilitation in comprehension of moderately degraded speech.

7.4 Experiment 2

7.4.1 Methods

Participants and Materials

We recruited another group of participant (n=101; (M age $\pm SD = 23.49 \pm 3.26$ years; age range = 18-30 years; 60 females, 1 preferred not to say) online via Prolific Academic. Same sentences were used as stimuli as in Experiment 1. Instead of compression, the auditory recordings were expanded by a factor of 1.35 to create slow speech. All other procedure, including the noise-vocoding through 4 channels, to create stimuli were the same as in Experiment 1. This resulted in two types of 4-channel noise-vocoded speech: slow speech and normal speech.

Procedure

Same procedure was followed as Experiment 1. Participants were asked to use earphones or headphones. They were instructed to report the entire sentence by typing in what they heard.

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

7.4.2 Analyses

The data analysis procedure was the same as Experiment 1. Accuracy was analyzed using Generalized Linear Mixed Models (GLMMs). We fit a model with maximal random effects structure. Treatment contrast was applied for both predictability and speech rate, mapping low predictability and normal speech to the intercept.

7.4.3 Results and discussion

Table 7.3: Response accuracy (mean and standard error of the mean) across all levels of speech rate and target word predictability in Experiment 2

Speed	Target word predictability	Mean	Standard error
Slow	Low	70.92	1.09
	High	94.25	.89
Normal	Low	73.09	1.02
	High	94.82	.70

Mean response accuracies (in percentage) for all experimental conditions aggregated across all participants and items are shown in Table 7.3 and Figure 7.2. It shows that accuracy increases with an increase in target word predictability, but it did not increase with a decrease is speech rate. The results of the statistical analysis confirmed these observations (Table 7.4.

We again found 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 ($\beta = 2.58$, SE = .30, $z = 8.65$, $p < .001$). In contrast to Experiment 1, we did not find a significant main effect of speech rate (β

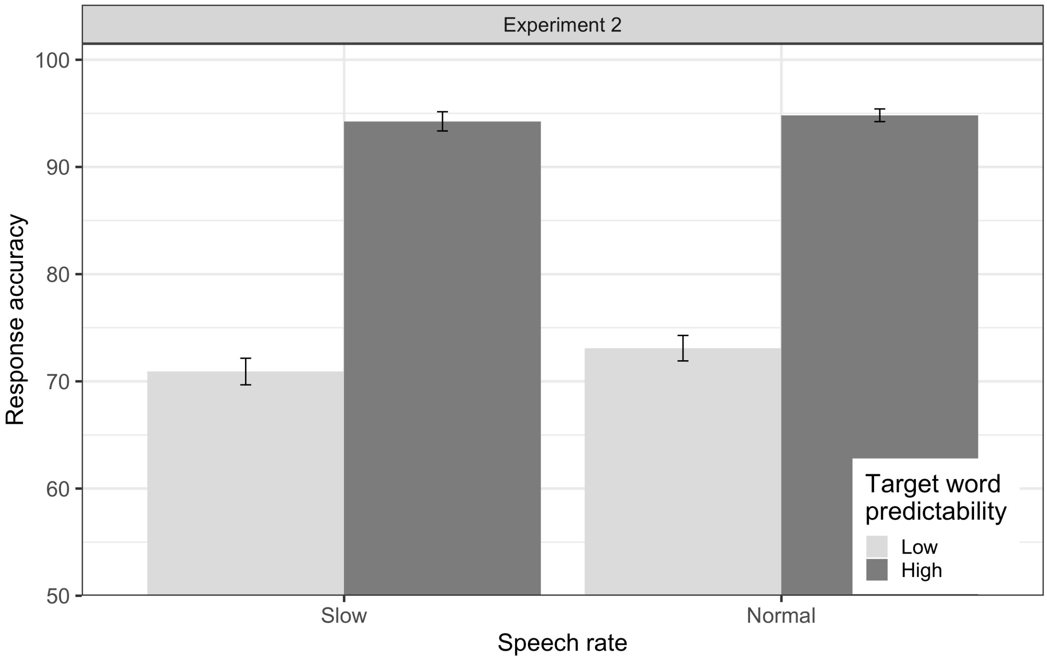

Figure 7.2: Mean response accuracy across all conditions in Experiment 2. Accuracy increased only with an increase in the target-word predictability, but a change in speech rate had no significant effect on accuracy. Error bars represent the standard error of the means.

Table 7.4: Estimated effects of the model accounting for the correct word recognition in Experiment 2

Fixed effects	Estimate	Std. Error	z value	p value
Intercept	1.41	.23	6.20	<.001
Speech rate (Slow)	-.08	.14	-57	.568
Target word predictability (High)	2.58	.30	8.65	<.001
Speech rate *times* Target word predictability	.44	.27	1.65	.099

= -.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.

In contrast to Experiment 1, the findings of Experiment 2 did not indicate a differential effect of speech rates in the comprehension of high and low predictability sentences. While 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 in both fast and slow speech than in normal speech, their ability to utilize context information was only impaired by the fast speech in the low predictability sentences.

7.5 Conclusion

The main goal of the present study was to examine whether the contextual facilitation (i.e., the facilitatory effect of semantic predictability) in comprehension of a moderately degraded speech is modulated by changes in speech rate. The results of two experiments revealed that fast speech selectively impedes the comprehension of low predictability sentences, while slow speech has no detrimental or beneficial effect on contextual facilitation.

In both experiments, our results showed a significant main effect of predictability at normal speech rate, i.e., we observed a facilitatory effect of semantic predictability at normal speech rate under moderate degradation level by noise-vocoding through 4 channels. This replicates the findings of earlier studies (e.g., Obleser & Kotz, 2010) and the findings of our own study presented in Chapter 6 in which participants were presented only with normal speech rate, and contextual facilitation was observed at the spectral degradation through 4-channel noise-vocoding. At this moderate degradation level, listeners could decode the context and form its meaning representation. Consequently, they generated predictions about the upcoming target word in a sentence even in low predictability conditions depending on the contextual constraint of the sentences (cf. 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 but not for high predictability sentences in fast speech. Listening to degraded speech is effortful and requires more attentional resources than 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 speech signal. The central auditory-language processing system does not rescale itself according to the speed of the fast speech presented in Experiment 1 (Lerner et al., 2014). It is then difficult to process the fast speech, decode the context information and form its meaning representation from the degraded speech to generate predictions about upcoming target words. This disproportionately affects low predictability sentences as fast speech interferes with the lexical processing of words, likely reducing the activation of target words in less constraining sentence contexts (Aydelott & Bates, 2004; Dambacher et al., 2012; cf. Goy et al., 2013). As a result, language comprehension in the low predictability condition is impaired more than in the high predictability condition in Experiment 1.

In contrast to Experiment 1, we did not find the expected interaction between speech rate and target word predictability in Experiment 2, i.e., a decrease in speech rate did not differentially affect the comprehension of high or low predictability sentences. Unlike Experiment 1, we did not observe a significant change in contextual facilitation in the slow speech at 4-channel noise-vocoding level in Experiment 2. Slowing down the speech gives listeners more time to process the information, including the context that is important to generate predictions. However, our findings show that the added time in slow speech does not benefit intelligibility and comprehension of sentences more than the normally available time at a normal speech rate. Comprehenders' lexical processing is optimal at the normal presentation rate (Dambacher et al., 2012). Although slow speech reduces effortful listening of degraded speech, the resources thus freed up by the slow speech are not allocated to enhance contextual facilitation. This argument is in line with the findings of M. B. Winn & Teece (2021), who reported that contextual facilitation does not increase when the speech is slowed down. Alternatively, it is plausible that the artificial expansion of speech introduced distortion in the speech signal. Although speech intelligibility and comprehension are increased by slow speech (Dincer D'Alessandro et al., 2018; Iwasaki et al., 2002), acoustic distortion due to artificial expansion reduces intelligibility and comprehension (Longster, 2003). As a result, we did not observe an overall amplification of contextual facilitation in the slow speech at the moderate degradation level in Experiment 2.

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

Similarly, in the "narrowed expectations" framework of degraded speech comprehension, StrauSS et al. (2013) argue that lexico-semantic cues are more robust to degradation and the target words can be processed faster and relatively more easily in a highly constraining context than in a less constraining context. They posit that listeners can activate a narrow range of most likely target words from the context in a high predictability sentence. Our results show that when the moderately degraded speech is speeded up, the context is still robust: There is enough processing time available in the fast speech to process the context and generate a small range of lexical predictions about an upcoming target word in a high predictability sentence. Therefore, the comprehension of high predictability sentences is not reduced due to an increase in speech rate. However, in a low predictability sentence, the range of probable sentence endings is too wide to generate enough lexical predictions that facilitate comprehension. When the processing time is short, it further reduces the activation of likely sentence endings, especially when the context is less constraining (Aydelott & Bates, 2004).

An alternative explanation of our findings of contextual facilitation could be that the listeners first identified the noun (e.g., Bälle), then integrated it with the verb (e.g., jongliert) instead of first identifying the verb and predicting the noun. Therefore, we conducted a set of complementary analyses (see the Supplementary Material) that supports the prediction-based explanation. We compared the effect (estimates) of forward-predictability (from the verb to noun) with that of backward integration (identifying the verb from the correct identification of the noun). In both experiments, the forward predictability effect was larger than the backward integration. This finding favours the explanation that the contextual facilitation we observed is due mainly to predictability rather than guessing or backward integration.

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

To conclude, we show that access (or restriction) to lexical processing is associated with the speed of information flow, and the constraints in attentional and cognitive resources are key factors that influence contextual facilitation of moderately degraded speech. Lexical processing is restricted in an effortful listening of the rapidly unfolding degraded speech; consequently, understanding the words that are not easily predictable from the context is difficult. On the contrary, auditory-language comprehension is optimal in the normal speech rate, and thus, slowing down the degraded speech does not necessarily amplify contextual facilitation.

7.6 Summary

Some summary statements

8
Discussion and conclusion

While there is no doubt that human readers an listeners use context information and form semantic predictions, there have been questions about the ubiquity of predictive nature of language processing, in particular, language comprehension. In this thesis, we examined predictive language processing while listeners are presented with degraded speech. By manipulating top-down and bottom-up processes, we examined the interactions of these two processes in understanding speech in adverse listening conditions. Our findings support the predictive processing account of language comprehension. Below we synthesize the results from the experiments in the preceding chapters, and put forth a nuanced argument in favor of the facilitatory effect of predictability in degraded speech comprehension.

8.1 Summary of the main findings

In Chapter 5, we investigated the role of attention to sentence context in semantic prediction. The results showed that listeners can use context information and form predictions about upcoming linguistic units only if they are attending to the context in the sentence. This finding highlights the critical role of attention in prediction in degraded speech comprehension. It also shows that prediction is not an automatic phenomenon; rather, prediction depends on listeners' attention to the context, and them forming its meaning representation.

In Chapter 6, we further examined the nature of contextual facilitation that was observed in Chapter 5. The results showed that semantic predictability facilitated comprehension of degraded speech at a moderate level of degradation. Importantly, such predictability effect was graded in nature. The latter finding refuted the claims made by 'narrowed expectations' views of StraSSe and colleagues (StrauSS et al., 2013).

In Chapter 7, we examined if the facilitatory effect of predictability at the moderate spectral degradation is influenced by bottom-up factor, like temporal degradation of speech (i.e., fast and slow speech). The findings show that fast speech incurs processing cost when the sentence endings are less predictable from the context, therefore, contextual facilitation is reduced in fast speech. In slow speech, on the other hand, no change is observed in the facilitatory effect, as the listeners are at their optimal level of performance already at the normal speech rate.

Results from Chapters 5 and 6 also showed that listeners do not adapt to degraded speech despite being exposed to it throughout the experiment. The change in target word predictability, i.e., semantic feature of the speech signal potentially interferes in adaptation to lower level perceptual property of the speech.

Taken together, the findings from these three chapters can be stated as: At a normal speech rate, when listeners attend to sentence context, semantic predictability facilitates language comprehension at a moderate level of spectral degradation.

8.2 Theoretical and practical implications

8.3 Limitation

- Iron law /Stanovic's law

8.4 Concluding remarks

Dont introduce new material, e.g., theories, results, data, etc. Reflect on the results, chapters. Create a satisfactory ending.

Sell your wares. You have made original contribution: "My original contribution to knowledge is ..."

Conclusion is more than a summary. Summary is just a part of the conclusion. Give 'what' happened in the thesis but also 'why'. Why do the chapters matter?

How you have contributed to the discipline? Different than contribution to the knowledge. Tell about local context. Position in the world.

What went wrong? What would you do differently?

Future work. It follows from 'what went wrong'.

Extendability. What / how to extend from your work.

Impact. Impact to the society; social, theoretical, financial, etc. This is different than contribution to knowledge and discipline.

Writing matters. Why should someone give a damn about your thesis?

Don't underclaim, dont't overclaim. About implications. What does the summary mean, what answers does it provide, does it not provide. Offer a series of pathways to future research. [My thesis has done X, it leads to Y; cite literature.] –> repeat recursively.

All in all, conclusions mean: Why does the thesis matter? Don't give new material. Give a link between chapters.

Don't say what you think, say what the data speaks.

9:57 1. Conclusions spring form the thesis 11:33 2. Sell your wares 12:52 3. A PhD conclusion is more than a summary 14:11 4. Significance to the discipline 15:39 5. What went wrong? 16:41 6. Future work 17:41 7. Impact, relevance and consequences 19:25 8. Writing matters 22:57 9. Don't underclaim and don't overclaim 25:29 10. Don't rely on the examiner's benevolence

Previously reported facilitatory effect of semantic predictability comes from studies conducted in laboratory setups. The current experiment was conducted online. There is a possibility that different participants used different hearing devices. Such variability in hearing devices could not be controlled for in our experiment although the participants were restricted to using only desktop/laptop computers. However, we have no reason to believe that these variance sources are systematically correlated with our between-group manipulation (global channel context) and the effects are constant within subjects. Moreover, the main finding of this study, i.e., the graded effect of semantic predictability, is observed in both the groups.

9
Ethics and funding

Ethics: The studies presented in this thesis involved human subjects. All subjects were recruited following the recommendations of the American Psychological Association. All subjects provided an informed consent in accordance with the Declaration of Helsinki. The ethics committee of the Deutsche Gesellschaft für Sprache (DGfS; EN: German Society for Language Science) provided ethical approval for the experiments conducted.

Funding: The research presented in this thesis was funded by the Deutsche Forschungsgemeinschaft (DFG; EN: German Research Foundation) under the research grant SFB1102 (Sonderforschungsbereiche; EN: Collaborative Research Center), Project ID 232722074.

Appendices

A

Experimental items

This is a list of high, medium and low predictability sentences used in all the experiments mentioned in Chapters 5, 6, and 7.

–> Insert table here <–

Bibliography

Aarts, A. A., Anderson, J. E., Anderson, C. J., Attridge, P. R., Attwood, A., Axt, J., Babel, M., Bahník, ., Baranski, E., Barnett-Cowan, M., Bartmess, E., Beer, J., Bell, R., Bentley, H., Beyan, L., Binion, G., Borsboom, D., Bosch, A., Bosco, F. A., ... Zuni, K. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 285–299. https://doi.org/10.1098/rstb.2008.0253

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaksi (Eds.), *Proceedings of the 2nd international symposium on information theory* (pp. 267–281). Akademiai Kaido.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/s0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Amichetti, N. M., Atagi, E., Kong, Y.-Y., & Wingfield, A. (2018). Linguistic Context Versus Semantic Competition in Word Recognition by Younger and Older Adults With Cochlear Implants. *Ear & Hearing*, *39*(1), 101–109. https://doi.org/10.1097/aud.0000000000000469

Ankener, C. S. (2019). *The influence of visual information on word predictability and processing effort* [Doctoral dissertation]. Saarland University; Saarländische Universitäts-und Landesbibliothek.

Ankener, C. S., Sekicki, M., & Staudte, M. (2018). The influence of visual uncertainty on word surprisal and processing effort. *Frontiers in Psychology*, *9*, 2387. https://doi.org/10.3389/fpsyg.2018.02387

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K.

(2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, *53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*, *19*(1), 29–56. https://doi.org/10.1080/01690960344000099

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. *arXiv*. https://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Berners-Lee, T., Cailliau, R., Groff, J. F., & Pollermann, B. (1992). World-wide web: The information universe. *Internet Research*, *2*(1), 52–58. https://doi.org/10.1108/eb047254

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341–345.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Bondell, H. D., Krishna, A., & Ghosh, S. K. (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*, *66*(4), 1069–1077.

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450

Brothers, T., & Kuperberg, G. R. (2021). Word predictability effects are linear, not logarithmic: Implications for probabilistic models of sentence comprehension. *Journal of Memory and Language*, *116*(September 2020), 104174. https://doi.org/10.1016/j.jml.2020.104174

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135–149. https://doi.org/10.1016/j.cognition.2014.10.017

Bruineberg, J., Dolega, K., Dewhurst, J., & Baltieri, M. (2021). The emperor's new markov blankets. *Behavioral and Brain Sciences*, 1–63. https://doi.org/10.1017/S0140525X21002351

Brunellière, A., Auran, C., & Delrue, L. (2019). Does the prosodic emphasis of sentential context cause deeper lexical-semantic processing? *Language, Cognition and Neuroscience*, *34*(1), 29–42. https://doi.org/10.1080/23273798.2018.1499945

Charpentier, F. J., & Stella, M. G. (1986). Diphone synthesis using an overlap-add technique for speech waveforms concatenation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2015–2018. https://doi.org/10.1109/icassp.1986.1168657

Chatterjee, S., & Hadi, A. S. (2012). Multiple linear regression. In *Regression analysis by example* (pp. 57–91). John Wiley & Sons.

Chen, F., & Loizou, P. C. (2011). Predicting the intelligibility of vocoded speech. *Ear and Hearing*, *32*(3), 331.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Christiansen, M. H., & Chater, N. (2015). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, 1–72. https://doi.org/10.1017/S0140525X1500031X

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/s0140525x12000477

Clark, H. H. (1973). The language-as-a-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Clendeninn, P. (1940). The vocoder. *Nature*, *145*(3665), 157. https://doi.org/10.1038/145157a0

Cockburn, A., Dragicevic, P., Besançon, L., & Gutwin, C. (2020). Threats of a replication crisis in empirical computer science. *Communications of the ACM*, *63*(8), 70–79. https://doi.org/10.1145/3360311

Cole, A. (2020). *The effects of prediction and speech rate on lexical processing* (p. 35) [Masters thesis, University of Maryland]. https://doi.org/10.13016/37my-jxp7

Coleman, E. B. (1964). Generalizing to a language population. *Psychological Reports*, *14*(1), 219–226. https://doi.org/10.2466/pr0.1964.14.1.219

Coltheart, M. (2004). Are there lexicons? *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *57*(7), 11531171. https://doi.org/10.1080/02724980443000007

Connolly, J. F., Stewart, S., & Phillips, N. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, *39*(2), 302–318.

Cooke, M., & Garcia Lecumberri, M. (2021). Estimating the performance gap between lab and remote speech perception experiment. *Acoustical Society of America Journal*, *149*(4), A111–A111.

Cooke, M., Scharenborg, O., & Meyer, B. T. (2022). The time course of adaptation to distorted speech. *The Journal of the Acoustical Society of America*, *151*(4), 2636–2646.

Corps, R. E., & Rabagliati, H. (2020). How top-down processing enhances comprehension of noise-vocoded speech: Predictions about meaning are more important than predictions about form. *Journal of Memory and Language*, *113*, 104114. https://doi.org/10.1016/j.jml.2020.104114

Dahan, D., & Magnuson, J. S. (2006). *Spoken word recognition* (pp. 249–283). Elsevier. https://doi.org/10.1016/b978-012369374-7/50009-2

Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia*, *50*(8), 1852–1870. https://doi.org/10.1016/j.neuropsychologia.2012.04.011

Darwiche, A. (2010). Bayesian networks. *Communications of the ACM*, *53*(12), 80–90. https://doi.org/10.1145/1859204.1859227

Darwin, C. (2005). *Praat scripts for producing shannon AM speech*. http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics*, *39*(4), 1025–1066.

Dincer D'Alessandro, H., Boyle, P. J., Ballantyne, D., De Vincentiis, M., & Mancini, P. (2018). The role of speech rate for Italian-speaking cochlear implant users: insights for everyday speech perception. *International Journal of Audiology*, *57*(11), 851–857. https://doi.org/10.1080/14992027.2018.1498139

Dudley, H. (1939). The vocoder. *Bell Laboratories Record*, *18*(4), 122–126.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 914–927. https://doi.org/10.1037/0096-1523.23.3.914

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012

Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. (2016). Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. *Ear & Hearing*, *37*(1), 101S–110S. https://doi.org/10.1097/aud.0000000000000300

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

Ellermeier, W., Kattner, F., Ueda, K., Doumoto, K., & Nakajima, Y. (2015). Memory disruption by irrelevant noise-vocoded speech: Effects of native language and the number of frequency bands. *The Journal of the Acoustical Society of America*, *138*(3), 1561–1569.

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

Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2013). The Brain Dynamics of Rapid Perceptual Adaptation to Adverse Listening Conditions. *Journal of*

Neuroscience, *33*(26), 10688–10697. https://doi.org/10.1523/jneurosci.4596-12.2013

Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1–20. https://doi.org/10.3758/s13428-021-01694-3

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

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149–161. https://doi.org/10.1016/j.bandl.2010.07.006

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*(1), 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Federmeier, K., Mclennan, D., Ochoa, E. de, & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(02), 133–146.

Federmeier, K., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Fernandez, L. B., Engelhardt, P. E., Patarroyo, A. G., & Allen, S. E. M. (2020). Effects of speech rate on anticipatory eye movements in the visual world paradigm: Evidence from aging, native, and non-native language processing. *Quarterly Journal of Experimental Psychology*, *73*(12), 2348–2361. https://doi.org/10.1177/1747021820948019

Ferreira, F., & Chantavarin, S. (2018). Integration and Prediction in Language Processing: A Synthesis of Old and New. *Current Directions in Psychological Science*, *27*(6), 443–448. https://doi.org/10.1177/0963721418794491

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368.

Ferreira, F., & Lowder, M. W. (2016). *Prediction, Information Structure, and Good-Enough Language Processing* (Vol. 65, pp. 217–247). Elsevier Ltd. https://doi.org/10.1016/bs.plm.2016.04.002

Fontan, L., Tardieu, J., Gaillard, P., Woisard, V., & Ruiz, R. (2015). Relationship Between Speech Intelligibility and Speech Comprehension in Babble Noise. *Journal of Speech, Language, and Hearing Research*, *58*(3), 977–986. https://doi.org/10.1044/2015_jslhr-h-13-0335

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, *33*(4), 465–495.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(5), 862–877. https://doi.org/10.1037/0278-7393.31.5.862

Friston, K. J., Parr, T., Yufik, Y., Sajid, N., Price, C. J., Holmes, E., & Square, Q. (2020). Generative models, linguistic communication and active inference. *Neuroscience & Biobehavioral Reviews*, *118*, 42–64. https://doi.org/10.1016/j.neubiorev.2020.07.005

Friston, K. J., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., & Holmes, E. (2020). Active listening. *Hearing Research*, *xxxx*, 107998. https://doi.org/10.1016/j.heares.2020.107998

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention - focusing the searchlight on sound. *Current Opinion in Neurobiology*, *17*(4), 437–455. https://doi.org/10.1016/j.conb.2007.07.011

Gadiraju, U., Möller, S., Nöllenburg, M., Saupe, D., Egger-Lampl, S., Archambault, D., & Fisher, B. (2017). Crowdsourcing versus the laboratory: Towards human-centered experiments using the crowd. In D. Archambault, H. Purchase, & T. HoSSfeld (Eds.), *Evaluation in the crowd. Crowdsourcing and human-centered experiments* (pp. 6–26). Springer, Cham. https://doi.org/10.1007/978-3-319-66435-4_2

Gagné, N., & Franzen, L. (2021). *How to run behavioural experiments online: Best practice suggestions for cognitive psychology and neuroscience.*

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110.

Garrido, M. I., Dolan, R. J., & Sahani, M. (2011). Surprise Leads to Noisier Perceptual Decisions. *I-Perception*, *2*(2), 112–120. https://doi.org/10.1068/i0411

Garvey, W. D. (1953). The intelligibility of speeded speech. *Journal of Experimental Psychology*, *45*(2), 102–108. https://doi.org/10.1037/h0054381

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*(1-2), 113–126. https://doi.org/10.1159/000208934

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Gold, J. I., & Watanabe, T. (2010). Perceptual learning. *Current Biology*, *20*(2), R46–R48. https://doi.org/10.1016/j.cub.2009.10.066

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585–612. https://doi.org/10.1146/annurev.psych.49.1.585

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

Gordon-Salant, S., & Fitzgibbons, P. J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *The Journal of the Acoustical Society of America*, *115*(4), 1808–1817. https://doi.org/10.1121/1.1645249

Goy, H., Pelletier, M., Coletta, M., & Pichora-Fuller, M. K. (2013). The Effects of Semantic Context and the Type and Amount of Acoustic Distortion on Lexical Decision by Younger and Older Adults. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1715–1732. https://doi.org/10.1044/1092-4388(2013/12-0053)

Greenberg, S. (1996). Auditory processing of speech. In N. J. Lass (Ed.), *Principles of experimental phonetics* (pp. 362–407). Mosby, St. Louis.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *Journal of the Acoustical Society of America*, *87*(6), 2592–2605. https://doi.org/10.1121/1.399052

Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology*, *24*(4), 699–711. https://doi.org/10.1111/j.14

20-9101.2010.02210.x

Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2014). Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience*, *7*(Jan), 1–16. https://doi.org/10.3389/fnsys.2013.00126

Hafter, E. R., Sarampalis, A., & Loui, P. (2007). Auditory Attention and Filters. In *Auditory perception of sound sources* (Vol. 29, pp. 115–142). https://doi.org/10.1007/978-0-387-71305-2_5

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*(3), 304–337.

Hakonen, M., May, P. J. C., Jääskeläinen, I. P., Jokinen, E., Sams, M., & Tiitinen, H. (2017). Predictive processing increases intelligibility of acoustically distorted speech: Behavioral and neural correlates. *Brain and Behavior*, *7*(9), e00789. https://doi.org/10.1002/brb3.789

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second Meeting of the North American Chapter of the Association for Computational Linguistics.*

Hartwigsen, G., Golombek, T., & Obleser, J. (2015). Repetitive transcranial magnetic stimulation over left angular gyrus modulates the predictability gain in degraded speech comprehension. *Cortex*, *68*, 100–110. https://doi.org/10.1016/j.cortex.2014.08.027

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), e2201968119. https://doi.org/10.1073/pnas.2201968119

Hervais-Adelman, A., Kumar, U., Mishra, R. K., Tripathi, V. N., Guleria, A., Singh, J. P., Eisner, F., & Huettig, F. (2019). Learning to read recycles visual cortical networks without destruction. *Science Advances*, *5*(9), eaax0262.

Heyselaar, E., Peeters, D., & Hagoort, P. (2021). Do we predict upcoming speech content in naturalistic environments? *Language, Cognition and Neuroscience*, *36*(4), 440–461. https://doi.org/10.1080/23273798.2020.1859568

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*(June 2017), 196–208. https:

//doi.org/10.1016/j.brainres.2018.11.013

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31. https://doi.org/10.1080/23273798.2015.1072223

Hunter, C. R., & Pisoni, D. B. (2018). Extrinsic cognitive load impairs spoken word recognition in high-and low-predictability sentences. *Ear and Hearing*, *39*(2), 378–389. https://doi.org/10.1097/AUD.0000000000000493

Husband, E. M., & Bovolenta, G. (2020). Prediction failure blocks the use of local semantic context. *Language, Cognition and Neuroscience*, *35*(3), 273–291. https://doi.org/10.1080/23273798.2019.1651881

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

Iwasaki, S., Ocho, S., Nagura, M., & Hoshino, T. (2002). Contribution of speech rate to speech perception in multichannel cochlear implant users. *Annals of Otology, Rhinology and Laryngology*, *111*(8), 718–721. https://doi.org/10.1177/000348940211100811

Jackendoff, R. (2002). *Foundations of language: How language connects to the brain, the world, evolution, and thinking*. Oxford University Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Janse, E. (2009). Processing of fast speech by elderly listeners. *The Journal of the Acoustical Society of America*, *125*(4), 2361–2373. https://doi.org/10.1121/1.3082117

Johnson, B. P., Dayan, E., Censor, N., & Cohen, L. G. (2021). Crowdsourcing in cognitive and systems neuroscience. *The Neuroscientist*, 10738584211017018. https://doi.org/10.1177/10738584211017018

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–188. https://doi.org/10.1017/S0140525X10003134

Kaiser, E., & Trueswell, J. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147. https://doi.org/10.1016/j.cognition.2004.01.002

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/s0749-596x(03)00023-8

Kaufeld, G. (2021). *Investigating spoken language comprehension as perceptual inference* (p. 183) [Doctoral dissertation]. Max Planck Research School (IMPRS) for Language Sciences.

Kemper, S., & Harden, T. (1999). Experimentally disentangling what's beneficial about elderspeak from what's not. *Psychology and Aging*, *14*(4), 656–670. https://doi.org/10.1037/0882-7974.14.4.656

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America*, *139*(4), 1618–1636. https://doi.org/10.1121/1.4944032

Kochari, A. R., & Flecken, M. (2019). Lexical prediction in language comprehension: a replication study of grammatical gender effects in Dutch. *Language, Cognition and Neuroscience*, *34*(2), 239–253. https://doi.org/10.1080/23273798.2018.1524500

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, *22*(9), 2197–2206. https://doi.org/10.1093/cercor/bhr310

Kuperberg, G. R. (2021). Tea with milk? A hierarchical generative framework of sequential event comprehension. *Topics in Cognitive Science*, *13*(1), 256–298. https://doi.org/10.1111/tops.12518

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lange, K. (2013). The ups and downs of temporal orienting: A review of auditory temporal orienting studies and a model associating the heterogeneous findings on the auditory N1 with opposite effects of attention and prediction. *Frontiers in Human Neuroscience*, *7*, 1–14. https://doi.org/10.3389/fnhum.2013.00263

Lange, K., & Röder, B. (2010). Temporal orienting in audition, touch, and across modalities. *Attention and Time*, 393–405.

Lecompte, D. C. (1995). An irrelevant speech effect with repeated and continuous background speech. *Psychonomic Bulletin & Review*, *2*(3), 391–397.

Leensen, M. C. J., & Dreschler, W. A. (2013). Speech-in-noise screening tests by internet, Part 3: Test sensitivity for uncontrolled parameters in domestic usage. *International Journal of Audiology*, *52*(10), 658–669. https://doi.org/10.3109/14992027.2013.803610

Lerner, Y., Honey, C. J., Katkov, M., & Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *Journal of Neurophysiology*, *111*(12), 2433–2444. https://doi.org/10.1152/jn.00497.2013

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Li, J., Xia, R., Ying, D., Yan, Y., & Akagi, M. (2014). Investigation of objective measures for intelligibility prediction of noise-reduced speech for Chinese, Japanese, and English. *The Journal of the Acoustical Society of America*, *136*(6), 3301–3312. https://doi.org/10.1121/1.4901079

Li, X., Zhang, Y., Li, L., Zhao, H., & Du, X. (2017). Attention is shaped by semantic level of event-structure during speech comprehension: An electroencephalogram study. *Cognitive Neurodynamics*, *11*(5), 467–481. https://doi.org/10.1007/s11571-017-9442-4

Lieberman, P. (2013). The unpredictable species. In *The unpredictable species*. Princeton University Press.

Liu, S., & Zeng, F.-G. (2006). Temporal properties in clear speech perception. *The Journal of the Acoustical Society of America*, *120*(1), 424–432. https://doi.org/10.1121/1.2208427

Loizou, P. C., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand speech. *The Journal of the Acoustical Society of America*, *106*(4), 2097–2103. https://doi.org/10.1121/1.427954

Longster, J. A. (2003). *Concatenative Speech Synthesis : A Framework for Reducing Perceived Distortion when using the TD-PSOLA Algorithm* [Doctoral dissertation]. Bournemouth University.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 149.

Love, T., Walenski, M., & Swinney, D. (2009). Slowed speech input has a differential impact on on-line and off-line processing in children's comprehension of pronouns. *Journal of Psycholinguistic Research*, *38*(3), 285–304. https://doi.org/10.1007/s10936-009-9103-9

Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies. *Language, Cognition and Neuroscience*, *31*(1), 73–79. https://doi.org/10.1080/23273798.2015.1036089

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60. https://doi.org/10.1016/j.cogpsych.2016.06.002

Lupyan, G., & Clark, A. (2015). Words and the World. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Malik, W. A., Marco-Llorca, C., Berendzen, K., & Piepho, H. P. (2020). Choice of link and variance function for generalized linear mixed models: a case study with binomial response in proteomics. *Communications in Statistics - Theory and Methods*, *49*(17), 4313–4332. https://doi.org/10.1080/03610926.2019.1599021

Markman, A. B., & Otto, A. R. (2011). Cognitive systems optimize energy rather than information. *Behav. Brain Sci*, *34*(207), 10–1017.

Marques, T., Nguyen, J., Fioreze, G., & Petreanu, L. (2018). The functional organization of cortical feedback inputs to primary visual cortex. *Nature Neuroscience*, *21*(5), 757–764. https://doi.org/10.1038/s41593-018-0135-z

Marrufo-Pérez, M. I., Eustaquio-Martn, A., & Lopez-Poveda, E. A. (2019). Speech predictability can hinder communication in difficult listening conditions. *Cognition*, *192*, 103992. https://doi.org/10.1016/j.cognition.2019.06.004

Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology*, *7*, 120.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7-8), 953–978. https://doi.org/10.1080/01690965.2012.705006

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McCullough, C. M. (1958). Context aids reading. *The Reading Teacher*, *11*(4), 225–229.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

McKeever, W. F., & VanDeventer, A. D. (1977). Visual and auditory language processing asymmetries: Influences of handedness, familial sinistrality, and sex. *Cortex*, *13*(3), 225–241.

Meng, Q., Wang, X., Cai, Y., Kong, F., Buck, A. N., Yu, G., Zheng, N., & Schnupp, J. W. H. (2019). Time-compression thresholds for Mandarin sentences in normal-hearing and cochlear implant listeners. *Hearing Research*, *374*, 58–68. https://doi.org/10.1016/j.heares.2019.01.011

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*(4), 545–567. https://doi.org/10.1016/j.jml.2012.01.001

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*(5), 329–335.

Minocher, R., Atmaca, S., Bavero, C., McElreath, R., & Beheim, B. (2021). Estimating the reproducibility of social learning research published between 1955 and 2018. *Royal Society Open Science*, *8*(9), 210450.

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye- movements are modulated by reading ability - Evidence from Indian low and high literates. *Journal of Eye Movement Research*, *5*(1), 1–10. https://doi.org/10.16910/jemr.5.1.3

Moon, I. J., & Hong, S. H. (2014). What is temporal fine structure and why is it important? *Korean Journal of Audiology*, *18*(1), 1–7. https://doi.org/10.7874/kja.2014.18.1.1

Moon, I. J., Won, J. H., Park, M. H., Ives, D. T., Nie, K., Heinz, M. G., Lorenzi, C., & Rubinstein, J. T. (2014). Optimal combination of neural temporal envelope and fine structure cues to explain speech identification in background noise. *Journal of Neuroscience*, *34*(36), 12145–12154. https://doi.org/10.1523/JNEUROSCI.1025-14.2014

Morton, J. (1964). The effects of context on the visual duration threshold for words. *British Journal of Psychology*, *55*(2), 165–180. https://doi.org/10.1111/j.2044-8295.1964.tb02716.x

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

Müller, J. A., Wendt, D., Kollmeier, B., Debener, S., & Brand, T. (2019). Effect of speech rate on neural tracking of speech. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00449

Musch, J., & Reips, U.-D. (2000). A brief history of web experimenting. In *Psychological experiments on the internet* (pp. 61–87). https://doi.org/10.1016/b978-012099980-4/50004-6

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: The case of speech in noise. *PLoS Biology*, *6*(5), 0978–0991. https://doi.org/10.1371/journal.pbio.0060126

Nejime, Y., & Moore, B. C. J. (1998). Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss. *The Journal of the Acoustical Society of America*, *103*(1), 572–576. https://doi.org/10.1121/1.421123

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, *142*, 107427. https://doi.org/10.1016/j.neuropsychologia.2020.107427

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

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Husband, E.

M., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., . . . Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a largescale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2018.0522

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., . . . Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, *7*, 1–24. https://doi.org/10.7554/elife.33468

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18. https://doi.org/10.1080/23273798.2015.1081703

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39.

Obleser, J. (2014). Putting the Listening Brain in Context. *Language and Linguistics Compass*, *8*(12), 646–658. https://doi.org/10.1111/lnc3.12098

Obleser, J., & Kotz, S. A. (2010). Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network. *Cerebral Cortex*, *20*(3), 633–640. https://doi.org/10.1093/cercor/bhp128

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*(2), 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020

Obleser, J., Wise, R. J. S., Alex Dresner, M., & Scott, S. K. (2007). Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions. *Journal of Neuroscience*, *27*(9), 2283–2289. https://doi.org/10.1523/jneurosci.4663-06.2007

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

Orena, A. J., & Colby, S. (2021). *Recognizing voices through a cochlear implant: A systematic review.*

Patro, C., & Mendel, L. L. (2020). Semantic influences on the perception of degraded speech by individuals with cochlear implants. *The Journal of the Acoustical Society of America*, *147*(3), 1778–1789. https://doi.org/10.1

121/10.0000934

Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. *Proceedings of the 7th Conference of the Cognitive Science Society*, 329–334.

Peelle, J. E. (2013). Cortical responses to degraded speech are modulated by linguistic predictions. *Proceedings of Meetings on Acoustics Ica2013*, *19*, 060108.

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

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1315–1330. https://doi.org/10.1037/0096-1523.31.6.1315

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pierce, A. G. J., & Ollason, J. G. (1987). Eight reasons why optimal foraging theory is a complete waste of time. *Oikos*, *49*(1), 111–117. https://www.jstor.org/stable/3565560

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, *95*(2), 201–236. https://doi.org/10.1016/j.cognition.2004.08.004

Poldrack, R., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., Temple, E., & Gabrieli, J. (1998). Auditory processing of temporally compressed speech: An fMRI study. *Journal of Cognitive Neuroscience*, *10*, 126–126.

Prolific. (2014). *Prolific academic*. https://www.prolific.co.

Pusse, F., Sayeed, A., & Demberg, V. (2016). LingoTurk: Managing crowdsourced tasks for psycholinguistics. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, 57–61. https://doi.org/10.18653/v1/n16-3012

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448.

Reips, U.-D. (2021). Web-based research in psychology. *Zeitschrift für Psychologie*.

Richards, S. A., Whittingham, M. J., & Stephens, P. A. (2011). Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework. *Behavioral Ecology and Sociobiology*, *65*(1), 77–89. https://doi.org/10.1007/s00265-010-1035-8

Roberts, B., Summers, R. J., & Bailey, P. J. (2011). The intelligibility of noise-vocoded speech: Spectral information available from acrosschannel comparison of amplitude envelopes. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1711), 1595–1600. https://doi.org/10.1098/rspb.2010.1554

Rodero, E. (2016). Influence of speech rate and information density on recognition: The moderate dynamic mechanism. *Media Psychology*, *19*(2), 224–242. https://doi.org/10.1080/15213269.2014.1002942

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(31), 1–17. https://doi.org/10.3389/fnsys.2013.00031

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, *106*(6), 3629–3636. https://doi.org/10.1121/1.428215

Ryskin, R., & Fang, X. (2021). The many timescales of context in language processing. In *Psychology of learning and motivation* (Vol. 75, pp. 201–243). Elsevier. https://doi.org/10.1016/bs.plm.2021.08.001

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, *181*(July 2017), 141–150. https://doi.org/10.1016/j.cognition.2018.08.018

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, *125*(1), 28.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207–1218. https://doi.org/10.3758/app.71.6.1207

Sanders, L. D., & Astheimer, L. B. (2008). Temporally selective attention modulates early perceptual processing: Event-related potential evidence. *Perception and Psychophysics*, *70*(4), 732–742. https://doi.org/10.3758/PP.70.4.732

Sanderson, S. K., & Roberts, W. W. (2008). The evolutionary forms of the religious life: A cross-cultural, quantitative analysis. *American Anthropologist*, *110*(4), 454–466. https://doi.org/10.1111/j.1548-1433.2008.00078.x

Sanford, A. J., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, *42*(2), 109–130.

Schlueter, A., Lemke, U., Kollmeier, B., & Holube, I. (2014). Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms. *The Journal of the Acoustical Society of America*, *135*(3), 1541–1555. https://doi.org/10.1121/1.4863654

Schneider, B. A., & Pichora-Fuller, M. K. (2001). Age-related changes in temporal processing: Implications for listening comprehension. *Seminars in Hearing*, *22*(3), 227–239.

Scholman, M. C., Demberg, V., & Sanders, T. J. (2020). Individual differences in expecting coherence relations: Exploring the variability in sensitivity to contextual signals in discourse. *Discourse Processes*, *57*(10), 844–861.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.

Seow, T. X. F., & Hauser, T. U. (2022). Reliability of web-based affective auditory stimulus presentation. *Behavior Research Methods*, *54*(1), 378–392. https://doi.org/10.3758/s13428-021-01643-0

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(4), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Shannon, R. V., Fu, Q.-J., & Galvin Iii, J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-Laryngologica*, *124*(0), 50–54. https://doi.org/10.1080/03655230410017562

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Sharit, J., Czaja, S. J., Nair, S., & Lee, C. C. (2003). Effects of age, speech rate, and environmental support in using telephone voice menu systems. *Human Factors*, *45*(2), 234–251. https://doi.org/10.1518/hfes.45.2.234.27245

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008a). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, *123*(1), 489–499. https://doi.org/10.1121/1.2783762

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008b). Effect of age, presentation method, and learning on identification of noise-vocoded words. *The Journal of the Acoustical Society of America*, *123*(1), 476–488. https://doi.org/10.1121/1.2805676

Simantiraki, O., & Cooke, M. (2020). Exploring listeners' speech rate preferences. *INTERSPEECH*, 1346–1350. https://doi.org/10.21437/Interspeech.2020-1832

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120.

Smith, N. J., & Levy, R. (2008). Optimal processing times in reading: A formal model and empirical investigation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *The Journal of the Acoustical Society of America*, *96*(3), 1314–1324. https://doi.org/10.1121/1.411453

Sommers, M. S., Spehar, B., Tye-Murray, N., Myerson, J., & Hale, S. (2020). Age differences in the effects of speaking rate on auditory, visual, and auditory-visual speech perception. *Ear and Hearing*, *41*(3), 549–560.

Stadler, W., Ott, D. V. M., Springer, A., Schubotz, R. I., Schütz-Bosbach, S., & Prinz, W. (2012). Repetitive TMS suggests a role of the human dorsal premotor cortex in action prediction. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00020

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass*, *9*(8), 311–327. https://doi.org/10.1111/lnc3.12151

Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin and Review*, *18*(2), 371–376. https://doi.org/10.3758/s13423-010-0046-9

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1–17. https://doi.org/10.1016/j.jml.2015.02.004

Stilp, C. (2020). Acoustic context effects in speech perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(1), 1–18. https://doi.org/10.1002/wcs.1517

StrauSS, A., Kotz, S. A., & Obleser, J. (2013). Narrowed Expectancies under Degraded Speech: Revisiting the N400. *Journal of Cognitive Neuroscience*, *25*(8), 1383–1395. https://doi.org/10.1162/jocn_a_00389

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*(5), 882–888.

Taleb, N. (2020). *Assessing the intelligibility and acoustic changes of time-processed speech* [Masters thesis, Case Western Reserve University; OhioLINK Electronic Theses; Dissertations Center]. http://rave.ohiolink.edu/etdc/view?acc_num=case1586637814204979

Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism Quarterly*, *30*(4), 415–433.

Thornton, A. R. D., Harmer, M., & Lavoie, B. A. (2007). Selective attention increases the temporal precision of the auditory N100 event-related potential. *Hearing Research*, *230*(1-2), 73–79. https://doi.org/10.1016/j.heares.2007.04.004

Tóth, B., Honbolygó, F., Szalárdy, O., Orosz, G., Farkas, D., & Winkler, I. (2020). The effects of speech processing units on auditory stream segregation and selective attention in a multi-talker (cocktail party) situation. *Cortex*, *130*, 387–400. https://doi.org/10.1016/j.cortex.2020.06.007

Tuthill, J. C., & Azim, E. (2018). Proprioception. *Current Biology*, *28*(5), R194–R203.

Vaden Jr, K. I., Kuchinsky, S. E., Ahlstrom, J. B., Teubner-Rhodes, S. E., Dubno, J. R., & Eckert, M. A. (2016). Cingulo-opercular function during word recognition in noise for older adults with hearing loss. *Experimental Aging Research*, *42*(1), 67–82.

Vaden, K. I., Kuchinsky, S. E., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2015). Cortical activity predicts which older adults recognize speech in noise and when. *Journal of Neuroscience*, *35*(9), 3929–3937.

Vaden, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2013). The cingulo-opercular network provides word-recognition benefit. *Journal of Neuroscience*, *33*(48), 18979–18986. https://doi.org/10.1523/

JNEUROSCI.1417-13.2013

Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., & Dehaene, S. (2012). A temporal bottleneck in the language comprehension network. *Journal of Neuroscience*, *32*(26), 9089–9102. https://doi.org/10.1523/JNEUROSCI.5685-11.2012

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

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Vasishth, S., Schad, D., Bürki, A., & Kliegl, R. (2022). Hypothetical repeated sampling and the t-test. In *Linear mixed models in linguistics and psychology: A comprehensive introduction (DRAFT)*.

Verhelst, W., & Roelands, M. (1993). Overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech. *IEEE International Conference on Acoustics, Speech and Signal Processing*, *2*, 554–557. https://doi.org/10.1109/icassp.1993.319366

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–393.

Welch, N., & Krantz, J. H. (1996). The World-Wide Web as a medium for psychoacoustical demonstrations and experiments: Experience and results. *Behavior Research Methods, Instruments, and Computers*, *28*(2), 192–196. https://doi.org/10.3758/bf03204764

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience*, *32*(40), 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012

Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & Cox, C. L. (2006). Effects of Adult Aging and Hearing Loss on Comprehension of Rapid Speech Varying in Syntactic Complexity. *Journal of the American Academy of Audiology*, *17*(07), 487–497. https://doi.org/10.3766/jaaa.17.7.4

Wingfield, A., Tun, P. A., Koh, C. K., & Rosen, M. J. (1999). Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech. *Psychology and Aging*, *14*(3), 380–389. https://doi.org/10.1037/0882-7974.14.3.380

Winn, M. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends in Hearing*, *20*, 1–17. https://doi.org/10.1177/2331216516669723

Winn, M. B., & Teece, K. H. (2021). Slower speaking rate reduces listening effort among listeners with cochlear implants. *Ear and Hearing*, *42*(3), 584. https://doi.org/10.1097/aud.0000000000000958

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, *49*(6), 770–785. https://doi.org/10.1111/j.1469-8986.2012.01366.x

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20–32. https://doi.org/10.1016/j.cortex.2015.03.014

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Wöstmann, M., & Obleser, J. (2016). Acoustic detail but not predictability of task-irrelevant speech disrupts working memory. *Frontiers in Human Neuroscience*, *10*, 538.

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*(6), 648–672. https://doi.org/10.1080/23273798.2014.995679