

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



Pratik Bhandari
Department of Psychology
Saarland University

A thesis submitted for the degree of

Doctor of Philosophy

September 2022

Dedicated to ...

Acknowledgements

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

Pratik Bhandari
Universität des Saarlandes,
Saarbrücken
10 March 2021

Abstract

Some abstract here

Contents

List of Figures	vi
List of Tables	vii
List of Abbreviations	viii
1 Introduction	1
1.1 Research goals	5
1.2 Research contributions	6
1.3 Overview of the thesis	8
1.4 Dissimination of research findings	11
2 Background	12
2.1 Speech distortion and degradation	12
2.2 Prediction and comprehension of degraded speech	14
2.2.1 Predictive language processing	14
2.2.2 Facilitatory effect of predictability	16
2.3 Summary	18
3 General methods	19
3.1 Experimental materials	19
3.1.1 Stimulus sentences	19
3.1.2 Speech processing	20
3.2 Data collection	23
4 General statistical approach	26
4.1 Linear regression	26
4.2 Binomial logistic regression	27
4.3 Mixed effects modeling	29
4.4 Binomial logistic mixed effects modeling	29
4.5 Running mixed effects models in R	30

5	Predictability effects of degraded speech are reduced as a function of attention	31
5.1	Introduction	32
5.2	Background	33
5.3	Experiment 1A	35
5.4	Methods	35
5.4.1	Participants	35
5.4.2	Materials	35
5.4.3	Procedure	36
5.5	Analyses	36
5.6	Results and discussion	37
5.7	Experiment 1B	40
5.8	Methods	40
5.8.1	Participants and Materials	40
5.8.2	Procedure	40
5.9	Analyses	40
5.10	Results and discussion	41
5.11	Conclusion	43
6	Semantic predictability facilitates comprehension of degraded speech in a graded manner	46
6.1	Introduction	47
6.2	Background	48
6.3	Methods	53
6.3.1	Participants	53
6.3.2	Materials	53
6.3.3	Procedure	55
6.4	Analyses	55
6.5	Results and discussion	56
6.6	Conclusion	58
7	Comprehension of degraded speech is modulated by the rate of speech	60
7.1	Introduction	60
7.2	Background	60
7.2.1	Comprehension of degraded speech	61

Contents

7.2.2	Comprehension of fast and slow speech	62
7.2.3	Comprehension of degraded speech at varying rates of pre- sentation	64
7.2.4	Study aims	66
7.3	Experiment 7A	67
7.4	Methods	67
7.4.1	Participants	67
7.4.2	Materials	67
7.4.3	Procedure	68
7.5	Analyses	69
7.6	Results and discussion	69
7.7	Experiment 7B	71
7.8	Methods	71
7.8.1	Participants and Materials	71
7.8.2	Procedure	71
7.9	Analyses	71
7.10	Results and discussion	72
7.11	Conclusion	73
8	Discussion and conclusion	77
8.1	Summary of the main findings	77
8.2	Theoretical and practical implications	77
8.3	Conclusion	77
9	Ethics and funding	78
Appendices		
A	Experimental items	80
Bibliography		81

List of Figures

3.1	Distribution of cloze probability ratings of target words in low, medium and high predictability sentences	20
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 löst die Aufgabe.’	22
5.1	Mean response accuracy across all conditions in Experiment 1A. Accuracy increased only with an increase in the number of noise vocoding channels. There is no change in accuracy with an increase or decrease in target-word predictability. Error bars represent standard error of the means.	38
5.2	Mean response accuracy across all conditions in Experiment 1B. Accuracy increased with an increase in number of noise vocoding channels and target-word predictability. Error bars represent standard error of the means.	41

List of Tables

3.1	Boundary frequencies (in Hz) for 1, 4, 6 and 8 channels noise-vocoding conditions	21
5.1	Mean response accuracy across all levels of speech degradation and target word predictability in Experiment 1.	38
5.2	Estimated effects of the best fitting optimal model accounting for the correct word recognition in Experiment 1A.	39
5.3	Mean response accuracy across all levels of speech degradation and target word predictability in Experiment 1B.	41
5.4	Estimated effects of the best fitting optimal model accounting for the correct word recognition in Experiment 1B.	42
5.5	Estimated effects of the best fitting optimal model accounting for the correct word recognition in both the experiments.	43

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

1

Introduction

One of the features that distinguishes us humans from other species is our ability to communicate using verbal language. We speak. We listen. We understand. This seemingly straight forward path of communication goes through plenty of hindrances. One of them is adverse listening condition: There is noise in the background during a conversation; poor signal transmission via a Zoom call renders a speaker’s speech distorted; a listener can have hearing loss, or he might be wearing hearing a prosthesis like cochlear implant or hearing aid. How does the cognitive system achieve the feat of understanding speech that is distorted by these internal and external factors? The answer lies in the ‘tactics’ that human language comprehenders use. We rely upon the preexisting linguistic knowledge, information about the world, and importantly, the context information to form predictions about linguistic events that are yet to be encountered.

Let’s take the following sentence, for example:

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

Most readers would expect the final word to be *a kite* in this sentence (DeLong et al., 2005; cf. Nieuwland et al., 2020). Here, the words preceding the ellipses provide a context. A reader can utilize their knowledge about what a *boy* would

1. Introduction

ideally do *outside* when a day is *breezy*. It leads the reader to predict that the sentence continuation is most likely *a kite* and not any other word like *a rocket*. Similar results are observed in auditory domain as well. Listeners use context information from what they have heard, and form predictions about upcoming word (e.g., [Altmann & Kamide, 2007](#); [Ankenier, 2019](#)).

Availability and usage of context information, and top-down prediction aid spoken language comprehension in adverse listening condition, for example when the speech signal is degraded. Throughout this thesis, one of the main questions that we grapple with is, what is the nature of predictability effect and contextual facilitation at different levels of speech degradation? Is prediction probabilistic, or is it an all-or-nothing phenomenon? Prior to this theoretical question, one may ask why study degraded speech comprehension in the first place — we do not come across degraded speech *per se* in our day-to-day conversation after all. There are two facets to the answer of this question. First, we cannot fully understand how speech and language comprehension system in humans operates unless we study how it responds to novel sounds. If predictive processing is the fundamental mechanism of language comprehension which it is claimed to be, then it is imperative we examine the claim in different listening conditions, with novel auditory stimuli. The extent to which acoustic-phonetic cues (temporal and spectral properties) of different forms of distorted speech interact with lexical-semantic properties (top-down processes) provides an insight into the predictive nature of language comprehension. Secondly, 35-40 adults per million inhabitants, and 10-12 children per 10000 newborns receive cochlear implant in Germany alone ([De Raeve et al., 2020](#)). Degraded speech mimics the auditory input of cochlear implants. The study of degraded speech comprehension also provides an insight of language processing in this section of population; and of the difficulties (e.g., listening effort, constraints in cognitive resources) they face compared to normal hearing population.

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

1. Introduction

(i) Replication crisis

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; e.g., Sanderson & Roberts, 2008). The results of an experiment do not hold up consistently when conducted again by another group of researchers. For example, in a multi-lab collaborative study, Nieuwland et al. (2020) did not find the same N400 effect at English article (a/an) that DeLong et al. (2005) had reported 15 years earlier. This thesis attempts to test if predictability effect reported by Obleser and colleagues in degraded speech comprehension (Obleser et al., 2007; Obleser & Kotz, 2010) is replicable.

(ii) Limitations of predictive processing

A number of authors claim that predictive processing is the fundamental nature of human cognition, and thus, by definition, also of language processing (A. Clark, 2013; Friston, Parr, et al., 2020; Friston, Sajid, et al., 2020; Kuperberg, 2021; Lupyan & Clark, 2015). At the same time, many studies have shown limitations in predictive language processing, for example, due to memory, aging, experimental paradigm, literacy, etc. (K. D. Federmeier et al., 2010; Huettig & Guerra, 2019; Huettig & Mani, 2016; Mishra et al., 2012). This thesis examines the factors like, auditory attention and speed of information flow, that can be prerequisites and limitations of semantic prediction in understanding degraded speech.

(iii) Debate on the nature of prediction

There are at least two schools of thought which argue that prediction is either all-or-nothing, or probabilistic in nature. In degraded speech comprehension, only one study has empirically investigated the theoretical postulation that prediction is restricted only for highly predictable sentence endings (StrauSS

1. *Introduction*

et al., 2013). This thesis attempts to undertake the challenge of teasing apart the nature of prediction in degraded speech comprehension.

(iv) **Multitudes of speech distortion**

Speech signal is distorted by multiple factors in real life. Our conversational environment is noisy, or the signal is distorted during transmission. Listeners can adapt to such distorted speech signal as shown by earlier studies; or, on the contrary, trying to focus on the ever-changing lexical task might not let them adapt at all even when only one form of distortion is applied to the speech signal. When loaded with more than one form of distortion, it is not clear how the context aids comprehension. This thesis undertakes this challenge of understanding how multiple forms of bottom-up processing difficulty interacts with the aid that top-down predictive processing provides.

(v) **Speed of information processing**

Unlike visual scene which opens in spatial dimension, speech signal flows in temporal dimension. This poses a challenge to listeners to process information at different speed and timescales. More time is available to process the information in slow speech while less time is available for fast speech. Listeners build up the meaning representation as they process the speech to predict upcoming linguistic unit. The facilitatory effect of predictability is reduced for fast speech; it is not clear how slow speech affects the effect. When the speech is degraded, listeners face additional difficulty. This thesis addresses the challenge of understanding how different speech rates of degraded speech moderate predictability effects.

The research goals of the thesis meet with these challenges. Studies addressing the challenges outlined above will primarily contribute in developing theories of predictive processing, and furthering the understanding of spoken language comprehension. In the next two sections, we present the research goals of this thesis, and the general overview of the all the chapters that follow.

1. Introduction

1.1 Research goals

The interaction between top-down and bottom-up processes has been extensively discussed and examined in the literature of reading studies and spoken language comprehension with clear speech (Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018; for reviews, Staub, 2015). In this thesis, our aim is to investigate such interaction, and the role of other top-down process (e.g., attention) in adverse listening condition, when speech is degraded. The goals of the research in correspondence to the central challenges of this thesis are outlined below.

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

Goal 2: The second goal of this thesis is to examine the nature of predictability effect. In the current literature of predictive processing, a debate exists about the nature of predictability effect: whether it is an all-or-nothing phenomenon or if it is probabilistic. With carefully designed experiments and materials, this thesis aims to tease apart the distinction between these two proposed nature of prediction.

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

Goal 4: The fourth goal of this thesis is to examine if variation in information flow affects facilitatory effect of predictability. The flow of information is

1. Introduction

not constant in a conversation as a speaker’s speech rate varies constantly. We therefore test if an increase or decrease in speech rate impedes overall intelligibility of degraded speech, and if it impedes or further aids the facilitatory effect of predictability.

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

Goal 6: The sixth goal of this thesis is to establish and consistently use a sensitive metric for the measurement of language comprehension. In the study of speech perception and language comprehension, different researchers have used different measurement metric. This becomes a problem specially when the effect of context in comprehension is under discussion. In this thesis, we advocate for the use of a metric that takes into account whether participants (in)correctly utilize the context evoking word in a sentence.

1.2 Research contributions

The research reported in this thesis examines theoretically interesting questions of predictive language processing and its limits when bottom-up auditory input is compromised. It contributes to the studies of speech perception, language comprehension, predictive coding, language science, audiology, psycholinguistics, psychology, and broadly to cognitive science. It opens a window towards understanding how human auditory and language processing mechanism work to help us comprehend speech in adverse listening conditions. In an applied setup,

1. *Introduction*

this informs translational/clinical researchers about language comprehension in cochlear implantees.

- **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. Rather, it considers the fact that in the study of effect of predictability, how well a context is recognized should also be taken into account. Thus it measures word recognition accuracy in the sentences in which context is correctly recognized. The use of such a metric improves interpretation of contextual facilitation across studies, which lacks in the extant literature. Goal 6 corresponds to this research contribution which is brought about by consistent use of this metric in [Chapters 6](#) and [7](#).

- **Nature of prediction**

First, we replicate previous findings that predictability indeed facilitates comprehension of degraded speech at a moderate level of degradation. This is of great value given the discussion on replication crisis in psychology and psycholinguistics. In the existing debate between all-or-nothing vs graded prediction, our findings clearly show that prediction in degraded speech comprehension is graded in nature. Goals 1 and 2 correspond to this research contribution which is brought about by the experiment described in [Chapter 6](#).

- **Limitation in predictive language processing**

We show that predictive processing all by itself cannot explain how listeners understand degraded speech. Although top-down predictions facilitate comprehension, we show that attention to context is a prerequisite for such contextual facilitation. Only when listeners attend to the context information and form its meaning representation, then top-down predictions can facilitate comprehension of degraded speech. Without proper attention to the context,

1. Introduction

predictability effects cannot be observed. Goal 3 corresponds to this research contribution which is brought about by the experiment described in [Chapter 5](#).

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

We show that different rates of information flow — increase or decrease in rate of speech — have different effects in language comprehension. It is brought about primarily by the difficulty in processing sentences with less predictable endings. This is one of the very few studies highlighting the role of speed of flow of information in contextual facilitation of degraded speech. Goal 4 corresponds to this research contribution which is brought about by the experiment described in [Chapter 7](#).

- **Adaptation to degraded speech**

We show that listeners do not adapt to degraded speech which is in contrast with some findings of speech perception experiments from other labs. 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 5 corresponds to this research contribution which is brought about by the experiment described mainly in [Chapter 6](#), and continued in [Chapter 7](#).

1.3 Overview of the thesis

The central theme of this thesis is the study of the interaction between top-down predictions and bottom-up auditory processes in degraded speech comprehension. Although there are studies showing reliance on prediction at a moderate level of speech degradation, this thesis replicated such findings focusing on theoretical question of the nature of such predictive processing. The studies in this thesis examined the factors like attention and rate of information flow that modulate semantic prediction, and also found evidence against adaptation and learning effect.

1. Introduction

Chapter 2 provides a background on the rest of the chapters. It provides an overview of degraded speech comprehension, and predictive language processing. 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, speech processing, and also provides an overview of online data collection.

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

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

Chapter 6 addresses the first, the second, the fifth, and the sixth research goals. The predictability effect that was 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

1. Introduction

the context. We observe predictability effects at a moderate level of speech degradation, thereby 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 whether or not the next-trial degradation level is certain, listeners do not adapt to degraded speech as 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 this 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 were presented with a moderately degraded speech at which we observed predictability effect in **Chapter 6**. In Experiment 7A, such degraded speech was presented at normal and fast speech rates. And in Experiment 7B, the speech rates were normal and slow. We show that although there is an overall decrease in intelligibility, predictability effects are not reduced for slow speech, i.e., when more time is available to process the incoming information. For fast speech however both intelligibility as well as predictability effects were reduced which were driven by the difficulty in processing words that were less predictable from the context. This chapter ends with the discussion on the limitations of predictive processing that are 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. General limitations of the studies are briefly discussed.

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

1.4 Dissimination of research findings

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

- Bhandari, P., Demberg, V., & Kray, J. (*in preparation*) Speaking fast and slow: How speech rate affects contextual facilitation in degraded speech comprehension.
- Bhandari, P., Demberg, V., & Kray, J. (*under review*) Predictability effects in degraded speech comprehension are reduced as a function of attention.
- Bhandari, P., Demberg, V., & Kray, J. (2022). Predictability effects are reduced as a function of attention. *Annual Convention of American Psychological Association*
- Bhandari, P., Demberg, V., & Kray, J. (2021). Semantic predictability facilitates comprehension of degraded speech in a graded manner. *Frontiers in Psychology*
- Bhandari, P., Demberg, V., & Kray, J. (2021). Predictability facilitates comprehension but not adaptation to degraded speech in a graded manner. *Conference of the Society for the Neurobiology of Language*
- Bhandari, P., Demberg, V., & Kray, J. (2021). Predictability facilitates comprehension of degraded speech in a graded manner. *Annual Meeting of Cognitive Neuroscience Society*

2

Background

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

2.1 Speech distortion and degradation

In an ideal situation, speech perception is a seamless process: a speaker produces an utterance, the speech signal transmits via some medium like air, and a listener perceives the signal as speech waves enter her ears and start a cascade of mechanical-neural processes of hearing. However, speech perception is hardly as smooth as it seems. There are primarily three sources distortion: Speech can be distorted by variability in speakers' production, like, accented speech, or soft and rapid speech.

2. Background

Distortion can arise from listener-related factors like, hearing loss, or auditory processing disorder. It can also be a result of noise that appears during the transmission, like ambient noise, or poor medium of transmission (e.g., distortion in the telephone line). All these sources of distortion make listening conditions adverse. For controlled scientific study, the effects of speech distortion, and the mechanism of listening in adverse listening condition are investigated using artificial distortion of speech. Different forms of distorted speech manipulate different properties of speech signal. For example, speeding or slowing a speech signal (i.e. speech compression or expansion) manipulates its temporal property. Similarly, noise vocoding manipulates its spectral property. Noise vocoding removes the spectral detail of the speech signal only leaving its temporal and periodicity cues (see Section 3.1.2). This method of speech degradation was initially developed as a means to reduce the information in speech signal to be transmitted through the telephone line (Clendeninn, 1940; Dudley, 1939). Shannon and colleagues later used the same technique as an analogue to cochlear implant such that number of channels used in a cochlear implant are similar to the number of noise vocoding channels in terms of their speech output and intelligibility (Loizou et al., 1999; Shannon et al., 1995; Shannon et al., 2004). Therefore, in addition to being a method of speech distortion to parametrically vary and control the quality of speech signal in a graded manner, noise vocoding also is a method of distortion that helps us understand the speech perception and language comprehension in cochlear implantees.

One of the main factors that determine the intelligibility of degraded speech is the number of noise vocoding channels.¹ The higher the number of noise vocoding channels, the more is the frequency specific information available in the degraded speech, consequently, higher is the intelligibility compared to the speech that is degraded with lesser number of noise vocoding channels. For example, listeners rate 8 channels noise vocoded speech to be more intelligible and less effortful compared to 2 channels noise vocoded speech (e.g., Sohoglu et al., 2012).

¹Throughout 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 signal, listeners rely also on the context information and form top-down predictions to understand speech in adverse listening condition. Below, we first review the role of predictions in language comprehension in general, 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, odor, and proprioception, has shown that perception and cognition primarily operate by predicting upcoming events (A. Clark, 2013; Marques et al., 2018; Seth, 2013; Stadler et al., 2012). Human language comprehension too has been claimed to be predictive in nature from as early as mid-twentieth century (McCullough, 1958; e.g., Miller et al., 1951; Morton, 1964) which in recent days has received overwhelming support from studies in psycholinguistics and cognitive neuroscience of language (e.g., DeLong et al., 2005; Lupyan & Clark, 2015; Pickering & Gambi, 2018). Empirical evidence from a number of studies suggests that readers and listeners predict upcoming words in a sentence when the words are predictable from the preceding context (Kuperberg & Jaeger, 2016; Nieuwland, 2019; for reviews, Staub, 2015). For instance, predictable words are read faster and are skipped compared to 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 the picture of the word that is predictable from the sentence context (Altmann & Kamide, 1999; Ankenier et al., 2018; Kamide et al., 2003). The sentence-final word in a highly constraining sentence (e.g., “*She dribbles a ball.*”) elicits a smaller N400 amplitude — a negative going EEG component that peaks around 400 ms

2. Background

post-stimulus and is considered as a neural marker of context-based semantic unexpectedness (Kutas & Federmeier, 2011) — than that in a less constraining sentence (K. D. Federmeier et al., 2007; e.g., “*She buys a ball.*”, 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 an event of “*travel*”, Metusalem et al., 2012). In sum, as the sentence context builds up, listeners make predictions about upcoming words in the sentence, and these in turn facilitate language comprehension. That is, individuals use the context available to them to generate predictions which aids understanding written and spoken language.

Limits of predictive language processing: It is important to note and acknowledge that the ubiquity and universality of predictive language processing has not gone unquestioned (Huettig & Mani, 2016). Apart from the debate on the nature of prediction, which we will come to later in this chapter, there are compelling evidence that question the necessity of prediction in language comprehension. For example, Mishra et al. (2012) showed that literacy is a key factor that limits listeners’ prediction about upcoming word. In visual word paradigm, they found that individuals with lower literacy showed less anticipatory eye movements compared to the individuals 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 is manifested in linguistic behavior. Similarly, cognitive aging has been shown to be a limiting factor in generating predictions. Smaller N400 amplitude and latency in older adults compared to younger adults have been shown as evidence of inability of older adults in predictive processing in language processing. Among older adults, those with lower working memory scores are shown to be further disadvantaged when it comes to the use of context information (K. Federmeier et al., 2002; K. D. Federmeier et al., 2010). Another line of argument that critiques the predictive processing comes from the observations of Huettig & Guerra (2019) (see also, Fernandez et al., 2020). They analyzed participants’

2. Background

anticipatory eye movements in the visual world paradigm and showed that listeners predict the target word only in an *artificial* set-up of long preview time coupled with slow speech. Alongside these critiques, there are also experiments demonstrating predictability effects that have been partly or fully replicated (Ankenier, 2019; e.g., Nieuwland et al., 2018). In this thesis, we study other factors (e.g., auditory temporal attention, speed of information processing) that can interact with, and potentially limit top-down semantic predictions.

2.2.2 Facilitatory effect of predictability

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

2. Background

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. Compared to highly constraining sentences, a higher number of channels (i.e., more bottom-up information) was required in less constraining sentences to achieve the same level of accuracy. They also concluded that when speech is degraded, word recognition is facilitated by predictability and sentence context. Taken together, these studies conclude that at moderate levels of degradation, participants rely more on the top-down predictions generated by the sentence context and less on the bottom-up perceptual processing of unclear, less reliable, and degraded speech signal (Obleser, 2014).

Nature of prediction: One of the debates in the literature of predictive language processing pertains this question: Is prediction probabilistic, or is it an all-or-nothing phenomenon? For instance, garden path phenomenon was explained as a parser’s irreversible prediction about the sentence structure which if fails or turns out to be incorrect then the parser reanalyzes the sentence and reformulates another prediction (e.g., Ferreira & Clifton Jr, 1986; see also, 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 N400 effect evoked by a word and its predictability measured by cloze probability (e.g., DeLong et al., 2005). The correlation between a word’s *surprisal* in its linguistic context and the processing effort associated with it also demonstrates the probabilistic nature of prediction (e.g., Hale, 2001; Smith & Levy, 2008).

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

2. Background

narrowed expectations from a restricted semantic space only when the sentence endings are highly predictable. They rule out the graded nature of predictability. In this thesis we take this theoretical account into consideration, and examine the nature of prediction in degraded speech comprehension.

2.3 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 Section 3.1.2. Importantly, we provided an overview of how predictive language processing aids in language comprehension, as well as its limitations. 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

In this chapter we provide an overview of the experimental materials that are used in the experiments described in Chapters 5, 6, and 7. Sentences used as experimental material were common to 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. 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 — 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

3. Methods

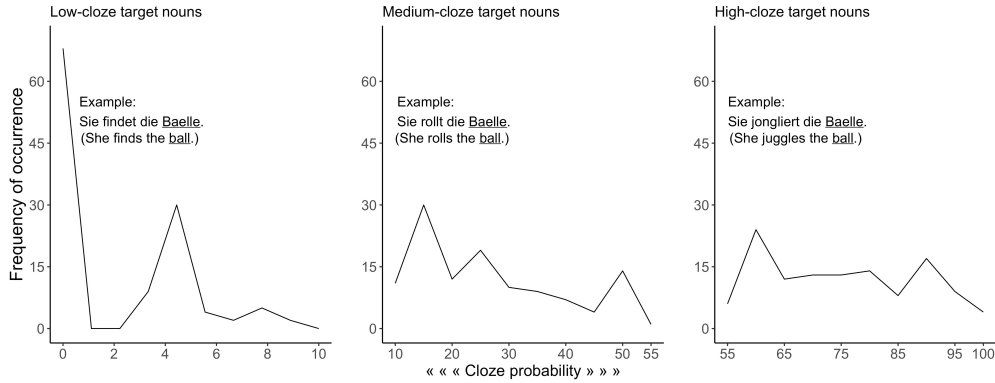


Figure 3.1: Distribution of cloze probability ratings of target words in low, medium and high predictability sentences

in 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 were 0.022 (SD = 0.027; range = 0.00 – 0.09) for low-predictability sentences, 0.274 (SD = 0.134; range = 0.1 – 0.55) for medium-predictability sentences, and 0.752 (SD = 0.123; range = 0.56 – 1.00) for high-predictability sentences. The distribution of cloze probability across low-, medium-, and high-predictability sentences is shown in Figure 3.1, and the cloze probability for individual sentence is shown in Appendix A.

3.1.2 Speech processing

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

Noise-vocoding

Noise vocoding is used to parametrically vary and control the quality of speech signal in a graded manner. It distorts a speech signal by dividing it into specific

3. Methods

frequency bands corresponding to the number of vocoder channels. The frequency bands are analogous to the electrodes of 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 makes the vocoded speech difficult to understand although temporal characteristics and periodicity of perceptual cues are preserved (Rosen et al., 1999).

The spectral degradation conditions of 1, 4, 6, and 8 channels were achieved for each of the 360 recorded sentences using a customized script originally written by Darwin (2005) in Praat software. The speech signal was divided into 1, 4, 6, and 8 frequency bands between 70 and 9,000Hz. The boundary frequencies were approximately logarithmically spaced following cochlear-frequency position functions (Erb, 2014; Greenwood, 1990). 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 3.1. Each of the modulated noise was then combined to produce degraded speech. Scaling was performed to equate the root-mean-square value of the original undistorted speech and the final degraded speech. This resulted into four levels of degradation: 1, 4, 6, and 8 channels noise vocoded speech.

Spectrograms of clear speech and noise-vocoded speech for the sentence *Er löst 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.

Table 3.1: 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

3. Methods

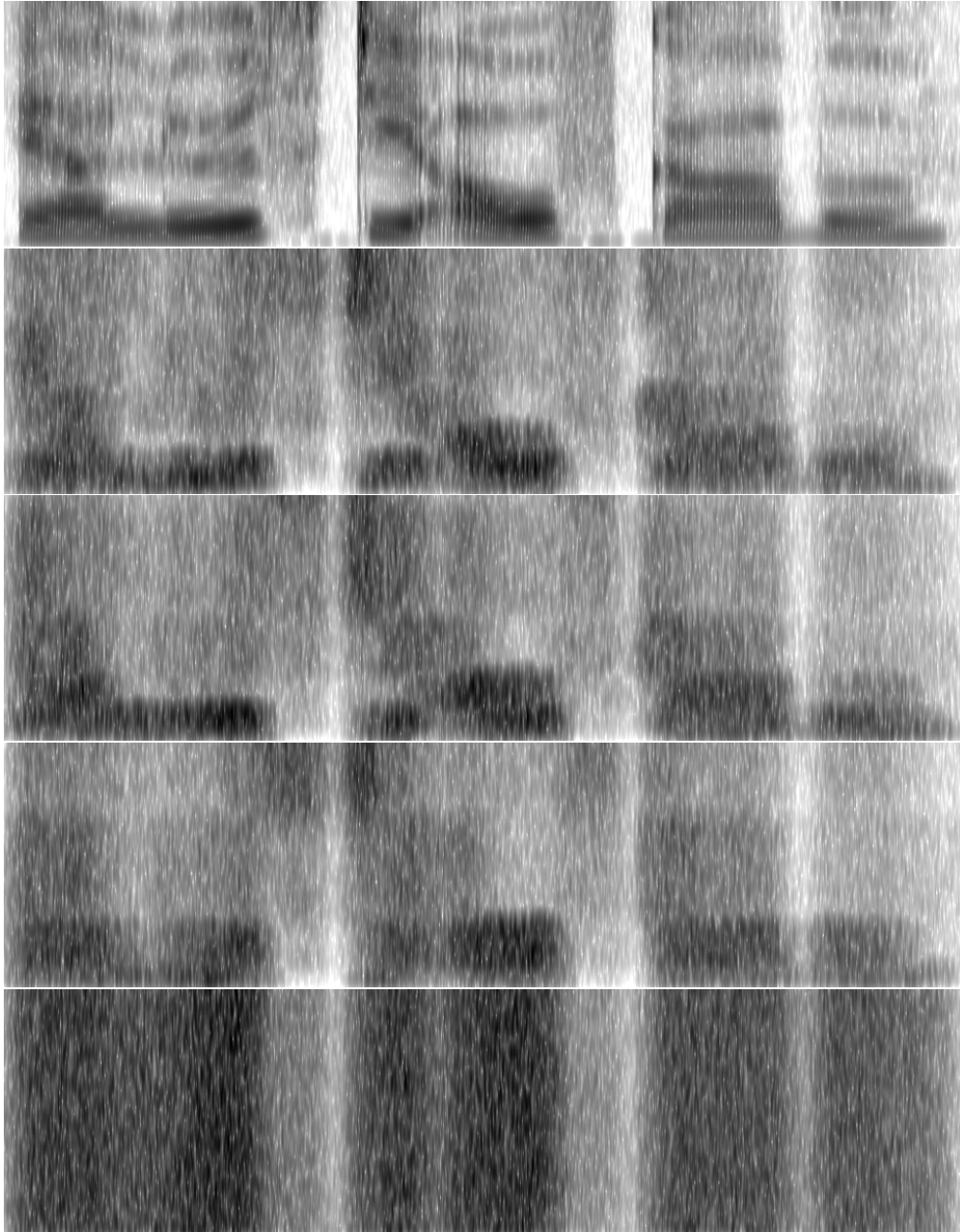


Figure 3.2: Spectrograms of clear speech, and degraded speech arranged with a decreasing number of noise vocoding channels (8, 6, 4 and 1 band) for the sentence ‘Er löst die Aufgabe.’

3. Methods

Speech compression and expansion

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 pitch-synchronous overlap-add technique (PSOLA, [Charpentier & Stella, 1986](#); [Moulines & Charpentier, 1990](#)) developed in the 1980s and later (overlap-add technique based on waveform similarity, [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 uniform time-compression algorithm (PSOLA) to create slow and fast speech with the compression factor of 1.35 and 0.65 respectively. PSOLA analyzes the pitch of an auditory signal in the time domain of its digital waveform to set pitch marks, and then segments the signal into successive analysis windows centered around those pitch marks. To create synthesized speech (i.e., fast or slow speech), a new set of pitch marks are calculated, and the analysis windows are rearranged. Depending on the time-compression factor, some analysis windows are deleted, and the remaining windows are concatenated by superimposing and averaging the neighboring analysis windows. The distortion of phonemic properties of speech signals are minimal when accelerating and slowing down within the range of factor 2 or below ([Moulines & Charpentier, 1990](#)).

In Chapter 7, we created fast and slow versions of 120 high-predictability sentences and 120 low-predictability sentences, but did not use medium-predictability sentences. These 480 recordings were then passed through 4 channels noise vocoding to use as experimental materials.

3.2 Data collection

The first generation of online experiments on human cognition began in the mid 1990s (for reviews, [Musch & Reips, 2000](#)) with the advent of the internet ([Berners-](#)

3. Methods

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 2 psycholinguistics experiments conducted online; one of which studied the effect of context in shallow vs. deep encoding of words. Despite the difficulty in conducting online experiments, and skepticism of journals towards publishing results of online experiments, Musch & Reips (2000) expressed optimism:

At the moment, the number of Web experiments is still small, but a rapid growth can be predicted on the basis of the present results. We would not be surprised if within the next few years, a fair proportion of psychological experiments will be conducted on the Web.

And by 2021, there has been a 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 get 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 those in laboratory experiments. Taking these advantages into consideration, psychologists and psycholinguists have conducted online experiments for almost 3 decades now. Scientists who only conducted laboratory experiments, or who conducted online experiments only occasionally were forced to conduct 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 auditory domain (Leensen & Dreschler, 2013; Os et al., 2021; Seow &

3. Methods

[Hauser, 2022](#); [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 laboratory and online. As the laboratory was shut down due to covid-19 pandemic, we moved the laboratory experiments online too. 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 to not have had any hearing loss, speech-language disorder, and 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 regression

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

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

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

4. Statistics

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

Below we briefly introduce binomial logistic regression and mixed effects model. Then we show a simple example of how binomial logistic mixed effects model is used in our data analyses.

4.2 Binomial logistic regression

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

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

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

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

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_n \cdot x_n + \varepsilon \quad (4.2)$$

4. Statistics

This is equivalent to,

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

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

where,

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

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

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

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

4.3 Mixed effects modeling

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

4.4 Binomial logistic mixed effects modeling

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

$$\ln\left(\frac{p}{1-p}\right) = \alpha + u_\alpha + w_\alpha + (\beta_1 + u_{\beta_1} + w_{\beta_1}) \cdot x_1 + (\beta_2 + u_{\beta_2} + w_{\beta_2}) \cdot x_2 + \dots + (\beta_n + u_{\beta_n} + w_{\beta_n}) \cdot x_n \quad (4.6)$$

where,

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

4. Statistics

In the next chapters of this thesis, we perform statistical analyses of the effect of predictability, speech degradation and speech rate on response accuracy. These variables are used in the fixed effects term. Subjects and items are used as random intercepts with by-subject and by-item slopes. The details of the models fitted to data from each experiment are given in Chapters 5, 6, and 7.

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

4.5 Running mixed effects models in R

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). Accuracy was analyzed with Generalized Linear Mixed Models (GLMMs) with `lmerTest` (Kuznetsova et al., 2017) and `lme4` (Bates, Mächler, et al., 2015) packages. Binary responses (correct responses coded as 1 and incorrect responses coded as 0) for all participants were fit with a binomial logistic mixed effects model.

On the data from each experiment, we fitted models with maximal random effects structure that included random intercepts for each participant and item (Barr et al., 2013). By-participant and by-item slopes included in the model are discussed in the Analysis sections of Chapters 5, 6, and 7. Model selection was based on Akaike Information Criterion (Grueber et al., 2011; Richards et al., 2011) unless otherwise stated. Random effects not supported by the data that explained zero variance according to singular value decomposition were excluded to prevent overparameterization (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 parameter, and iii) both item- and participant-related correlation parameters. The best fitting model among the parsimonious and extended models was then selected as the optimal model for our data.

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 the context information, and form top-down semantic predictions. This provides contextual facilitation in understanding the degraded speech. Importantly, it is moderated by top-down attentional allocation to the context. The aim of this study was to examine the role of attention for understanding linguistic information in an adverse listening condition, i.e., when the speech was degraded. To assess the role of attention we varied task instructions in two experiments in which participants were instructed to listen to short sentences and thereafter to type in the last word they have heard, or to type in the whole sentence. We were interested in how these task instructions influence the interplay between top-down prediction and bottom-up perceptual processes during language comprehension. As described in the previous chapter, these sentences varied in the degree of predictability (low, medium, high) as well as in the degree of speech degradation (1, 4, 6 and 8 noise vocoding channels). Results indicated better word recognition for highly predictable sentences at moderate levels of degradation only when attention was directed to the whole sentence. This underlines the important role of attention in language comprehension.

5.1 Introduction

In optimal listening conditions, understanding speech is highly automatized and seemingly easy. But conditions are often far from optimal in our day-to-day communication. As simple as a weak internet connection can create an adverse listening condition, for example, in an online meeting. Although the speech signal gets distorted or degraded at times, we do not frequently fail to understand such degraded speech. Listeners overcome the difficulty and successfully understand the speech by using context information. It contains information in a given situation about a topic of conversation, semantic and syntactic information of a sentence structure, world knowledge, visual information, etc. (Altmann & Kamide, 2007; Kaiser & Trueswell, 2004; Knoeferle et al., 2005; Xiang & Kuperberg, 2015; for reviews, Ryskin & Fang, 2021; Stilp, 2020). To utilize the context information, however, listeners must attend to it and build up a meaning representation of what has been said. Processing and comprehending degraded speech is more effortful and requires more attentional resources than that for clear speech (Eckert et al., 2016; Peelle, 2018; Wild et al., 2012).

In this chapter we examine how attention modulates the predictability effects brought about by contextual information at different levels of spectral degradation of speech. We address the existing unclarity in the literature regarding how listeners distribute their attentional resources in adverse listening conditions: On the one hand, listeners can attend throughout the whole stream of speech and may thereby profit from the context information to predict sentence endings. On the other hand, listeners can focus their attention on linguistic material at a particular time point in the speech stream and, as a result, miss critical parts of the sentence context. If the goal is to understand a specific word in an utterance, there is a trade-off between allocating attentional resources to the perception of that word vs. allocating resources also to the understanding of the linguistic context and generating predictions.

5. *Attention-prediction interplay*

This study reported in this chapter was conducted with an aim to investigate how the allocation of attentional resources induced by different task instructions influence language comprehension and, in particular, the use of context information under adverse listening conditions. To examine the role of attention on predictive processing under degraded speech, we ran two experiments in which we manipulated task instructions. In [Experiment 1A](#), participants were instructed to only repeat the final word of the sentence they heard, while in [Experiment 1B](#), they were instructed to repeat the whole sentence, and by this drawing attention to the entire sentence including the context. In both experiments we varied the degree of predictability of sentence endings as well as the degree of speech degradation.

5.2 Background

As we have discussed earlier in [Chapter 1](#) and [Chapter 2](#), it is generally agreed upon that human language processing is predictive in nature, such that comprehenders generate expectations about upcoming linguistic materials based on context information ([Kuperberg & Jaeger, 2016](#); [Nieuwland, 2019](#); [Pickering & Gambi, 2018](#); for reviews, see [Staub, 2015](#)). When the bottom-up speech signal is less informative in an adverse listening condition, listeners rely more on the context information to support language comprehension ([Amichetti et al., 2018](#); [Obleser & Kotz, 2010](#); [Sheldon et al., 2008b](#)). However, it is not just the quality of speech signal that determines and influences the reliance and use of predictive processing; attention to auditory input is important as well. Auditory attention allows a listener to focus on the speech signal of interest (for reviews, see [Fritz et al., 2007](#); cf. [Lange, 2013](#)). For instance, it has been shown that a listener can attend to and derive information from one stream of sound among many competing streams as demonstrated in the well-known *cocktail party effect* ([Cherry, 1953](#); [Haft 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

5. *Attention-prediction interplay*

varying in pitch, pure tones of different harmonics) but attends to any one of the tones appearing at a specified time point, this is reflected in a larger amplitude of N1 (e.g., [Lange & Röder, 2010](#); see also, [Sanders & Astheimer, 2008](#)) which is the first negative going ERP component peaking around 100 ms post-stimulus considered as a marker of auditory selective attention ([Näätänen & Picton, 1987](#); [Thornton et al., 2007](#)). Hence, listeners can draw attention to and process one among multiple competing speech streams.

So far, most previous studies have investigated listeners' attention within a single speech stream by 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 accented or de-accented critical word. The N1 amplitude was larger for words with such attention probe than for words without a probe. These findings support the view that attention can be flexibly directed either by instructions towards a specific signal or by linguistic probes ([X. Li et al., 2017](#); see also, [Brunellière et al., 2019](#)). Thus, listeners are able to select a part or segment of stream of auditory stimuli to pay attention to.

The findings on the interplay of attention and prediction mentioned above come from studies most of which used a stream of clean speech or multiple streams of clean speech in their experiments. They cannot tell us about the attention-prediction interplay in degraded speech comprehension. Specifically, we do not know what role attention to a segment of speech stream plays in the contextual facilitation of degraded speech comprehension, although separate lines of research show that listeners attend to most informative portion of speech stream (e.g., [Astheimer & Sanders, 2011](#)), and semantic predictability facilitates comprehension of degraded speech (e.g., [Obleser & Kotz, 2010](#)). In two experiments, we therefore examined whether context-based semantic predictions are automatic during effortful listening to degraded speech, when participants are instructed to report only the final word of the sentence, or the entire sentence. We varied the task instructions to the listeners from [Experiment 1A](#) to [Experiment 1B](#) which required

5. *Attention-prediction interplay*

them to differentially attend to the target word (not binding the context), or to the target word including the context. We hypothesized that when listeners pay attention only to the contextually predicted target word, they do not form top-down predictions, i.e., there should not be a facilitatory effect of target word predictability. In contrast, when listeners attend to the whole sentence, they do form expectations such that the facilitatory effect of target word predictability will be observed.

5.3 Experiment 1A

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

5.4 Methods

5.4.1 Participants

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

5.4.2 Materials

Materials used in the experiment were created by the method described in Chapter 3 (Section 3.1). That is, there were 360 sentences with 120 sentences in each of these 3 categories: low predictability, medium predictability and high predictability. The mean cloze probabilities of target words for low, medium and high predictability sentences were 0.022 ± 0.027 ($\bar{x} \pm \text{SD}$; range = 0.00 - 0.09), 0.274 ± 0.134 ($\bar{x} \pm \text{SD}$; range = 0.1 - 0.55), and 0.752 ± 0.123 ($\bar{x} \pm \text{SD}$; range = 0.56 - 1.00)

5. Attention-prediction interplay

respectively. All the 360 sentences were then noise vocoded through 1, 4, 6, and 8 channels to create degraded speech.

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

5.4.3 Procedure

Participants were asked to use headphones or earphones. A sample of noise vocoded speech not used in the practice trial and the main experiment was provided so that the participants could adjust the loudness to a preferred level of comfort at the beginning of the experiment. They were instructed to listen to the sentences and to type in the target word (noun) by using the keyboard. The time for typing in the response was not limited. They were also informed at the beginning of the experiment that some of the sentences would be ‘noisy’ and not easy to understand, and in these cases, they were encouraged to guess what they might have heard. Eight practice trials with different levels of speech degradation were given to familiarize the participants with the task before presenting all 120 experimental trials with an inter-trial interval of 1000 ms. The experiment was approximately 40 minutes long.

5.5 Analyses

Out of XYZ trials from 49 participants, there were only 5 correct responses at 1 channel speech degradation condition. Therefore, the 1 channel speech degradation condition was excluded from the analysis.

5. *Attention-prediction interplay*

Response accuracy was analyzed with Generalized Linear Mixed Models (GLMMs) following the procedure described in Chapter 4 (Section 4.5). Binary responses (correct/incorrect) for all participants were fit with a [binomial logistic mixed-effects model](#) (Jaeger, 2006, 2008). Noise condition (categorical; 4, 6, and 8 channels noise vocoding), target word predictability (categorical; high, medium, and low), and the interaction of number of channels and target word predictability were included in the fixed effects.

We first fitted a model with maximal random effects structure that included random intercepts for each participant and item (Barr et al., 2013). Both, by-participant, and by-item random slopes were included for number of channels, target word predictability, and their interaction. Non-significant higher-order interactions were excluded from the fixed-effects structure and from the random-effects structure in a stepwise manner. Random effects not supported by the data that explained zero variance were excluded and a more parsimonious model was obtained (Bates, Kliegl, et al., 2015). Such a model was then extended separately with i) item-related correlation parameters, ii) participant-related correlation parameters, and iii) both item- and participant-related correlation parameters when applicable. Aiming for model parsimony, the best fitting model among the parsimonious and extended models was then selected as the optimal model for our data. Model selection was based on AIC (Burnham & Anderson, 2002; Grueber et al., 2011; Richards et al., 2011).

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

5.6 Results and discussion

Mean response accuracies for all experimental conditions are shown in Table 5.1 and Figure 5.1. It shows that accuracy increases with an increase in the number

5. Attention-prediction interplay

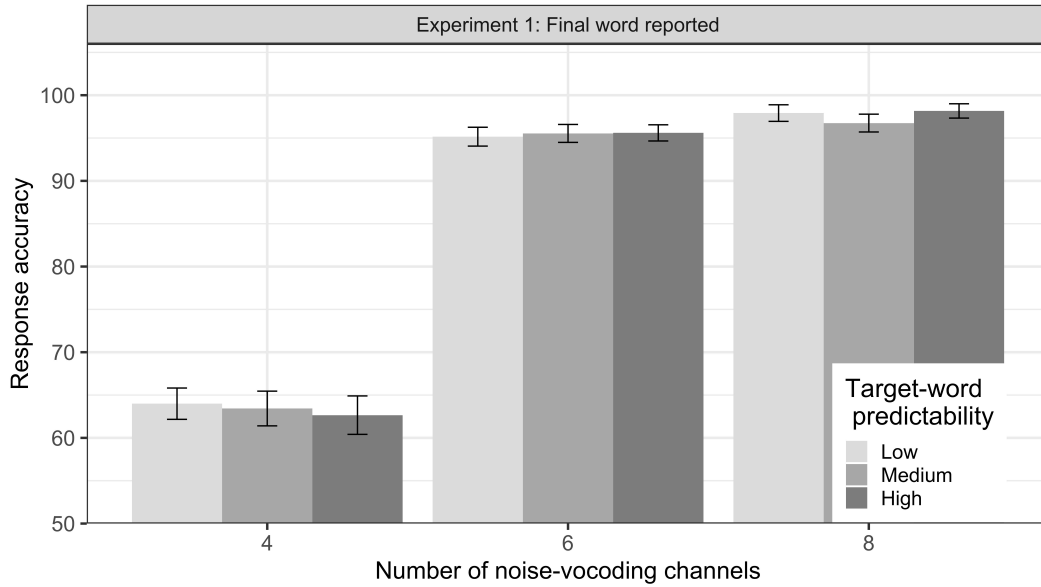


Figure 5.1: Mean response accuracy across all conditions in Experiment 1A. Accuracy increased only with an increase in the number of noise vocoding channels. There is no change in accuracy with an increase or decrease in target-word predictability. Error bars represent standard error of the means.

of noise vocoding channels, i.e., with the decrease in speech degradation. However, accuracy does not increase with an increase in target word predictability. The results of statistical analyses confirmed these observations (Table 5.2).

Table 5.1: Mean response accuracy 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

We found that there was a significant main effect of number of channels, indicating that response accuracy in the 8 channels noise vocoded speech was higher

5. Attention-prediction interplay

than in both 4 channels ($\beta = -3.49$, $SE = .23$, $z(4246) = -15.30$, $p < .001$) and 6 channels noise vocoded speech ($\beta = -.69$, $SE = .22$, $z(4320) = -3.12$, $p = .002$). That is, when the number of channels increased to 8, listeners made more correct responses (see Figure 5.1). However, there was no significant main effect of target word predictability ($\beta = -.07$, $SE = .17$, $z(4246) = -.42$, $p = .68$, and $\beta = -.003$, $SE = .16$, $z(4246) = -.02$, $p = .98$), and no significant interaction between number of noise vocoding channels and target word predictability (all $ps > .05$).

The results of Experiment 1A indicated a decrease in response accuracy with an increase in speech degradation from 8 channels to 6 channels noise vocoding condition, and from 8 channels to 4 channels noise vocoding condition. However, response accuracy did not increase with an increase in target word predictability, and the interaction between number of noise vocoding channels and target word predictability was also absent, in contrast to previous findings (Obleser et al., 2007; Obleser & Kotz, 2011; see also Hunter & Pisoni, 2018). These results suggest that the task instruction, which asked participants to only report the final word, indeed lead to neglecting the context, and therefore the facilitatory effect of prediction was not observed. However, to further test the hypothesis — as mentioned in the beginning of this chapter — that predictability effect is dependent on attentional effect, we conducted a second experiment. In the second experiment, we changed the task instruction to draw participants' attention to the entire sentence such that they could attend and decode the whole sentence including the context.

Table 5.2: Estimated effects of the best fitting optimal model accounting for the correct word recognition in Experiment 1A.

Fixed effects	Estimate	Std. Error	z value	p value
Intercept	4.20	.24	17.52	<.001
Noise condition (4 channels)	-3.49	.23	-15.30	<.001
Noise condition (6 channels)	-.69	.22	-3.12	.002
Target word predictability (Low-Medium)	.07	.17	.42	.67
Target word predictability (High-Low)	.003	.16	.02	.98

5.7 Experiment 1B

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

5.8 Methods

5.8.1 Participants and Materials

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

5.8.2 Procedure

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

5.9 Analyses

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

5. Attention-prediction interplay

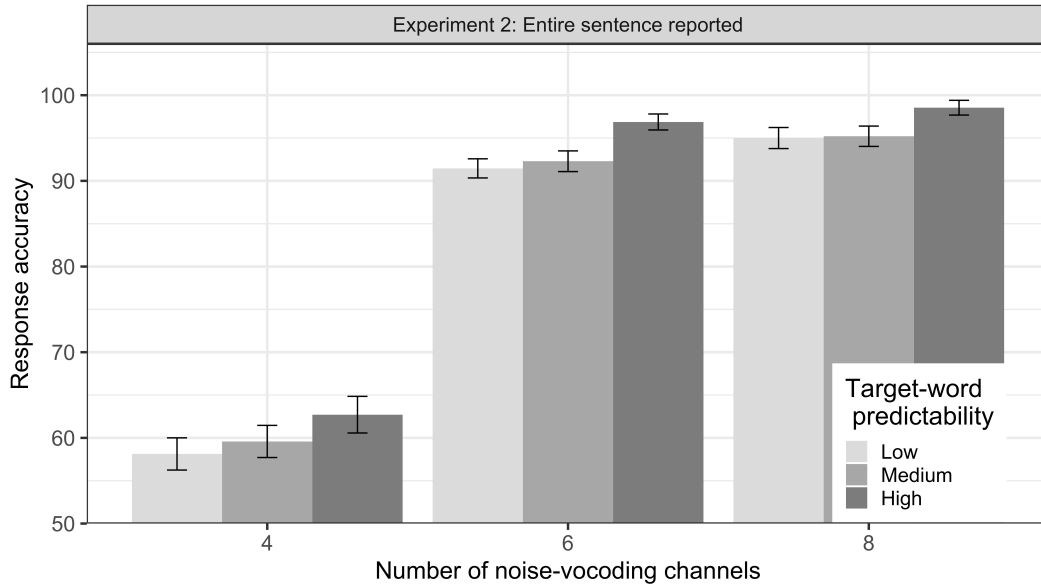


Figure 5.2: Mean response accuracy across all conditions in Experiment 1B. Accuracy increased with an increase in number of noise vocoding channels and target-word predictability. Error bars represent standard error of the means.

5.10 Results and discussion

Mean response accuracy for different conditions are shown in Table 5.3 and are displayed in Figure 5.2. It shows that the accuracy increased with an increase in both the number of noise vocoding channels, and the target word predictability.

Table 5.3: Mean response accuracy across all levels of speech degradation and target word predictability in Experiment 1B.

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

These observations are confirmed by the results of statistical analyses (Table

5. Attention-prediction interplay

5.4): We again found a main effect of number of noise vocoding channels such that response accuracy at 8 channels was higher than both 4 channels ($\beta = -3.49$, $SE = .23$, $z(4320) = -15.29$, $p < .001$), and 6 channels noise vocoding ($\beta = -0.61$, $SE = .20$, $z(4320) = -3.07$, $p = .002$). In contrast to Experiment 1A, there was also a main effect of target word predictability: Response accuracy in high predictability sentences was significantly higher than in low predictability sentences ($\beta = 1.25$, $SE = .28$, $z(4320) = 4.50$, $p < .001$). We also found a statistically significant interaction between speech degradation and target word predictability ($\beta = -.95$, $SE = .30$, $z(4320) = -3.14$, $p = .002$). Subsequent subgroup analyses of each channel condition showed that the interaction was driven by the difference in response accuracy between high predictability sentences and low predictability sentences at 8 channels ($\beta = 1.42$, $SE = .62$, $z(1440) = 2.30$, $p = .02$), and 6 channels noise vocoding conditions ($\beta = 1.14$, $SE = .34$, $z(1440) = 3.31$, $p < .001$).

Table 5.4: Estimated effects of the best fitting optimal model accounting for the correct word recognition in Experiment 1B.

Fixed effects	Estimate	Std. Error	z value	p value
Intercept	4.07	.23	17.51	<.001
Noise condition (4 channels)	-3.49	.23	-15.29	<.001
Noise condition (6 channels)	-.61	.20	-3.07	.002
Target word predictability (Low-Medium)	-.10	.16	-.60	.55
Target word predictability (High-Low)	1.25	.28	4.50	<.001
Noise condition \times Target word predictability	-.95	.30	-3.14	<.001

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

5. Attention-prediction interplay

To test whether the difference between experimental manipulations is statistically significant, we combined the data from both the experiments in a single analysis. We ran another binomial linear mixed-effects model on response accuracy and followed the same procedure as Experiment 1A and Experiment 1B to obtain the optimal model. The model summary is shown in Table 5.5.

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

Fixed effects	Estimate	Std. Error
Intercept	4.09	.23
Noise condition (4 channels)	-3.49	.23
Noise condition (6 channels)	-.58	.22
Target word predictability (Low-Medium)	.15	.27
Target word predictability (High-Low)	1.09	.35
Experimental group	.19	.28
Noise condition (4 channels) \times Target word predictability (Low-Medium)	-.15	.30
Noise condition (6 channels) \times Target word predictability (Low-Medium)	-.30	.35
Noise condition (4 channels) \times Target word predictability (High-Low)	-.73	.36
Noise condition (6 channels) \times Target word predictability (High-Low)	-.12	.40
Noise condition (4 channels) \times Experimental group	-.11	.27
Noise condition (6 channels) \times Experimental group	-.11	.30
Target word predictability (High-Low) \times Experimental group	-.45	.18

The model revealed that the critical interaction between experimental manipulation and target word predictability was indeed statistically significant ($\beta = -.45$, $SE = .18$, $z(8566) = -2.55$, $p = .011$), i.e., the effect of predictability was larger in the group that was asked to type in the whole sentence. Together, these findings suggest that the change in task instruction, which draws attention either to the entire sentence or only to the final word, is critical for making use of the context information under degraded speech.

5.11 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

5. *Attention-prediction interplay*

critical target word. The results of two experiments revealed that attentional processes clearly modulate the use of context information for predicting sentence endings when the speech signal is degraded.

In contrast to the first experiment, the results of the second experiment show an interaction between target word predictability and degraded speech. This is generally in line with existing studies that found a facilitatory effect of predictability at different levels of speech degradation when the participants were instructed to pay attention to the entire sentence (e.g., at 4 channels or 8 channels noise vocoded speech, [Obleser et al., 2007](#); [Obleser & Kotz, 2010](#)). The important new finding that our study adds to the present literature is that this predictability effect may be weakened or even lost, when listeners are instructed to report only the final word of the sentence that they heard, like in Experiment 1A. The lack of predictability effect and contextual facilitation can most likely be attributed to listeners not successfully decoding the meaning of the verb of the sentence, as the verb is the primary predictive cue for the target word (noun) in our stimuli. Hence, this small change in task instructions from Experiment 1A to Experiment 1B sheds light on the role of top-down regulation of attention on using context for language comprehension in adverse listening conditions. In adverse listening conditions, language comprehension is generally effortful so that focusing attention to only a part of the speech signal seems much beneficial in order to enhance stimulus decoding. However, the results of this study also show that this comes at the cost of neglecting the context information that could be beneficial for language comprehension. Our findings hence demonstrate that there is a trade-off between the use of context for generating top-down predictions vs. focusing all attention on a target word. Specifically, the engagement in the use of context and generation of top-down predictions may change as a function of attention (see also, [J. Li et al., 2014](#)). This claim is also corroborated by the significant change in predictability effects (or contextual facilitation) from Experiment 1A to Experiment 1B, in the combined dataset.

5. *Attention-prediction interplay*

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

In conclusion, this study provides a novel insight into the modulatory role of attention regulation in the interaction between top-down predictive and bottom-up auditory processes. We show that task instructions affect distribution of attention to the degraded speech signal. This, in turn, means that when insufficient attention is given to the context, top-down predictions cannot be generated, and the facilitatory effect of predictability is substantially reduced. The findings of this study indicate limitations to predictive processing accounts of language comprehension.

6

Semantic predictability facilitates comprehension of degraded speech in a graded manner

In the previous chapter we concluded that predictability facilitated comprehension of degraded speech only when listeners attended to the context, at least when the speech was not extremely degraded by 1 channel noise vocoding. A few lacunae in the study in Chapter 5 warrant further examination and exploration of its the conclusion. Firstly, there was an implicit assumption that all the noun-correct responses were borne out of correct identification of the context evoking words (i.e., verbs). Instructions to the participants and the experimental design were made to test the effect of attention to context vs attention only to the target word such that we could not test if participants identified the context correctly in [Experiment 1A](#). Secondly, we also did not consider if listeners adapted to the degraded speech, and how such an adaptation, if present, could modulate or interact with contextual facilitation. And most importantly, we only showed that there is a difference between high and low predictability sentences — the granularity of predictability was not tested. The theoretical question was focused on the role of attention to context rather than on the nature of prediction. The aim of the study presented in this chapter is to address these limitations and examine if

6. *Graded effect of predictability*

listeners form narrowed expectations or whether predictions are generated across a wide range of probable sentence endings in a graded manner. We also take into account the accuracy of context identification (not just the accuracy of target word identification) as well as the possible learning effect due to adaptation to degraded speech. The results showed that in contrast to the “narrowed expectations” view postulated for predictive processing of degraded speech, listeners probabilistically preactivate upcoming words from a wide range of semantic space, not limiting only to highly probable sentence endings. We also did not find any learning effects. We speculate that when there is a trial-by-trial variation in semantic feature (e.g., sentence predictability), listeners do not adapt to low-level perceptual property (e.g., speech quality).

6.1 Introduction

In the literature of speech perception and sentence processing, studies have argued that prediction is either probabilistic and graded, or it is all-or-nothing [CITE]. Very few studies have investigated such theoretical questions within the domain of adverse listening conditions, specifically in degraded speech comprehension (e.g. [StrauSS et al., 2013](#); see also [Corps & Rabagliati, 2020](#); [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 ([Hunter & Pisoni, 2018](#); [Obleser & Kotz, 2010](#); e.g., [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 replicate the previous findings of these predictability effects, and extend them further by testing if listeners form narrowed

6. Graded effect of predictability

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 context, and hence form a probabilistic prediction. We also test the presence of perceptual adaptation and its effect on contextual facilitation. We set a metric of measurement of language comprehension that considers whether listeners correctly identified the context information.

6.2 Background

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. Obleser and colleagues, for example, compared high and low predictability sentences and observed contextual facilitation, in terms of a 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. These studies, however, 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 strong-context, low-typical words and weak-context, low-typical words were largest. N400 responses at the latter two were not statistically different from each other. However, the N400 response were smallest at highly predictable (strong-context and high-typical) words. 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

6. Graded effect of predictability

narrowed expectations from a restricted semantic space when the sentence endings are highly predictable. This is contrary to the view that readers and listeners form a probabilistic prediction of upcoming word in a sentence. When the sentence endings are less predictable, listeners cannot preactivate those less predictable sentence endings in an adverse listening condition. For example, Nieuwland et al. (2018) 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. (2021) 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 to 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 those studies that examined degraded speech comprehension, used only two levels of semantic predictability (high and low). The granularity and the nature of prediction remain yet to be tested in degraded speech comprehension.

Adaptation to degraded speech

Listeners quickly adapt to novel speech with artificial acoustic distortions (Dupoux & Green, 1997). With a repeated exposure to degraded speech, listeners' comprehension improves over time (Guediche et al., 2014; Samuel & Kraljic, 2009). When the noise condition 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 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

6. Graded effect of predictability

level (6 or 4 channels noise vocoded speech) was presented in one block. There was no uncertainty about the next-trial speech degradation within a block from the participants' perspective, i.e., the *global channel context* was certain or predictable. Additionally, 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, then a listener is uncertain about any upcoming trials' signal quality. If such multiple levels of degradation are due to the presentation of multiple channels of noise vocoded speech, then the *global channel context* is unpredictable or uncertain. This can influence perceptual adaptation such that it might be totally absent with the change in the characteristics of auditory signal 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). It can, thus, be speculated that if the noise vocoded speech varies from one trial to the next, then the adaptation to noise in this scenario might be different from the earlier case in which spectral degradation is constant throughout the experiment. Perceptual adaptation, however, is not limited to trial-by-trial variability of stimulus property. Listeners can adapt to auditory signal at different time courses or time scales (Atienza et al., 2002; see also, Whitmire & Stanley, 2016). In addition to the differences in intrinsic trial-by-trial variability (that is involved in short timescale trial-by-trial adaptation), the global differences in the presentation of vocoded speech can result in the general adaptation at a longer timescale which can differ between predictable and unpredictable channel contexts.

Only a limited number of studies has looked at how next-trial noise-uncertainty and global context of speech property influence adaptation. For example, words were presented at +3dB SNR and +10dB SNR in a word-recognition task in a pseudorandom order (Vaden et al., 2013). The authors wanted to minimize the certainty about the noise conditions in the block. They proposed that an adaptive control system (cingulo-opercular circuit) might be involved to optimize task performance when listeners are uncertain about an upcoming trial (Eckert et

6. Graded effect of predictability

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 change in performance over the course of 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. In the above-mentioned studies, the authors did not compare participants' task performance in a blocked design against the presentation in a pseudo-randomized block of different degradation levels to make an inference about general adaptation to degraded speech at a longer timescale and adaptation at a shorter timescale on a trial-by-trial level.

Measurement of language comprehension

How we measure language comprehension has rarely been guided by any specific theoretical motive in the existing literature. There is a discrepancy across studies in how language comprehension in degraded speech is quantified. Some studies that reported contextual facilitation in degraded speech comprehension used proportion of correctly reported *final* words only (Hunter & Pisoni, 2018; Obleser & Kotz, 2010; e.g., 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. Additionally, none of these measures take into account if listeners have correctly identified the context (cf. Amichetti et al., 2018; see also, Marrufo-Pérez et al., 2019) which should have been the most important factor to be considered in the first place.

6. Graded effect of predictability

The present study

Stemming from the results of Chapter 5, and from the motivation driven by the open questions outlined above, the goals of the study in this chapter were threefold: The first goal was to replicate the facilitatory effect of predictability, and examine the nature of predictability, i.e., to test if listeners form narrowed expectations. Obleser and colleagues have shown predictability effects (or contextual facilitation) to appear only at a moderate level of speech degradation by using only two levels of sentence 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 by also taking into account the accuracy of context. If the listeners form narrowed predictions only for high-cloze target words, then the facilitatory effect of semantic prediction will be observed only at these highly predictable sentence endings. Listeners' response to medium-cloze target words and low-cloze target words would be quite similar as 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. The addition 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

6. Graded effect of predictability

sentences of different levels of predictability blocked by each channel conditions (*predictable channel context*), and pseudo-randomized across all channels (*unpredictable channel context*). Based on previous findings, we expected that in the unpredictable channel context (i.e., when sentences are presented in a random order of spectral degradation) participants' word recognition performance will be worse than in the predictable channel context (Garrido et al., 2011; i.e., when the sentences are blocked by noise-vocoding, Sommers et al., 1994; Vaden et al., 2013). To further examine perceptual adaptation, we also considered the effect of trial number in the analyses.

We also aimed at measuring language comprehension with a metric that reflects 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 and assigned them to the *predictable channel context* ($n = 50$; \pm SD = 23.6 ± 3.2 years; age range = 18 – 30 years; 14 females). We created *unpredictable channel context* group from participants ($n = 48$; $\bar{x} \pm$ SD = 24.44 ± 3.5 years; age range = 18 – 31 years; 16 females) from Experiment 1B. All participants were native speakers of German residing in Germany. Exclusion criteria for participating in this study were self-reported hearing disorder, speech-language disorder, or any neurological disorder.

6.3.2 Materials

We used the same stimuli described in Chapter 3 (Section 3.1). They were digital recordings of 360 German sentences spoken by a female native speaker of German in a normal rate of speech. All sentences were in present tense consisting of pronoun, verb, determiner, and object (noun) in the Subject-Verb-Object form. We used 120

6. Graded effect of predictability

sentences each for low predictability, medium predictability, and high predictability sentences that differed in the cloze probability of sentence final target words. Their mean cloze probabilities were 0.022 ± 0.027 ($\bar{x} \pm \text{SD}$; range = 0.00 – 0.09) for low predictability sentences, 0.274 ± 0.134 ($\bar{x} \pm \text{SD}$; range = 0.1 – 0.55) for medium predictability sentences, and 0.752 ± 0.123 ($\bar{x} \pm \text{SD}$; range = 0.56 – 1.00) for high predictability sentences.

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

The same set of stimuli and experimental lists were used in the predictable channel context. Each participant was presented with 120 unique sentences blocked by channel conditions, i.e., blocked by noise vocoding channels. There were four blocks of stimuli, 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. Graded effect of predictability

6.3.3 Procedure

Participants were asked to use earphones or headphones. A prompt to adjust loudness was displayed at the beginning of the experiment: A noise vocoded sound not used in the main experiment was presented, and participants were asked to adjust the loudness at their level of comfort. One spoken sentence was presented in each trial with an inter-trial interval of 1000ms. Eight practice trials were presented before presenting 120 experimental trials. They were asked to enter what they had heard (i.e., to type in the entire sentence) *via* keyboard. Guessing was encouraged. The response was not timed. The experiment was about 40 minutes long.

6.4 Analyses

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

We analyzed response accuracy with generalized linear mixed models following the procedure described in Chapter 4 (Section 4.5). Binary responses (correct/incorrect) for all participants in both groups (predictable channel context and unpredictable channel context) were fit with a binomial logistic mixed-effects model (Jaeger, 2006, 2008). Noise condition (categorical; 4, 6, and 8 channels noise

6. Graded effect of predictability

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 noise condition and target word predictability were included in the fixed effects.

We first fitted a model with maximal random effects structure that included random intercepts for each participant and item (Barr et al., 2013). Both, by-participant and by-item random slopes were included for noise condition, target word predictability and their interaction. To find the optimal model for the data, non-significant higher-order interactions were excluded from the fixed-effects structure (and from the random-effects structure) in a stepwise manner. Model selection was based on Akaike Information Criterion (Grueber et al., 2011; Richards et al., 2011).

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

6.5 Results and discussion

We observed that the mean response accuracy increased with an increase in number of noise vocoding channels from 4 to 6 to 8, and with an increase in target word predictability from low to medium to high (see Figure X.X). This trend was consistent across both the channel contexts; Figure X.X and Figure Y.Y show this trend for predictable channel context (i.e., blocked design) and unpredictable channel context (i.e., randomized design) respectively. Mean accuracies across all conditions are given in Table X and Y, and Figure X.X.

These observations are confirmed by the results of statistical analyses (Table XYZ). We found a significant main effect of channel condition indicating that the

6. Graded effect of predictability

response accuracy was higher in the 8 channels than in the 4 channels ($\beta = -2.87$, $SE = .22$, $z(6917) = -13.1$, $p < .001$) and 6 channels ($\beta = -.66$, $SE = .19$, $z(6917) = -3.42$, $p < .001$). There was a significant main effect of target word predictability suggesting that response accuracy was lower at low predictability sentences than both high predictability sentences ($\beta = 2.18$, $SE = .3$, $z(6917) = 7.2$, $p < .001$) and medium predictability sentences ($\beta = -.52$, $SE = .27$, $z(6917) = -1.97$, $p = .049$). We also found a significant interaction between channel condition and target word predictability ($\beta = -.71$, $SE = .29$, $z(6917) = -2.44$, $p = .015$).

We performed a subgroup analyses on each noise channel condition which revealed that the interaction was driven by the effect of predictability at 4 channels: The accuracy at high predictability sentences was higher than medium predictability sentences ($\beta = 1.14$, $SE = .37$, $z(1608) = 3.1$, $p < .001$), which in turn was also higher than low predictability sentences ($\beta = 1$, $SE = .24$, $z(1608) = 4.2$, $p < .001$). There was no significant difference in response accuracy between low predictability and high predictability sentences at both 6 channels ($\beta = .33$, $SE = .32$, $z(2590) = 1.04$, $p = .3$) and 8 channels ($\beta = -.014$, $SE = .32$, $z(2719) = -.04$, $p = .97$).

We also found a significant main effect of global channel context which showed that the response accuracy was higher in predictable channel context than in unpredictable channel context ($\beta = -.27$, $SE = .14$, $z(6917) = -2.02$, $p = .04$).

Further, to test the effect of practice on adaptation to degraded speech, we added trial number as a fixed effect in the maximal model. Note that there were 30 trials in each block in the predictable channel context (i.e., blocked design). For comparability, we divided unpredictable channel context (i.e., randomized design) into four blocks 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 ($\beta = -.0004$, $SE = .01$, $z(6917) = -.05$, $p = .97$). It remained constant within each block in the predictable channel context ($\beta = -.02$, $SE = .01$, $z(3291) = -1.43$, $p = .15$) as well as in the unpredictable channel context ($\beta = .01$, $SE = .01$, $z(3291) = 1.05$, $p = .29$).

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 they have listened to.

Results of our study showed the expected interaction between predictability and degraded speech, that is, language comprehension was better for high-cloze than for low-cloze target words when the speech signal was moderately degraded by noise vocoding through 4 channels, while the effect of predictability was absent when speech was not intelligible (noise vocoding through 1 channel). These results are fully in line with Obleser & Kotz (2010); we partly included identical sentences from their study in the present study (see Appendix A). Importantly, in contrast to their study, we had also created sentences with medium-cloze target words (which were intermediate between high-cloze and low-cloze target words) and found that the effect of predictability was also significant when comparing sentences with medium-cloze target words against sentences with low-cloze target words at 4 channels noise vocoding condition. Recognition of a target word was dependent on its level of predictability (measured by cloze probability), and correct recognition was not just limited to high-cloze target words. These significant differences in response accuracy between medium-cloze and low-cloze target words, and between medium-cloze and high-cloze target words at noise-vocoding through 4 channels show that the sentence-final word recognition is facilitated by semantic predictability in a graded manner. This is in line with the findings from the ERP literature where it has been observed that semantic predictability, in terms of cloze probability of target word of a sentence, modulates semantic processing, indexed by N400, in a graded manner (DeLong et al., 2005; Nieuwland et al., 2018; Wlotko & Federmeier, 2012).

6. *Graded effect of predictability*

The interpretation of the observed graded effect of semantic predictability at the moderate level of spectral degradation (i.e., at noise-vocoding through 4 channels) provides a novel insight into how listeners form prediction when the bottom-up input is compromised. That is, in an adverse listening condition, listeners rely more on top-down semantic prediction than on bottom-up acoustic-phonetic cues. However, such a reliance on top-down prediction is not an all-or-nothing phenomenon; instead, listeners form a probabilistic prediction of the target word. The effect of target word predictability on comprehension is not sharply focused solely on high-cloze target words like a ‘searchlight’ as proposed by Strauss and colleagues. But rather it is spread across a wide range including low-cloze and medium-cloze target words. As the cloze probability of the target words decreases from high to low, the focus of the searchlight becomes less precise.

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

7

Comprehension of degraded speech is modulated by the rate of speech

7.1 Introduction

7.2 Background

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

7. Speech rate and predictability effects

7.2.1 Comprehension of degraded speech

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

Listeners rely more on top-down predictions when the speech signal is less intelligible due to spectral degradation. Hence, they use the context information of the sentence to narrow down their predictions to a smaller set of semantic categories or words (StrauSS et al., 2013; see also, Corps & Rabagliati, 2020). However, it is important that the context itself is ‘intelligible enough’ which is the case when the speech is only moderately degraded. For example, Obleser and colleagues (Obleser et al., 2007; Obleser & Kotz, 2010; Obleser & Kotz, 2011) found that at moderate levels of speech degradation, target words (the sentence final words) were better recognized when it was predictable from the sentence context than when it was unpredictable. When the speech signal is clear or only very mildly degraded, there is typically no effect of predictability on comprehension, as even unpredictable words can be understood well in this condition (intelligibility is at ceiling). In contrast, when the speech signal is extremely degraded (for instance at 1 channel noise vocoding), no facilitation from the context can be observed as the context itself cannot be understood and hence it cannot help with comprehension (Bhandari et al., 2021; Obleser & Kotz, 2010).

7. *Speech rate and predictability effects*

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

7.2.2 **Comprehension of fast and slow speech**

A change in speech rate manipulates the speech signal without producing any spectral degradation (Charpentier & Stella, 1986; Moulines & Charpentier, 1990; Schlueter et al., 2014). Understanding fast speech is more effortful compared to normal and slow speech, (e.g., Müller et al., 2019; Winn & Teece, 2021; see also, Simantiraki & Cooke, 2020), and its intelligibility and comprehension are reduced (Fairbanks & Kodman Jr., 1957; Pelle & Wingfield, 2005; Schlueter et al., 2014). The comprehension deficit in fast speech has been linked to speed of processing (Gordon-Salant & Fitzgibbons, 1995; Tun, 1998; see also, Rönnberg et al., 2013) given the limited time available to decode and understand the information in the fast speech. It has also been suggested that the decoding and identification of incoming information in the fast speech puts a high demand on available cognitive resources (e.g., Rodero, 2016) such that processing rapidly flowing information exhausts the cognitive resource required for language processing (Gordon-Salant & Fitzgibbons, 2004; Janse, 2009). Hence, intelligibility and comprehension of fast speech is reduced compared to normal speech. In contrast, the central auditory-language comprehension system is shown to be flexible to process slow speech without reducing intelligibility (Lerner et al., 2014; see also, Vagharchakian et al., 2012). However, some earlier studies have casted doubt on the processing advantage of slow speech (Kemper & Harden, 1999; e.g., Nejime & Moore, 1998; see also, Liu & Zeng, 2006; Love et al., 2009). In sum, compression and expansion of speech have differential effects on speech intelligibility and language comprehension. Intelligibility and comprehension of fast speech is reduced while that of slow speech is generally increased compared to normal speech.

7. Speech rate and predictability effects

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

The effects of predictability at varying rates of presentation have been also investigated with self-paced reading studies. For example, Wlotko & Federmeier (2015) presented participants with context evoking sentences followed by sentences

7. Speech rate and predictability effects

containing a target word that was either expected (mean cloze probability of 74%) or unexpected (either same or different semantic category, both with cloze probability of approximately 0%). They found that the facilitation effect (as reflected in the N400 amplitude) was reduced at the sentences that were presented fast compared to the ones that were presented slow. They suggested that at fast rate of presentation, predictive preactivation of words was not common: There was not enough time to activate proper representation for processing of upcoming word. In the same study, however, when the fast presentation was followed by slow presentation in separate blocks, semantic facilitation effect was not reduced. That is, increase in the flow of information did not always impair the ability to predict (see also, [Cole, 2020](#)). They argued that once the brain is engaged in predictive comprehension mode, for example, first in the slow presentation rate, it can then continue to allocate resources in the same mode under faster presentation rate.

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

7.2.3 Comprehension of degraded speech at varying rates of presentation

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

7. Speech rate and predictability effects

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

So far, the evidence for the detrimental effects of combination of spectral degradation and speech compression is sparse. Most of the evidence on this comes from cochlear implant users (e.g., Y. Li et al., 2011; Su et al., 2016) as they are the group of listeners whose auditory input is spectrally degraded. Iwasaki et al. (2002) found that a change in speech rate from slow to fast reduced word recognition accuracy in cochlear implant users. Similarly, their speech perception was impaired with increased rate of speech, and it was improved when the speech rate was decreased (e.g., Dincer D'Alessandro et al., 2018). Meng et al. (2019) used both clean speech and spectrally degraded speech from MSP (Mandarin speech perception) and MHINT (Mandarin hearing in noise test) corpora in normal-hearing listeners and cochlear implant users. They found that an increase in speech rate had much severe effect at spectrally degraded speech (4 channels sine-wave vocoded) than at clean speech. To achieve the same level of accuracy, listeners required degraded speech to be much slower than the normal speech. All of these studies indicated that when speech rate is increased, intelligibility and comprehension of degraded speech is reduced. However, these studies did not

7. Speech rate and predictability effects

examine the utility of context information when speech is distorted by more than one means, although the role of context and predictability in any one of these multiple distortions has been extensively discussed in the literature (Aydelott & Bates, 2004; Bhandari et al., 2021; C. Clark et al., 2021; Corps & Rabagliati, 2020; Goy et al., 2013; Obleser & Kotz, 2010; Os et al., 2021; Sheldon et al., 2008a, 2008b).

Taken together, the utility of semantic predictability in comprehension of degraded speech is fairly established whereas its effects in comprehension of both fast speech as well as slow speech is inconsistent. Whether semantic predictability is beneficial even for degraded speech presented at different rates (fast and slow) has not been investigated yet.

7.2.4 Study aims

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

To achieve the goal, we conducted two experiments in which listeners were required to listen to the sentences and type in the entire sentence they hear. Sentence comprehension (word recognition accuracy) for high and low predictability sentences were assessed in fast speech (Experiment 1), and slow speech (Experiment 2). Because the processing demand increases, and a limited time is to be available

7. Speech rate and predictability effects

to process the context and form predictions (e.g., [Aydelott & Bates, 2004](#); [Wlotko & Federmeier, 2015](#)), we expected that the contextual facilitation (i.e., the difference between high and low predictability sentences) will be reduced for fast speech compared to normal speech (Experiment 3A). However, for slow speech due to abundance of time to process the degraded speech and the context, and reduction in effortful processing (e.g., [Winn & Teece, 2021](#)), we expected contextual facilitation to be increased compared to normal speech (Experiment 3B). We expected that both increase and decrease in contextual facilitation will be primarily driven by the ease of processing high predictability sentences as compared to low predictability sentences ([Aydelott & Bates, 2004](#); [Goy et al., 2013](#)).

7.3 Experiment 7A

7.4 Methods

7.4.1 Participants

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

7.4.2 Materials

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

7. Speech rate and predictability effects

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

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

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

7.4.3 Procedure

Participants were asked to use headphones or earphones. A sample of noise vocoded speech not used in the practice trial and the main experiment was provided so that the participants could adjust the loudness to a preferred level of comfort at the beginning of the experiment. The participants were instructed to listen to the sentences and to type in the entire sentence by using the keyboard. The time for typing in the response was not limited. They were also informed at the beginning of the experiment that some of the sentences would be ‘noisy’ and not easy to understand, and in these cases, they were encouraged to guess what they might have heard. They were not informed about the speed of speech being slow/fast or normal. Eight practice trials with different levels of speech degradation were given to familiarize the participants with the task before presenting all 120 experimental trials with an inter-trial interval of 1000 ms.

7.5 Analyses

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

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

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

We fitted a model with maximal random effects structure that included random intercepts for each participant and item (Barr et al., 2013). Both, by-participant and by-item random slopes were included for target word predictability, speed and their interaction.

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

7.6 Results and discussion

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

7. *Speech rate and predictability effects*

4.16, $p < .001$). These suggested that participants' response accuracy was higher for the high predictability sentences than for the low predictability sentences, and for normal speech than for fast speech. We also found a significant interaction between target word predictability and speech rate ($\beta = 1.06$, $SE = .42$, $z = 2.50$, $p = .01$). As it can be seen in Figure 1, the effect of target word predictability was reduced at fast speech. These results are shown in Table 3.

Separate planned analyses of each predictability level were performed following the same procedure described above in the Analysis section. The results shown in Table 4 and Table 5 revealed that there was no significant main effect of speech rate at high predictability condition ($\beta = .02$, $SE = .34$, $z = .05$, $p = .96$). At low predictability condition, in contrast, we found a significant main effect of speech rate ($\beta = -.99$, $SE = .27$, $z = -3.72$, $p < .001$). Hence, response accuracy decreased at fast speech only for the low predictability condition.

Separate planned analyses of each speech rate shown in Table 6 and Table 7 revealed that there was significant main effect of predictability in both normal speech ($\beta = 1.98$, $SE = 0.28$, $z = 7.05$, $p < .001$) and fast speech conditions ($\beta = 2.67$, $SE = .37$, $z = 7.14$, $p < .001$), but the effect appeared to be higher for fast speech ($\beta = 2.67$) than for normal speech ($\beta = 1.98$). This, however, is a result of significant reduction in accuracy at low predictability condition at fast speech rather than due to an increase in accuracy at high predictability condition. This can also be seen in Table 2 and Figure 1.

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

7. Speech rate and predictability effects

sentences and increases the contextual facilitation in comprehension of moderately degraded speech.

7.7 Experiment 7B

7.8 Methods

7.8.1 Participants and Materials

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

We used the same sentences that were used in Experiment 1. However, the auditory recordings were expanded by a factor of 1.35 to create slow speech. All other procedure to create stimuli were identical to Experiment 1 resulting in two sets (one set of normal and one set of slow speech) of 240 sentences (120 high and 120 low predictability sentences) which were passed through 4 channels noise vocoding.

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

7.8.2 Procedure

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

7.9 Analyses

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

7. *Speech rate and predictability effects*

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

7.10 Results and discussion

Mean response accuracies for all experimental conditions are shown in Table 8. There was a significant main effect of target word predictability indicating that participants' response accuracy was higher for the high predictability condition than for the low predictability condition ($\beta = 2.58$, $SE = .30$, $z = 8.65$, $p < .001$). In contrast to Experiment 1, we did not find a significant main effect of speech rate ($\beta = -.08$, $SE = .15$, $z = .57$, $p = .568$), nor there was a significant interaction between speech rate and target word predictability ($\beta = .44$, $SE = .27$, $z = 1.65$, $p = .099$). These suggested that there was no change in participants' response accuracy with a reduction in speech rate, nor did the contextual facilitation significantly increase or decrease with slowing down of the speech rate. It can be seen in Figure 2 that the effect of target word predictability did not change with speech rate. These results are shown in Table 9.

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

7. *Speech rate and predictability effects*

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

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

7.11 Conclusion

The main goals of the present study were to examine if semantic predictability facilitates comprehension of moderately degraded speech at different speech rates given that the ease of processing the sentences varies with their predictability and speed. The results of the two experiments revealed that fast speech selectively impedes the comprehension of low predictability sentences, while slow speech has no effect on contextual facilitation at a moderate level of degradation.

In both the experiments, our results showed a significant main effect of predictability at normal speech rate, i.e., we observed a facilitatory effect of semantic predictability at normal speech rate under moderate degradation level of 4 channels noise vocoding. This replicates the findings from earlier studies like Obleser & Kotz (2010) and Bhandari et al. (2021) in which participants were presented only

7. Speech rate and predictability effects

with normal speech rate, and contextual facilitation was observed at 4 channels noise vocoding. At this moderate degradation level, listeners were able to decode the context and form its meaning representation. Consequently, they generated predictions about the upcoming target word in a sentence even in low predictability condition depending on the contextual constrain of the sentences (Bhandari et al., 2021; see also, StrauSS et al., 2013).

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

In contrast to Experiment 1, we did not find the expected interaction between speech rate and target word predictability in Experiment 2, i.e., decrease in speech rate did not differentially affect the comprehension of high or low predictability sentences although the comprehension tended to decrease at slow speech rate compared to normal speech rate. As opposed to Experiment 1, we did not observe a significant change in contextual facilitation at slow speech rate at a moderate degradation level in Experiment 2. Slowing down the speech provides listeners more time to process the information, including the context that is important to generate

7. Speech rate and predictability effects

predictions. However, our findings show that this ease of processing does not benefit intelligibility and comprehension of sentences with highly predictable target words any more than the sentences with less predictable target words. This claim is in line with the results of Winn & Teece (2021)] who reported that intelligibility does not increase when the speech is slowed down even though there is a difference in the intelligibility between high and low predictability sentences at both normal and slow speech rates.

Accounts from speech perception, and predictive language processing point to a common expectation: contextual facilitation is enhanced when listeners have more time to process the incoming information [CITE]. Additionally, there are conflicting empirical evidence whether increase or decrease of speech rate provides benefit in intelligibility, comprehension, and contextual facilitation. Given these state of affairs, our findings add new theoretical insights with empirical support into the interplay among spectral degradation, speech rate, and semantic prediction. Although reducing the speech rate provides time to process the information (including the context) in the degraded speech, this eases the processing of both high and low predictability to a similar extent. And thus, no increased facilitatory effect is observed at slow speech rate. In contrast, increasing the speech rate adds more cognitive load on the top of the effort required to process degraded speech. This results in difficulty in processing and understanding the rapidly unfolding sentences among which this difficulty is further increased when the target words are not easily predictable.

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

To conclude, we show that processing speed and constraints in attentional and cognitive resources are key factors that influence contextual facilitation of moderately degraded speech. When enough time is available to process information, i.e.,

7. Speech rate and predictability effects

at slow speech, contextual facilitation does neither increase nor decrease. However, at a time crunch of information processing, i.e., at fast speech, contextual facilitation is reduced such that the fast speech is detrimental to understanding words that are not easily predictable from the context.

8

Discussion and conclusion

8.1 Summary of the main findings

8.2 Theoretical and practical implications

8.3 Conclusion

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



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>
- Adams, E. M., Gordon-Hickey, S., Morlas, H., & Moore, R. (2012). Effect of rate-alteration on speech perception in noise in older adults with normal hearing and hearing impairment. *American Journal of Audiology*, *21*(1), 22–32. [https://doi.org/10.1044/1059-0889\(2011/10-0023\)](https://doi.org/10.1044/1059-0889(2011/10-0023))
- Adams, E. M., & Moore, R. E. (2009). Effects of speech rate, background noise, and simulated hearing loss on speech rate judgment and speech intelligibility in young listeners. *Journal of the American Academy of Audiology*, *20*(1), 28–39. <https://doi.org/10.3766/jaaa.20.1.3>
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. [https://doi.org/10.1016/s0010-0277\(99\)00059-1](https://doi.org/10.1016/s0010-0277(99)00059-1)
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. <https://doi.org/10.1016/j.jml.2006.12.004>
- Amichetti, N. M., Atagi, E., Kong, Y.-Y., & Wingfield, A. (2018). Linguistic Context Versus Semantic Competition in Word Recognition by Younger and Older Adults With Cochlear Implants. *Ear & 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>

A. Experimental items

- 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>
- Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, 49(12), 3512–3516. <https://doi.org/10.1016/j.neuropsychologia.2011.08.014>
- Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning and Memory*, 9(3), 138–150. <https://doi.org/10.1101/lm.46502>
- 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>
- Bhandari, P., Demberg, V., & Kray, J. (2021). Semantic predictability facilitates comprehension of degraded speech in a graded manner. *Frontiers in Psychology*, 12, 3769. <https://doi.org/10.3389/fpsyg.2021.714485>
- 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>

A. Experimental items

- Burnham, K. P., & Anderson, D. R. (2002). Basic Use of the Information-Theoretic Approach. In *Model selection and multimodel inference* (2nd ed., pp. 98–148). https://doi.org/10.1007/978-0-387-22456-5_3
- 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.
- 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>
- 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, C., Guediche, S., & Lallier, M. (2021). Compensatory cross-modal effects of sentence context on visual word recognition in adults. *Reading and Writing*, 34(8), 2011–2029. <https://doi.org/10.1007/s11145-021-10132-x>
- 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](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.
- 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>
- 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., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech and Language*, 28(2), 543–571. <https://doi.org/10.1016/j.csl.2013.08.003>

A. Experimental items

- 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>
- 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>
- De Raeve, L., Archbold, S., Lehnhardt-Gorjany, M., & Kemp, T. (2020). Prevalence of cochlear implants in Europe: Trend between 2010 and 2016. *Cochlear Implants International*, 21(5), 275–280. <https://doi.org/10.1080/14670100.2020.1771829>
- 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>
- 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

A. Experimental items

- 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](https://doi.org/10.1016/S0022-5371(81)90220-6)
- 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., 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, 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.
- Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25(3), 348–368.
- 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>

A. Experimental items

- Friston, K. J., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., & Holmes, E. (2020). Active listening. *Hearing Research*, *xxx*, 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*.
- 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>
- 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\(2013/12-0053\)](https://doi.org/10.1044/1092-4388(2013/12-0053))
- 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.1420-9101.>

A. Experimental items

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
- 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*.
- Harris, D. J. (1960). Combinations of distortion in speech. *Archives of Otolaryngology*, 72(2), 227–232.
- 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>
- Heilbron, M., Armeni, K., Schoffelen, S., Jan-Mathijs, Hagoort, P., & de Lange, F. P. (2021). A hierarchy of linguistic predictions during natural language comprehension. *bioRxiv*, 2020–2012. <https://doi.org/10.1101/2020.12.03.410399>
- 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.
- 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*,

A. Experimental items

- 39(2), 378–389. <https://doi.org/10.1097/AUD.0000000000000493>
- 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>
- Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech* (p. 231) [Doctoral dissertation]. Stanford University.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446. <https://doi.org/10.1016/j.jml.2007.11.007>
- 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>
- 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](https://doi.org/10.1016/s0749-596x(03)00023-8)
- 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>

A. Experimental items

- 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>
- Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, 115(1), 362–378. <https://doi.org/10.1121/1.1635842>
- 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, 263. <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.
- 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>
- Li, J., Xia, R., Ying, D., Yan, Y., & Akagi, M. (2014). Investigation of objec-

A. Experimental items

- tive 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>
- Li, Y., Zhang, G., Kang, H., Liu, S., Han, D., & Fu, Q.-J. (2011). Effects of speaking style on speech intelligibility for Mandarin-speaking cochlear implant users. *The Journal of the Acoustical Society of America*, 129(6), EL242–EL247. <https://doi.org/10.1121/1.3582148>
- 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>
- 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>
- Lupyan, G., & Clark, A. (2015). Words and the World. *Current Directions in Psychological Science*, 24(4), 279–284. <https://doi.org/10.1177/0963721415570732>
- 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-Martín, 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, D. W., Murphy, R. L., & Meyer, A. (1956). Articulation reduction by combined distortions of speech waves. *The Journal of the Acoustical Society of America*, 28(4), 597–601. <https://doi.org/10.1121/1.1918111>
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*,

A. Experimental items

- 27(7-8), 953–978. <https://doi.org/10.1080/01690965.2012.705006>
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- McCullough, C. M. (1958). Context aids reading. *The Reading Teacher*, 11(4), 225–229.
- 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>
- Meteyard, L., & Davies, R. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092. <https://doi.org/10.31234/osf.io/h3duq>
- 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>
- 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

A. Experimental items

- 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>
- 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., Vasisht, 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>
- Obleser, J. (2014). Putting the Listening Brain in Context. *Language and Linguis-*

A. Experimental items

- tics 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*.
- Os, M. van, 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.
- 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>
- 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: Demon-*

A. Experimental items

- strations, 57–61. <https://doi.org/10.18653/v1/n16-3012>
- 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>
- 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>
- Rönnerberg, 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., & 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>
- 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>

A. Experimental items

- 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, 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

A. Experimental items

- 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>
- 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>
- 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
- Su, Q., Galvin, J. J., Zhang, G., Li, Y., & Fu, Q. J. (2016). Effects of within-talker variability on speech intelligibility in Mandarin-speaking adult and pediatric cochlear implant patients. *Trends in Hearing*, 20, 1–16. <https://doi.org/10.1177/2331216516654022>
- Th. Gries, S. (2015). The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora*, 10(1), 95–125. <https://doi.org/10.3366/cor.2015.0068>
- 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>
- Tun, P. A. (1998). Fast noisy speech: Age differences in processing rapid speech with background noise. *Psychology and Aging*, 13(3), 424–434. <https://doi.org/10.1037/0882-7974.13.3.424>

A. Experimental items

- 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>
- 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>
- 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>
- Whitmire, C. J., & Stanley, G. B. (2016). Rapid Sensory Adaptation Redux: A Circuit Perspective. *Neuron*, 92(2), 298–315. <https://doi.org/10.1016/j.neuron.2016.09.046>
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience*, 32(40), 14010–14021. <https://doi.org/10.1523/JNEUROSCI.1528-12.2012>
- Winn, M. B., & 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 im-

A. *Experimental items*

fact 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>

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>