Investigating Spoken Language Comprehension as Perceptual Inference

The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond.
More information can be found at www.mpi.nl/imprs.

This research was conducted at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands with support from the Max Planck Society for the Advancement of Science, Munich, Germany.

Cover design by Annika Tritschler.

ISBN: 978-94-92910-22-6

Printed and bound by Ipskamp Drukkers, b.v.

Investigating Spoken Language Comprehension as Perceptual Inference

Proefschrift
ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus J.H.J.M. van Krieken,
volgens besluit van het college van de decanen
in het openbaar te verdedigen op dinsdag 19 januari 2021
om 10.30 uur precies

door
Greta Kaufeld
geboren op 30 maart 1989
te Nienburg/Weser (Duitsland)

Promotor:

Prof. dr. Antje S. Meyer

Copromotoren:

Dr. Andrea E. Martin (MPI)

Dr. Hans Rutger Bosker (MPI)

Manuscriptcommissie:

Prof. dr. Peter Hagoort

Prof. dr. Yiya Chen (Universiteit Leiden)

Prof. dr. Sonja Kotz (Universiteit Maastricht)

Contents

1 | General introduction

Animals share the ability to perceive the world around them, and to adjust their rich and diverse behaviors to cope efficiently with their environment (Fetsch, DeAngelis, & Angelaki, 2013; Olshausen, 2014). In the Tunisian desert, ants will travel for hundreds of meters in search for food – yet they reliably find their way back to their nest, using visual and tactile information such as skylight, wind, and landmarks in their surroundings (Buehlmann, Mangan, & Graham, 2020; Huber & Knaden, 2015; Wehner, 2003). In the sea, dolphins can obtain a sense of their three-dimensional environment through echolocation (e.g., Au & Hastings, 2008). As diverse as animal behaviors can be, they all require the combination of multiple sources of information – cues – into reliable percepts of the external world.

In humans, one of the most awe-inspiring and widely studied behaviors is communication. Just like the diverse behaviors seen in animals, human communication is also grounded in perception: From multiple sources of sensory information, humans are able to understand complex meanings, articulate novel thoughts and ideas, and communicate those ideas and meanings to others around them. The sensory cues we draw on for communication are diverse, including visual, auditory, and spatio-temporal information. For spoken language comprehension (which is the focus of this thesis), the sensory information available typically consists of an auditory signal, manifesting itself as a series of quasi-periodic fluctuations of air pressure. Crucially, these air pressure fluctuations do not, intrinsically, carry any meaning by themselves. It is only in combination with our learned knowledge of a language that understanding can arise. How humans do this – how we generate *meaning* from air pressure fluctuations – is the central question of this thesis.

1.1 Perception as an inference problem

A wealth of research has discussed perception as an *inference* problem (e.g., Olshausen, 2014; Wei & Körding, 2011). Sensory cues, such as sights, sounds, odors, and tactile stimuli, do not yield sufficient information by themselves to generate a coherent understanding of the outside world (Wei & Körding, 2011). Even worse, they can be noisy and unstable, or even incomplete or partially absent. As such, perception can be seen as an ill-posed problem (Ernst & Bülthoff, 2004; Olshausen, 2014), where the task at hand (generating a reliable percept) cannot be solved by simply combining all pieces of sensory information. Instead, animals have to draw on both sensory cues from their environment and their prior knowledge about the world, *inferring* meaningful percepts by combining information from these two sources.

One of the earliest formalizations of perception as an inference problem dates back to Helmholtz (Hatfield, 1990), who argued that “sensations are only signs for the properties of the external world, whose interpretation must be learned through experience” (von Helmholtz, 1896, cited in Hatfield, 1990). In other words, meaningful percepts can only arise as the result of an inference process, combining external, sensory stimuli and internal, learned knowledge.

Perceptual inference has since been investigated in great detail from psychological, mathematical and neurophysiological angles. One particularly interesting line of research formalizes perceptual inference within the theoretical framework of *cue integration*.

1.2 Formalizing perceptual inference as cue integration

The basic building block in models of cue integration is the *cue*. Definitions of what constitutes a cue are notoriously vague (for brief discussions of this problem, see Ernst and Bülthoff, 2004; Martin, 2016) – cues are usually defined as “any sensory information that gives rise to a sensory estimate” (Ernst & Bülthoff, 2004) or, even more generally, “any signal or piece of information bearing on the state of some property of the environment” (Fetsch et al., 2013). For example, in visual processing, a cue can be shading, linear perspective, or binocular disparity (Landy, Banks, & Knill, 2011).

Models of cue integration posit that organisms *combine* and *integrate* multiple cues in order to arrive at robust estimates, or percepts, of the world. They are assumed to do this in an ideal-observer fashion, where the goal of the organism is to arrive at the single most reliable estimate (Landy et al., 2011).

There are several ways of mathematically formalizing cue integration, depending on the specific assumptions the modeler makes about the distribution and independence of cues (Landy et al., 2011). One way to determine the estimate with the smallest possible variance is Maximum Likelihood Estimation (MLE), where an integrated estimate is computed by summing the estimates derived from all individual cues, weighted by their reliability (Equation 1.1, from Landy et al., 2011):

$$\hat{x} = \sum_{i=1}^{n} w_i x_i \tag{1.1}$$

In the above equation, $\hat{x}$ denotes the integrated estimate; x_i is an individual estimate based on cue i, and w_i is the weight associated with cue i. Cue weights are proportional to their corresponding cue's *reliability*, which is defined as the cue's inverse variance, $r_i = 1/\sigma_i^2$ (Landy et al., 2011). Weights from all available cues are usually constrained to sum to 1. The reliability of the integrated estimate $\hat{x}$ is simply the sum of all individual reliabilities

$$r = \sum_{i=1}^{n} r_i, \tag{1.2}$$

from which it becomes clear that the reliability of the integrated estimate $\hat{x}$ will always be greater than (or at least equal to) that of the most reliable individual cue (Oruç, Maloney, & Landy, 2003). This also means that the variance of the integrated estimate $\hat{x}$ will always be equal to or lower than that of the individual cue with the smallest variance (Landy et al., 2011). Combining and integrating multiple cues is thus a useful strategy in two ways: 1) it maximizes the information content in a given situation, and 2) it minimizes the variance and therefore increases the robustness of the percept (Ernst & Bülthoff, 2004).

Any given cue might be very reliable in one situation, but fairly unreliable in another. To account for this variability, the weights associated with specific cues are not fixed, but can be adjusted depending on the cue's reliability (and variance) in a given situation (Ernst & Bülthoff, 2004). Related to this, prior knowledge and top-down influences on perception can also be incorporated into the cue integration model. In the simplest way, prior knowledge could be expressed as an additional summand (with an associated reliability and weight) in Equation 1.1.

Another, perhaps even more intuitive way of formalizing cue integration is by means of Bayesian Inference. In the Bayesian framework, the posterior probability distribution of a percept p given some sensory data d can be calculated as the product of the prior (*P(p)* – the probability of observing a given percept in the world) and the likelihood (*P(d|p)* – the probability of observing the sensory data arising from a specific percept). Notably, the Bayesian formalization includes an explicit prior term, which captures the organism's previous knowledge about the world and how likely it is to observe a given percept.

$$P(p|d) = \frac{P(d|p)P(p)}{P(d)} \tag{1.3}$$

Since there are usually several sources of sensory information (or in other words, several cues), and since the denominator *P(d)* is a constant that can usually be ignored (Landy et al., 2011), Equation 1.3 can be expressed as:

$$P(p|d_1, ..., d_n) \propto \prod_{i=1}^{n} P(d_i|p)P(p). \tag{1.4}$$

Intriguingly, these mathematical formalizations of cue integration rely on only two core computations: summation and normalization, which have been shown to arise both in individual neurons as well as between populations of neurons (Carandini & Heeger, 2012; Fetsch et al., 2013). As such, models of cue integration are particularly promising because not only have they been shown to accurately predict behaviour (see, e.g., Ernst & Bülthoff, 2004; Fetsch et al., 2013, for overviews), they are also neurophysiologically plausible.

1.3 Investigating spoken language comprehension as cue integration

Martin (2016, 2020) suggested cue integration as a framework to conceptualize language comprehension as perceptual inference. In this view, exogenous, acoustic cues are weighted and integrated through Bayesian inference with endogenous, linguistic cues in order for robust linguistic percepts to emerge. These inferred linguistic percepts (e.g., phonemes, syllables, words, phrases, sentences, and higher-level structures) can, themselves, act as endogenous cues for processing further downstream. The system is thus capable of supporting language comprehension across all levels of linguistic hierarchy in an iterative fashion.

Previous research has usually divided cues relevant for spoken language comprehension into two broad categories: *signal-based* and *knowledge-based* cues. Signal-based cues (sometimes also referred to as "acoustic cues") are all cues that can be measured as qualities of the acoustic signal. Examples include voice onset time, formant values, vowel length, speech rate, and so on (e.g., Bosker, 2017a; Lisker & Abramson, 1967; Maslowski, Meyer, & Bosker, 2019b; Reinisch & Sjerps, 2013; Toscano & McMurray, 2012). Knowledge-based cues (sometimes referred to as "linguistic cues" or "memory-based cues"), on the other hand, are usually considered to have been learned through experience, and they are not necessarily measurable from the properties of the acoustic signal. Examples include morphosyntactic, lexical, contextual and semantic information (Gwilliams, Linzen, Poeppel, & Marantz, 2018; Huettig & Janse, 2016; Martin, Monahan, & Samuel, 2017; Mattys, Melhorn, & White, 2007; Mattys, White, & Melhorn, 2005; Tuinman, Mitterer, & Cutler, 2014).

Surprisingly, relatively little is known about how listeners combine different cues during spoken language comprehension, especially when these cues are noisy and conflicting. Similarly, our knowledge is still limited about how exactly cue integration and perceptual inference could be instantiated in the brain during language comprehension, and more specifically, the computations and algorithms by which listeners combine different cues are still fairly elusive (Martin, 2016, 2020). The cue integration framework offers a novel way of investigating the types of signal-based and knowledge-based cues that listeners draw on, because it makes predictions about how cues might be – iteratively and flexibly – combined into robust linguistic percepts and meanings.

1.4 Summary and thesis outline

The main goal of this thesis is to investigate the cognitive and neural mechanisms underlying spoken language comprehension through the theoretical lens of cue integration and perceptual inference. Specifically, the aim is to investigate how cues from distinct levels of linguistic hierarchy are combined and integrated in order to arrive at meaningful linguistic percepts, especially in situations where conflicting cues might be available, or where cues are not equally reliable. Investigating this question in detail will help us understand more precisely the types of information, or cues, that the brain draws on when inferring meaning from sound, and how these might be combined into robust percepts during spoken language comprehension.

In the remainder of this thesis, I present results from three studies designed to investigate spoken language comprehension as perceptual inference, through the lens of cue integration. The studies in Chapters 2 and 3 used behavioral and eye-tracking measures to investigate how listeners combine and integrate knowledge-based and signal-based cues during online language comprehension.

Both chapters used contextual speech rate as the acoustic, signal-based cue (e.g., Baese-Berk et al., 2013; Baese-Berk, Dilley, Henry, Vinke, & Banzina, 2019; Bosker, 2017a; Dilley & Pitt, 2010; Maslowski, Meyer, & Bosker, 2019a) and morphosyntactic information as the linguistic, knowledge-based cue (e.g., Martin et al., 2017; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

Specifically, the experiment reported in Chapter 2 tested whether listeners use acoustic information to draw inferences about morphosyntactic gender and make predictions about upcoming lexical items and referents. This was probed using the feminine/neuter gender-marked determiner *ein/eine* in German, where the two variants only differ in the presence or absence of a single schwa phoneme. The hypotheses were that 1) the acoustic cue of contextual speech rate would influence the perception of the presence or absence of the (gender-marking) schwa phoneme, and 2) listeners — in turn — would use this acoustic information to infer morphosyntactic gender and, by extension, predict the gender of the upcoming lexical item. Crucially, the reliability of the acoustic cue was variable, thus allowing the experiment to probe whether listeners draw inferences and make predictions even in the presence of uncertainty.

Chapter 3 used the same signal-based and knowledge-based cues of contextual speech rate and morphosyntactic knowledge, this time asking a complementary question. Having found that listeners draw on both contextual speech rate and morphosyntactic information, even in the presence of uncertainty, the question was now how exactly these two cues are combined and weighted in an online fashion. Several models of spoken language comprehension posit that knowledge-based cues “outweigh” signal-based cues (e.g., Mattys et al., 2005), while cue integration frameworks predict that the weighting of different cues depends on their reliability in a given situation. The aim for this experiment was to investigate in more detail how signal-based and knowledge-based cues are weighted against each other in situations of uncertainty.

Chapter 4 presents data from an electroencephalography (EEG) experiment, which aimed to examine the neural responses to knowledge-based and signal-based cues in more detail. Specifically, the experiment contrasted Dutch sentences with jabberwocky (pseudo-sentence) items and word lists, thus probing the contributions of sentence-level prosody (using the jabberwocky control), lexical semantics (using the word list control), and acoustic fluctuations in the modulation spectra of the speech envelope (using backward controls for all three conditions).

Chapter 5 presents results from a spectral power analysis of the EEG data from the previous chapter. The aim of this chapter was to bridge two lines of previous research that have investigated the cortical response to spoken language comprehension using different techniques (e.g., Ding, Melloni, et al., 2017; Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Keitel, Gross, & Kayser, 2018).

In Chapter 6, I conclude this thesis with a broader discussion and summary of the experimental chapters. The studies in all four experimental chapters showed that listeners use signal-based and knowledge-based cues to infer meaning from sound. I discuss these experimental results within a wider context, paying particular attention to the broader questions that arise from the research in this thesis, and how they might be addressed in future experiments. I also outline the hypotheses for a planned experiment that aims to combine our insights from the previous chapters. Unfortunately, this experiment could not be conducted due to the testing restrictions related to COVID-19.

Note that Chapters 2 to 4 were written as independent journal articles. As such, they overlap to some extent in their literature reviews and discussions.

2 | Contextual speech rate influences morphosyntactic prediction and integration[1]

Abstract

Understanding spoken language requires the integration and weighting of multiple cues, and may call on cue integration mechanisms that have been studied in other areas of perception. In the current study, we used eye-tracking (visual-world paradigm) to examine how contextual speech rate (a lower-level, perceptual cue) and morphosyntactic knowledge (a higher-level, linguistic cue) are iteratively combined and integrated. Results indicate that participants used contextual rate information immediately, which we interpret as evidence of perceptual inference and the generation of predictions about upcoming morphosyntactic information. Additionally, we observed that early rate effects remained active in the presence of later conflicting lexical information. This result demonstrates that (1) contextual speech rate functions as a cue to morphosyntactic inferences, even in the presence of subsequent disambiguating information; and (2) listeners iteratively use multiple sources of information to draw inferences and generate predictions during speech comprehension. We discuss the implication of these demonstrations for theories of language processing.

[1] Adapted from Kaufeld, G., Naumann, W., Meyer, A. S., Bosker, H. R., & Martin, A. E. (2020). Contextual speech rate influences morphosyntactic prediction and integration. *Language, Cognition & Neuroscience, 35*(7), 933-948.

2.1 Introduction

Speech is an important part of human behaviour. From energy fluctuations in the air, we are able to infer complex meaning, acquire novel information, and experience rich emotions. Doing so requires us to minimally map the properties of the acoustic signal onto more abstract units, such as phonemes, morphemes, syllables, words, and sentences. Establishing this mapping between perception and meaning is, however, rarely straightforward, because the acoustic speech signal does not carry unambiguous, physically quantifiable markers for abstract, hierarchical linguistic units and structures. On top of that, it can contain multiple sources of noise, variation and uncertainty.

How does the brain accomplish this ill-posed task of mapping the acoustic signal onto linguistic units and structures? One branch of speech perception models aiming to help answer this question is tightly linked to psychophysiological models of cue integration. The goal of the current study is to examine how signal-based, perceptual (relative duration) cues and knowledge-based, linguistic cues (morphosyntactic cues to gender) are iteratively combined within such a framework of cue integration.

Cue integration as a mechanistic model for perception

Cue integration as a psychophysiological mechanism has been researched in depth in the fields of vision and multisensory perception. The underlying idea is that our perceptual experience of the world emerges from drawing inferences based on the synthesis of multiple incoming pieces of sensory information, or cues (Ernst & Bülthoff, 2004; Fetsch et al., 2013). A cue can, in principle, be "any signal or piece of information bearing on the state of some property of the environment" (Fetsch et al., 2013, p. 12) or "any sensory information that gives rise to a sensory estimate" (Ernst & Bülthoff, 2004, p. 163; see also their brief discussion of why defining a cue is so hard). Multiple cues to a specific percept are combined by means of summation and, to alleviate the sampling uncertainty arising from the fact that different cues may not be equally reliable in any given situation, integrated (or weighted) by means of normalisation. A cue's reliability in a given situation is thus encoded as its weight during the integration process. This can be formalised both as a linear operation (Equation 2.1), or in terms of Bayesian inference (see, e.g., Fetsch et al. (2013), or Landy et al. (2011), for a more detailed overview of the underlying computations). One of the most attractive aspects of cue integration as a model of perception is the neurophysiological

$$\hat{x} = \sum_{i=1}^{n} w_i x_i \tag{2.1}$$

Equation 2.1: from Landy et al., 2011. $\hat{x}$ is the estimate of the percept, and x_i is an individual cue with its associated weight w_i.

plausibility of its underlying computations: summation and normalisation have been proposed as canonical neural computations that the brain uses to solve problems across different brain regions, modalities and contexts (Carandini & Heeger, 1994, 2012).

Speech perception as cue integration

Models related to cue integration have been proposed for phoneme categorisation as early as the 1970s (e.g., Oden & Massaro, 1978; Sawusch & Pisoni, 1974). More recently, C-CuRE ("Computing Cues Relative to Expectations"; e.g., McMurray, Cole, & Munson, 2011; McMurray & Jongman, 2011), a model of speech perception that takes context into account, has been proposed and investigated extensively (e.g., Apfelbaum, Bullock-Rest, Rhone, Jongman, & McMurray, 2014; McMurray et al., 2011; Toscano & McMurray, 2015). In C-CuRE, acoustic cues are encoded relative to specific values that the listener expects in a given situation. Crucially, these expectations can be established and adjusted based on other acoustic cues. The basic computation behind C-CuRE is linear regression: Initial regression equations predicting specific cue values are established based on previous knowledge and contextual information. These regression functions are a formalisation of what McMurray and Jongman (2011) term "expectations". Newly perceived cues are interpreted relative to these expectations by computing the variance of the perceived cue from its predicted value. Note that this notion of "computing cues relative to expectations" bears striking similarities to the concept of computing prediction errors within a predictive coding framework (Toscano & McMurray, 2015).

Models such as C-CuRE propose different types of acoustic cues that are involved in making categorisation decisions on a phonemic level, and they make some predictions about how these cues interact amongst each other (e.g., McMurray & Jongman, 2011; Toscano & McMurray, 2015). However, they do not go beyond acoustic cues, and they do not make predictions about how phoneme categorisation might tie into a framework of speech comprehension that takes higher-level language comprehension as the goal of the perceptual system. There is widespread evidence that phoneme perception can be influenced by higher-level non-acoustic cues (e.g., Connine & Clifton, 1987; Fox, 1984; Ganong, 1980; Martin et al., 2017; Pitt & Samuel, 1993; Rohde & Ettlinger, 2012; van Alphen & McQueen, 2001), so any comprehensive model of speech comprehension has to account for the ways in which sensory, signal-based cues interact with morphosyntactic, lexical, pragmatic, and other knowledge-based information online.

Notably, a cue-based model of word segmentation was proposed by Mattys et al. (2005): Based on a series of word detection experiments, they suggested a hierarchy of cues for word segmentation, where both signal-based and knowledge-based cues are taken into account by the language comprehension system. The model is organised into three tiers (Tier I: lexical tier; Tier II: segmental tier; Tier III: metrical prosodic tier), with cues from higher levels of the tier hierarchy (corresponding to lexical and contextual information) taking precedence over cues from lower levels (Mattys et al., 2005). Based on a further set of experiments (Mattys et al., 2007), the authors later updated their model to include a more "graded" relationship between cues from different tiers. Especially this later model is very similar in idea to models of cue integration, where cues can be dynamically combined across levels of perceptual hierarchy. However, the model suggested by Mattys et al. (2007, 2005) focuses on word segmentation, leaving open the important question of how the comprehension system achieves *understanding* above and beyond segmenting the acoustic signal into words.

Language comprehension as cue integration

Martin (2016) proposed cue integration as a general mechanism for language processing on all levels, outlining how such a model can begin to explain all stages of language comprehension and production, from sensory processing to dialogue. In this model, functional equivalents to formal linguistic representations and higher-level meaning are inferred from sensory information by iteratively extracting, combining, and integrating relevant linguistic cues (cf. Figure 2.1). Martin (2016) suggests a cascading architecture where cues can be combined and integrated across different levels of language comprehension through a process called sensory resampling. By resampling the input across different levels of processing, multiple cues can be derived from the same sensory input. Linguistic representations that have been inferred from sensory cues can thus, in turn, be cues for higher levels of representations. For example, acoustic cues can give rise to abstract percepts such as phonemes and morphemes; phonemes and morphemes can, in turn, act as cues towards the percept of a word; words

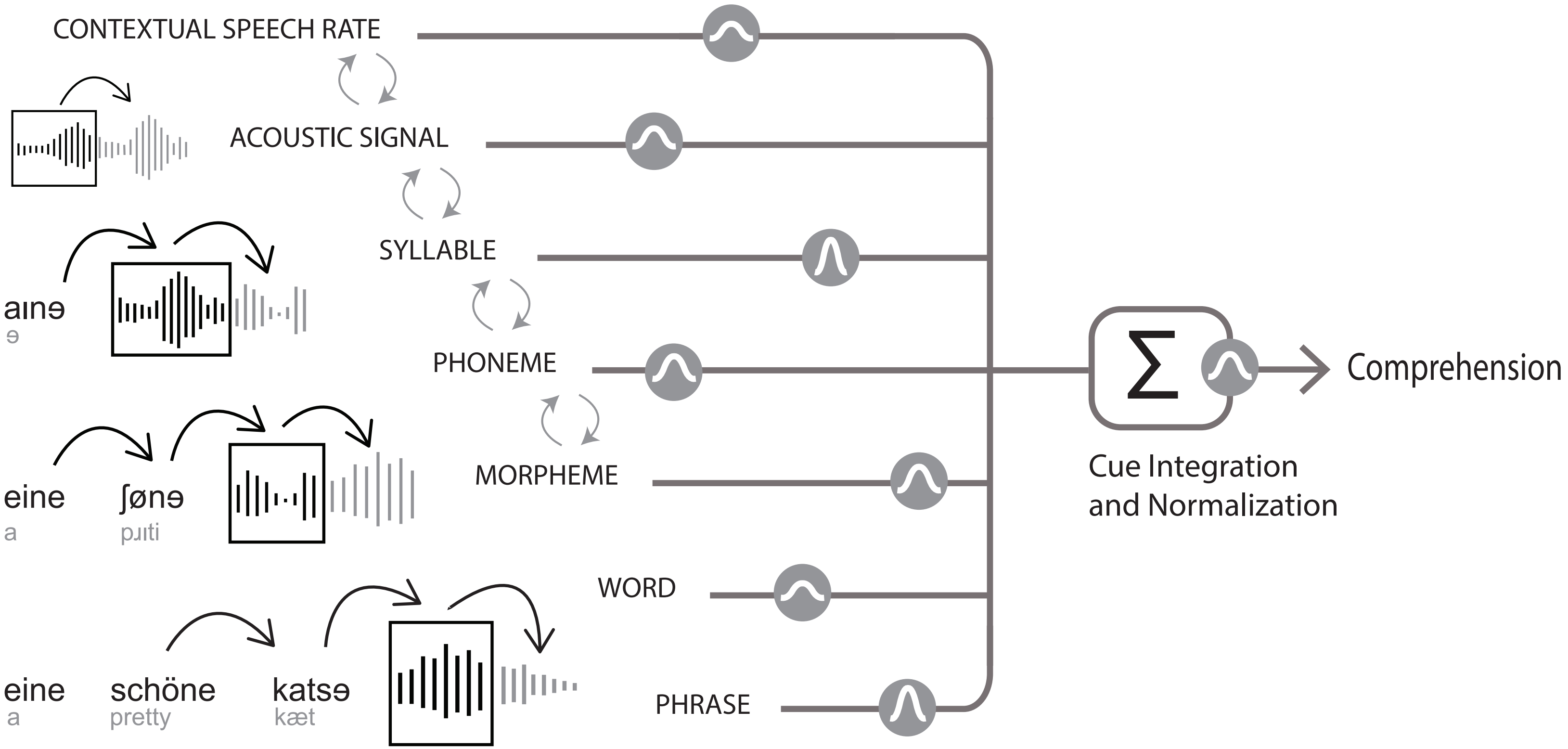

Figure 2.1: Simplified graphical representation of the cue integration architecture for speech comprehension (adapted from Martin, 2016). Cues and their corresponding reliabilities, represented by Gaussian icons, are integrated across different levels of linguistic hierarchy. Predictions about upcoming linguistic information are visualised by black arrows pointing forward. Grey arrows represent sensory resampling, such that cues from different linguistic levels of representations can influence each other.

can be cues to phrasal representations; and so on. In other words, cues are not only representations of the linguistic input, they also form the link between representations from different levels of linguistic hierarchy (Martin, 2016). Note how this differs from the notion of cues in most connectionist frameworks, such as the Competition Model (e.g., E. Bates & MacWhinney, 1987), where cues and their weights arise from inherent properties and features of a language. Within the model of cue integration suggested by Martin (2016), cues mark the transform of the sensory signals of speech and sign into structured linguistic representations. A significant part of this neural transform is performed by internally generated representations that have been generalised into linguistic knowledge after learning – potentially, but not exclusively, from language-inherent features.

More generally, within a framework of cue integration, a psycholinguistic cue can be any source of information that is relevant for language processing, including endogenously generated representations and predictions (Martin, 2016). In the following section, we will briefly discuss how cue integration as a model of language comprehension can speak to the current debate about the role of prediction and anticipatory language processing (e.g., Huettig, 2015; Nieuwland et al., 2018).

Cue integration and prediction during language processing

The role of our expectations about upcoming linguistic information in language comprehension has been investigated extensively in the last two decades (see Huettig, 2015; Nieuwland et al., 2018, for comprehensive reviews). Anticipatory language processing has been shown to occur for features on multiple levels of language processing, including semantic (Altmann & Kamide, 1999; Federmeier & Kutas, 1999; Federmeier, McLennan, Ochoa, & Kutas, 2002; Szewczyk & Schriefers, 2013), orthographic (Laszlo & Federmeier, 2009), morphosyntactic (Kamide, Scheepers, & Altmann, 2003; Van Berkum et al., 2005; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003, 2004), and specific visual features (Rommers, Meyer, Praamstra, & Huettig, 2013). Based on these findings, several psycholinguistic models have been built on the assumption that prediction is one of the fundamental mechanisms of language processing (e.g., Dell & Chang, 2013; Pickering & Garrod, 2007). These models are in line with more general models of cognition where brains are seen as “prediction machines” that are “constantly engaged” in the task of minimising the prediction error between incoming sensory information and previously established expectations (Clark, 2013). However, as Huettig and Mani (2016) and others (e.g., Huettig, 2015; Nieuwland et al., 2018; Rabagliati & Bemis, 2013) have pointed out, these “strict prediction models” fail to explain how we understand language in situations where upcoming linguistic information cannot (or need not) be predicted. In order to account for all of the available empirical findings, psycholinguistic models are needed that allow listeners to make predictions when they can (because it might be helpful for further language processing), but to avoid doing so when they can’t (because the input might be too noisy or not informative enough).

As Martin (2016) points out, this optional capacity to make predictions can be implemented within a framework of cue integration. Bottom-up activity corresponds to integrated cues and their reliabilities, which are compared against top-down predictive activations. The potential mismatch between integrated cues and predictions is fed forward as a subset of cue reliabilities, corresponding to the notion of a “prediction error”. Note that this ties in directly with the iterative nature of cue integration: The predictive activation itself acts as a cue for further processing and is therefore associated with a specific cue reliability (and thus weight) of its own, which is normalised against the reliability of all other available and relevant cues. Crucially, predictive activation does not necessarily have to occur: If the available lower-level cues to base predictions on are not reliable enough, or simply too sparse, no anticipatory language processing will be initiated.

Current study

In the current experiment, we asked how the speech comprehension system takes up and integrates cues from different levels of linguistic hierarchies, aiming to test predictions of the cue integration model as suggested by Martin (2016). More specifically, we asked three questions: First, does the system immediately use lower-level perceptual cues online in order to infer higher-level cues, even in the presence of subsequent disambiguating information? Second, are inferential gender cues immediately deployed to make predictions about upcoming linguistic information? Third, how does the system handle incoherence between inferences made based on an early perceptual cue and subsequent lexical information?

To address these questions, we conducted an eyetracking experiment using the visual world paradigm. In the following two sections, we will briefly discuss two cues which will form the basis of our experiment. Contextual speech rate is a perceptual cue that has been argued to influence the earliest stages of phoneme categorisation; gender morphology is a linguistic cue that has been shown to influence linguistic prediction and integration. These two cues occur on different levels of linguistic hierarchy, so they will allow us to investigate cross-level integration online.

Contextual speech rate: an early perceptual cue

Contextual rate manipulations have been shown to influence duration-based phoneme perception: For instance, the perception of a vowel that is ambiguous between short /ɑ/ and long /a:/ in Dutch is biased towards /a:/ when embedded in a fast context sentence, but biased towards /ɑ/ when presented after a slow context sentence (e.g., Bosker, 2017a, 2017b; Bosker & Reinisch, 2017; Bosker, Reinisch, & Sjerps, 2017; Maslowski, Meyer, & Bosker, 2018; Maslowski et al., 2019a). Similar findings have been reported for other (duration-cued) segmental distinctions, such as /b-p/ (Gordon, 1988), /b-w/ (Miller & Baer, 1983; Wade & Holt, 2005), /p-p#p/ (Pickett & Decker, 1960), and singleton-geminate (Mitterer, 2018). In fact, reduced highly coarticulated linguistic units can even be missed entirely by listeners when presented in slow contexts. For instance, a reduced "terror" can be perceived as "tear", omitting the second unstressed syllable "-or", when embedded in a slow sentence (Baese-Berk et al., 2019). Similarly, the function word "or" in a phrase such as "leisure (or) time" can be perceived as present or absent depending on contextual speech rate (Dilley & Pitt, 2010), and the determiner "a" in a sentence such as "The Petersons are looking to buy (a) brown hen(s) soon" can perceptually "appear" or "disappear" when embedded in fast or slow contexts (Brown, Dilley, & Tanenhaus, 2012).

These effects of contextual speech rate are referred to by different names, such as "rate normalisation" (adopted here), "disappearing word effect", "distal rate effect", and "lexical rate effect" – but always involve rate-dependent speech perception. Rate normalisation effects have been observed to arise very early during perception, and they appear to modulate the uptake and weighting of other acoustic cues.

Reinisch and Sjerps (2013) investigated the time course of the uptake and interplay of spectral, durational, and contextual cues for rate normalisation. Native speakers of Dutch were asked to categorise minimal word pairs such as /tɑk/ (*branch*) and /ta:k/ (*task*), where the vowel had been manipulated to be both spectrally and durationally ambiguous between /ɑ/ and /a:/, embedded in fast and slow contexts. They found that contextual rate cues were used very rapidly, influencing perception and categorisation of the target word at the same point in time as vowel-internal durational cues. These findings are in line with accounts of speech rate effects arising at early stages of lexical processing, potentially involving general auditory mechanisms (see also Bosker, 2017a; Bosker and Ghitza, 2018; Maslowski et al., 2019b; Miller and Dexter, 1988; Sawusch and Newman, 2000; Wade and Holt, 2005; but see Pitt, Szostak, and Dilley, 2016).

Toscano and McMurray (2015) investigated the interplay of contextual rate effects with voice onset time (VOT) in an eye tracking experiment. English-speaking participants were asked to categorise minimal word pairs such as *beach* and *peach*, where the VOT of the initial plosive had been manipulated to be temporally ambiguous between the voiced and voiceless tokens. Eye gaze data indicated that contextual rate cues were used simultaneously with VOT cues, again suggesting that rate effects occur early during perception, and that contextual speech rate can be seen as a cue that modulates other acoustic cues.

Gender morphology: a linguistic cue

There is plenty of evidence showing that listeners rapidly make use of morphological information during speech comprehension. Bölte and Connine (2004) showed that gender-marked determiners can facilitate subsequent language processing in German, and gender priming effects have been reported for a multitude of languages, including German (e.g., Hillert and Bates, 1994; see Friederici and Jacobsen, 1999, for a comprehensive review of the gender priming literature). Importantly, gender information has been shown to be involved in both the prediction of upcoming referents (e.g., Szewczyk and Schriefers, 2013; Van Berkum et al., 2005; Wicha et al., 2004; but see Guerra, Nicenboim, and Helo, 2018; Kochari and Flecken, 2019), and the perception of following ambiguous phonemes (Martin et al., 2017).

Current experiment

In the current experiment, we examined the influence of contextual speech rate on the perception of the presence or absence of the morphosyntactic inflectional suffix /-ə/ (schwa), marking gender on indefinite determiners (feminine *eine* vs. neuter *ein*) in German. Participants were presented with two pictures on a screen, corresponding to a neuter and a feminine target noun (e.g. *Katze*$_{FEMININE}$,“cat” vs. *Reh*$_{NEUTER}$, “deer”), while listening to auditory instructions at fast or slow rates, asking them to look at one of the two pictures (*Schauen*

Sie jetzt sofort auf eine$_{FEMININE}$/*ein*$_{NEUTER}$ *außergewöhnlich liebe*$_{FEMININE}$ *Katze*$_{FEMININE}$/*liebes*$_{NEUTER}$ *Reh*$_{NEUTER}$, "Now look at once at an$_{FEMININE/NEUTER}$ exceptionally friendly$_{FEMININE/NEUTER}$ cat$_{FEMININE}$/deer$_{NEUTER}$"). We had manipulated the indefinite determiner, *ein?*, to be ambiguous between perceived presence and absence (perceived either as *ein*, marking neuter, or as *eine*, marking feminine gender). Crucially, the indefinite determiner containing this ambiguous schwa phoneme was the earliest morphosyntactic cue indicating the gender (and, by proxy, lexical identity) of the target, thus allowing participants to make predictions about upcoming referents.

The cue integration model predicts that listeners rapidly use perceptual cues to draw inferences that are, in turn, deployed as cues for higher levels of processing. Previous findings reported by Brown et al. (2012) suggest that speech rate is, indeed, used by listeners to draw inferences about higher-level linguistic features, such as number. Brown and colleagues used a visual world paradigm to investigate listeners' perception of the singular indefinite determiner "a" in a sentence such as "The Petersons are looking to buy (a) brown hen(s) soon", where the carrier sentence surrounding the determiner region was manipulated to be either slow or fast. Overall, listeners were more likely to perceive the determiner as being "present" in fast as opposed to slow contexts, as evidenced by preferential looks towards pictures corresponding to a singular (plural) interpretation in fast (slow) contexts during the target time window. From a cue-integration perspective, this suggests that listeners used the acoustic cues from contextual speech rate and vowel duration to infer higher-level linguistic information about the number of the target noun.

In line with the findings reported by Brown et al. (2012) and the predictions from cue integration theory, we thus hypothesised that listeners would rapidly use lower-level contextual speech rate cues in order to infer higher-level morphosyntactic gender and lexical information. Specifically, when embedded in a fast context sentence, the ambiguous schwa phoneme should appear relatively long in contrast to the preceding phonemes – similar to more long /a:/ vowel responses after fast speech in Reinisch and Sjerps (2013). Participants should therefore be more likely to perceive the schwa as being present in a fast context, leading them to interpret the determiner as *eine*. This would, in turn, allow them to infer feminine gender based on the presence of the gender morpheme, and make predictions about the lexical identity of the target picture. Conversely, the ambiguous schwa phoneme should sound relatively short when embedded in a slow context sentence, possibly making the schwa perceptually disappear.

Participants should thus be more likely to perceive the indefinite determiner as being *ein* in a slow context, allowing them to infer neuter gender and make predictions about the target picture's gender and lexical identity. Crucially, if participants used contextual rate cues to infer morphosyntactic information and then operationalised this information to make predictions about the target noun, we should find anticipatory looks to the relevant picture well before the onset of the target noun. Analysing a time window immediately after the onset of the ambiguous schwa phoneme and before target onset thus allowed us to address both the temporal (question 1) and the predictive aspect (question 2) of cue integration.

The cue integration framework further predicts that cues can interact across levels of linguistic hierarchy – that is, signal-based, acoustic cues can influence the expectation and perception of knowledge-based, inferential cues, and vice versa (see also Mattys et al., 2007, and Chapter 3 of this thesis). However, to our knowledge, previous eye-tracking studies that investigated speech rate as a possible signal-based cue towards morphosyntactically relevant information have exclusively investigated it in combination with other ambiguous acoustic or morphosyntactic cues. That is, in Brown et al. (2012), the sibilant ("hen[*s* *s*]oon" vs. "hen [*s*]oon") was, itself, ambiguous. In fact, the authors specifically designed their stimuli to "increase participants' reliance on the determiner [...] as a cue to number" (Brown et al., 2012, p. 1375), and their analyses confirm that listeners based their judgements on the perception of the determiner, rather than a combination of both number cues (Brown et al., 2012, p. 1377). As such, their experiment does not readily speak to how potentially mismatching cues are *combined* across distinct levels of linguistic hierarchy online, and whether morphosyntactic inferences are computed in the presence of subsequent disambiguating information. This is different from the current experiment: Here, listeners heard an acoustic cue (the ambiguous schwa), based on which gender and, consequentially, the lexical identity of the target word could be inferred. Crucially, this inference-based lexical preselection could either match or mismatch the identity of the target noun. In contrast to Brown et al. (2012), the subsequent gender cues from adjective and target item in our experiment were always reliable and could, in principle, entirely disambiguate the ambiguous schwa (but importantly, only "in retrospect"). To summarise, our experiment investigates how contextual speech rate (which is an early perceptual, signal-based cue), gender morphology (which is an inferred knowledge-based cue), and lexical information are iteratively combined online during spoken language comprehension.

Analysing a time window after the onset of the disambiguating adjective and target noun allowed us to address our third question: How does the system handle incoherence between early perceptual and higher-level linguistic cues when integrating lexical information? By this point in time, participants had already encountered the "unreliable" schwa gender cue ("unreliable" because perception of the ambiguous schwa phoneme should be affected by our rate manipulations), as well as the relatively "reliable" gender cue carried in the adjective and the target word itself. There are three plausible scenarios for how these two cues could be integrated: First, it is possible that the earliest cue completely dominates the later cues as soon as it enters the system. If that were the case, we should observe clear rate effects, and no potential revision based on cues in the target time window. Second, it is possible that participants perceive the first cue as so unreliable that it is immediately overridden as soon as more reliable target cues become available. If that were the case, we should observe no effects of contextual speech rate during the target window. Third, it is possible that both cues are active in the target window to a certain extent. After all, taking all the available information into account would seem to be the best protection against fallibility. If that were the case, the early perceptual cue should remain active in the system for as long as it is relevant for linguistic processing, and we may observe rate effects even after the onset of the disambiguating target information. This is especially interesting given that phoneme-level contextual rate effects have been claimed to be "fragile" Baese-Berk et al. (2019). As such, our experiment offers novel insights into how the brain infers linguistic cues from the acoustic signal, and how these inferential cues might be combined with information from higher levels of linguistic hierarchy during online sentence comprehension.

2.2 Methods

Our aim was to test whether and how contextual speech rate influences morphosyntactic and lexical prediction and integration. We used eye-tracking (visual world paradigm) in order to obtain online measures of the influence of contextual rate on the perception of the presence or absence of the morphosyntactic inflectional suffix /-ə/, marking gender on indefinite determiners (feminine *eine* /aɪnə/ vs. neuter *ein* /aɪn/) in German.

Participants

Native German speakers (N = 35, 26 females, M_{age} = 22 years) with normal hearing were recruited from the Max Planck Institute (MPI) participant pool, with informed consent as approved by the Ethics Committee of the Social Sciences Department of Radboud University (Project Code: ECSW2014-1003-196). Participants were paid for their participation. We excluded five participants from the analysis due to calibration failures, leaving us with N = 30 (23 females, M_{age} = 23 years).

Materials and design

Auditory stimuli consisted of 25 German sentences (e.g. *Schauen Sie jetzt sofort auf ein(e) außergewöhnlich liebe(s) Katze*$_{FEM}$*/Reh*$_{NEU}$, "Now look at an exceptionally friendly cat/deer"; see Appendix for a complete list of all the stimuli), all sharing the same sentence frame but ending in either a feminine (e.g. *liebe Katze*) or a neuter target reference (*liebes Reh*). Feminine-neuter target pairs were selected that did not have any phonological overlap between the two target nouns (see Appendix). We recorded a female native speaker of German, who was naïve to the purpose of the experiment, reading all sentences with either target reference, but always with the determiner *eine*. Recordings were made in a sound-attenuated booth and digitally sampled at 44,100 Hz on a computer located outside the booth with Audacity software (Audacity Team, 2019).

For each sentence, the lead-in carrier sentence (*Schauen Sie jetzt sofort auf*) was compressed or expanded in order to yield a fast (66% original duration), a neutral (100% original duration), and a slow (1 / 66% = 150% original duration) syllable rate using PSOLA in Praat (Boersma & Weenink, 2020). Moreover, the duration of the suffix /-ə/ on all determiners *eine* was manipulated. Specifically, 5-step duration continua were created for each recorded *eine* by compressing the word-final schwa using PSOLA in Praat, ranging from perceived absence (40% original duration) to perceived presence (52% original duration) of the schwa phoneme, in steps of 3% (based on piloting). This resulted in a total of 750 unique stimuli (25 sentences × 2 target references × 3 rates × 5 schwa durations).[2]

[2] Note that a distinction is commonly made between *distal* and *proximal* speech rate manipulations (see Heffner, Newman, & Idsardi, 2017, for an in-depth discussion of this distinction), where *proximal* context refers to the context directly adjacent to the ambiguous region of interest, whereas *distal* context refers to linguistic material that is further away (i.e., non-adjacent from the ambiguous region of interest). In the current experiment, we are manipulating context that is not directly adjacent to the ambiguous schwa phoneme. That is, the syllable *ein-* inter-

A categorisation pretest was conducted in order to (1) verify that the duration continua systematically shifted perception from absence to presence of the schwa phoneme; and (2) verify that faster speech rates would bias listeners to explicitly report hearing *eine* (instead of *ein*). Native speakers of German who did not participate in any of the other experiments ($N = 6$, 3 females, $M_{age} = 26$) listened to excerpts (i.e. incomplete sentences) of 250 randomly selected manipulated sentences. Specifically, these excerpts included all the speech up to the disambiguating adjective (e.g. *Schauen Sie jetzt sofort auf ein(e) außergewöhnlich*), thus avoiding biasing influences from the target references on determiner categorisation. Listeners indicated via button press whether they had heard *ein* or *eine*. The categorisation curves (Figure 2.2) clearly showed that (1) higher steps on the duration continua (i.e. longer schwa) led to more *eine* responses (i.e. fewer *ein* responses); and (2) faster rates (indicated by the different coloured lines in Figure 2.2) clearly shifted perception towards more *eine* responses. Note that in the eye-tracking experiment, only stimuli from the fast and slow condition were used (no neutral rate condition). Visual stimuli consisted of pictures taken from the MultiPic database (Duñabeitia et al., 2018) presented in 300 × 300 pixel resolution.

In order to minimise the duration of the experiment, participants were randomly allocated to one of two groups: one group was presented with 13 sentences in all possible conditions (13 sentences × 2 target references × 2 rates × 5 duration steps = 260 trials total), the other group with the remaining 12 sentences in all possible conditions (240 trials total). The presentation of the stimuli was randomised in each block, such that all sentences were presented to the participant once before a repetition occurred.

Procedure

Participants were tested individually in a sound-conditioned booth. They were seated at a distance of approximately 60 cm in front of a 50,8 cm by 28,6 cm screen with a tower-mounted Eyelink 1000 eye-tracking system (SR Research) and listened to stimuli at a comfortable volume through headphones. Stimuli were delivered using Experiment Builder software (SR Research). Eye movements were recorded using right pupil-tracking at a sampling rate of 1000 Hz.

Each experimental session started with a nine-point calibration procedure followed by a validation procedure. Participants' task was to listen to the stimuli

vened between the rate-manipulated context and the ambiguous schwa phoneme; as such, our rate manipulation can be considered *distal*.

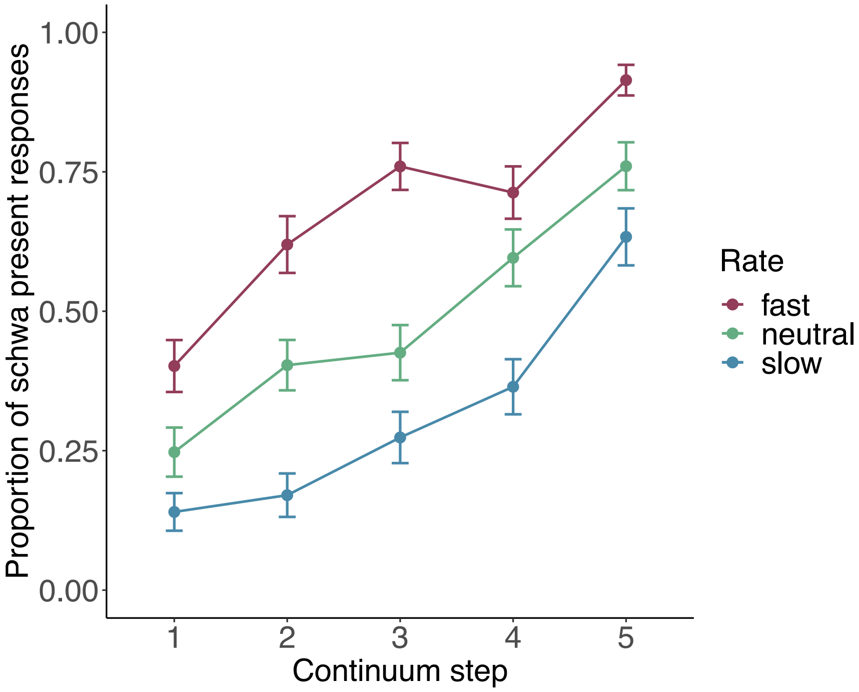

Figure 2.2: Categorisation curves from the pretest of the proportion of schwa present (i.e. eine) *responses as a function of duration continuum step, split for three different contextual speech rates (red: fast rate; green: neutral rate; blue: slow rate).* Participants in the pretest only heard short excerpts from the stimulus sentences and indicated whether they heard *ein* or *eine*. Longer schwa durations (e.g. step 5) led to more *eine* responses (i.e. fewer *ein* responses) and faster speech rate biased listeners to report more *eine* responses. Error bars represent the standard error of the mean.

and click with the computer mouse on one of two pictures corresponding to the two possible sentence-final target references. Note that participants were thus not making any explicit judgment about whether or not they perceived a schwa. In fact, they were ignorant about the intent of the schwa duration and speech rate manipulations. The visual stimuli were presented centred in the left and right halves of the screen. The side of the neuter and female option on the screen was counterbalanced.

On each trial, participants first had to click with the computer mouse on a blue rectangle in the middle of the screen to centre their eye gaze and mouse position. This screen was immediately followed by two pictures. After one second of preview, the auditory stimulus was presented. Participants could only respond by clicking on one of the presented pictures after sound offset. The pictures stayed on the screen until the participant responded by clicking on one of the presented pictures. After an inter-trial interval of one second following the mouse click, the next trial started automatically. Participants first completed a practice session with four trials to become familiarised with the task. Every 80 trials, participants were allowed to take a self-paced break. The experiment took about 35 minutes to complete.

2.3 Results

Prior to the analyses, blinks and saccades were excluded from the data. We divided the screen into two sections (left and right) and coded fixations on either half as a look toward that particular picture. The eye fixation data were down-sampled to 100 Hz. Participants were very accurate at performing the task: less than 0.2% of the mouse responses were incorrect ($n = 10$). Since the number of incorrect responses was so low, and because we were primarily interested in eye movements prior to and shortly after target onset rather than mouse clicks, we did not exclude any trials from the analyses. Mixed effects logistic regression models (GLMMs: Quené & van den Bergh, 2008) with a logistic link function (Jaeger, 2008) as implemented in the MixedModels package version 2.1.2+ (D. Bates, Mächler, Bolker, & Walker, 2015) in Julia version 1.2.0 (Bezanson, Edelman, Karpinski, & Shah, 2017) evaluated participants' eye fixations. The eye fixation data were evaluated in two time windows: one pre-target time window following the offset of the ambiguous schwa token, and one post-target time window following the onset of the earliest disambiguating target cue. Note that, in cases of a feminine target, the earliest reliably disambiguating cue was the onset of the target noun itself, whereas for a neuter target, the earliest cue was the onset of the morpheme *–s* on the adjective, marking neuter gender.

Pre-target window

The analysis of the data in the pre-target time window tested whether participants showed an anticipatory target preference – well before the target reference – triggered by the schwa duration in the determiner and the contextual speech rate. The time window of interest was defined as starting from 200 ms after the offset of the ambiguous schwa phoneme, because the offset is the earliest time point at which participants have access to the duration cues on the schwa (note that 200 ms corresponds approximately to the time it takes to launch a saccade; Matin, Shao, and Boff, 1993) and lasting until the onset of the earliest disambiguating cue. For feminine target references, this is the onset of the target word itself; for neuter targets, it is the onset of the morpheme *-s* on the adjective preceding the target word. Figure 2.3 shows fixation proportions to the feminine picture depending on the context rate (slow vs. fast rate), with the time window of interest shaded grey.

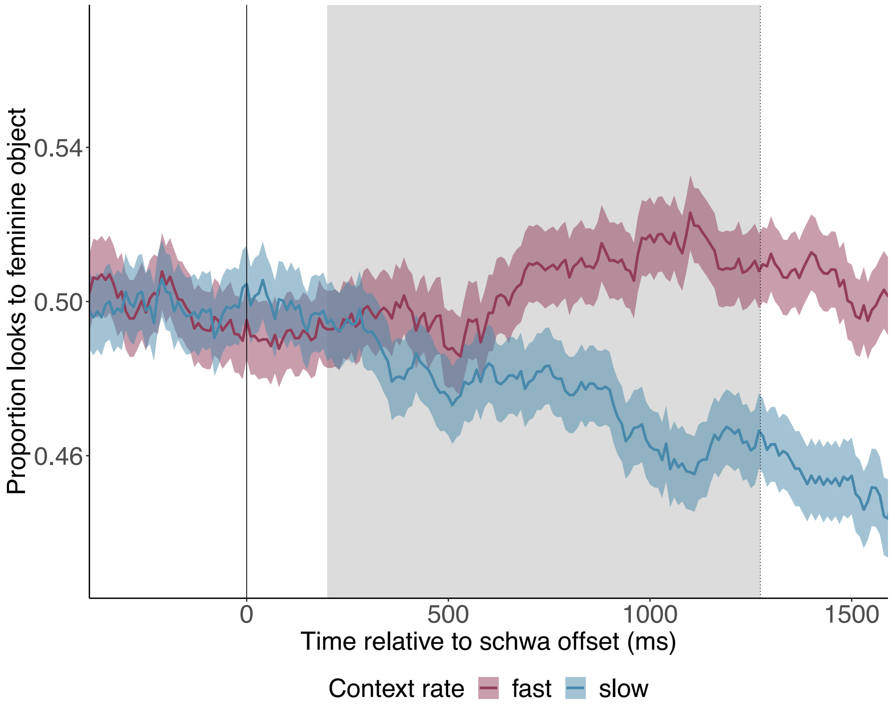

Figure 2.3: Proportion of looks to feminine object across time in fast (red) and slow contexts (blue). Time point 0 marks the offset of the ambiguous schwa phoneme, indicated by the solid vertical line. The dotted vertical line indicates the mean onset of the disambiguating sound: for feminine target references, this is the onset of the target word itself; for neuter targets, it is the onset of the morpheme *-s* on the adjective preceding the target word. Shown in grey is the area of interest, spanning from 200ms after schwa offset until the mean onset of the disambiguating cue. Overall, the proportion of looks to the feminine object was higher in fast as opposed to slow contexts. Shading around the coloured lines represents the standard error of the mean.

We predicted that a fast speech rate would bias the perception of the ambiguous determiner *ein[?]* towards *eine* (and away from *ein*) and would trigger more looks to the feminine picture well before the target referent had been heard. Conversely, the slow speech rate would bias perception towards *ein* and, as a consequence, would induce more looks to the neuter picture. Since no phonetic information about the target was available to the listener in the pre-target time window, we analysed participants' looks to just one of the two objects (the feminine object, instead of looks to the target), coded binomially.

A generalised linear mixed model with a logistic linking function tested the binomial looks to the feminine picture (1 = yes, 0 = no) for fixed effects of Rate (categorical predictor with two levels: fast coded as 0.5, slow as -0.5), Time (continuous predictor; z-scored around the mean within the analysis window), Step (continuous predictor; centred: schwa duration continuum Step 1 coded as -2, Step 3 as 0, Step 5 as 2), and their interactions. Additionally, the model included a fixed effect of Lag, capturing the binomial looks to the feminine picture at the previous sample (1 = yes, 0 = no). The Lag predictor addresses the autocorrelated nature of eye gaze data (cf. Cho, Brown-Schmidt, & Lee, 2018). The random effects structure contained random intercepts for Participants and Items and by-participant and by-item slopes for all fixed factors including Lag (but not their interactions).

The model revealed a significant effect of Rate ($\beta = 0.114$, SE = 0.046, $z = 2.481$, $p = 0.013$), demonstrating that upon hearing an ambiguous phoneme, participants were more likely to look at the feminine object during trials that included a fast context rate. Crucially, this happened before the onset of any further disambiguating cues. We also found a significant interaction between Time and Step ($\beta = 0.021$, SE = 0.007, $z = 3.128$, $p = 0.002$), indicating that higher continuum steps led to an increasingly higher proportion of looks to the feminine object as time passed. Finally – and unsurprisingly –, the model revealed a significant main effect of Lag, indicating that looks to the feminine object were, indeed, dependent on the gaze at the previous sample ($\beta = 8.215$, SE = 0.090, $z = 91.678$, $p < 0.001$). Overall, these results support our hypotheses: Participants were more likely to look at the picture corresponding to the feminine object in the fast rate, thus indicating that they were more likely to have perceived a schwa phoneme in the fast as opposed to the slow context, and that they used that percept as a morphological gender cue towards the target picture.

Post-target window

The analysis of the data in the post-target time window tested whether the effects of contextual rate and schwa duration manipulations persisted even after the perception of disambiguating phonological cues (i.e. after target onset). The time window of interest was defined as starting from 200 ms after the onset of the earliest disambiguating cue (target word onset for feminine, *-s* morpheme onset for neuter targets) and lasting until 200 ms after the offset of the target word's initial syllable. As noted above, there was no phonological overlap between target and competitor images, so the earliest target-specific acoustic cues can, in principle, entirely disambiguate between the two. Evidence of this can be seen in Figure 2.4, where we observe preferential looks towards the target picture well before the offset of the first syllable of the target.

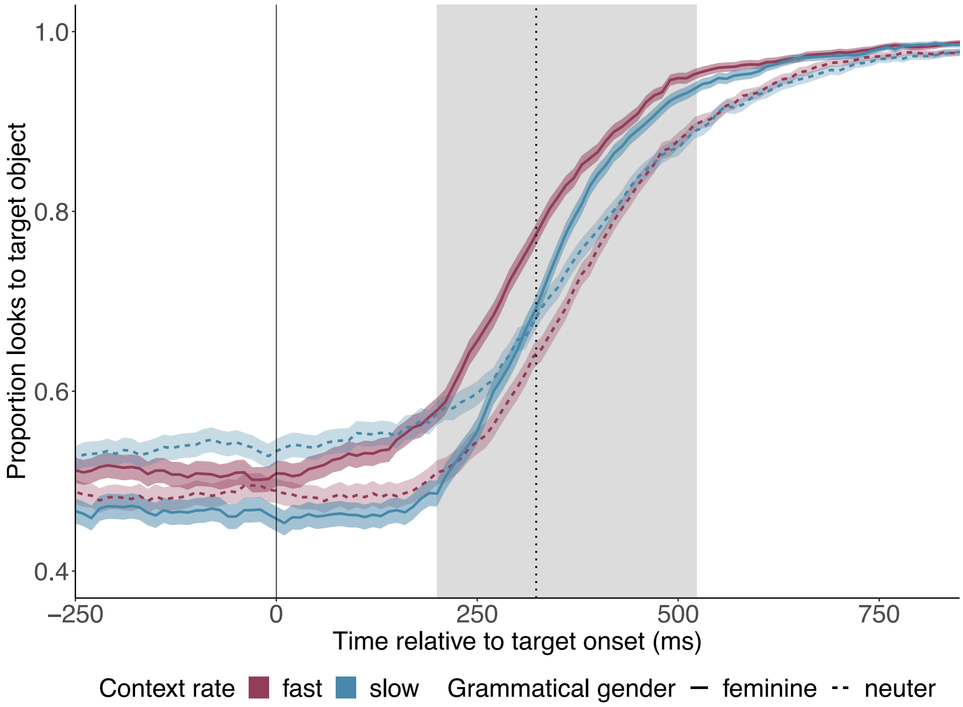

Figure 2.4: Proportion of looks to target object across time for feminine targets (solid) and neuter targets (dashed) in fast (red) and slow contexts (blue). The feminine-fast (solid red line) and neuter-slow (dashed blue line) conditions represent the Congruent conditions; the feminine-slow and neuter-fast conditions represent the Incongruent conditions. Time point 0 marks the onset of the earliest disambiguating cue (onset of the target word for feminine targets, morpheme -*s* on the preceding adjective for neuter targets), indicated by the vertical solid line. The vertical dotted line indicates the mean offset of the first target word syllable. Shown in grey is the area of interest, spanning from 200 ms after onset of the disambiguating cue until 200 ms after the mean offset of the initial target word syllable. Shading around the coloured lines represents the standard error of the mean.

We had crossed the factors rate and target gender. According to our predictions (and as shown in the pretest), an ambiguous /-ə/ token presented in a *fast* context is more likely to be perceived as *present*. In terms of our experimental manipulation, the perceived presence of a schwa phoneme corresponds to the perception of the determiner *eine*, marking feminine gender. Fast context rates should therefore bias participants' looking preference towards the picture corresponding to the feminine object. We therefore refer to trials with a feminine target presented in a fast context sentence as *rate-gender congruent trials*. Similarly, an ambiguous /-ə/ token presented in a *slow* context is more likely to be perceived as *absent*, thus corresponding to the perception of the neuter determiner *ein* and eliciting more looks towards the picture corresponding to the neuter object. Therefore, trials with a neuter target presented in a slow context sentence are also referred to as *rate-gender congruent trials*. Conversely, *feminine+slow* and *neuter+fast* trials are referred to as *rate-gender incongruent*. The use of this congruency coding allowed us to specifically test for potentially facilitating effects of *congruent* contextual speech rate on target looks, independent of the speech rate in a given trial. As can be seen in Figure 2.4, participants seemed to be faster to look at the correct target picture in congruent as opposed to incongruent trials.

A GLMM with a logistic linking function tested the binomial looks to the target picture (1 = yes, 0 = no) for fixed effects of Congruency (categorical predictor with two levels: congruent coded as 0.5; incongruent as -0.5), Step (continuous predictor; centred: schwa duration continuum Step 1 coded as -2, Step 3 as 0, Step 5 as 2), and Time (continuous predictor; z-scored around the mean within the analysis window), and all their interactions. Again, we also included a Lag predictor (categorical predictor coding looks to the target picture at the previous sample: 1 = yes, 0 = no) in order to alleviate the autocorrelation problem (Cho et al., 2018). The random effects structure contained random intercepts for Participants and Items and by-participant and by-item random slopes for all fixed factors including Lag (but not their interactions).

The model revealed a significant effect of Time (β = 1.635, SE = 0.120, z = 13.611, p < 0.001), indicating, unsurprisingly, that participants increasingly looked at the target picture as time passed. Crucially, a significant effect of Congruency was found (β = 0.124, SE = 0.060, z = 2.081, p = 0.038), indicating that participants showed more looks to the target referent if the preceding morphological cue, inferred from the perceived presence or absence of the schwa phoneme based on contextual speech rate, was "congruent" with the target gender (e.g. fast with feminine targets; slow with neuter targets). No effect of Step could be established (β = -0.024, SE = 0.020, z = -1.193, p = 0.233). This is not surprising, considering that low Steps would have biased participants towards perceiving a schwa as not being present (thus leading to a neuter interpretation), and high Steps would have biased participants toward perceiving a schwa as being present (thus leading to a feminine interpretation); since half of the targets were neuter and the other half were feminine, any biasing effect of Step simply averages out between the two target genders.

Moreover, several interactions were observed. An interaction between Congruency and Time (β = -0.245, SE = 0.029, z = -8.330, $p < 0.001$) indicated that the beneficial effect of a congruent speech rate diminished with time. However, a positive three-way interaction (β = 0.109, SE = 0.020, z = 5.324, $p <$ 0.001) between Congruency, Step and Time indicated that this only held for the lower continuum steps. The model also found an interaction between Congruency and Step (β = 0.103, SE = 0.019, z = 5.324, $p < 0.001$), indicating that the effect of Congruency was smaller for lower continuum steps (i.e. shorter schwa durations). This may be interpreted in light of the pretest: The rate effect was smaller at lower continuum steps (cf. Figure 2.2), and as such the effect of congruency would also be expected to be smaller. Finally, we found an interaction between Time and Step (β = 0.046, SE = 0.010, z = 4.358, $p < 0.001$); although we currently lack an interpretation for this interaction, note that the estimate is very small. Finally – and again as expected – the model revealed a significant main effect of Lag, indicating that looks to the target object were, indeed, dependent on the gaze at the previous sample (β = 6.971, SE = 0.070, z = 99.391, $p < 0.001$).

2.4 Discussion

The goal of the current study was to investigate three main questions. First, we asked whether we could observe early perceptual cues being rapidly used online in order to infer higher-level linguistic cues, even in the presence of subsequent disambiguating information. Second, we asked whether these inferential cues that were based on perceptual cues are deployed to make predictions about upcoming linguistic information. Third, we asked how the language comprehension system handles incoherence between early perceptual and higher-level linguistic cues when integrating lexical information. We addressed these questions by experimentally inducing contextual rate normalisation effects on the phoneme /-ə/, which can act as a morphosyntactic gender cue on indefinite determiners in German. In the following, we will discuss our results in light of these three questions.

Contextual speech rate is rapidly used as a cue for speech processing

We found evidence for contextual speech rate acting as an early and robust cue for speech comprehension. Listeners' perception of the morpheme /-ə/ in German was significantly influenced by the rate of the preceding context. We observed these rate normalisation effects immediately after the presentation of the ambiguous schwa token (200 ms after schwa offset), and well before any acoustic information about the target referent itself was available to the listeners. These results support previous accounts of rate normalisation effects arising during early stages of lexical processing and influencing phoneme perception almost immediately (Bosker, 2017a; Maslowski et al., 2019b; Newman & Sawusch, 2009; Reinisch & Sjerps, 2013; Toscano & McMurray, 2015).

Our findings are novel in two ways. First, to our knowledge, previous eye-tracking studies on contextual rate normalisation have mostly investigated minimal word pairs (e.g. *tak* vs. *taak* (Reinisch & Sjerps, 2013); *tear* vs. *terror* (Baese-Berk et al., 2019); *eens speer* vs. *een speer* (Reinisch, Jesse, & McQueen, 2011)), where the interpretation of the ambiguous phoneme had implications on a lexical level, but did not affect further linguistic processing on the sentence level (although see Brown et al., 2012). In contrast, the rate manipulation in the current experiment affected the perception of a purely morphosyntactic minimal pair (*ein* vs. *eine*). Here, we show for the first time how contextual speech rate – an acoustic, signal-based cue – interacts online with subsequent gender information from a lexical, knowledge-based cue, which occurs on a higher level of the linguistic hierarchy and was potentially conflicting with the earlier cue. As such, this is the first eye-tracking experiment to our knowledge where the rate manipulation carried implications for further inference-based morphosyntactic prediction and integration of subsequent lexical material.

Second, previous research has mostly used experimental tasks that involved explicit identification or categorisation of the ambiguous word. In contrast to that, our design allowed us to tap perception of the ambiguous determiner *ein?*, crucially without explicitly asking participants for a categorisation decision between *ein* and *eine*. This contrasts with earlier eye-tracking studies of rate normalisation (e.g., Reinisch & Sjerps, 2013; Toscano & McMurray, 2012, 2015), where participants did make explicit categorisation decisions about the ambiguous target sounds under study. Notably, this is also different from the experiment reported by Brown et al. (2012), where participants decided between singular or plural targets and thus made explicit judgments about the informational content of the phoneme affected by the rate manipulation. As such, our results suggest that rate normalisation operates automatically, even when attention is not drawn to the ambiguous target sounds tested. This corroborates recent findings from Maslowski et al. (2019b), who showed evidence that listeners normalise for speech rate even without an explicit recognition task (using repetition priming). In light of these two aspects, our findings demonstrate that (1) rate normalisation affects a large set of duration-cued distinctions, including morphosyntactic minimal pairs, and (2) rate normalisation impacts incremental spoken language processing, even when the task does not require participants to make explicit judgments. As such, rate normalisation observed in lab-based psycholinguistic experiments appears to be a perceptual process that likely also contributes to the comprehension of natural and spontaneous conversation.

Inferences that were made based on perceptual cues can be used as higher-level cues to make predictions about upcoming referents

As stated above, our experiment went beyond mere phonemic or lexical identification: The indefinite determiner containing the ambiguous schwa token was the first cue towards the gender of the target picture, so it was a crucial building block for subsequent steps of language processing. Our eye gaze analysis in a time window after the ambiguous schwa token showed that participants not only immediately made use of contextual information upon perceiving the ambiguous token, but also rapidly used that information to draw inferences about the gender of the target referent. This was reflected in participants looking more towards the picture that corresponded to the gender that the rate manipulation biased them towards.

Our experiment contributes to the current debate around prediction during language comprehension (cf. Huettig, 2015; Nieuwland et al., 2018). Several studies have found effects of anticipatory language processing with regard to gender information (e.g., Szewczyk & Schriefers, 2013; Van Berkum et al., 2005; Wicha, Bates, et al., 2003; Wicha, Moreno, & Kutas, 2003; Wicha et al., 2004), while others have failed to replicate these findings (Guerra et al., 2018; Kochari & Flecken, 2019). Why is it that we find evidence for anticipatory language processing in our current experiment, while others did not? One reason might be that we provided participants with the same fixed sentence frame on every trial, making the *ein/eine* distinction relatively informative – possibly more so than it would be in more naturalistic settings. Moreover, language comprehension occurred within a very small referential "world" in our experiment: Participants were presented with two pictures at a time, thus limiting their choices for possible predictions considerably. Presumably, these two factors facilitated the predictive processing observed. Nevertheless, the fact that rate normalisation induces the kind of predictive behaviour that we observe with our paradigm is strong evidence for the utility of contextual rate cues in speech processing.

As stated earlier, prediction is a possibility, but not a necessity, for language comprehension within a cue integration framework. We therefore do not take our findings as evidence in favour of, or against anticipatory language processing, per se; rather, we believe that our results can be seen as step towards a more comprehensive account of language processing where predictions *can* be part of the processing architecture.

Early perceptual cues remain active in the speech and language comprehension system during subsequent processing

Even after hearing the disambiguating beginning of the target referent, participants were significantly slower to look at the target object in *rate-gender incongruent* trials (i.e. in trials where the actual target gender did not match the gender corresponding to the schwa perception induced by the preceding context rate manipulation). We believe that this finding – a robust effect of a low-level perceptual cue, even in the presence of the reliably unambiguous first syllable of the target word – indicates that the early perceptual cue does, indeed, remain active in the system, until it can (or cannot) be integrated with additional incoming information.

These observations are in line with previous behavioural studies (Heffner, Newman, Dilley, & Idsardi, 2015; Morrill, Baese-Berk, Heffner, & Dilley, 2015), where rate effects also persisted even in the presence of constraining higher-level linguistic information. Crucially, using the visual-world paradigm allowed us to measure responses to the rate manipulation without asking for explicit categorisation of *ein* vs. *eine*, so in contrast to previous studies, no task-driven attention was drawn to the ambiguous sounds. Taken together, these findings suggest that phoneme-level rate effects are not "fragile", as has previously been suggested, but rather that they are robust and persist even in the presence of higher-level linguistic (in our case lexical) information. Interestingly, results reported by Morrill et al. (2015), as well as the results in Chapter 3 of this thesis, suggest that listeners are flexible in the way that they weigh specific cues, depending on the context and listening situation. Models of cue integration can accommodate these results, given that cue weights can be updated dynamically depending on the cue's reliability within a given situation.

Our observations also speak to recent findings by Gwilliams et al. (2018). They reported online MEG evidence showing that sensitivity to phoneme ambiguity occurs at the earliest sensory stages of speech processing, and that this sensitivity to ambiguity, along with other fine-grained acoustic features such as VOT, appeared to be maintained throughout later processing stages, even as further lexical information entered the system. The authors suggest that this reflects a reassessment of the ambiguous speech sound as additional input is being perceived. We believe that these findings can also be explained within a cue integration architecture: The early perceptual cue remains active for as long as it is relevant for linguistic processing, and its validity and reliability are "reassessed" incrementally as part of sensory resampling, as it is integrated with cues from higher levels of linguistic processing.

Our experiment is not the first to examine contextual speech rate as an early perceptual cue within a cue integration framework. Toscano and McMurray (2012, 2015) have argued that contextual speech rate can modulate the uptake of other phonological cues, such as VOT. They elegantly explain this within the C-CuRE framework: expected values are established based on contextual speech rate, and new cues are computed relative to those expectations. In fact, Toscano and McMurray (2015) suggest that adjusting these expectations can be explained within C-CuRE "as a form of predictive coding", and they point out that cue integration models of speech perception have to be linked to lexical processes. Their observations thus fit seamlessly into a more general framework of cue integration for language processing as suggested by Martin (2016), where the system makes use of all relevant pieces of information across different levels of linguistic hierarchies in order to reduce fallibility.

Based on our findings, new questions for future research can be formulated. For example, an iterative model of cue integration would suggest that lower-level perceptual ambiguity would carry through to even higher levels of linguistic processing that go beyond morphosyntax. Future experiments could therefore investigate whether rate normalisation effects induced by contextual speech rate also affect semantic prediction and integration. If so, do early perceptual cues even remain active within a larger discourse? It seems plausible that there would be at least some temporal limit regarding how long early ambiguous cues remain active in the system. If so, it would be desirable to test where that cut-off point might be, or whether it can be dynamically adjusted depending on the reliability of a specific cue in a given situation.

In the current experiment, we investigated two cues, specifically: contextual speech rate and grammatical gender. As Martin (2016) and others have pointed out, one of the hardest definitions to provide within a cue integration framework is what can constitute a cue. Future experiments are thus needed in order to determine an inventory of psycholinguistic cues and examine which other (lower- and higher-level, knowledge- and signal-based) pieces of information the brain draws on to arrive at robust linguistic units and structures.

Finally, with regard to our third question, it might be interesting to investigate in more detail *why* it took participants longer to look at the target picture in *rate-gender incongruent* trials, that is, which sub-mechanisms of cue integration and/or oculomotor control might have caused this delay. One possible explanation would be integration difficulty of the second cue in the presence of the earlier, incongruent cue. This integration difficulty could arise from participants generally taking longer to integrate the mismatching cue, but it is also possible that participants attempted the integration process multiple times and therefore took longer to converge on the target. Another possible explanation might be a "spill-over" effect, where participants were slower to look at the target in incongruent trials because of the additional time it took them to first shift their gaze, either by cancelling a previously planned saccade, or by initiating an entirely new saccade (see Altmann, 2011, for a general discussion of language-mediated eye movements). Though this was not the focus of our current experiment, investigating the subroutines at play during the integration of incongruent cues in more detail may be an interesting objective for further research.

Taken together, our results show that contextual rate effects rapidly influence not only lexical processing, but also subsequent morphosyntactic prediction and integration. Linguistic models of cue integration offer a promising step towards a mechanistic explanation for how the brain accomplishes the task of inferring complex meaning from a noisy acoustic signal by operationalising both lower-level, perceptual and higher-level, linguistic cues.

Appendix

Stimuli. Spoken sentences were manipulated to include tokens of the indefinite determiner *ein[?]* that were ambiguous between *ein* and *eine* (schwa manipulated between 40-52% original duration). Furthermore, we introduced rate manipulations (slow vs. fast) in the preceding context (underlined). Sentences 1-13 were used in group A of the experiment, sentences 14-25 in group B. The first target noun is feminine, the second neuter.

(1) Schauen Sie jetzt sofort auf ein? außerordentlich zahme(s) Ziege/Pferd.
Now look at an exceptionally tame goat/horse.

(2) Schauen Sie jetzt sofort auf ein? außergewöhnlich liebe(s) Katze/Reh.
Now look at an exceptionally darling cat/deer.

(3) Schauen Sie jetzt sofort auf ein? außerordentlich nette(s) Frau/Kind.
Now look at an exceptionally friendly woman/child.

(4) Schauen Sie jetzt sofort auf ein? außergewönlich schlichte(s) Kirche/Dach.
Now look at an exceptionally plain church/roof.

(5) Schauen Sie jetzt sofort auf ein? außerordentlich schicke(s) Krone/Geschenk.
Now look at an exceptionally pretty crown/present.

(6) Schauen Sie jetzt sofort auf ein? außergewöhnlich dicke(s) Spinne/Walross.
Now look at an exceptionally fat spider/walrus.

(7) Schauen Sie jetzt sofort auf ein? außerordentlich schwere(s) Robbe/Nilpferd.
Now look at an exceptionally heavy seal/hippo.

(8) Schauen Sie jetzt sofort auf ein? außergewöhnlich lange(s) Angel/Flugzeug.
Now look at an exceptionally long fishing rod/airplane.

(9) Schauen Sie jetzt sofort auf ein? außerordentlich neue(s) Bluse/Fahrrad.
Now look at an exceptionally new blouse/bicycle.

(10) Schauen Sie jetzt sofort auf ein? außerordentlich süße(s) Orange/Eis.
Now look at an exceptionally sweet orange/ice cream cone.

(11) Schauen Sie jetzt sofort auf ein? außergewöhnlich teure(s) Perle/Schloss.
Now look at an exceptionally expensive pearl/castle.

(12) Schauen Sie jetzt sofort auf ein? außergewöhnlich hübsche(s) Fee/Kleid.
Now look at an exceptionally pretty fairy/dress.

(13) Schauen Sie jetzt sofort auf ein? außerordentlich weiche(s) Matratze/Sofa.
Now look at an exceptionally soft mattress/couch.

(14) Schauen Sie jetzt sofort auf ein? außerordentlich scharfe(s) Säge/Messer.
Now look at an exceptionally sharp saw/knife.

(15) Schauen Sie jetzt sofort auf ein? außergewöhnlich schlaue(s) Maus/Baby.
Now look at an exceptionally smart mouse/baby.

(16) Schauen Sie jetzt sofort auf ein? außergewöhnlich schöne(s) Stadt/Mädchen.
Now look at an exceptionally beautiful town/girl.

(17) Schauen Sie jetzt sofort auf ein? außerordentlich braune(s) Eule/Kamel.
Now look at an exceptionally brown owl/camel.

(18) Schauen Sie jetzt sofort auf ein? außergewöhnlich große(s) Burg/Schiff.
Now look at an exceptionally big fortress/ship.

(19) Schauen Sie jetzt sofort auf ein? außerordentlich gute(s) Wurst/Bier.
Now look at an exceptionally good sausage/beer.

(20) Schauen Sie jetzt sofort auf ein? außerordentlich böse(s) Wespe/Krokodil.
Now look at an exceptionally mean wasp/crocodile.

(21) Schauen Sie jetzt sofort auf ein? außergewöhnlich schnelle(s) Bahn/Auto.
Now look at an exceptionally fast train/car.

(22) Schauen Sie jetzt sofort auf ein? außerordentlich schlanke(s) Nase/Knie.
Now look at an exceptionally skinny nose/knee.

(23) Schauen Sie jetzt sofort auf ein? außergewöhnlich wilde(s) Giraffe/Zebra.
Now look at an exceptionally wild giraffe/zebra.

(24) Schauen Sie jetzt sofort auf ein? außergewöhnlich alte(s) Zeitung/Buch.
Now look at an exceptionally old newspaper/book.

(25) Schauen Sie jetzt sofort auf ein? außerordentlich frische(s) Tomate/Brot.
Now look at an exceptionally fresh tomato/bread.

3 Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension[1]

Abstract

During spoken language comprehension, listeners make use of both knowledge-based and signal-based sources of information, but little is known about how cues from these distinct levels of representational hierarchy are weighted and integrated online. In an eye-tracking experiment using the visual world paradigm, we investigated the flexible weighting and integration of morphosyntactic gender marking (a knowledge-based cue) and contextual speech rate (a signal-based cue). We observed that participants used the morphosyntactic cue immediately to make predictions about upcoming referents, even in the presence of uncertainty about the cue's reliability. Moreover, we found speech rate normalization effects in participants' gaze patterns even in the presence of preceding morphosyntactic information. These results demonstrate that cues are weighted and integrated flexibly online, rather than adhering to a strict hierarchy. We further found rate normalization effects in the looking behavior of participants who showed a strong behavioral preference for the morphosyntactic gender cue. This indicates that rate normalization effects are robust and potentially automatic. We discuss these results in light of theories of cue integration and the two-stage model of acoustic context effects.

[1] Adapted from Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (2020). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(3), 549–562.

3.1 Introduction

When comprehending spoken language, listeners make use of multiple cues from different information sources and across several hierarchical levels of linguistic representations. A distinction is commonly made between cues from at least two sources: acoustic, or “signal-based” cues, and linguistic, or “knowledge-based” cues. Signal-based cues include the spectral and temporal properties of the acoustic speech signal, such as voice onset time (VOT; e.g., Lisker & Abramson, 1967; Toscano & McMurray, 2015) and contextual speech rate (Bosker, 2017a; Maslowski et al., 2019a; Reinisch & Sjerps, 2013). Knowledge-based cues, on the other hand, include knowledge about phonotactic and syntactic constraints (e.g., Huettig & Janse, 2016; McQueen, 1998; Tuinman et al., 2014), as well as semantic context (Altmann & Kamide, 1999; Wicha et al., 2004). Consequently, many models of spoken word and language comprehension incorporate at least some degree of interaction between information from both knowledge-based and signal-based information sources (Marslen-Wilson, 1987; McClelland & Elman, 1986), but few of them make predictions about how the brain computationally integrates this available information from different levels of linguistic hierarchy (although see, e.g., Norris & McQueen, 2008, for a Bayesian implementation of lexical recognition).

The goal of the current study is to contribute to our understanding of language comprehension by investigating how signal-based and knowledge-based cues are integrated and weighted against each other during online speech comprehension. Using eyetracking within the visual world paradigm, we investigate two questions: (a) Are knowledge-based, morphosyntactic cues toward grammatical gender immediately used to generate predictions about upcoming referents, even in the presence of uncertainty? (b) Are signal-based, contextual speech rate cues used even in the presence of preceding morphosyntactic information? We also investigate, for the first time, variations in the strategies that participants employ when integrating cues with each other by mapping participants’ behavioral responses to their eye-tracking data. We discuss the implications of our findings within the framework of cue integration (Martin, 2016) and the two-stage model of acoustic context effects (Bosker et al., 2017).

Language processing as hierarchical cue integration

Drawing on principles from perception, speech processing, and neurophysiology, Martin (2016) suggested a framework of cue integration for language processing, offering a general mechanism of how the brain utilizes cues across multiple levels of hierarchy to comprehend and produce language (see, e.g., Ernst & Bülthoff, 2004; Fetsch et al., 2013, for detailed descriptions of cue integration for visual and multisensory perception). Within cue integration frameworks, relevant cues are *combined* by means of summation and *integrated* by normalization against all other available cues. Each cue has an associated *weight*, which is a formalization of how reliable the cue is in a given situation and in combination with all other cues. Cue weights can be dynamically updated, which gives the system the flexibility to generate robust percepts even in the presence of uncertainty, noise, and variability.

Models related to cue integration have previously been suggested for phoneme categorization (e.g., McMurray & Jongman, 2011) and lexical recognition (e.g., Norris & McQueen, 2008). Martin (2016) suggested a cascading cue integration architecture across all levels of language processing, where functional equivalents of formal linguistic representations can emerge from sensory cues, and can in turn act as cues for higher-level representations. For speech comprehension, this means extracting and integrating relevant cues from signal-based and knowledge-based sources in order to infer higher-level linguistic information and meaning (Martin, 2016).

Establishing a hierarchical inventory of cues for spoken language comprehension remains a challenging objective for psycholinguistic research. Based on a series of experiments in which the amount and reliability of information from cues at different levels of representation was systematically manipulated, Mattys et al. (2005) proposed a hierarchically organized model of lexical segmentation. According to the original version of their model, cues are organized into three hierarchical tiers consisting of lexical (Tier I), segmental (Tier II), and metrical prosodic (Tier III) cues. Crucially, cues from Tier I, which can include contextual, syntactic, semantic, and morphological information, form the highest level of the hierarchy and can override cues from the lower two levels of representation (Mattys et al., 2005). However, in a subsequent set of experiments, Mattys et al. (2007) found that effects of syntactic knowledge on lexical segmentation could be attenuated and modulated by conflicting acoustic cues. Using a word monitoring task, they assessed how participants processed the combination of a morphosyntactic cue (singular vs. plural lexical information; e.g., *those women* vs. *that woman*) with a subsequent acoustic cue (pivotal /s/, e.g., *take#spins* vs. *takes#pins*). In a neutral listening situation without preceding syntactic information, listeners made use of acoustic cues for segmentation, as evidenced by faster target detection times for *pins* in *takes#pins*, and *spins* in *take#spins*. When preceded by a plural noun phrase, the syntactic cue took precedence over the acoustic cue (i.e., faster target detection for *spins* in *"those women take#spins"* and *"those women takes#pins"*). This result is in line with a hierarchical model of speech processing, where syntactic cues can "override" acoustic cues. For singular noun phrases, however, no effect of superiority for the syntactic cue was found, showing the same pattern of results as for the neutral condition (i.e., faster target detection for *pins* in *"that woman takes#pins"*, and *spins* in *"that woman take#spins"*). Mattys et al. (2007) therefore proposed a graded, dynamic relationship between knowledge-based and signal-based cues. The concept of a dynamic link between cues from different levels of hierarchy, although not mathematically formalized in the model by Mattys et al. (2007), bears striking similarities to cue weighting and normalization as suggested by linguistic models of cue integration (Martin, 2016).

Integrating and weighting knowledge- and signal-based cues

A growing body of research has investigated the interplay between signal-based and knowledge-based cues. Most relevant for our purposes are studies investigating contextual speech rate cues. The speech rate in a lead-in sentence can change the perception of a following target word: For instance, a vowel ambiguous between short /ɑ/ and long /a:/ in Dutch is perceived as /a:/ in the context of a fast speech rate because it sounds relatively long compared with the short vowels in the fast context, but as /ɑ/ in the context of a slow speech rate (Bosker, 2017a; Bosker & Reinisch, 2017; Maslowski et al., 2019a; Reinisch & Sjerps, 2013). This process, known as rate normalization, influences many duration-cued phonemic contrasts, such as singleton-geminate (Mitterer, 2018), /b/-/p/ (Gordon, 1988), /b/-/w/ (Wade & Holt, 2005), and recognition of unstressed syllables (*form* vs. *forum*; Baese-Berk et al., 2019) and words (*silver jewelry* vs. *silver or jewelry*; Dilley and Pitt, 2010; *cease* vs. *see us*; Baese-Berk et al., 2019). Importantly, contextual rate effects have been shown to arise very rapidly during spoken word comprehension, and have thus been hypothesized to occur at the earliest stages of perception (e.g., Bosker & Ghitza, 2018; Reinisch & Sjerps, 2013; Toscano & McMurray, 2015).

How exactly contextual speech rate cues interact with knowledge-based cues during online speech processing is unclear. For example, Morrill et al. (2015) examined the interacting effects of contextual speech rate and linguistic knowledge on reduced word recognition using a transcription task. They presented participants with utterances that included highly reduced function words, such as “or” in the sentence “*Don must see the har*bor [or] b*oats*.” Depending on the perception of the reduced function word “or” (in square brackets), this sentence could be interpreted as either “Don must see the harbor boats” or “Don must see the harbor or boats.” Crucially, the rate of the surrounding context (italics in the example) was manipulated to be either slowed or unaltered. Morrill et al. (2015) observed that slowing down the speech rate in the context made the reduced function word “or” perceptually disappear: Participants transcribed the sentence without the critical function word (e.g., “harbor boats” rather than “harbor or boats”). Moreover, even when the reduced function word was syntactically obligatory (e.g., “*Conner knew that bread and butt*er [are] b*oth in the pantry*,” where the sentence is only grammatical if the function word “are” is perceived as being present), participants still transcribed the sentence without the function word if it was embedded in slow speech. In fact, the effect of contextual speech rate was even observed to be comparable across syntactically optional and syntactically obligatory sentences, and no significant interaction was found between speech rate and syntactic obligatoriness, suggesting that the weighting of contextual speech rate was not modulated by conflicting syntactic cues. Contrasting older and younger speakers, Heffner et al. (2015) reported similar results: Presented with similar stimuli as used in Morrill et al. (2015), participants in both age groups were less likely to report a critical word if it was (a) presented in a slow context, and (b) syntactically optional. Again, the interaction between the two predictors was nonsignificant, suggesting that participants made use of knowledge-based and signal-based cues independently.

The observation in Morrill et al. (2015) and Heffner et al. (2015) that the weighting of contextual speech rate as a cue to lexical recognition is not modulated by conflicting syntactic cues seems to clash with Mattys et al.’s (2005) proposal that syntactic knowledge operates at the highest tier of lexical recognition. At the same time, the findings raise several questions. First, Morrill et al. (2015) and Heffner et al. (2015) used a transcription task, where participants were asked to transcribe the auditory stimuli after having heard the entire utterance. The results therefore reflect participants’ explicit decision-making about the nature of the stimuli and do not offer direct insights into *when* during comprehension signal- and knowledge-based cues are extracted, combined, and weighted. Second, the critical target region in Morrill et al.’s (2015) and Heffner et al.’s (2015) stimuli always preceded the syntactic cue. That is, in a sentence like “*Conner knew that bread and butt*er [are] b*oth in the pantry*,” where perception of the word “are” was obligatory for the sentence’s grammaticality, participants only discovered that the verb was syntactically obligatory *after* presentation of the critical region. This is especially interesting because Mattys et al. (2007) suggested that the time-course of knowledge-based and signal-based cues might play a crucial role in the way in which these two sources of information are integrated. If that is the case, it is possible that the absence of an interaction between acoustic and syntactic cues in Morrill et al. (2015) and Heffner et al. (2015) was due to their order in the stimuli. Finally, Morrill et al. (2015) and Heffner et al. (2015) reported group averages, but they did not investigate individual variation in cue weighting. Assuming a relative degree of flexibility in cue weighting as suggested by Martin (2016) and Mattys et al. (2007), as well as the results reported by Morrill et al. (2015) and Heffner et al. (2015), the question emerges whether individual participants also employed different strategies during the experiment, or whether cue-weighting effects generally arise on a group level.

Current Study

In the current experiment, we aimed to examine the flexible interplay between signal- and knowledge-based cues during online spoken language comprehension. More specifically, we used eyetracking within the visual world paradigm to test the robustness of signal-based contextual rate cues in the presence of earlier knowledge-based cues to grammatical gender. This allowed us to investigate how the system integrates potentially conflicting cues from different levels of linguistic hierarchy. We manipulated minimal word pairs in Dutch to contain vowel tokens that were ambiguous between short /ɑ/ and long /a:/ (e.g., vat_{NEUTER} “barrel,” $vaat_{COMMON}$ “dishes”), embedded in carrier sentences at slow or fast speech rates. Participants were presented with two pictures on a screen, corresponding to the short /ɑ/ or long /a:/ noun (e.g., a picture of a barrel and a picture of dishes), while listening to auditory instructions at fast or slow rates asking them to look at one of the two pictures (e.g., *Kijk nu eens naar* de_{COMMON}/het_{NEUTER} *ontzettend vuile* vat_{COMMON}/$vaat_{NEUTER}$, *alsjeblieft*, “Now look once at $the_{COMMON/NEUTER}$ terribly dirty $barrel_{COMMON}$/$dishes_{NEUTER}$, please”). Participants then clicked on the picture which they thought corresponded to the target. Crucially, the carrier sentences contained a preceding morphosyntactic cue in the form of the definite article de_{COMMON} or het_{NEUTER}, which has previously been shown to elicit anticipatory language processing within the visual world paradigm (Huettig & Janse, 2016). However, Huettig and Janse (2016) found that only about half of the participants showed the expected anticipatory looking behavior (see their Figure 5) in an experimental paradigm that only targeted morphosyntactic prediction. Similarly, using the same experimental paradigm, Huettig and Guerra (2019) reported evidence for anticipatory language processing being attenuated by factors such as shorter preview time, implicit versus explicit participant instructions, and faster speech rate of the carrier sentence. As such, it remains unclear whether morphosyntactically driven anticipatory looking behavior can be observed in an experiment with additional signal-based cues to target perception. In our experimental manipulation, the article could act as an early cue towards grammatical gender, and thus bias participants' perception towards one of the two vowel interpretations. There were thus two cues towards the "target" picture in our experiment: (a) the gender of the article preceding the noun, and (b) the contextual speech rate and its consequences for the relative perception of the temporal properties of the ambiguous vowel. Which of the two nouns participants considered the "target" was entirely up to them, depending on whether they preferred the information conveyed by the knowledge-based gender cue or the signal-based speech rate cue.

The cue integration model predicts that the system rapidly extracts and integrates signal-based and knowledge-based cues during spoken language comprehension. Our experimental manipulation allowed us to investigate the relative contribution of these cues from distinct levels of linguistic representation as a function of participants' looking preferences as the information in the sentence unfolded. We hypothesized that listeners would rapidly use the morphosyntactic gender cue conveyed by the article in order to make predictions about the ambiguous noun, which would be reflected in participants looking more toward the picture corresponding to the gender of the article. Crucially, this would occur well before the onset of the noun. The rate manipulation introduced a potential mismatch between the gender of the article and the gender of the (perceived) noun, making both cues somewhat unreliable for participants. Analyzing a time window immediately after the offset of the article, but before the onset of the ambiguous vowel, thus allowed us to address our first research question: Are knowledge-based, morphosyntactic cues toward grammatical gender immediately used to generate predictions about upcoming referents, even in the presence of uncertainty?

Second, we asked whether listeners would take signal-based contextual rate cues into account even in the presence of preceding disambiguating, potentially conflicting, articles. This would be reflected in participants shifting their gaze toward the picture corresponding to the vowel perception elicited by the rate manipulation after hearing the ambiguous vowel. Specifically, when embedded in a slow context sentence, the ambiguous vowel should appear relatively short in contrast to the preceding speech sounds, thus biasing participants toward perceiving the vowel as /ɑ/ and looking at the corresponding picture. Conversely, fast context rates should bias participants toward looking more at the picture corresponding to an /a:/ vowel interpretation (Reinisch & Sjerps, 2013). Analyzing a time window immediately after the offset of the ambiguous vowel (i.e., the earliest moment in time when participants could access the duration cues on the vowel) until the end of the utterance thus allowed us to answer our second research question: Are signal-based, contextual speech rate cues used even in the presence of preceding, potentially conflicting, morphosyntactic information?

Regarding both of our questions, it is possible that one of the two cues is entirely overwritten by the other, and that participants base their choice of response only on the cue that they perceive as more reliable. For example, the signal-based, speech rate induced cue might be entirely overwritten by the preceding knowledge-based, morphosyntactic cue. If, as Mattys et al. (2005) suggested, syntactic cues are generally weighted more strongly than acoustic cues, we should thus not find significant changes to eye fixations as a function of contextual speech rate during the vowel window, because participants would simply weigh the syntactic cue more heavily and ignore the contextual rate manipulation. Observing more looks toward the picture corresponding to the gender of the article in the carrier sentence, but no effect of the speech rate manipulation during the vowel window, would thus be in line with Mattys et al.'s (2005) original model. Conversely, observing rate normalization effects in the noun window, even in the presence of preceding morphosyntactic information, would be in line with the later model suggested by Mattys et al. (2007), and with more general, computationally formalized models of cue integration (Martin, 2016).

Using eye-tracking within the visual world paradigm allowed us to investigate this potentially flexible weighting of signal-based and knowledge-based cues while participants were processing the sentences online. However, it is possible that individual participants employ different strategies during cue integration and sentence comprehension (cf. Van Bergen & Bosker, 2018). If, for instance, half of our participants weighed the knowledge-based cue more heavily, while the other half relied more strongly on the signal-based cue, then standard analysis of average behavior across all participants would not be very insightful.

Therefore, we also mapped participants' offline behavioral responses (target categorization mouse-clicks) to their online cue weighting behavior as evidenced in their gaze patterns. Specifically, we created a measure of each individual's preference for the knowledge-based versus the signal-based cue based on their categorization responses, which we then linked to participants' eye fixations in the vowel window. Rather than drawing conclusions about each cue's relative weight based solely on average behavioral measures (e.g., Heffner et al., 2015; Morrill et al., 2015), we were thus able to investigate whether individual strategies were reflected in different eye-tracking patterns.

Moreover, mapping participants' behavioral responses to their eye-tracking data also allowed us to test whether we could find online evidence for rate manipulation effects in the eye fixation data for participants whose behavioral responses principally followed the knowledge-based, syntactic cue. This question is relevant in light of debate about the robustness of phoneme-level rate effects, which some have proposed to be "fragile" (Baese-Berk et al., 2019), while others have argued that they are robust and potentially automatic (Bosker et al., 2017; Reinisch & Sjerps, 2013). Observing phoneme-level rate effects even for participants who behaviorally favored the preceding morphosyntactic cue would be strong evidence for rate normalization effects arising very early during perception, unmodulated by other information sources.

3.2 Method

We aimed to test (a) whether knowledge-based, morphosyntactic cues toward grammatical gender were immediately used to generate predictions about upcoming referents, and (b) whether signal-based, contextual speech rate cues persisted even in the presence of preceding, potentially conflicting, morphosyntactic information. We used eye-tracking (visual world paradigm) in order to obtain online measures of the influence of these knowledge-based and signal-based cues on the perception of the phonemic vowel contrast /ɑ/ versus /a:/ in Dutch. We further mapped online eye-tracking data to offline behavioral data in order to investigate the strategies that individuals employ while combining different sources of information.

Participants

Native speakers of Dutch ($N = 36$, 19 females, $M_{age} = 22$ years) with self-reported normal hearing were recruited from the Max Planck Institute for Psycholinguistics (MPI) participant pool, with informed consent as approved by the Ethics Committee of the Social Sciences Department of Radboud University (Project Code: ECSW2014-1003–196). Participants were paid for their participation.

Materials and Design

Stimuli consisted of seven Dutch sentences, each containing a unique /ɑ/-/a:/ minimal pair that differed in grammatical gender (common vs. neuter; e.g., *(de) as*$_{COMMON}$ “ash” - *(het) aas*$_{NEUTER}$ “bait”). Each sentence followed a specific sentence frame, for instance, *Kijk nu eens naar [de|het] ontzettend vieze [as|aas] alsjeblieft*; “Look now once to [the$_{COMMON}$|the$_{NEUTER}$] very dirty [ash$_{COMMON}$|bait$_{NEUTER}$] please” (see Appendix for a complete list of stimuli). We recorded a female native speaker of Dutch, who was naïve to the purpose of the experiment, reading the sentences in two syntactic conditions (with *de* and *het*) and with both nouns. Recordings were made in a sound-attenuated booth and digitally sampled at 44,100 Hz on a computer located outside the booth with Audacity software (Audacity Team, 2019).

For the various speech rate and syntactic conditions, we manipulated, using PSOLA in Praat (Boersma & Weenink, 2020), the speech rate of the lead-in fragment (*Kijk nu eens naar*), adjectival phrase (e.g., *ontzettend vieze*) and the final fragment (*alsjeblieft*) of each sentence in a combined fashion through linear compression with a factor of 0.66 (fast condition) and 1.5 (1/0.66; slow condition) of the original recording. We created two syntactic conditions of each sentence by replacing the article *het* from each sentence with the article *de* from a recording of the same sentence.

For the nouns, we required vowels that were both spectrally and durationally ambiguous. The /ɑ/-/a:/ vowel contrast in Dutch is cued by both spectral (lower formant values for /ɑ/, higher formant values for /a:/) and temporal cues (shorter duration for /ɑ/, longer duration for /a:/; Escudero, Benders, and Lipski, 2009). We created two-dimensional spectral and durational vowel continua for each vowel by first creating a linear 9-point duration continuum (1 = original duration of /ɑ/; 9 = original duration of /a:/; in steps of 12.5% of the duration difference; using PSOLA in Praat). Then, for each duration step, we used sample-by-sample linear interpolation (9-point continuum; 1 = 100% /ɑ/ + 0% /a:/; 5 = 50% /ɑ/ + 50% /a:/; 9 = 0% /ɑ/ + 100% /a:/) to create different spectral versions of the durationally matched vowels (i.e., changing vowel quality).

We then conducted a pretest in order to choose the most suitable (i.e., the most ambiguous) combinations of duration and interpolation steps for each item pair. Participants who were naïve to the purpose of the experiment and did not participate in the main experiment (N = 20, 15 females, M_{age} = 23.8 years) listened to short excerpts of the created stimuli, consisting of only the adjectival phrase, noun, and outro, thus avoiding any biasing information from the article (e.g., *ontzettend vieze* as_{COMMON}/aas_{NEUTER} *alsjeblieft*). They indicated via button press whether they had heard the word corresponding to the vowel /ɑ/ or /a:/ (e.g., *as* or *aas*). Based on the results of the pretest, we selected a unique set of five different duration steps from one and the same interpolation step for each item pair. These five steps spanned a perceptual range of relatively few long /a:/ responses (mean long /a:/ categorization of Step 1 = 22%) to relatively many long /a:/ responses (mean long /a:/ categorization of Step 5 = 65%). This resulted in seven unique five-step duration continua with fixed vowel qualities.

The resulting 140 stimuli (2 Syntactic Conditions × 2 Speech Rates × 7 Pairs × 5 Continuum Steps) formed an experimental block. Participants were presented with two blocks in an experimental session, so that each participant was exposed to 280 sentences in total. The pictures for the visual-world paradigm were selected from the MultiPic database (Duñabeitia et al., 2018) if available, or retrieved from copyright-free online resources. All pictures were scaled to a dimension of 300 pixels at the longest side.

Procedure

Participants were tested individually in a sound-conditioned booth. They were seated at a distance of approximately 60 cm in front of a 50.8 cm × 28.6 cm screen with a tower-mounted Eyelink 1000 eye-tracking system (SR Research) and listened to stimuli at a comfortable volume through headphones. Stimuli were delivered using Experiment Builder software (SR Research). Eye movements were recorded using right pupil-tracking at a rate of 1000 Hz.

Each trial started with a blue fixation rectangle in the middle of the screen to center the mouse and the participant's gaze position. The rectangle disappeared when the participant clicked on it. The fixation screen was immediately followed by the presentation of the visual stimuli. After a 1-s preview interval, the auditory stimulus was played. Participants were instructed to listen to the complete auditory stimulus (no response possible before audio offset) and to click on the corresponding picture on the screen. We did not instruct participants about the (non-)grammaticality of some article plus vowel combinations (i.e., hearing het_{NEUTER} in combination with as_{COMMON} with a short vowel sounds ungrammatical). Thus, participants were free to choose whichever cue to base their categorization responses on. The trial ended when participants had clicked on a picture. The positioning of the visual stimuli, centered in the left or right half of the screen, was counterbalanced across participants, and the order of trials within a block was randomized across participants.

In order to get familiarized with the task, participants completed four practice trials before the experiment. After half of the experiment, participants were allowed to take a self-paced break. Including instructions, calibration, and debriefing, the experimental procedure took approximately 50 to 60 min to complete.

3.3 Results

Behavioral categorization data

Due to the nature of our stimuli and the morphosyntactic regularities of the Dutch language, we could not simply include the gender of the definite article in our statistical analyses, because there is no 1:1-mapping between each noun's gender and its associated vowel length. In other words, hearing the article het_{NEUTER} might bias participants toward looking at the picture corresponding to a long vowel for some item pairs (e.g., aas_{NEUTER} "bait" and as_{COMMON} "ash"), but to a short vowel for others (e.g., $vaat_{COMMON}$ "dishes" and vat_{NEUTER} "barrel"). In order to capture this variability for further statistical analyses, we decided to include a binomial article bias variable in our statistical analyses. To further illustrate, when presented with two pictures (e.g., aas_{NEUTER} "bait" and as_{COMMON} "ash"), hearing the article het_{NEUTER} would be a cue toward a *long* vowel interpretation, whereas hearing the article de_{COMMON} would be a cue toward a *short* interpretation. Depending on the trial-specific combination of article and pictures, each trial can thus be considered to introduce a *long* or *short* article bias.

Figure 3.1 shows participants' categorization responses (calculated as the proportion of long responses) split by the two article biases, collapsed across all nouns, for slow and fast context rates.

We used GLMMs with a logistic link function (Jaeger, 2008) as implemented in the lme4 library (D. Bates et al., 2015) in R (R Development Core Team, 2012) in order to evaluate the binomial response corresponding to a long vowel interpretation (1 = yes, 0 = no) for fixed effects of article bias (categorical predictor with two levels: article bias toward long vowel coded as 0.5; short as -0.5), continuum (continuous predictor; centered: Step 1 coded as -2, Step 3 as 0, Step 5 as 2), and rate (categorical predictor with two levels: fast coded as 0.5; slow as -0.5), and all their interactions. The random effects structure contained random intercepts for participants and items, because adding additional random slopes resulted in nonconvergence.

Effects	Estimate	*SE*	z	*p*	Variance	*SD*
Fixed Effects						
Intercept	.103	.241	.429	.668		
Rate^a	1.759	.057	30.855	<.001		
Continuum^b	.383	.020	19.377	<.001		
Article bias^a	2.479	.059	42.229	<.001		
Rate:Continuum	.076	.039	1.939	.052		
Rate:Article bias	.166	.111	1.486	.137		
Continuum:Article bias	.007	.039	.173	.863		
Rate:Continuum:Article bias	.246	.078	3.142	.002		
Random Effects						
Intercept|Participant					.629	.793
Intercept|Item					.279	.528

Table 3.1: Mixed-effects logistic regression results of the behavioral responses corresponding to a long vowel. ^{a}Contrast coded (slow = -0.5, fast = +0.5; article bias towards short vowel = -0.5, article bias towards long vowel = +0.5); ^{b}centered.

The complete model outputs are summarized in Table 3.1. The model revealed a significant effect of continuum ($p < .001$), indicating that participants were more likely to select the "long vowel" picture at higher continuum steps. This shows that our experimental vowel manipulation was successful. The model also revealed a significant effect of article bias ($p < .001$), indicating that participants were more likely to select the "long vowel" in trials in which the preceding article was congruent with the picture corresponding to a long vowel interpretation.

Crucially, there was also a significant effect of rate ($p < .001$), indicating that participants were more likely to respond with the picture corresponding to the long vowel in fast contexts. This indicates that rate effects occurred even in the presence of earlier morphosyntactic information.

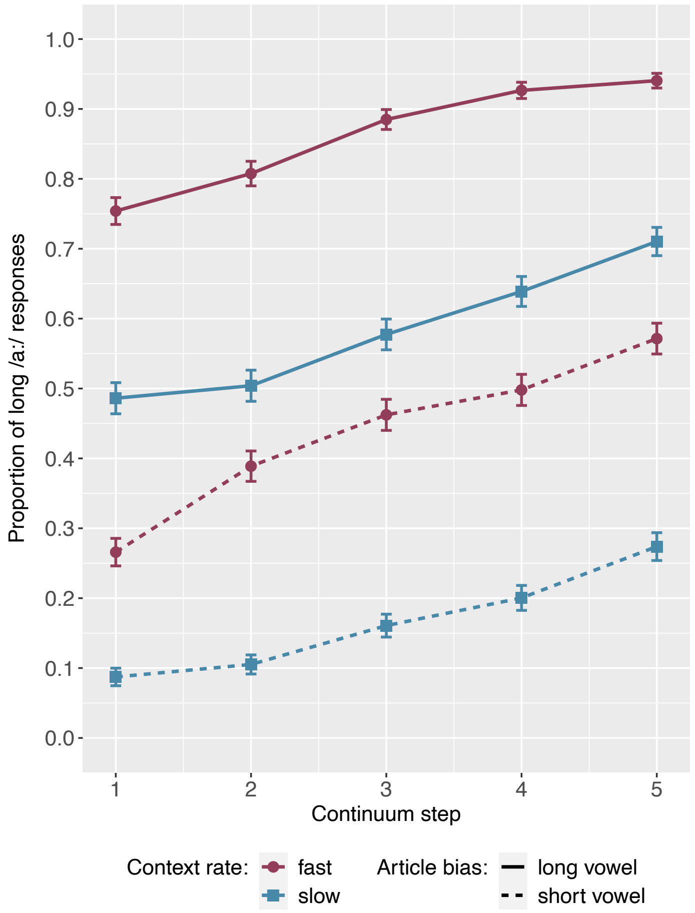

Figure 3.1: Categorization curves showing the proportion of long vowel responses. Long vowel responses are plotted as a function of duration continuum step, split for the two speech rates (red: fast rate; blue: slow rate) and the two article biases (solid: article biases towards "long vowel" interpretation; dashed: article biases towards "short vowel" interpretation). Error bars represent the standard error of the mean.

We also found a significant three-way interaction between rate, continuum, and article bias ($p = .002$), indicating that the effect of article bias was slightly more pronounced at higher duration continuum steps in fast contexts.[2]

Investigating attenuating effects of the morphosyntactic cue

If rate effects are easily modulated and overridden by higher-level information, it is possible that the rate effects we observe here are attenuated by the presence of the earlier morphosyntactic cue and thus smaller than they would be in

[2] In order to investigate whether these effects changed as a function of experimental block, we also tested a model including an additional fixed effect of Block and all possible interactions. This model revealed no main effect of Block and no interactions with Block.

isolation. We investigated this question by comparing the behavioral responses from the experiment to those from the pretest, where participants heard the manipulated vowel embedded in a fast or slow context, but without any additional morphosyntactic information (i.e., sentence excerpts excluding the article; see Materials and Design section).

A GLMM tested the binomial responses corresponding to a long vowel interpretation (1 = yes, 0 = no) for fixed effects of continuum (continuous predictor; centered: Step 1 coded as -2, Step 3 as 0, Step 5 as 2), rate (categorical predictor with two levels: fast coded as 0.5; slow as -0.5), and experiment (categorical predictor with two levels: main experiment coded as 0.5; pretest as -0.5), and the interaction between rate and experiment. The model contained random intercepts for participants and items.

The complete model outputs are summarized in Table 3.2. The model revealed significant main effects of rate ($p < .001$) and continuum ($p < .001$), but no main effect of experiment ($p = .338$). No interaction between rate and experiment was observed ($p = .252$), indicating that the rate effect was not attenuated by the presence of preceding morphosyntactic information in the main experiment. We take this to suggest that the rate effect was robust against modulation by higher-level information.

Effects	Estimate	*SE*	z	*p*	Variance	*SD*
Fixed Effects						
Intercept	-.034	.179	-.192	.848		
Rate^a	1.196	.088	13.535	<.001		
Continuum^b	.291	.015	18.831	<.001		
Experiment^a	.201	.209	.959	.338		
Rate:Continuum	.031	.043	.711	.477		
Rate:Experiment	.202	.176	1.146	.252		
Random Effects						
Intercept|Participant					.466	.683
Intercept|Item					.148	.385

Table 3.2: Mixed-effects logistic regression results of the behavioral responses corresponding to a long vowel during pretest and main experiment. ^{a}Contrast coded (slow = -0.5, fast = +0.5; pretest = -0.5, main experiment = +0.5); ^{b}centered.

Eye-tracking data

Prior to the analyses, blinks and saccades were excluded from the data. We divided the screen into two sections (left and right) and coded fixations on either half as a look toward that particular picture. The eye fixation data were down-sampled to 100 Hz for simplicity. We used GLMMs with a logistic link function (Jaeger, 2008) as implemented in the lme4 library (D. Bates et al., 2015) in R (R Development Core Team, 2012) in order to evaluate participants' eye fixations as the meaning of each sentence unfolded across time.

Article window analysis

In order to investigate our first question, we analyzed a time window spanning from 200 ms after article onset, accounting for the time it takes to launch a saccade (Matin et al., 1993) until the onset of the ambiguous word. This allowed us to test whether anticipatory language processing based on the gender information carried in the article occurred even when the article was not a univocally reliable cue toward the noun. We expected to find that items containing an article that biased toward the object corresponding to a long vowel interpretation would elicit more looks to the long object, well before the onset of the noun. Figure 3.2 shows participants' eye fixations (calculated as the proportion of looks to the pictures of long vowel interpretation) split by article bias (to either the long or the short vowel interpretation) and rate (fast vs. slow) in the article window, with the analysis window shaded in gray. Note that we do not illustrate the article analysis window in its entirety here, because the onset of the ambiguous noun was earlier in the fast compared with the slow speech rate condition and the length of the analysis window thus differed between the two rate conditions.

A GLMM tested the binomial looks to the object corresponding to a long vowel interpretation (1 = yes, 0 = no) for fixed effects of article bias (categorical predictor with two levels: long coded as 0.5; short as -0.5), time (z-scored around the mean of the analysis window), and their interaction. The random effects structure contained random intercepts for participants and items and by-participant and by-item random slopes for both fixed factors and their interactions. Note that we did not include rate as a predictor in this model because participants had not yet heard the ambiguous vowel at this point in time.

The model (see Table 3.3) revealed a significant effect of article bias ($p <$ .001), indicating that participants were more likely to look at the picture corresponding to a long vowel interpretation if the article corresponded to that interpretation. We also found a significant interaction between article bias and time, indicating that the effect of article bias grew over time (i.e., we observed a larger effect in later parts of the time window; $p = .002$).

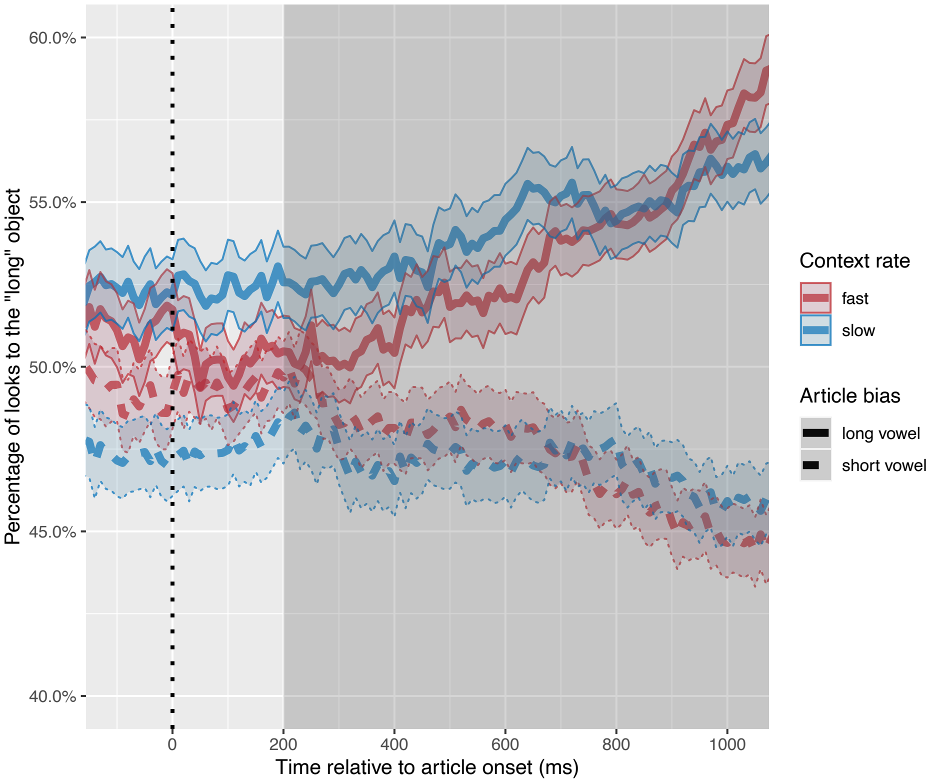

Figure 3.2: Percentage of looks to the "long vowel" object across time in the article window. Time point 0 marks the onset of the article. Looks towards the long item are plotted across time following the onset of the article in two syntactic categories (trials with a "long vowel" article bias contained an article that biased participants towards the object corresponding to the long interpretation: solid line; trials with a "short vowel" article bias contained an article that biased participants towards the short interpretation: dashed line) when embedded in contexts of distinct rates (fast rate: red line, slow rate: blue line). The area shaded in grey indicates an illustration of the window analyzed in the article window analysis, spanning from 200 ms after article onset until the onset of the ambiguous word. Red and blue shading indicates standard error of the mean.

Vowel window analysis

In order to investigate whether the effects of the rate manipulation on eye fixations could still be observed after the presentation of preceding morphosyntactic information (in our case, the article encoding the gender of the noun), we analyzed a vowel window ranging from 200 ms after the offset of the manipulated vowel until speech offset. Again, this time window was chosen in order to account for the 200 ms that it takes to launch a saccade (Matin et al., 1993). We selected vowel offset, rather than vowel onset, to be the starting point of

Effects	Estimate	*SE*	z	*p*	Variance	*SD*
Fixed Effects						
Intercept	.019	.094	.203	.839		
Article bias[a]	.299	.085	3.519	<.001		
Time[b]	.011	.016	.704	.481		
Time:Article bias	.142	.045	3.162	.002		
Random Effects						
Intercept|Participant					.012	.110
Article bias|Participant					.192	.438
Time|Participant					.004	.062
Time:Article bias|Participant					.054	.233
Intercept|Item					.063	.251
Article bias|Item					.018	.136
Time|Item					.001	.031
Time:Article bias|Item					.005	.069

Table 3.3: Mixed-effects logistic regression results of the looks to the long object across time in the article window. [a]Contrast coded (article bias towards short vowel = -.5, article bias towards long vowel = .5); [b]z-scored.

the time window, because listeners only had access to the critical duration cues on the vowel after hearing it in its entirety. Figure 3.3 shows participants' eye movements, calculated as the proportion of looks to the pictures of long vowel interpretation, split by article bias to either the long or the short vowel interpretation and rate (fast vs. slow) in the vowel window, with the vowel analysis window shaded in gray. Figure 3A.1 (Appendix) illustrates the effect of continuum reported below.

A GLMM tested the binomial looks to the object corresponding to a long vowel interpretation (1 = yes, 0 = no) for fixed effects of article bias (categorical predictor with two levels: long coded as 0.5; short as -0.5), continuum (continuous predictor; centered: Step 1 coded as -2, Step 3 as 0, Step 5 as 2), rate (categorical predictor with two levels: fast coded as 0.5; slow as -0.5), time (z-scored around the mean of the analysis window), and all their interactions. The random effects structure contained random intercepts for participants and items and by-participant and by-item random slopes for all fixed factors (but not their interactions, as the model failed to converge if they were also added to the random effects structure).

The complete model outputs are summarized in Table 4 (see "Base Model" column). The model revealed significant main effects of continuum (p < .001; see Figure 3A.1 in the Appendix) and rate (p < .001). These results indicate that (a) participants were more likely to look at the picture corresponding to a long vowel at higher continuum steps, and (b) participants were more likely to look at the "long" picture in fast as opposed to slow contexts. Note that the main effect of rate indicates that rate manipulations have an effect on vowel perception

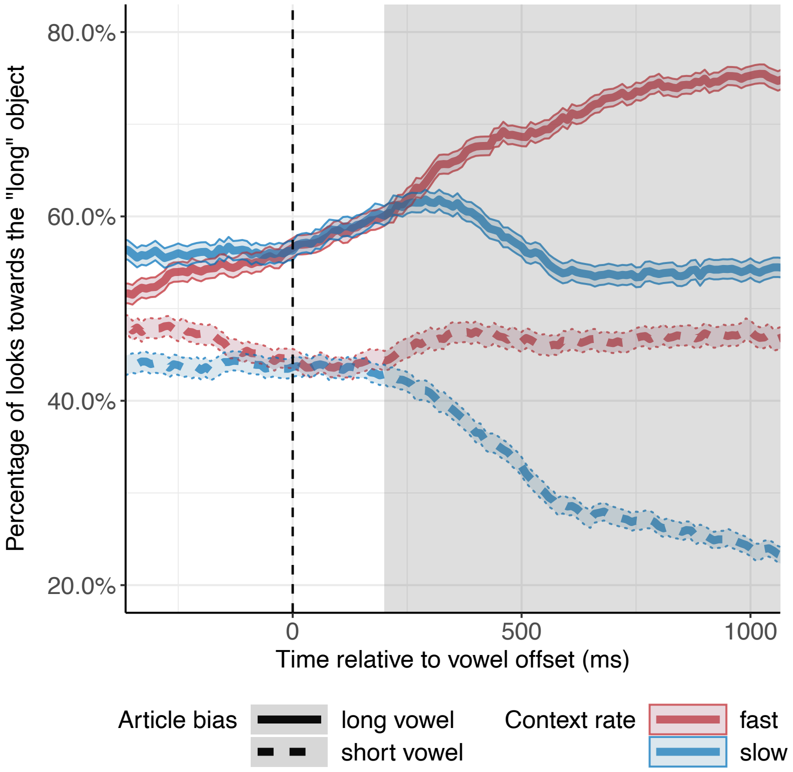

Figure 3.3: Percentage of looks to the "long" object across time in the vowel window. Time point 0 marks the offset of the ambiguous vowel. Looks towards the long item are plotted across time in two syntactic categories (trials with a "long vowel" article bias contained an article that biased participants towards the object corresponding to the long interpretation: solid line; trials with a "short vowel" article bias contained an article that biased participants towards the short interpretation: dashed line) when embedded in contexts of distinct rates (fast rate: red line, slow rate: blue line). The area shaded in grey indicates the window analyzed in the vowel window analysis, spanning from 200 ms after vowel offset until stimulus offset. Red and blue shading indicates standard error of the mean.

independently of any article bias. The model also revealed a significant effect of article bias ($p < .001$), indicating that participants were still more likely to look at the "long vowel" picture in the vowel window if the preceding article was congruent with a "long vowel" interpretation. It thus appears that, generally, participants did not entirely dismiss the morphosyntactic cue upon hearing the acoustic cue, nor the other way around.

On top of these main effects, the model also revealed two-way interactions between time and rate ($p < .001$), time and continuum ($p < .001$), and time and article bias ($p < .001$). These interactions indicate that the effects of rate,

	Base model						Extended model					
Effects	Est.	*SE*	Z	*p*	var.	*SD*	Est.	*SE*	Z	*p*	var.	*SD*
Fixed Effects												
Intercept	.073	.145	.504	.614			.061	.135	.448	.654		
Time[b]	-.030	.067	-.453	.651			-.038	.006	-6.679	<.001		
Rate[a]	.791	.109	7.265	<.001			.701	.009	71.215	<.001		
Continuum[c]	.237	.041	5.792	<.001			.221	.003	63.308	<.001		
Article bias[a]	1.251	.220	5.679	<.001			1.141	.009	116.252	<.001		
Individual strategy[c]							-.064	.208	-.308	.758		
Time:Rate	.444	.012	37.868	<.001			.358	.011	31.274	<.001		
Time:Continuum	.152	.004	36.821	<.001			.141	.004	34.904	<.001		
Time:Article bias	.332	.012	28.474	<.001			.252	.011	22.119	<.001		
Rate:Continuum	.115	.007	16.069	<.001			.111	.007	15.853	<.001		
Rate:Article bias	-.023	.020	-1.149	.251			-.061	.019	-3.093	.002		
Rate:Individual strategy							-1.150	.017	-66.537	<.001		
Continuum:Article bias	.017	.007	2.362	.018			.012	.007	1.775	.076		
Article bias:Individual strategy							2.237	.016	138.433	<.001		
Time:Rate:Continuum	.103	.008	12.513	<.001			.115	.008	14.254	<.001		
Time:Rate:Article bias	.064	.023	2.757	.006			.066	.023	2.892	.004		
Time:Continuum:Article bias	-.012	.008	-1.419	.156			-.013	.008	-1.660	.097		
Rate:Continuum:Article bias	.050	.014	3.533	<.001			.036	.014	2.576	.010		
Time:Rate:Continuum:Article bias	.128	.017	7.768	<.001			.124	.016	7.694	<.001		
Random Effects												
Intercept|Participant					.275	.525					.225	.474
Time|Participant					.066	.256						
Rate|Participant					.186	.431						
Continuum|Participant					.021	.146						
Article bias|Participant					1.006	1.003						
Intercept|Item					.114	.338					.098	.314
Time|Item					.021	.143						
Rate|Item					.053	.229						
Continuum|Item					.008	.088						
Article bias|Item					.207	.455						

Table 3.4: Mixed-effects logistic regression results of looks to the long object across time in the vowel window. The “Base model” includes the analysis described in the section *Vowel window analysis*, the “Extended model” additionally includes effects of (and interactions with) Individual Bias reported in section *Investigating individual strategies*. [a]Contrast coded (slow = -0.5, fast = +0.5; article bias towards short vowel = -0.5, article bias towards long vowel = +0.5); [b]z-scored; [c]centered.

continuum, and article bias all grew stronger over time. The absence of an interaction between rate and article bias ($p = .251$) further suggests that participants used knowledge- and signal-based cues independently during the vowel window (but see the *Investigating individual strategies* section). There were also small significant interactions between rate and continuum ($p < .001$), indicating that rate effects were slightly stronger at higher ends of the vowel continuum; and between continuum and article bias ($p = .018$), indicating that the effect of article bias was more pronounced at higher continuum steps. However, because these have relatively small effect sizes and we did not have specific predictions regarding interactions with continuum, we do not discuss these further.

Furthermore, the model also revealed several significant three-way interactions and even a four-way interaction. Note, however, that all these interactions had very small estimates, contributing only modestly to the observed patterns. The model revealed an interaction between time, rate, and continuum ($p < .001$), indicating that the rate effect grew slightly stronger across time at higher continuum steps; a three-way interaction between time, rate, and article bias ($p = .006$), indicating that the rate effect grew slightly stronger across time for trials that were biasing participants toward the long interpretation; and a three-way interaction between rate, continuum, and article bias ($p < .001$), suggesting a slightly diminished effect of rate for trials with a long-vowel-congruent article bias at lower continuum steps. Finally, the model revealed a significant four-way interaction between rate, article bias, continuum, and time ($p < .001$); we currently lack an explanation for this, but note that the effect is very small.[3]

Investigating individual strategies

Taken together with previous findings by Morrill et al. (2015), Mattys et al. (2007), and Heffner et al. (2015), our results point toward a mechanism of spoken language comprehension that integrates cues from both knowledge- and signal-based levels. However, all previous studies reported averages, so it cannot be ruled out that individual participants showed strong preferences for one of the two cues. Specifically, in our study, the knowledge-based effect of article bias and the signal-based effect of rate could be driven by different participants. This is especially interesting in light of our second research question: Investigating whether rate effects in online gaze patterns persist even for participants who

[3] In order to investigate whether this effect changed as a function of experimental block, we also tested a model including an additional fixed effect of block and interactions between article bias*block and rate*block. This model revealed no additional main effect of block and no interactions with block.

behaviorally favor the syntactic cue will allow us to gain new insights into the robustness of phoneme-level contextual rate effects. In the following section, we report an analysis in which we map participants' behavioral responses to their eye-tracking data.

Participants' behavior on the categorization task can be classified to fall in between two extremes, depending on which of the cues the participants weighted more strongly during the experiment. "Syntax-followers" would attribute a higher weight to the morphosyntactic information carried in the definite article, while "acoustics-followers" would weigh the acoustic information induced by the contextual speech rate more strongly. Participants' behavioral categorization responses offer insights into which of the two cues they preferred in explicit categorization, and thus by proxy into which cue they weighted more strongly. Investigating each participant's eye-tracking behavior while taking their behavioral preference into account thus yields further insights into how different participants combined the two (possibly competing) cues in an online fashion, and whether the dispreferred cue was still considered by individuals that behaviorally favored the other cue.

For further analyses, we first created an individual strategy variable, which captured each participant's ratio of syntax-following responses. Specifically, we calculated the proportion of each participant's "long" responses after hearing an article biasing toward a "long" response, and subtracted from this the proportion of "long" responses after hearing an article biasing toward a "short" response. This resulted in an individual strategy score between 0 and 1 for each participant ($M = 0.42$, $SD = 0.36$, $min = 0.03$, $max = 1$; complete data given in the Appendix). Participants that weighted the morphosyntactic cue on the article very strongly, and the contextual speech rate less so, would be expected to have an individual strategy score approaching 1. In contrast, participants that weighted the contextual speech rate cue more strongly would have an individual strategy score around 0. Generally, participants appeared to behaviorally favor a mixture of the two cues. This is reflected in the group mean individual strategy score of 0.42, as well as in the observation that no participant had an individual strategy score of exactly 0, and only one participant had an individual strategy score of 1.

For plotting purposes, we split participants into two groups based on their individual strategy scores. Participants with an individual strategy higher than the mean (0.42) were considered syntax-followers, participants with an individual strategy score of 0.42 or lower were considered acoustics-followers. Figure 3.4 shows the eye-tracking responses split by group.

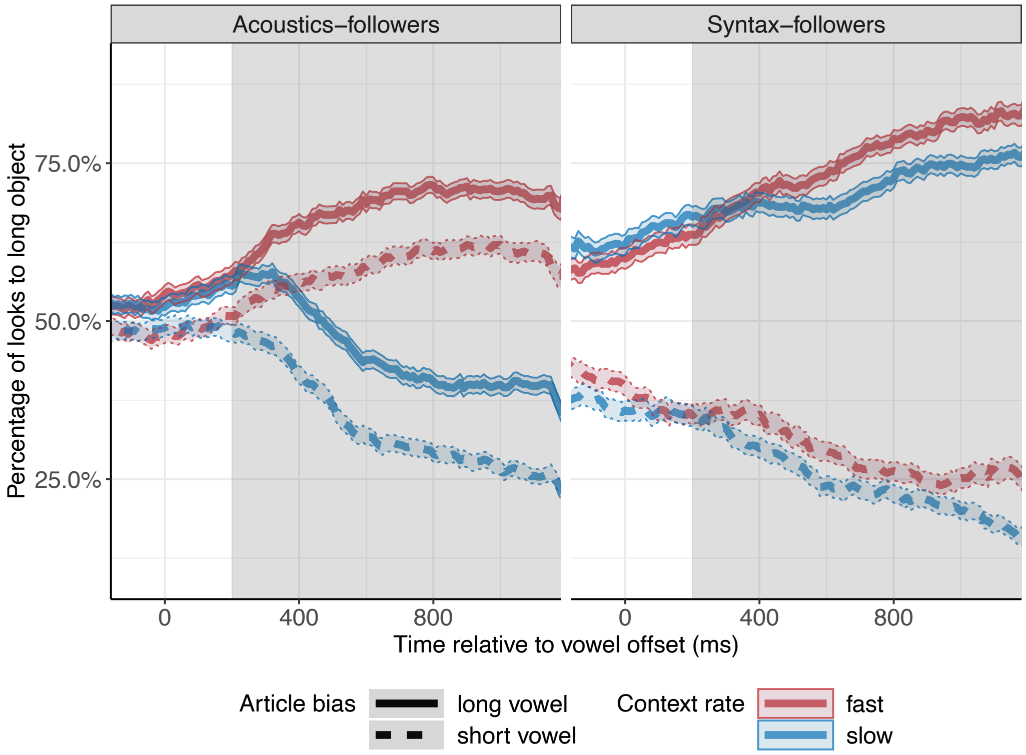

Figure 3.4: Percentage of looks to the object corresponding to a long vowel interpretation across time in the vowel window, split by acoustics-following (left panel) vs. syntax-following (right panel) participants. Time point 0 marks the offset of the manipulated vowel. Looks towards the long item are plotted across time following the onset of the noun in trials with a long article bias (solid line) and trials with a short article bias (dashed line). Ambiguous vowels were embedded in contexts at a fast (red line) or slow rate (blue line). The area shaded in grey indicates the analysis time window, ranging from 200 ms after vowel offset until the end of the stimulus. Red and blue shading indicates standard error of the mean.

We extended the GLMM which analyzed the vowel time window (described in the *Vowel window analysis* section) to include a main effect of individual strategy (continuous predictor, centered), an interaction term between article bias and individual strategy, and an interaction term between rate and individual strategy.

The complete model outputs are summarized in Table 3.4 (see “Extended Model” column). The model revealed the same significant effects as the simpler base model for the vowel time window, with main effects of rate, article bias, and continuum. In addition, we observed a main effect of time ($p < .001$)

and a small interaction between rate and article bias (p = .002), which were not significant in the simpler model. These additional effects indicated that participants looked more toward the long object as time progressed, and that the overall rate effect was slightly more pronounced in trials in which the article biased listeners toward a short vowel interpretation.

Crucially, the extended model revealed an additional significant interaction between rate and individual strategy (p < .001), indicating that rate effects were stronger for participants with a lower individual strategy score (i.e., acoustics-followers), as well as a significant interaction between article bias and individual strategy (p < .001), indicating that the effect of article bias was more pronounced for participants with higher individual strategy scores (i.e., syntax-followers).

Taken together, these findings indicate that, while it appears to be the case that different participants employed different strategies during the experiment, participants were unlikely to rely exclusively on either of the two cues. Crucially, the effect of individual strategy modulated the rate effect only to a limited extent. in fact, based on the estimates of the predictor rate and the interaction between rate and individual strategy, one learns that the model still predicts a small rate effect for participants with an individual strategy score of 1 (i.e., participants that exclusively gave syntax-following responses). Specifically, recall that the estimate of rate of 0.70 reflects the rate effect at the mean individual strategy score (i.e., 0.42). The estimate of the interaction between rate and individual strategy (-1.15) allows calculation of the predicted rate effect at an extreme individual strategy score of 1: 0.70 + (-1.15 [1 - 0.42]) = 0.033. Although this value is small, it still reflects a positive rate effect, as predicted.

In order to further illustrate this, we linked each participant's rate effect size in their looking behavior to their behavioral individual strategy score. Specifically, we calculated individual eyetracking rate effect sizes in the vowel window by subtracting each participant's mean proportion of looks to the long object in slow contexts from their mean proportion of looks to the long object in fast contexts. The resulting measure thus captures the difference in that participant's eye fixation behavior between fast and slow contexts, and thus their individual rate effect. Figure 3.5 shows each individual's rate effect size plotted against their individual strategy score. We color-coded participants with an individual strategy score equal to or higher than 0.43 as syntax-followers (purple dots) and those with a lower score as acoustics-followers (yellow dots) for illustration purposes.

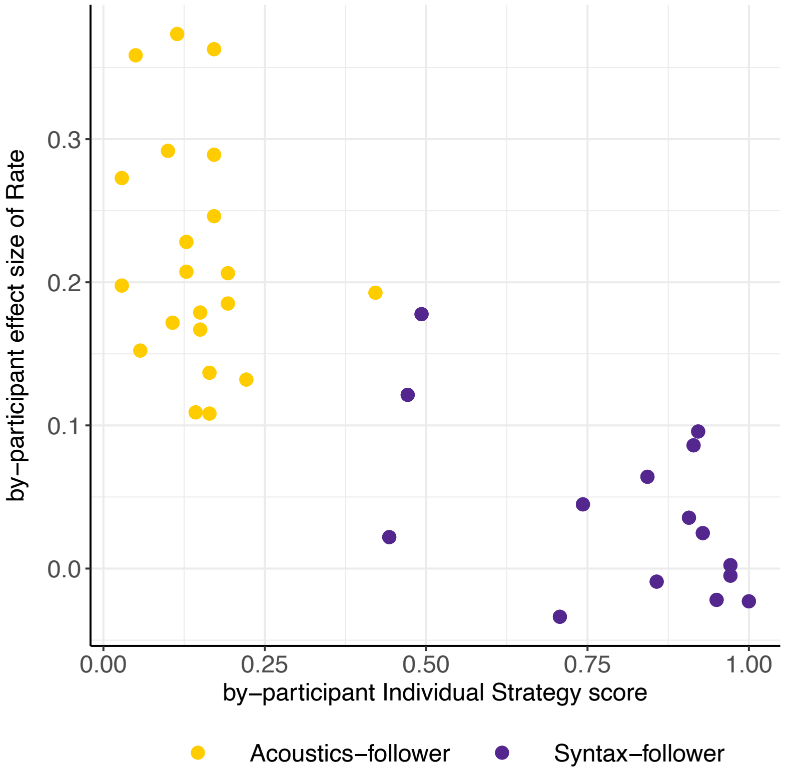

Figure 3.5: Individual participants' rate effect size plotted against their individual strategy score. Individual strategy scores were calculated as the proportion of syntax-adhering responses for each participant, and rate effect sizes were calculated as each participant's difference in looks to the long object between slow and fast context sentences in the vowel window. For illustration purposes only, participants are color-coded as "syntax-followers" (purple dots; individual strategy score > 0.42) or "acoustics-followers" (yellow dots; individual strategy score <= 0.42).

3.4 Discussion

The aim of the current study was to investigate two main questions. First, we asked whether knowledge-based, morphosyntactic cues toward gender are immediately used to generate predictions about upcoming referents. Second, we asked whether signal-based, contextual speech rate effects persist even in the presence of earlier disambiguating morphosyntactic information. We addressed these questions by experimentally inducing contextual rate normalization effects on an ambiguous vowel between short /ɑ/ and long /a:/ in Dutch minimal pairs, while at the same time providing an earlier morphosyntactic gender cue. In the following, we discuss our results in light of these two questions.

Knowledge-based cues are rapidly taken up and used to make predictions about upcoming referents, even in the presence of uncertainty

Our analyses of the article window show that participants were more likely to look at the object corresponding to the vowel interpretation that was consistent with the morphosyntactic gender information conveyed in the definite article. These results indicate that participants rapidly use knowledge-based cues in order to make predictions about the gender of the upcoming noun.

Huettig and Janse (2016) reported results from a similar eyetracking experiment in which participants were presented with auditory stimuli containing an article that matched only one of four possible objects on the participant's screen (e.g., *Kijk naar de*$_{COMMON}$ *afgebeelde piano*$_{COMMON}$, "Look at the displayed piano"). They found anticipatory looks to the target picture well before target onset, suggesting that participants made predictions about the upcoming target noun. Note, however, that their experimental manipulation always included a 1:1 correspondence between the article and the following auditory target noun. The article was thus an extremely salient and reliable cue that univocally pointed to the upcoming target noun. Here, we report evidence for anticipatory language processing based on the gender of the article even though it is not necessarily a reliable cue toward the noun that followed it.

Our results, and those obtained by Huettig and Janse (2016) and others (e.g., Martin et al., 2017; Szewczyk & Schriefers, 2013; Van Berkum et al., 2005; Wicha, Moreno, & Kutas, 2003; Wicha et al., 2004) indicate that listeners use knowledge-based cues to predict upcoming words. As such, our results add to the recent debate about the role of predictions in language processing (e.g., Nieuwland et al., 2018). Importantly for our work, Kochari and Flecken (2019) and others (e.g., Guerra et al., 2018) reported evidence suggesting that listeners do not necessarily predict the gender of an upcoming noun based on knowledge-based (semantic) cues. Moreover, Huettig and Guerra (2019) recently showed that the prediction of a target noun based on the gender of a preceding article could be attenuated by factors such as shorter preview time and faster speech rate of the carrier sentence. Specifically, they only observed anticipatory looks toward a target object in situations where auditory targets were preceded by a sentence presented at a slow rate. For "normal" (faster) contextual speech rates, participants appeared to only predict the upcoming material if they had ample time to preview the potential targets (long preview: 4 s; short preview: 1 s), or if they were specifically instructed to make predictions. Huettig and Guerra (2019) take these findings to indicate that prediction is not a necessity during language processing. In our current experiment, we do find anticipatory looks toward the target picture based on the gender of the preceding article, both for slow *and* fast speech rates, showing that faster speech rates do not necessarily “eliminate” predictions all together. Taking the present results together with those reported in Huettig and Guerra (2019), we conclude that listeners are flexible in their use of cues and their weighting, a point which we return to below. Note that this is entirely in line with Huettig and Guerra’s (2019, p. 200) conclusion that prediction is “contingent on the situation the listener finds herself in”; from a cue integration perspective, we would argue that prediction is contingent on the *reliability and weighting of the available cues*.

As Kochari and Flecken (2019) mention, an undoubtedly important objective of future research will be to investigate the content and extent of lexical predictions in more detail. For the present research, concerning the integration of different types of cues, it was important to demonstrate that the knowledge-based cue of gender marking was indeed utilized by the participants in our experimental paradigm.

Signal-based, contextual speech rate effects persist even in the presence of preceding knowledge-based, morphosyntactic cues

We observed more looks toward the picture corresponding to a long vowel interpretation in the vowel window for items embedded in fast context sentences. Crucially, this effect arose *in spite of* preceding morphosyntactic cues, which participants demonstrably made use of to make predictions earlier on (see previous section). This result suggests that contextual speech rate acts as a salient cue for language processing. Moreover, we did not find evidence for a differential effect size of contextual speech rate on participants’ categorization decisions in the pretest (i.e., without preceding articles) versus eye-tracking experiment (with preceding articles). This absence of an interaction between rate effects and morphosyntactic constraints corroborates earlier work (Heffner et al., 2015; Morrill et al., 2015), together highlighting the automaticity of contextual rate effects.

In addition, the rate effect observed in the eye-tracking data arose very rapidly in time (around 200-250 ms after vowel offset; cf. Figure 3.3), which is about the earliest time point at which effects can be expected to emerge in eye-tracking data (Matin et al., 1993). This is in line with previous studies observing very early evidence for acoustic context effects in eye-tracking experiments (Reinisch & Sjerps, 2013; Toscano & McMurray, 2015). Thus, our results add to a growing body of literature showing that effects of contextual speech rate are robust and arise very early during perception (e.g., Bosker, 2017a; Maslowski et al., 2019a).

We interpret these findings with reference to the two-stage model of acoustic context effects, introduced in Bosker et al. (2017). In this model, acoustic context effects, including rate normalization, are suggested to arise at two distinct processing stages. The first stage encompasses early and automatic perceptual normalization processes, while a second stage involves later cognitive adjustments, for instance driven by indexical speech properties. The early time point of the rate effects and the robustness across individuals together suggest that the rate effects observed here arose at the first stage of contextual processing.

For the first time, we also report individual variation in weighting the signal-based and knowledge-based cues by mapping eyetracking onto behavioral results. We showed that even participants who had a clear behavioral preference for the knowledge-based cue for the most part still exhibited small rate effects in their eye fixations. In fact, all participants except for some individuals at the extreme end of the individual strategy scale showed an effect of rate.

These findings are difficult to integrate within models of speech comprehension that posit a stronger influence of syntactic, knowledge-based cues compared to signal-based cues such as contextual speech rate (e.g., Mattys et al., 2005). In fact, our observation that individual participants employed different behavioral strategies during the experiment challenges speech comprehension models that propose a fixed hierarchy of cues. Instead, our results suggest that cues can be weighted flexibly during the comprehension process, both on a group level and by individual participants. Language comprehension models of cue integration (Martin, 2016) offer a promising formalization that can accommodate these results.

Our data also speak to the question how the system handles uncertainty across time. We found that participants immediately used both the morphosyntactic and the acoustic cue as soon as they became available, rather than delaying looks to either of the pictures until all the information about the entire sentence was available, or disregarding one cue entirely. Instead, as suggested by cue integration frameworks, participants appeared to immediately combine the available cues in order to arrive at a robust percept.

Based on our findings, we can formulate new questions for future research. Most notably, the question arises which general factors determine the weighting and reliability of cues. Our results indicate that listeners weighted knowledge-based and signal-based cues flexibly, but what drives this cue weighting in the presence of uncertainty and across different experimental settings? It is possible that cue reliabilities are strongly modulated by exogenous, situational cues. For example, Martin (2016) suggested that nonlinguistic percepts such as gaze, facial expression, or joint-action contexts might modulate the reliability of certain linguistic cues in dialogue settings. Investigating these additional factors behind cue weighting and the interplay between cue reliabilities and their underlying modulators will be an exciting objective of future research.

Our experiment investigated two specific cues: gender information conveyed by a definite article, and contextual speech rate. It is unclear how specific our findings are to precisely these two cues, and whether a different, or more nuanced, picture might emerge for other combinations of cues. As such, caution should be taken when making claims about the integration and weighting of knowledge-based and signal-based cues in general. Rather, we believe that our results are a first step toward establishing a set of cues that the system can draw on during language processing, and how it can combine them (Martin, 2016). Further experiments could investigate different combinations of cues in more detail in order to observe whether similar effects arise.

Our findings are particularly interesting in light of results reported by Mattys et al. (2007). They found that attenuating effects of conflicting acoustic cues on the reliability of syntactic cues were contingent on the acoustic cue being realized *before* the syntactic one, suggesting that cue reliability and weighting can be modulated by the time course and order in which different pieces of information enter the system. For our current experiment, we would argue that the realization of the acoustic cue occurred *after* the morphosyntactic cue. Although the contextual rate information was available from the beginning of the sentence, it only became meaningful upon perception of the duration of the ambiguous vowel due to the continuum manipulation. This information always occurred after the article. In our experiment, the influence of acoustic cues on target perception were thus not contingent on their time course within the stimulus. We also showed that individual participants employed different strategies when weighting and integrating the acoustic and syntactic cues with each other – this is clear evidence against a strict hierarchy of cue weights. Further, our observation that small rate effects still arose for many participants with a clearly syntax-driven individual strategy also demonstrates that contextual speech rate is a robust acoustic cue that is not easily "overwritten" by conflicting syntactic information.

On a related note, Reinisch et al. (2011) conducted a series of experiments in which they investigated the use of distal and proximal contextual speech rate cues. While listeners generally appeared to rely more strongly on proximal than on distal context, the results also suggested that effects of distal speech rate grew stronger with the amount of context that listeners were presented with (i.e., "longer" contexts elicited more pronounced rate effects than "shorter" contexts). Reinisch et al. (2011) interpret this as a "cumulative effect." An interesting question for future research would be to investigate in more detail which modulators cause listeners to weigh certain cues more strongly than others.

Language comprehension usually takes place in settings that are more natural and flexible than our experimental setup. Further experiments could therefore investigate the interplay of knowledge-based and signal-based cues in a more naturalistic setting, for example during dialogue. Given that we find rate effects to be robust, even in the presence of disambiguating morphosyntactic information, it would be interesting to investigate to which extent they persist in situations that more closely resemble "real life" language use, where a lot more variability exists.

Taken together, our findings indicate that listeners rapidly extract and integrate both morphosyntactic, knowledge-based cues conveyed by a definite article and signal-based, acoustic cues conveyed by contextual speech rate. Rather than processing these cues separately in a strictly hierarchical fashion, listeners appear to take all available sources of information into account and update their beliefs about the incoming speech material depending on the reliability that they assign to the available cues.

Appendix

Stimuli. Spoken sentences were manipulated to include tokens of the noun that contained a vowel that was both spectrally and durationally ambiguous between /ɑ-a:/ (see Methods section for details). Sentences fell into one of two syntactic conditions, depending on the article (de/het). Furthermore, we introduced rate manipulations (slow vs. fast) in the preceding context, the adjectival phrase, and the outro (underlined).

(1) Kijk nu eens naar het/de ontzettend vieze as$_{COM}$/aas$_{NEU}$ alsjeblieft.
Look now once at the incredibly dirty ashes/bait please.

(2) Kijk nu eens naar het/de geweldig sjieke graf$_{COM}$/graaf$_{NEU}$ alsjeblieft.
Look now once at the immensely elegant tomb/count please.

(3) Kijk nu eens naar het/de ontzettend rode hart$_{COM}$/haard$_{NEU}$ alsjeblieft.
Look now once at the incredibly red heart/fireplace please.

(4) Kijk nu eens naar het/de geweldig sterke span$_{COM}$/spaan$_{NEU}$ alsjeblieft.
Look now once at the immensely strong yoke/spade please.

(5) Kijk nu eens naar het/de ontzettend bruine raam$_{COM}$/ram$_{NEU}$ alsjeblieft.
Look now once at the incredibly brown window/buck please.

(6) Kijk nu eens naar het/de ontzettend vuile vat$_{COM}$/vaat$_{NEU}$ alsjeblieft.
Look now once at the incredibly dirty barrel/dishes please.

(7) Kijk nu eens naar het/de geweldig grote rad$_{COM}$/raad$_{NEU}$ alsjeblieft.
Look now once at the immensely large wheel/council please.

Individual strategy score per participant. Individual strategy scores were calculated as the proportion of each participant's "long" responses after hearing an article biasing towards either a "long" or a "short" interpretation. Values closer to 1 indicate more "syntax-following" responses (i.e., responding with the long (short) vowel interpretation after hearing an article biasing toward that interpretation); values closer to 0 indicate more "acoustics-following" responses (i.e., responding with the long (short) vowel interpretation after hearing the ambiguous vowel embedded in a fast (slow) context).

Participant	Score	Group
28	0.03	Acoustics
10	0.03	Acoustics
7	0.05	Acoustics
19	0.06	Acoustics
21	0.10	Acoustics
29	0.11	Acoustics
16	0.11	Acoustics
5	0.13	Acoustics
13	0.13	Acoustics
20	0.14	Acoustics
17	0.15	Acoustics
8	0.15	Acoustics
34	0.16	Acoustics
15	0.16	Acoustics
24	0.17	Acoustics
2	0.17	Acoustics
36	0.17	Acoustics
25	0.19	Acoustics
35	0.19	Acoustics
33	0.22	Acoustics
6	0.42	Acoustics
8	0.44	Syntax
3	0.47	Syntax
11	0.49	Syntax
1	0.71	Syntax
22	0.74	Syntax
32	0.84	Syntax
30	0.86	Syntax
27	0.91	Syntax
12	0.91	Syntax
26	0.92	Syntax
4	0.93	Syntax
23	0.95	Syntax
14	0.97	Syntax
31	0.97	Syntax
18	1.0	Syntax

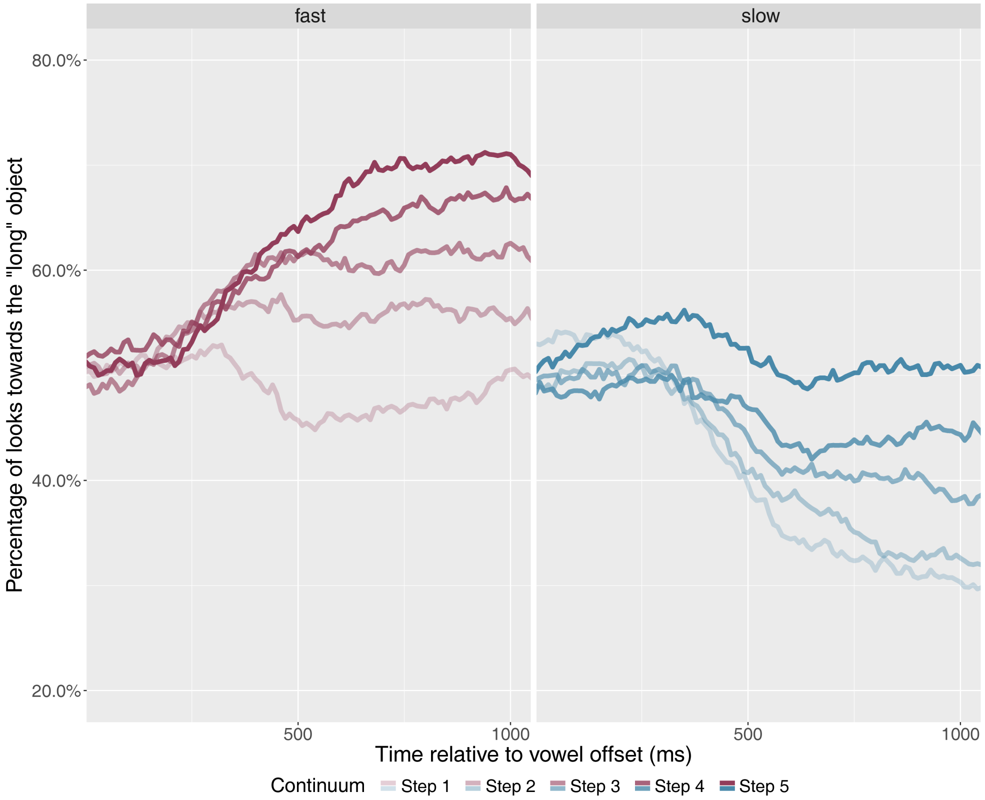

Figure 3A.1: Percentage of looks to the object corresponding to a long vowel interpretation across time for each vowel continuum step, split by fast (left panel) vs. slow (right panel) context rate. Time point 0 marks the offset of the manipulated vowel.

4 Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy[1]

Abstract

Neural oscillations track linguistic information during speech comprehension (e.g., Ding et al., 2016; Keitel et al., 2018), and are known to be modulated by acoustic landmarks and speech intelligibility (e.g., Doelling, Arnal, Ghitza, & Poeppel, 2014; Zoefel & VanRullen, 2015). However, studies investigating linguistic tracking have either relied on non-naturalistic isochronous stimuli or failed to fully control for prosody. Therefore, it is still unclear whether low frequency activity tracks linguistic structure during natural speech, where linguistic structure does not follow such a palpable temporal pattern. Here, we measured electroencephalography (EEG) and manipulated the presence of semantic and syntactic information apart from the timescale of their occurrence, while carefully controlling for the acoustic-prosodic and lexical-semantic information in the signal. EEG was recorded while 29 adult native speakers listened to naturally-spoken Dutch sentences, jabberwocky controls with morphemes and sentential prosody, word lists with lexical content but no phrase structure, and backwards acoustically-matched controls. Mutual information (MI) analysis revealed sensitivity to linguistic content: MI was highest for sentences at the phrasal (0.8-1.1 Hz) and lexical timescale (1.9-2.8 Hz), suggesting that delta-band activity is modulated by lexically-driven combinatorial processing beyond prosody, and that linguistic content (i.e., structure and meaning) organizes neural oscillations beyond the timescale and rhythmicity of the stimulus. This pattern is consistent with neurophysiologically inspired models of language comprehension (Martin, 2016; Martin & Doumas, 2017) where oscillations encode endogenously generated linguistic content over and above exogenous or stimulus-driven timing and rhythm information.

[1] Adapted from Kaufeld, G., Bosker, H. R., ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience,* Advance online publication.

4.1 Introduction

How the brain maps the acoustics of speech onto abstract structure and meaning during spoken language comprehension remains a core question across cognitive science and neuroscience. A large body of research has shown that neural populations closely track the envelope of the speech signal, which correlates with the syllable rate (e.g., Kösem et al., 2018; Peelle & Davis, 2012; Zoefel & VanRullen, 2015), yet much less is known about the degree to which neural responses encode higher-level linguistic information such as words, phrases and clauses. While previous studies suggest a crucial role for delta-band oscillations in the top-down generation of hierarchically structured linguistic representations (e.g., Ding et al., 2016; Keitel et al., 2018), they have so far either relied on non-naturalistic stimuli or failed to fully control for prosody. Here, we employ a novel experimental design that allows us to investigate how structure and meaning shape the tracking of higher-level linguistic units, while using naturalistic stimuli and carefully controlling for prosodic fluctuations.

The strongest evidence for tracking of linguistic information so far are studies by Ding et al. (Ding, Melloni, et al., 2017; Ding et al., 2016), who found enhanced activity in the delta frequency range for sentences compared to word lists. They investigated this using isochronous, synthesized stimuli devoid of prosodic information. Yet phrases, clauses, and sentences usually do have acoustic-prosodic correlates (e.g., pauses, intonational contours, final lengthening, fundamental frequency reset; cf. Eisner and McQueen, 2018). These might not be as prominent in the modulation spectrum of speech as syllables (Ding, Patel, et al., 2017), but listeners draw on them during language comprehension and learning (e.g., Soderstrom, Seidl, Nelson, & Jusczyk, 2003). As such, Ding et al. cannot clearly distinguish between the generation of linguistic structure and meaning vs. inferred prosody, and it is unclear whether their results generalize to naturalistic stimuli, where the timing of linguistic units is more variable.

Almost orthogonally to Ding et al. (2017; 2016), Keitel et al. (2018) used naturalistic stimuli and found enhanced tracking (compared to reversed controls) in the delta-theta frequency range. However, as they did not include a systematic control for linguistic content, it is unclear whether their results are driven by tracking of prosodic information in the acoustic signal, rather than linguistic information.

In the current study, we bridge this gap by contrasting these two core sources of linguistic representations: prosodic structure, which can, but does not always, correlate with syntactic and information structure, and lexical semantics, which arises in isolated words and concepts. Participants listened to naturally spoken, structurally homogenous sentences, jabberwocky items (containing sentence-like prosody, but no lexical semantics), and word lists (containing lexical semantics, but no sentence-like structure and prosody; see Table 4.1 for examples). Additionally, we used reversed speech as the core control of our experiment, because it has an identical modulation spectrum for each forward condition.

Using electroencephalography (EEG), we analyzed tracking at linguistically relevant timescales as quantified by Mutual Information (MI) – a typical measure of neural tracking that captures the informational similarity between two signals (Cogan & Poeppel, 2011; Gross et al., 2013; Kayser, Ince, Gross, & Kayser, 2015; Keitel et al., 2018; Keitel, Ince, Gross, & Kayser, 2017). Figure 4.1 shows an overview of the experimental design and analysis pipeline.

We hypothesize that neural tracking ("entrainment in the broad sense", as defined by Obleser & Kayser, 2019) will be stronger for stimuli containing higher-level linguistic structure and meaning, above and beyond the acoustic-prosodic (jabberwocky) and lexical-semantic (word list) controls. This may reflect a process of perceptual inference (Martin, 2016, 2020), whereby biological systems like the brain encode their environment not only by reacting in a series of stimulus-driven responses, but by combining stimulus-driven information with endogenous, internally-generated, inferential knowledge and meaning (Meyer, Sun, & Martin, 2019). In sum, our study offers novel insights into how structure and meaning influence the neural response to natural speech *above and beyond* prosodic modulations and word-level meaning.

4.2 Materials and Methods

Participants

35 native Dutch speakers (26 females, 9 males; age range 19-32; $M_{age} = 23$) participated in the experiment. They were recruited from the Max Planck Institute for Psycholinguistics' participant database with written consent approved by the Ethics Committee of the Social Sciences Department of Radboud University (Project code: ECSW2014-1003-196a). All participants in the experiment reported normal hearing and were remunerated for their participation. Six participants were excluded from the analysis due to excessive artifact contamination, leaving us with $N = 29$.

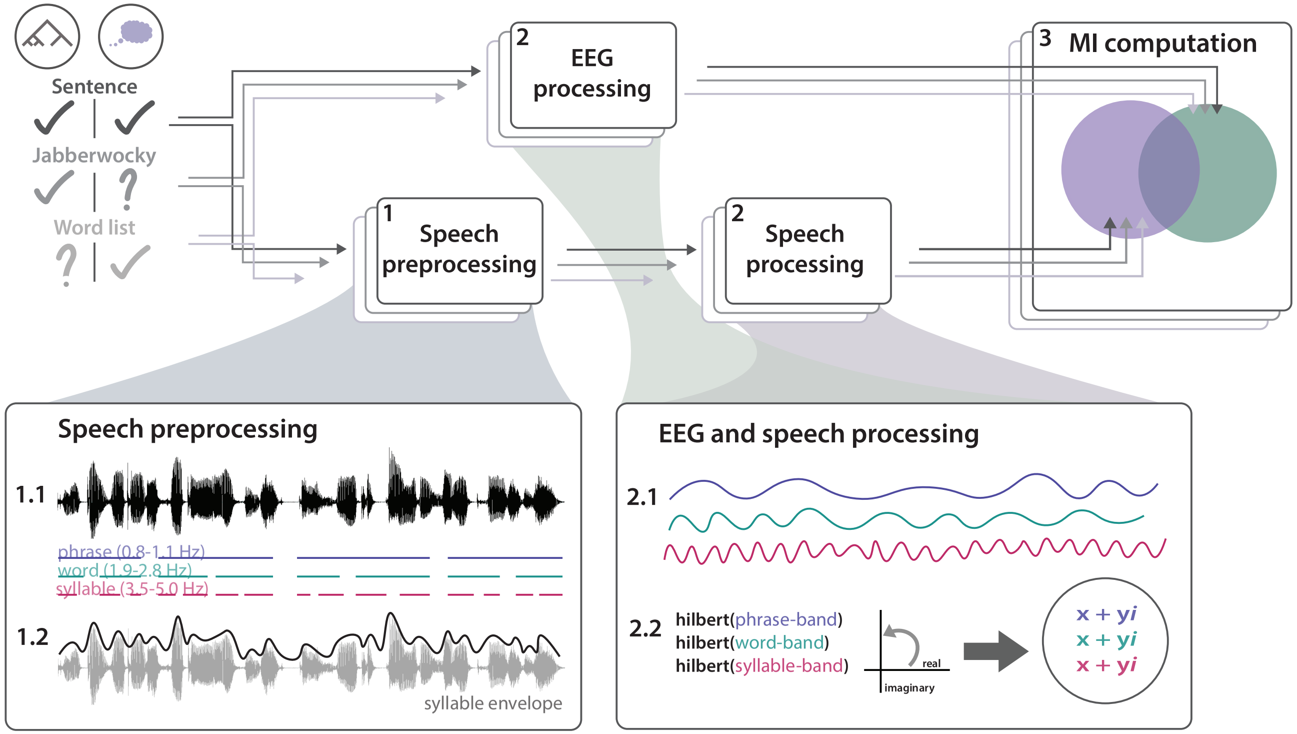

Figure 4.1: Experimental design and analysis pipeline. Participants listened to sentences, jabberwocky items, and word lists while their brain response was recorded using EEG. Step 1: Speech Processing. 1.1) The speech signal is annotated for the occurrence of phrases, words, and syllables in the stimuli. Based on this, frequency bands of interest for each of the linguistic units can be identified. 1.2) A cochlear filter is applied to the speech stimuli and the amplitude envelope is extracted. Step 2: Further processing is identical for both speech and EEG modalities. 2.1) Broadband filters are applied in the previously identified frequency bands of interest. 2.2) Hilbert transforms are computed in each filtered signal, and real and imaginary parts of the Hilbert transform output are used for further analysis. Step 3: MI Computation. Mutual information is computed between the pre-processed speech and EEG signal in each of the three conditions and their respective backward controls.

Materials

The experiment used three conditions: Sentence, Jabberwocky, and Wordlist. Eighty sets (triplets) of the three conditions (Sentence, Jabberwocky, Wordlist) were created, resulting in 240 stimuli. In addition to one "standard" forward presentation of each stimulus, participants also listened to a version of each of the stimuli played backwards, thus resulting in a total of 480 stimuli.

Dutch stimuli consisted of 10 words, which were all disyllabic except for "de" (*the*) and "en" (*and*), thus resulting in 18 syllables in total. Sentences all consisted of two coordinate clauses, which followed the structure *[Adj N V N Conj Det Adj N V N]*. Word lists consisted of the same 10 words as in the Sentence condition, but scrambled in syntactically implausible ways (either *[V V Adj Adj Det Conj N N N N]*, or *[N N N N Det Conj V V Adj Adj]*, in order to avoid any plausible internal combinations of words). Jabberwocky items were created using the wuggy pseudoword generator (Keuleers & Brysbaert, 2010), following the same syntactic structure as the sentences. Specifically, standard wuggy parameters were set to match 2 out of 3 subsyllabic segments wherever possible, as well as letter length, transition frequencies, and length of subsyllabic segments. Wuggy's lexicality feature was used to ensure that none of the generated pseudowords were existing lexical items in Dutch. In addition, all pseudowords were proof-read by native Dutch speakers in order to ensure that none of their phonetic forms matched that of an existing word in Dutch. Inflectional morphemes (e.g., plural morphemes) as well as function words ("de" - *the* and "en" - *and*) were kept unchanged. Table 4.1 shows an example of stimuli in each condition. (See the Appendix for a list of all 480 stimuli and their translations.)

Forward stimuli were recorded by a female native speaker of Dutch in a sound-attenuating recording booth. All stimuli were recorded at a sampling rate of 44.1 kHz (mono), using the Audacity sound recording and analysis software (Audacity Team, 2019). After recording, pauses were normalized to ∼150 ms in all stimuli, and the intensity was scaled to 70 dB using the Praat voice analysis software (Boersma & Weenink, 2020). Stimuli from all three conditions were then reversed using Praat. Figure 4.2 shows modulation spectra for forward and backward conditions.

Procedure

Participants were tested individually in a sound-attenuating and Faraday-cage enclosed booth. They first completed a practice session with 4 trials (one from

Sentence	Jabberwocky	Wordlist
[Bange helden] [plukken bloemen] en de [bruine vogels] [halen takken].	[Garge ralden] [spunken drijmen] en de [druize gomels] [paven mukken].	[helden bloemen] [vogels takken] de en [plukken halen] [bange bruine]
[Timid heroes] [pluck [flowers] and the [brown birds] [gather branches].	*[Flimid lerops] [bruck clowters] and the [trown plirds] [shmather blamches].*	*[heroes flowers] [birds branches] the and [pluck gather] [timid brown]*

Table 4.1: Example items in Sentence, Jabberwocky, and Wordlist conditions. Sentences consisted of 10 words (disyllabic, except for “de” (“the”) and “en” (“and”)) and carried sentence prosody. Jabberwocky items consisted of 10 pseudo-words with morphology; they also carried sentence-like prosody. Word lists consisted of the same 10 words as the Sentence condition, but scrambled so as to be syntactically implausible. They had list-prosody. Marked with square brackets are “phrases” in all three conditions. Note that the (pseudo-)words in all three conditions had the same stress patterns. See Appendix for a complete list of all stimuli.

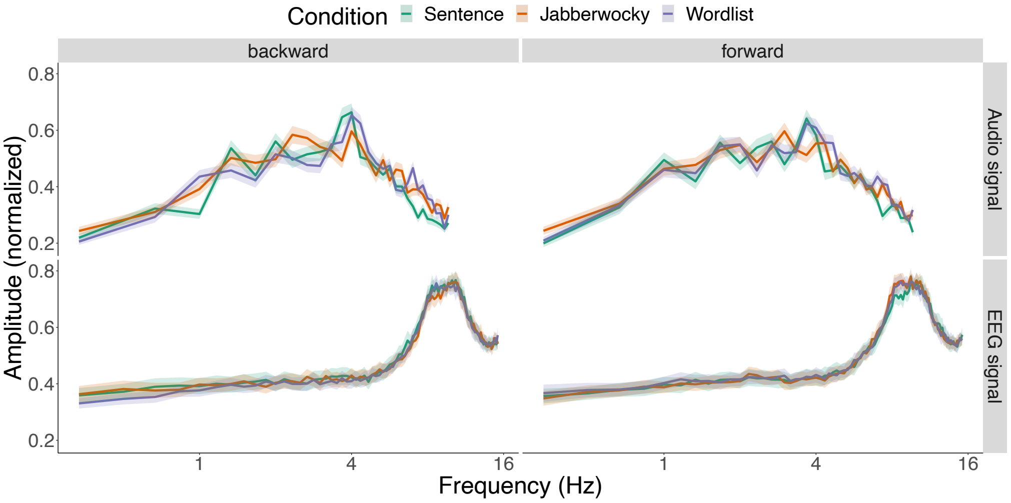

Figure 4.2: Modulation spectra of forward and backward stimuli. Green: Sentence; Orange: Jabberwocky; Purple: Wordlist. Modulation spectra were calculated following the procedure and Matlab script described in Ding, Patel, et al. (2017). Note that a cochlear filter is applied to the acoustic stimuli, but not the brain data. Small deviations between the modulation spectrum of each forward condition and its backward counterpart are due to numerical inaccuracy; mathematically, the frequency components of forward and backward stimuli are identical.

each forward condition and one backward example) to become familiarized with the experiment. All 80 stimuli from each condition were presented to the participants in separate blocks. The order of the blocks was pseudo-randomized across listeners, and the order of the items within each block was randomized. During each trial, participants were instructed to look at a fixation cross which was displayed at the center of the screen (to minimize eye movements during the trial), and listen to the audio, which was presented to them at a comfortable level of loudness. The audio recording was presented 500 ms after the fixation cross appeared on the screen, and the fixation cross remained on the screen for the entire duration of the audio recording. Fifty ms after the end of each recording, the screen changed to a transition screen (a series of hash symbols (#####) indicating that participants could blink and briefly rest their eyes), after which participants could advance to the next item via a button-press. After each block, participants were allowed to take a self-paced break. The experiment was run using the Presentation software (Neurobehavioral Systems) and took about 50-60 minutes to complete. EEG was continuously recorded with a 64-channel EEG system (MPI equidistant montage) using BrainVision Recorder software, digitized at a sampling rate of 500 Hz and referenced to the left mastoid. The time constant for the hardware high-pass filter was 10s (0.016 Hz; first-order Butterworth filter with 6 dB/octave), the high-cutoff frequency was 249 Hz. The impedance of electrodes was kept below 25 kΩ. Data was re-referenced offline to the average reference.

EEG data preprocessing

The analysis steps were carried out using the FieldTrip analysis toolkit revision 20180320 (Oostenveld, Fries, Maris, & Schoffelen, 2011) on MATLAB version 2016a (MathWorks, Inc.). The raw EEG signal was segmented into a series of variable length epochs, starting at 200 ms before the onset of the utterance and lasting until 200 ms after its end. The signal was low-pass filtered to 70 Hz, and a band-stop filter centered around 50 Hz (±2 Hz) was applied in each epoch to exclude line noise (both zero-phase FIR filters using Hamming windows). Channels contaminated with excessive noise were excluded from the analysis. Independent component analysis was performed on the remaining channels, and components related to eye movements, blinking, or motion artifacts, were subtracted from the signal. Epochs containing voltage fluctuations exceeding ±100 μV or exceeding a range of 150 μV were excluded from further analysis. We selected a cluster of 22 electrodes for all further analyses based on previous studies that found broadly-distributed effects related to sentence processing (e.g., Kutas, Van Petten, and Kluender, 2006; Kutas and Federmeier, 2000; see also Ding, Melloni, et al., 2017). Specifically, the electrode selection included the following electrodes: 1, 2, 3, 4, 5, 8, 9, 10, 11, 28, 29, 30, 31, 33, 34, 35, 36, 37, 40, 41, 42, 43 (electrode names based on the MPI equidistant layout). We note that our results also hold for all electrodes, as described in the *Results* section below.

Speech preprocessing

For each stimulus, we computed the wideband speech envelope at a sampling rate of 150 Hz following the procedure reported by Keitel et al. (2018) and others (e.g., Gross et al., 2013; Keitel et al., 2017). We first filtered the acoustic waveforms into 8 frequency bands (100-8,000 Hz; third-order Butterworth filter, forward and reverse), equidistant on the cochlear frequency map (Smith, Delgutte, & Oxenham, 2002). We then estimated the wide-band speech envelope by computing the magnitude of the Hilbert transformed signal in each band and averaging across bands.

The timescales of interest for further Mutual Information analysis were identified in a similar fashion as described in Keitel et al. (2018). We first annotated the occurrence of linguistic units (phrases, words, and syllables) in the speech stimuli. Here, phrases were defined as adjective-noun/noun-verb combinations (e.g., in the Sentence condition: “bange helden” – *timid heroes*; “plukken bloemen” – *pluck flowers*, and so on; in the Jabberwocky condition: “garge ralden” – *flimid lerops* etc.; in the Wordlist condition, a “pseudo-phrase” corresponds to adjacent noun-noun, verb-verb and adjective-adjective pairs, e.g., “helden bloemen” – *heroes flowers)*. Unit-specific bands of interest were then identified by converting each of the rates into frequency ranges across conditions. This resulted in the following bands: 0.8-1.1 Hz (phrases); 1.9-2.8 Hz (words); and 3.5-5.0 Hz (syllables). Note that the problem the brain faces during spoken language comprehension is even more complex than this, because the timescales of linguistic units can highly overlap, even within a single sentence (Obleser, Herrmann, & Henry, 2012). Populations of neurons that “entrain” to words will thus also have to be sensitive to information that occurs outside of these – rather narrow – frequency bands.

For an additional, exploratory annotation-based MI analysis (section *Results*, subsection *Tracking of abstract linguistic units*), we further created linguistically abstracted versions of our stimuli. Specifically, our aim was to create annotations that captured linguistic information at the phrase frequency entirely independent of the acoustic signal. Based on the word-level annotations of our stimuli, we created dimensionality-reduced arrays for further analysis (cf. the “Semantic composition” analyses reported by Brodbeck, Presacco, and Simon (2018)).

Specifically, we identified all time points in the spoken materials where words could be integrated into phrases, and marked each of these words associated with phrase composition (e.g., in a sentence such as "bange helden plukken bloemen en de bruine vogels halen takken" (*timid heroes pluck flowers and the brown birds gather branches*), the words "helden" (*heroes*), "bloemen" (*flowers*), "vogels" (*birds*) and "takken" (*branches*) were marked). All these critical words were coded as 1 for their entire duration, while all other time points (samples) were marked as 0 (cf. Brodbeck et al., 2018). This resulted in an abstract "spike train" array of phrase-level structure building that is independent of the acoustic envelope. We repeated this procedure for all items individually in all three conditions, since our stimuli were naturally spoken and thus differed slightly in duration and time course. Note that, consequently, this "phrase-level composition array" is somewhat arbitrary for the Wordlist condition as there are, per definition, no phrases in a word list. We annotated "pseudo-phrases" the same way as shown in Table 4.1. The procedure is visualized in Figure 4.3.

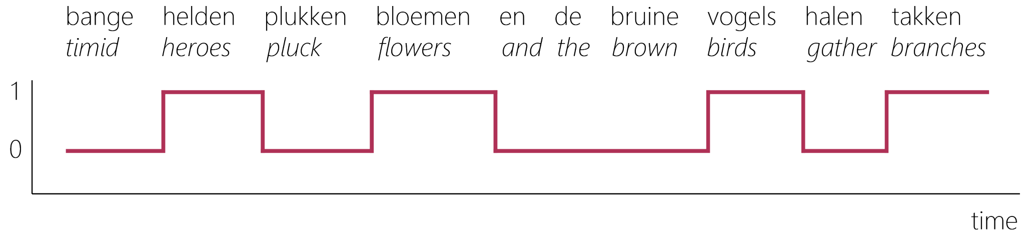

Figure 4.3: Visualization of the phrase-level annotations (inspired by Fig.2 in Brodbeck et al., 2018). Across time, the response array takes value 0 for words that cannot (yet) be integrated into phrases, and value 1 for words that can, resulting in a "pulse train" array.

Mutual Information analysis

We used Mutual Information (MI) in order to quantify the statistical dependency between the speech envelopes and the EEG recordings according to the procedure described in Keitel et al. (2018; see also Gross et al., 2013; Kayser et al., 2015; Keitel et al., 2017). Based on the previously identified frequency bands of interest (see subsection *Speech preprocessing* above), we filtered both speech envelopes and EEG signals in each band (third-order Butterworth filter, forward and reverse). We then computed the Hilbert transform in each band, which resulted in two sets of two-dimensional variables (one for speech signals and one for EEG responses) in each condition (forward and backward; see Ince et al., 2017, for a more in-depth description). To take brain-stimulus lag into account, we computed MI at 5 different lags, ranging from 60 to 140 ms in steps of 20 ms, and to exclude strong auditory evoked responses to the onset of auditory stimulation in each trial we excluded the first 200 ms of each stimulus-signal pair. MI values from all five lags were averaged for subsequent statistical evaluation. We further concatenated all trials from speech and brain signals in order to increase the robustness of MI computation (Keitel et al., 2018). In addition to computing "general" MI (containing information about both phase and power), we also isolated the part of the Hilbert transform corresponding to phase and computed "phase MI" values, separately.

Statistical analysis

In order to test whether the statistical dependency between the speech envelope and the EEG data as captured by MI was modulated by the linguistic structure and content of the stimulus, we compared MI values in all three frequency bands separately. Linear mixed models were fitted to the log-transformed, trimmed (5% on each end of the distribution) MI values in each frequency band using lme4 (D. Bates et al., 2015) in R (R Core Development Team, 2012). Models included main effects of Condition (three levels: Sentence, Wordlist, Jabberwocky) and Direction (two levels: Forward, Backward), as well as their interaction. All models included by-participant random intercepts and random slopes for the Condition*Direction interaction. For model coefficients, degrees of freedom were approximated using Satterthwaite's method as implemented in the package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). We used treatment coding in all models, with Sentence being the reference level for Condition, and Forward the reference level for Direction. We then computed all pairwise comparisons within each direction using estimated marginal means (Tukey correction for multiple comparisons) with emmeans (Length, Singmann, Love, Buerkner, & Herve, 2018) in R (i.e., comparing Sentence Forward to Jabberwocky Forward and Wordlist Forward, but never Sentence Forward to Jabberwocky Backward, because we had no hypotheses about these comparisons). The same statistical analyses, including identical model structures, were further applied to MI values computed on the isolated phase coefficients.

For the exploratory dimensionality-reduced MI analysis, we performed the same set of statistical analyses (but only in one single frequency band). Specifically, we fitted a linear mixed model including main effects of Condition (three levels: Sentence, Wordlist, Jabberwocky) and Direction (two levels: Forward, Backward), as well as their interaction and by-participant random intercepts and random slopes for the Condition*Direction interaction to the log-transformed, trimmed MI values. We then computed estimated marginal means precisely as described in the previous section.

4.3 Results

Speech tracking

We computed Mutual Information (MI) between the Hilbert-transformed EEG time series and the Hilbert-transformed speech envelopes within three frequency bands of interest that corresponded to the occurrence rates of phrases (0.8-1.1 Hz), words (1.9-2.8 Hz), and syllables (3.5-5.0 Hz) in a cluster of central electrodes.

Specifically, we designed our experiment to assess whether the brain response is driven by the (quasi-)periodic temporal occurrence of linguistic structures and prosody, or whether it is modulated as a function of the linguistic content of those structures. Using MI allowed us to quantify and compare the degree of speech tracking across sentences, word lists, and jabberwocky items.

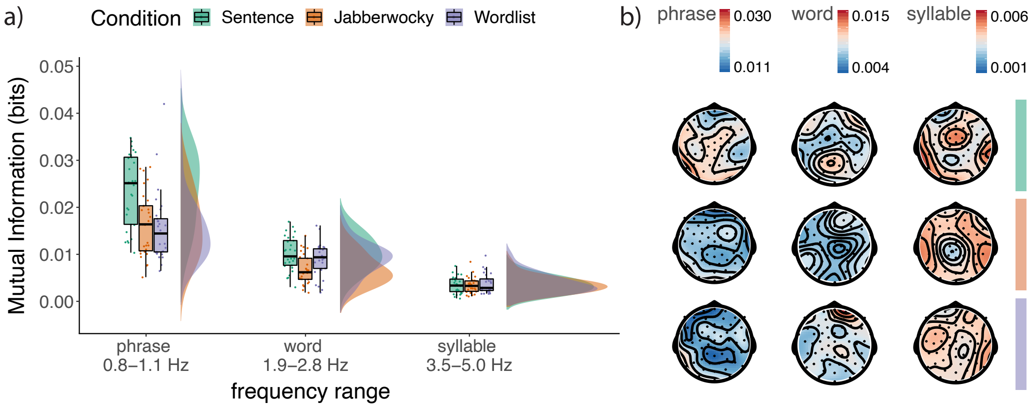

Figure 4.4: MI between speech signals and brain responses. Panel a) shows MI for Sentence (green), Jabberwocky (orange), and Wordlist items (purple) for phrase, word and syllable timescales across central electrodes (each dot represents one participant's mean MI response averaged across electrodes). Panel b) shows the average scalp distribution of MI per condition and band, averaged across participants. Raincloud plots were made using the Raincloud package in R (Allen et al., 2019).

Our analyses revealed condition-dependent enhanced MI at distinct timescales for the forward conditions (see Figure 4.4). In the phrase frequency band (0.8-1.1 Hz), the mixed effects model revealed a significant effect of Condition (Sentence = treatment level; Jabberwocky: β = -0.452, SE = 0.096, $p < 0.001$; Wordlist: β = -0.491, SE = 0.116, $p < 0.001$) and Direction (Forward = treatment level; Backward: β = -0.885, SE = 0.117, $p < 0.001$), as well as Condition*Direction interactions (Jabberwocky*Backward: β = 0.429, SE = 0.152, $p = 0.008$; Wordlist*Backward: β = 0.523, SE = 0.185, $p = 0.009$). The estimated marginal means corroborated these results, revealing significant pairwise effects only between the Forward conditions (Sentence – Jabberwocky: Δ = 0.452, SE = 0.098, $p < 0.001$; Sentence – Wordlist: Δ = 0.491, SE = 0.118, $p < 0.001$; all results Tukey corrected for multiple comparisons), but not the backward controls. The observation that none of the effects was present in the backward speech controls demonstrates that they were not driven by the acoustic properties of the stimuli (see Tables 4.2 and 4A.1 (in the Appendix) for complete model outputs).

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.45	0.10	30.0	4.61	< 0.01
Sentence - Wordlist	0.49	0.12	30.0	4.17	< 0.01
Jabberwocky - Wordlist	0.04	0.10	30.1	0.38	0.93
Direction = Backward					
Sentence - Jabberwocky	0.02	0.11	30.0	0.20	0.98
Sentence - Wordlist	-0.03	0.14	30.0	-0.23	0.97
Jabberwocky - Wordlist	-0.06	0.14	30.0	-0.40	0.92

Table 4.2: Estimated marginal means for MI between speech signals and brain responses in the phrase frequency band (0.8-1.1 Hz). P-value adjustment: tukey method for comparing a family of 3 estimates.

In the word frequency band (1.9-2.8 Hz), the mixed effects model revealed a significant effect of Condition (Sentence = treatment level; Jabberwocky: β = -0.484, SE = 0.121, $p < 0.001$) and Direction (Forward = treatment level; Backward: β = -0.499, SE = 0.136, $p < 0.001$). The pair-wise contrasts further revealed that this Sentence – Jabberwocky difference was only significant for the forward conditions (Δ = 0.484, SE = 0.123, $p = 0.001$), not for the backward controls. Again, this finding indicates that the differences we observed were not driven by differences in the acoustic signals, themselves (see Tables 4.3 and 4A.2 for complete model outputs).

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.48	0.12	30.0	3.94	< 0.01
Sentence - Wordlist	0.16	0.08	29.9	1.96	0.14
Jabberwocky - Wordlist	-0.33	0.14	30.0	-2.40	0.06
Direction = Backward					
Sentence - Jabberwocky	0.25	0.11	30.1	2.31	0.07
Sentence - Wordlist	0.08	0.12	30.1	0.62	0.81
Jabberwocky - Wordlist	-0.18	0.10	30.0	-1.72	0.21

Table 4.3: Estimated marginal means for MI between speech signals and brain responses in the word frequency band (1.9-2.8 Hz). P-value adjustment: tukey method for comparing a family of 3 estimates.

In the syllable frequency range (3.5-5.0 Hz), the mixed effects model revealed no significant effects of Condition or Direction, and no interaction between the two (see Table 4A.3 in the Appendix for complete model output).

Taken together, these findings indicate that neural tracking is enhanced for linguistic structures at timescales specific to that structure's role in the unfolding meaning of the sentence, consistent with neurophysiologically inspired models of language comprehension (Martin, 2016, 2020; Martin & Doumas, 2017).

An almost identical pattern of results emerged when computing MI over all electrodes (rather than a cluster of central ones). In the phrase frequency range, the mixed effects model revealed significant effects of Condition (Jabberwocky: β = -0.401, SE = 0.075, p < 0.001; Wordlist: β = -0.418, SE = 0.088, p < 0.001) and Direction (Backward: β = -0.743, SE = 0.087, p < 0.001), as well as significant Condition*Direction interactions (Jabberwocky*Backward: β = 0.296, SE = 0.099, p = 0.006; Wordlist*Backward: β = 0.332, SE = 0.134, p = 0.019).

In the word frequency range, the model revealed significant effects of Condition (Jabberwocky: β = -0.407, SE = 0.093, p < 0.001; Wordlist: β = -0.179, SE = 0.052, p = 0.002) and Direction (β = -0.316, SE = 0.090, p = 0.002), but not their interaction. For the forward conditions, the pair-wise comparisons further confirmed significantly higher MI for sentences compared to jabberwocky items (Sentence Forward – Jabberwocky Forward: Δ = 0.407, SE = 0.095, p < 0.001) and sentences compared to word lists (Sentence Forward – Wordlist Forward: Δ = 0.179, SE = 0.053, p = 0.006). Surprisingly, we also found significantly enhanced MI for sentences compared to jabberwocky items in the backward conditions in the word frequency (Sentence Backward – Jabberwocky Backward: Δ = 0.288, SE = 0.083, p = 0.005), so we cannot exclude the possibility that this effect is driven to some extent by differences in the acoustic signal.

Note, however, that the estimate of this effect is smaller for the backward than the forward differences.

Again, there were no significant effects in the syllable frequency range when computing MI over all electrodes. (See Appendix for complete model outputs.)

Phase MI

When computing MI on the isolated phase values from the Hilbert transform, we again found condition-dependent differences at distinct timescales (see Figure 4.5).

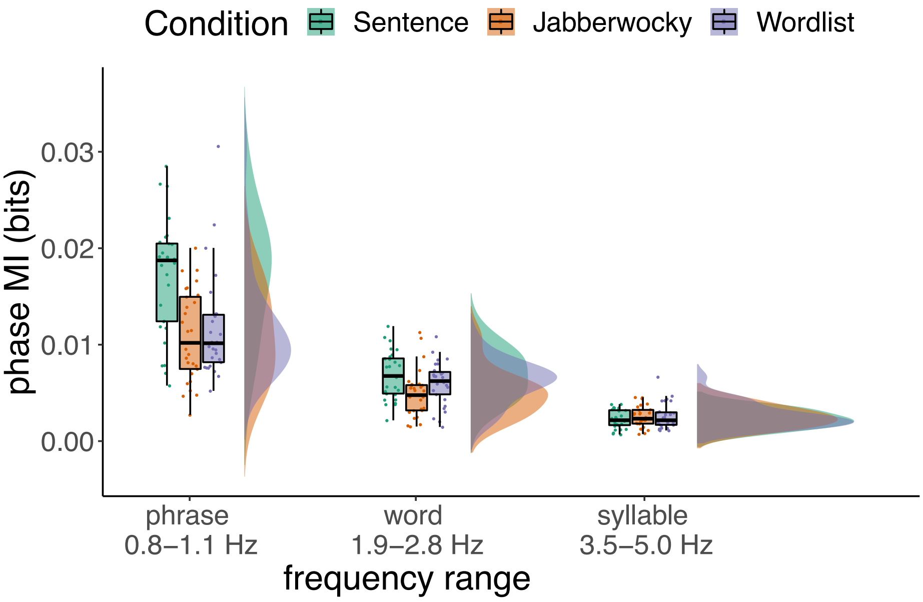

Figure 4.5: MI between the isolated phase of speech signals and brain responses. Sentences (green), Jabberwocky items (orange), and Word lists (purple) for phrase, word and syllable timescales across central electrodes (each dot represents one participant's mean MI response averaged across electrodes).

In the phrase frequency band (0.8-1.1 Hz), the models revealed significant effects of Condition (Sentence = treatment level; Jabberwocky: β = -0.497, SE = 0.097, $p < 0.001$; Wordlist: β = -0.402, SE = 0.118, $p = 0.002$) and Direction (Forward = treatment level; Backward: β = -0.805, SE = 0.106, $p < 0.001$), as well as their interaction (Jabberwocky*Backward: β = 0.368, SE = 0.150, p = 0.020). For the forward conditions, the pair-wise contrasts further corroborated these results, with sentences eliciting higher phase MI than jabberwocky items (Sentence Forward – Jabberwocky Forward: $\Delta = 0.497$, SE = 0.099, $p = 0.001$) and sentences eliciting higher phase MI than word lists (Sentence Forward –

Wordlist Forward: $\Delta = 0.402$, SE = 0.120, $p = 0.006$; again, all results Tukey corrected for multiple comparisons).

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.50	0.10	30.0	5.05	< 0.01
Sentence - Wordlist	0.40	0.12	30.0	3.36	< 0.01
Jabberwocky - Wordlist	-0.10	0.10	30.1	-0.93	0.63
Direction = Backward					
Sentence - Jabberwocky	0.13	0.12	30.0	1.06	0.55
Sentence - Wordlist	0.07	0.13	30.0	0.51	0.87
Jabberwocky - Wordlist	-0.06	0.14	30.0	-0.47	0.89

Table 4.4: Estimated marginal means for phase MI in the phrase frequency band. P-value adjustment: tukey method for comparing a family of 3 estimates.

In the word frequency band (1.9-2.8 Hz), the mixed effects model revealed a significant effect of Condition (Sentence = treatment level; Jabberwocky: β = -0.380, SE = 0.121, $p = 0.004$) and Direction (Forward = treatment level; Backward: β = -0.474, SE = 0.126, $p < 0.001$). The pair-wise contrasts further revealed significantly higher MI for forward sentences compared to forward jabberwocky items (Sentence Forward – Jabberwocky Forward: $\Delta = 0.380$, SE = 0.123, $p = 0.012$), but not their backwards controls. Again, this result demonstrates that the effect is not driven by the acoustic properties of the stimuli (see section *Materials and Methods* for details about our mixed effects models structures; see Appendix for detailed model outputs).

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.38	0.12	30.1	3.09	0.01
Sentence - Wordlist	0.12	0.07	29.7	1.63	0.25
Jabberwocky - Wordlist	-0.26	0.14	30.1	-1.86	0.17
Direction = Backward					
Sentence - Jabberwocky	0.21	0.12	30.0	1.73	0.21
Sentence - Wordlist	0.06	0.12	30.0	0.54	0.85
Jabberwocky - Wordlist	-0.14	0.10	30.0	-1.42	0.35

Table 4.5: Estimated marginal means for phase MI in the word frequency band. P-value adjustment: tukey method for comparing a family of 3 estimates.

Computing phase MI over all electrodes (rather than a cluster of central ones) revealed a similar pattern of results (see Tables 4A.12-4A.16). In the phrase frequency range, the mixed model revealed significant effects of Condition (Sentence = treatment level; Jabberwocky: β = -0.356, SE = 0.075, $p < 0.001$;

Wordlist: β = -0.309, SE = 0.089, p = 0.002), Direction (Forward = treatment level; Backward: β = -0.662, SE = 0.076, p < 0.001) and their interaction (Jabberwocky*Backward: β = 0.185, SE = 0.089, p = 0.047.) The estimated marginal means showed significant pair-wise comparisons only in forward conditions, with forward sentences showing higher phase MI than forward jabberwocky items and forward word lists (Sentence Forward – Jabberwocky Forward: Δ = 0.356, SE = 0.076, p < 0.001; Sentence Forward – Wordlist Forward: Δ = 0.309, SE = 0.091, p = 0.005) and no significant effects for the backward comparisons.

In the word frequency band, the mixed effects model revealed significant effects of Condition (Sentence = treatment level; Jabberwocky: β = -0.329, SE = 0.089, p < 0.001; Wordlist: β = -0.139, SE = 0.045, p = 0.005) and Direction (Forward = treatment level; Backward: β = -0.351, SE = 0.091, p < 0.001). The estimated marginal means further corroborated this finding only in the forward conditions (Sentence Forward – Jabberwocky Forward: Δ = 0.329, SE = 0.091, p = 0.003; Sentence Forward – Wordlist Forward: Δ = 0.139, SE = 0.046, p = 0.014). In contrast to the "general" MI values, we found no significant differences between the backward controls when computing the isolated phase MI over the entire head. Again, these findings are consistent with neurophysiologically inspired models of language comprehension (Martin, 2016, 2020; Martin & Doumas, 2017).

Tracking of abstract linguistic units

Inspecting the modulation spectra of our stimuli (Figure 4.2), it is apparent that – although carefully designed – the acoustic signals are not entirely indistinguishable between conditions based on their spectral properties. Most notably, Sentence stimuli appear to exhibit a small peak at around 0.5 Hz (roughly corresponding to the phrase timescale in our stimuli) compared to the other two conditions. It is important to note that (1) differences between conditions are not surprising, given that our stimuli were naturally spoken, and (2) we specifically designed our experiment to include backward versions of all conditions to control for slight differences between the acoustic envelopes of the forward stimuli. That being said, we conducted an additional, exploratory analysis of the phrase frequency band in order to further reduce the potential confound of differences between the acoustic modulation spectra, and to disentangle the distribution of linguistic phrase representations and the acoustic stimulus even further. Specifically, we computed MI in the delta-theta range (0.8-5 Hz) between the brain response and abstracted, dimensionality-reduced annotations of all stimuli, containing only information about when words could be integrated into phrases (Brodbeck et al., 2018; see section Materials and Methods for detailed descriptions of how these annotations were created).

These annotation-based analyses revealed significant effects of Condition (Sentence = treatment level; Jabberwocky: β = -0.326, SE = 0.112, p = 0.007; Wordlist: β = -0.521, SE = 0.120, $p < 0.001$), Direction (Forward = treatment level; Backward: β = -0.754, SE = 0.115, $p < 0.001$) and their interaction (Jabberwocky*Backward: β = 0.352, SE = 0.164, p = 0.040; Wordlist*Backward: β = 0.621, SE = 0.156, $p < 0.001$). The estimated marginal means further revealed increased MI for forward sentences compared to forward jabberwocky items and forward word lists (Sentence Forward – Jabberwocky Forward: Δ = 0.326, SE = 0.114, p = 0.021; Sentence Forward – Wordlist Forward: Δ = 0.521, SE = 0.123, $p < 0.001$; all results Tukey corrected for multiple comparisons) and no significant difference among the backward controls (see Table 4.6 and Appendix for complete model outputs).

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.33	0.11	29.8	2.85	0.02
Sentence - Wordlist	0.52	0.12	30.0	4.25	< 0.01
Jabberwocky - Wordlist	0.20	0.13	30.0	1.54	0.29
Direction = Backward					
Sentence - Jabberwocky	-0.03	0.12	30.0	-0.22	0.97
Sentence - Wordlist	-0.10	0.10	30.1	-0.99	0.59
Jabberwocky - Wordlist	-0.07	0.11	30.1	-0.69	0.77

Table 4.6: Estimated marginal means for MI (log-transformed) calculated over abstract phrase representations. P-value adjustment: tukey method for comparing a family of 3 estimates.

Again, the same pattern of results also emerged when computing MI over all electrodes: The mixed effects model revealed significant effects of Condition (Jabberwocky: β = -0.365, SE = 0.087, $p < 0.001$; Wordlist: β = -0.611, SE = 0.098, $p < 0.001$), Direction (β = -0.813, SE = 0.090, $p < 0.001$) and their interaction (Jabberwocky*Backward: β = 0.390, SE = 0.148, p = 0.014; Wordlist*Backward: β = 0.678, SE = 0.131, $p < 0.001$). The pair-wise contrasts were, again, only significant between the forward conditions (Sentence Forward – Jabberwocky Forward: Δ = 0.365, SE = 0.088, $p < 0.001$; Sentence Forward – Wordlist Forward: Δ = 0.611, SE = 0.100, $p < 0.001$). These results support our previously reported findings, showing that neural tracking is influenced by the presence of abstract linguistic information. In other words, this exploratory analysis supports our earlier finding that the brain's "sensitivity" to linguistic structure and meaning goes above and beyond the acoustic signal and both word-level semantic and prosodic controls.

4.4 Discussion

The current experiment tested how the brain attunes to linguistic information. Contrasting sentences, word lists and jabberwocky items, we analyzed, by proxy, how the brain response is modulated by sentence-level prosody, lexical semantics, and compositional structure and meaning. Our findings show that 1) the neural response is driven by compositional structure and meaning, beyond both acoustic-prosodic and lexical information; and 2) the brain most closely tracks the most structured representations on the timescales we analyzed. To our knowledge, this is the first study to systematically disentangle the contribution of linguistic content from its timing and rhythm in natural speech by employing linguistically-informed controls. Additionally, our data demonstrates cortical tracking of naturalistic language without a non-linguistic task such as syllable counting and outlier trial or target-detection tasks. We show that oscillatory activity attunes to structured and meaningful content, suggesting that neural tracking reflects computations related to inferring linguistic representations from speech, and not merely tracking of rhythmicity or timing. We discuss these findings in more detail below.

Using Mutual Information analysis, we quantified the degree of speech tracking in frequency bands corresponding to the timescales at which linguistic structures (phrases, words, and syllables) could be inferred from our stimuli. On the phrase timescale, we found that sentences had the most shared information between stimulus and response. Crucially, this is not merely a chunking mechanism (e.g., Bonhage, Meyer, Gruber, Friederici, & Mueller, 2017; Ghitza, 2017) – participants could have "chunked" the word lists (which have their own naturally produced non-sentential prosody) into units of adjacent words, and the jabberwocky items into prosodic units. This is especially interesting given recent work by Jin, Lu, and Ding (2020), showing that enhanced delta-band activity can be "induced" in listeners by teaching them to chunk a sequence of (synthesized) words according to different sets of artificial grammar rules. Conversely, the observed patterns of activity cannot exclusively be driven by the lexico-semantic content of our stimuli (see Frank & Yang, 2018) – sentences and word lists contained the same lexical items, yet MI was enhanced for Sentence stimuli, where words could be combined into phrases and higher-level representations. As such, we argue that the dominating process we observe appears to be processing of compositional semantic structure, above and beyond prosodic chunking and word-level meaning. We show that the brain aligns more to periodically occurring units when they contain meaningful information and are thus relevant for linguistic processing.

On the word timescale, the emerging picture is somewhat more diverse than on the phrase timescale. Specifically, we found enhanced tracking for sentences compared to jabberwocky items. We tentatively take this finding to indicate that, at the word timescale, the dominant process appears to be context-dependent word recognition – perhaps based in perceptual inference. This is further corroborated by the results of computing MI over all electrodes, rather than a subset, with sentences eliciting higher MI than both jabberwocky items and word lists. Note, however, that we also found enhanced MI on the word timescale for word lists compared to jabberwocky items in the backward controls when computing MI over all electrodes. Here, listeners could not have processed words within the context of phrases or sentences, which makes it somewhat difficult to integrate these results.

There continues to be a vibrant debate about whether language-related cortical activity in the delta-theta range is truly oscillatory in nature, or whether the observed patterns of neural activity arise as a series of evoked responses (e.g., Haegens & Zion Golumbic, 2018; Obleser & Kayser, 2019; Rimmele, Morillon, Poeppel, & Arnal, 2018; Zoefel, ten Oever, & Sack, 2018). Our current results cannot speak to this question; in fact, we have been careful to refer to our results as “tracking” rather than “entrainment” throughout this paper. To be clear, we do not take the observed increased MI for sentences compared to jabberwocky items and word lists as evidence for an intrinsic “phrase- ”, or “word-level oscillator”. Rather, we interpret our findings as a manifestation of the cortical computations that may occur during language comprehension. Here, we observe them in the delta frequency range because that is the timescale on which higher-level linguistic units occur in our stimuli.

Many previous studies have shown that attention can modulate neural entrainment (e.g., Calderone, Lakatos, Butler, & Castellanos, 2014; Ding & Simon, 2013; Haegens, Handel, & Jensen, 2011; Lakatos et al., 2013; Zion Golumbic et al., 2013). Importantly, Ding et al. (2018) found that tracking beyond the syllable envelope requires attention to the speech stimulus. In our current experiment, participants were instructed to attentively listen to the audio recordings in all conditions, but it is possible that “attending to sentences” might be easier than “attending to jabberwocky items”, and that listeners pay closer attention to higher-level structures in intelligible and meaningful speech. As such, we cannot rule out the possibility that our effects might be influenced by a mechanism based on attentional control. It is, however, difficult to disentangle “attention” from “comprehension” in this kind of argument – meaningful information within a stimulus can arguably only lead to increased attention if it is comprehensible. We plan to investigate these questions in future experiments.

Overall, the pattern of results is consistent with cue-integration-based models of language processing (Martin, 2016, 2020), where the activation profile of different populations of neurons over time encodes linguistic structure as it is inferred from sensory correlates in real-time (Martin & Doumas, 2017). Martin’s (2016, 2020) model of language processing builds on and extends neurophysiological models of cue integration (e.g., Ernst and Bülthoff, 2004; Fetsch et al., 2013; Landy et al., 2011; see McMurray and Jongman, 2011; Toscano and McMurray, 2010; and Norris and McQueen, 2008, for cue-integration-based models of speech and word recognition). The underlying mechanism of cue integration relies on only two core computations: summation and normalization, both of which have been proposed as canonical neural computations (e.g., Carandini & Heeger, 1994, 2012). Multiple cues (which can, in principle, be any piece of sensory information that is available in a given situation) are combined via summation and integrated via normalization in order to arrive at a robust percept (e.g., Ernst & Bülthoff, 2004). Cues are associated with corresponding weights, which can be dynamically updated in order to account for the fact that not all cues are equally reliable (or even available) in any given situation (see also E. Bates & MacWhinney, 1989, for a model of sentence processing that posits competition between different linguistic percepts as a result of cue validity and ranking). As such, the process of integrating multiple cues into a percept can be thought of as an inference problem (e.g., Landy et al., 2011). Martin (2016, 2020) proposed that, during all stages of language processing, the brain might draw on these same neurophysiological computations.

Crucially, inferring linguistic representations from speech sounds requires not only bottom-up, sensory information, but also top-down, memory-based cues (e.g., Marslen-Wilson, 1987). Martin (2016, 2020) therefore suggested that cue integration during language comprehension is an iterative process, where cues that have been inferred from the acoustic signal can, in turn, become cues for higher levels of processing. The pattern of findings in our current experiment strongly speaks to cue-integration-based models of language comprehension: We observe that tracking of the speech signal is enhanced when meaningful linguistic units can be inferred, suggesting that alignment of populations of neurons might, indeed, encode the generation of inference-based linguistic representations (Martin & Doumas, 2017).

There are, of course, many open questions that arise from our results. Perhaps most obviously (although presumably limited by the resolution of time-frequency analysis), it would be interesting to investigate how “far” cue integration can be traced during even more natural language comprehension situations (cf. Alday, 2019; Alexandrou, Saarinen, Kujala, & Salmelin, 2020). To what degree are higher-level linguistic cues, such as sentential, contextual, or pragmatic information, encoded in the neural response? Another interesting avenue for future research would be to investigate whether similar patterns can be observed during language production. Martin (2016, 2020) suggested that not only language comprehension, but also language production draws on principles of cue integration. Finally – and consequentially, if cue integration underlies both comprehension and production processes –, we would be curious to learn more about cue integration “in action”, specifically during dialogue settings, where interlocutors comprehend and plan utterances nearly simultaneously.

In summary, this study showed that speech tracking is sensitive to linguistic structure and meaning, above and beyond prosodic and lexical-semantic controls. In other words: Content determines tracking, not just timescale. This extends previous findings and advances our understanding of spoken language comprehension in general, because our experimental manipulation allows us, for the first time, to disentangle the influence of linguistic structure and meaning on the neural response from word-level meaning and prosodic regularities occurring in naturalistic stimuli. Cue-integration-based models of language processing (Martin, 2016, 2020; Martin & Doumas, 2017) offer a neurophysiologically plausible, mechanistic explanation for our results.

Appendix

Stimuli. Sentence, Jabberwocky and Wordlist versions of all stimuli, as well as the English translation of the Sentence stimuli.

1. *Sentence:* Vlotte meesters schenken wijsheid en een aardig kindje schildert sterren.
   (Easygoing teachers offer wisdom and a nice child paints stars.)
   Jabberwocky: Snatte waasters scharken wielheid en een aallig wundje schurdert sperben.
   Word list: schildert schenken vlotte aardig wijsheid kindje meesters sterren en een

2. Gekke meisjes snijden uien en de scherpe messen maken wondjes.
   (Crazy girls cut onions and the sharp knives cause wounds.)
   Gelpe muikjes floeden euer en de strerbe letsen lapen wouwses.
   uien wondjes meisjes messen de en snijden maken gekke scherpe

3. Kleine obers tapten biertjes en de domme gasten breken borden.
   (Little waiters poured beers and the stupid guests break plates.)
   Spiene abels pipten beeltjes en de lolme gonten flepen varden.
   tapten breken kleine domme de en borden obers biertjes gasten

4. Lange mannen bouwen huisjes en de lieve honden brengen planken.
   (Tall men build houses and the sweet dogs bring boards.)
   Lalve wanzen botren raasjes en de reeve rorden brargen sponken.
   planken mannen huisjes honden de en bouwen brengen lange lieve

5. Trotse moeders hebben baby’s en de lieve oma’s geven snoepjes.
   (Proud mothers have babies and the sweet grandmas give sweets.)
   Pletse hijders rabben obis en de rieze bawun beben vliepjes.
   hebben geven trotse lieve de en snoepjes moeders baby’s oma’s

6. Goede sporters renden rondjes en de grote wolken bieden schaduw.
   (Good athletes ran laps and the big clouds provide shade.)
   Vijde spenters rarden rouwses en de spode delken vuiden scharub.
   schaduw sporters rondjes wolken de en renden bieden goede grote

7. Stoute muizen knagen gaten en de boze huurders haten dieren.
   (Naughty mice gnaw holes and the angry tenants hate animals.)

Stemte mieven snamen vaden en de vone hinkders doten weiren.
knagen haten stoute boze de en gaten muizen huurders dieren

8. Slimme eekhoorns vinden nootjes en de groene kikkers vangen vliegjes.
(Smart squirrels find nuts and the green frogs catch flies.)
Plemme oekboorns ganden zietjes en de broeze wokkers gongen snoegjes.
eekhoorns nootjes kikkers vliegjes de en vinden vangen slimme groene

9. Bange ridders zoeken toevlucht en de gouden sleutel opent deuren.
(Frightened knights seek refuge and the golden key opens doors.)
Garge ludders nijken toepricht en de gatden speetel ogens weiren.
zoeken opent bange gouden de en sleutel ridders toevlucht deuren

10. Blauwe visjes zwemmen baantjes en de grijze kippen horen piepjes.
(Blue fish swim laps and the grey chickens hear beeps.)
Braube bispes knimmen gaantres en de brijne dappen lolen peugjes.
kippen visjes baantjes piepjes de en zwemmen horen blauwe grijze

11. Grote leeuwen vinden voedsel en de jonge schapen blijken geitjes.
(Big lions find food and the young sheep turn out to be goats.)
Spode loorden ginten baadsel en de jarge straben ploeken gaukjes.
vinden lijken grote jonge de en voedsel leeuwen schapen geitjes

12. Kwade jongens breken glazen en de strenge juffen schrijven regels.
(Angry boys break glasses and the strict teachers write rules.)
Smate jargens drepen flaven en de strelle ceffen schroezen lemels.
jongens juffen glazen regels de en breken schrijven kwade strenge

13. Zieke kindjes krijgen appels en de kalme zusters breien sokken.
(Sick children get apples and the calm nurses knit socks.)
Neike wundjes spijmen atsels en de malge nutters pleuen senken.
krijgen breien zieke kalme de en kindjes appels zusters sokken

14. Warme landjes hebben strandjes en de korte dagen brengen vreugde.
(Warm countries have beaches and the short days bring joy.)
landjes strandjes dagen vreugde de en hebben brengen warme korte
Marle lerkjes mobben strastpes en de warte lapen spelgen fleufde.

15. Zwarte geiten proefden suiker en de rotte tanden hebben gaten.
(Black goats tasted sugar and the rotten teeth have cavities.)
Flakte beuten praasden feeker en de hatte palden mabben voten.
proefden hebben zwarte rotte de en gaten geiten suiker tanden

16. Rode mieren dragen takken en de wilde katten vangen vogels.
(Red ants carry branches and the wild cats catch birds.)
Lote keeren tramen tenken en de kelde lutten gargen valmen.
takken mieren katten vogels de en dragen vangen rode wilde

17. Grauwe wolken brengen regen en de zware buien breken takken.
(Grey clouds bring rain and the heavy showers break branches.)
Kraube louken pletgen lepen en de plave gijen smesen tonken.
brengen breken grauwe zware de en buien wolken regen takken

18. Blije artsen helpen mensen en de oude tantes hebben nichtjes.
(Happy doctors help people and the old aunts have nieces.)
Ploeie alfjen hospen miksen en de aide paltes labben zechtjes.
artsen mensen tantes nichtjes de en helpen hebben oude blije

19. Houten tafels hebben laatjes en de ronde knikkers lijken druiven.
(Wooden tables have drawers and the round marbles look like grapes.)
Hemten pacels libben raakjes en de dande vlokkers woeken driezen.
lijken houten hebben ronde de en tafels laatjes knikkers druiven

20. Zwarte laarzen trekken aandacht en de vreemde mannen schrobben vloeren.
(Black boots attract attention and the strange men scrub floors.)
Knorte raarnen grikken aangucht en de smijmde lonnen schrimben knijren.
vloeren laarzen aandacht mannen de en trekken schrobben zwarte vreemde

21. Snelle jagers volgen spoortjes en de tamme hazen leggen keutels.
(Fast hunters follow tracks and the tame hares lay turds.)
Flolle cavers valmen vleertjes en de torme lamen lelmen weitels.
volgen leggen snelle tamme de en jagers spoortjes keutels hazen

22. Leuke otters zoeken visjes en de grote leeuwen bijten mensen.
(Nice otters look for fish and the big lions bite people.)
Rauke akters nijken vaspes en de plode loorten gijden molsen
leeuwen otters visjes mensen de en zoeken bijten leuke grote

23. Stille meisjes mengen sapjes en de rijke zeilers slurpen koffie.
(Silent girls mix juices and the rich sailors slurp coffee.)
Stimpe muikjes lelgen sekjes en de lijse neulers plunpen wiffie.
mengen slurpen stille rijke de en zeilers meisjes sapjes koffie

24. Trotse slagers snijden biefstuk en de blije klanten kopen worstjes.
(Proud butchers cut steak and the happy customers buy sausages.)
Bletse tramers proeden vaafstuk en de knoeie sponten women wuchtjes.
biefstuk worstjes klanten slagers de en snijden kopen trotse blije

25. Knappe schilders winnen prijsjes en de wilde paarden aten peren.
(Handsome painters win prizes and the wild horses ate pears.)
Flippe scharders dinzen proekjes en de kelde deurden usen remen.
winnen aten knappe wilde de en schilders prijsjes paarden peren

26. Stompe messen snijden broodjes en de paarse pennen schrijven woorden.
(Blunt knives cut sandwiches and the purple pens write words.)
Starpe retsen knoeden braaljes en de waadse parnen schroezen moorten.
messen broodjes pennen woorden de en snijden schrijven stompe paarse

27. Bruine apen zoeken vruchten en de witte schapen aten blaadjes.
(Brown monkeys seek fruit and the white sheep ate leaves.)
Driene onen nijken smechten en de kette straven oken bleegjes.
zoeken aten bruine witte de en apen schapen blaadjes vruchten

28. Oude opa's snoeien heggen en de lieve oma's bakten koekjes.
(Old grandpas trim hedges and the sweet grandmas baked cookies.)
Adde obos knooien relgen en de ruive onis borten moefjes.
koekjes opa's heggen oma's de en snoeien bakten oude lieve

29. Trieste zwemmers schrijven brieven en de lompe zangers zingen liedjes.
(Sad swimmers write letters and the rude singers sing songs.)
Breeste knimmers schroezen pleiven en de laspe zallers zannen riefjes.
schrijven zingen trieste lompe de en brieven liedjes zangers zwemmers

30. Stoere vaders prikken gaten en de vlotte moeders koken uien.
(Tough fathers poke holes and the easygoing mothers cook onions.)
Stijne gaters drekken vaden en de knette hijders mosen auer.
moeders vaders gaten uien de en prikken koken stoere vlotte

31. Mooie vogels zingen wijsjes en de gekke meiden wassen kleren.
(Beautiful birds sing tunes and the crazy girls wash clothes.)
Woeie govels zanpen waadjes en de gesse kieden pansen pleven.
zingen wassen mooie gekke de en vogels wijsjes kleren meiden

32. Knappe zangers geven feestjes en de stoere werklui kopen biertjes.
(Handsome singers give parties and the tough workers buy beers.)

Smippe zalpers beben feursjes en de steepe werfmui rogen booltjes.
zangers werklui biertjes feestjes de en geven kopen knappe stoere

33. Vieze kwallen prikken duikers en de dunne vissers huurden bootjes.
(Dirty jellyfish sting divers and the thin fishermen rented boats.)
Beeze flollen kwokken keekers en de murne bitsers lutsden boepjes.
prikken huurden vieze dunne de en duikers vissers bootjes kwallen

34. Sterke vaders dragen dochters en de mooie meisjes zoenden jongens.
(Strong fathers carry daughters and the beautiful girls kissed boys.)
Sperre goders tramen wichters en de moene miekjes nienden jorlens.
dochters meisjes vaders jongens de en dragen zoenden sterke mooie

35. Lompe kappers knippen haren en de vlugge klussers bouwen muren.
(Rude hairdressers cut hair and the quick handymen build walls.)
Lolle mippers vrappen lalen en de snigge spessers botren luven.
knippen bouwen lompe vlugge de en muren kappers klussers haren

36. Leuke vrouwen spelen cello en de dikke drummers poetsten trommels.
(Nice women play cello and the fat drummers cleaned drums.)
Mauke smouven pleren jeldo en de wokke plurmers peursten spolmels.
vrouwen drummers cello trommels de en spelen poetsten leuke dikke

37. Arme vrouwen poetsten schoenen en de trage laptops brengen spanning.
(Poor women polished shoes and the slow laptops cause tension.)
Orle vrulwen poonsten scheemen en de drame lanteps spelgen klanzing.
poetsten brengen arme trage de en vrouwen spanning schoenen laptops

38. Lieve meisjes plukten appels en de schuwe jongens vrezen hoogtes.
(Sweet girls picked apples and the shy boys are afraid of heights.)
Reeve muipjes slunten atjels en de schine jargens flenen haaites.
meisjes hoogtes appels jongens de en plukten vrezen lieve schuwe

39. Drukke winkels lokten klanten en de lange mannen kopen schoenen.
(Busy stores attracted customers and the tall men buy shoes.)
Spunke linsels lurten spalten en de lalve wanzen loben scheegen.
lokten kopen drukke lange de en winkels klanten schoenen mannen

40. Gekke jongens pesten eenden en de zieke meiden poetsten tanden.
(Crazy boys harass ducks and the sick girls brushed teeth.)
Gelse jormens tetten oelden en de neike kieden peugsten palden.
tanden jongens meiden eenden de en pesten poetsten gekke zieke

41. Vreemde vrouwen hebben heimwee en de trouwe buren zenden brieven.
(Strange women are homesick and the loyal neighbors send letters.)
Smuimde flouven rabben heipwij en de blouve guven zarden pleiven.
hebben zenden vreemde trouwe de en heimwee vrouwen brieven buren

42. Kleine baby's horen liedjes en de wijze kerels lezen kranten.
(Small babies hear songs and the wise guys read newspapers.)
Speune bawus ronen riefjes en de moeze lenels remen sponten.
baby's kerels liedjes kranten de en horen lezen kleine wijze

43. Saaie buren kopen borden en de jonge kindjes pakten snoepjes.
(Boring neighbors buy plates and the young children grabbed sweets.)
Siere gulen loben girden en de jelge wirtjes penten vloesjes.
pakten saaie kopen jonge de en buren borden kindjes snoepjes

44. Snelle schaatsers vinden gaatjes en de kleine jongens spelen voetbal.
(Fast skaters find holes and the little boys play soccer.)
Flolle schijnsers ginten geekjes en de speene jargens sleren boenbel.
schaatsers voetbal gaatjes jongens de en vinden spelen snelle kleine

45. Witte paarden trekken koetsen en de saaie vorsten wenkten burgers.
(White horses pull carriages and the boring monarchs beckoned to civilians.)
Kette peenden drakken dietsen en de soene viksten lankten vurmers.
trekken wenkten witte saaie de en paarden vorsten burgers koetsen

46. Stille schilders belden vrienden en de roze scooter levert pizza.
(Quiet painters called friends and the pink scooter delivers pizza.)
Stimpe schadders benten fleunden en de lone sjaater rezert pixta.
schilders scooter vrienden pizza de en roze stille levert belden

47. Oude mensen rijden bussen en de toffe oma's maken grapjes.
(Old people drive busses and the cool grandmas make jokes.)
Amde molsen mieden gossen en de puffe onos lapen spipjes.
rijden maken oude toffe de en bussen grapjes oma's mensen

48. Brede tantes zoenden wangen en de malle neven roken jointjes.
(Plump aunts kissed cheeks and the silly nephews smoke joints.)
Krete pastes noonden largen en de kolle zeben kosen jarstjes.
wangen neven jointjes tantes de en zoenden roken brede malle

49. Jonge bakkers maken broden en de nette klanten ruiken taartjes.
(Young bakers make bread and the polite customers smell cakes.)

Jorle bassers hapen smoten en de zutte spalten hieken toordjes.
maken ruiken jonge nette de en broden taartjes klanten bakkers

50. Klamme handen voelen muren en de knusse kachels drogen kleren.
(Damp hands feel walls and the cozy stoves dry clothes.)
Klorme londen vijren luven en de vresse maspels blomen spemen.
handen muren kachels kleren de en voelen drogen klamme knusse

51. Snelle zwemmers slurpen ijsthee en de vieze kwallen zoeken water.
(Fast swimmers slurp iced tea and the dirty jellyfish look for water.)
Knille knummers plunpen ijfkwee en de bieve smollen nijken lates.
slurpen zoeken vieze snelle de en zwemmers ijsthee kwallen water

52. Bange helden plukken bloemen en de bruine vogels halen takken.
(Frightened heroes pick flowers and the brown birds fetch branches.)
Garge ralden spunken drijmen en de druize gomels paven mukken.
helden bloemen vogels takken de en plukken halen bange bruine

53. Vlotte otters bouwen dammen en de snelle hazen doden kevers.
(Smooth otters build dams and the fast hares kill beetles.)
Zwitte olders botren lemmen en de vralle lamen zoten mezers.
bouwen doden vlotte snelle de en otters dammen hazen kevers

54. Saaie meesters geven lessen en de vele brieven worden stapels.
(Boring teachers give lessons and the many letters become piles.)
Soene waasters beben hussen en de bene pleeven rarden stagelt.
meesters lessen stapels brieven de en geven worden saaie vele

55. Luikse wafels stillen honger en de rotte appels krijgen schimmel.
(Liège waffles satisfy hunger and the rotten apples are getting moldy.)
Ruipre lafelt stimpen morger en de hatte ampelt spijmen schurmel.
krijgen stillen luikse rotte de en wafels honger appels schimmel

56. Enge slangen eten muizen en de grote kippen leggen eitjes.
(Scary snakes eat mice and the big chickens lay eggs.)
Elme spalgen eber mienen en de vlode dappen relgen eutres.
slangen muizen kippen eitjes de en eten leggen enge grote

57. Scherpe scharen knippen blaadjes en de snelle auto’s rijden rondjes.
(Sharp scissors cut leaves and the fast cars drive laps.)
Strerbe stranen smeppen bleegjes en de flalle euvos loeden lortjes.
knippen rijden scherpe snelle de en scharen blaadjes auto’s rondjes

58. Luie tieners dekken tafels en de dikke dames brengen koffie.
(Lazy teenagers set tables and the fat ladies bring coffee.)
Reie teezers lenken mabels en de wokke lapes brargen wiffie.
tieners tafels dames koffie de en dekken brengen luie dikke

59. Zure bessen maken vlekken en de zachte bijen maken honing.
(Sour berries make stains and the soft bees make honey.)
Nume betjen lasen zwokken en de nochte guien lapen moping.
maken maken zure zachte de en bessen bijen vlekken honing

60. Vieze mannen smeren olie en de toffe ouders kopen kaartjes.
(Dirty men smear oil and the cool parents buy tickets.)
Bieve wanzen fleven oree en de piffe amders homen koostjes.
mannen olie ouders kaartjes de en smeren kopen vieze toffe

61. Franse schilders verven muren en de Vlaamse bakkers kneden brooddeeg.
(French painters paint walls and the Flemish bakers knead bread dough.)
Flanje schunders bernen lunen en de knuimse garkers dreten slooddieg.
verven kneden Franse Vlaamse de en schilders muren bakkers brooddeeg

62. Vlotte lopers maken meters en de boze tieners slopen ruiten.
(Fast runners make meters and the angry teenagers wreck windows.)
Snette rogeres dapen peders en de vone teezers spomen hieten.
lopers meters tieners ruiten de en maken slopen vlotte boze

63. Gele bloemen lokken bijen en de brakke mensen drinken water.
(Yellow flowers attract bees and the hungover people drink water.)
Bese drijmen lunken veien en de plokke moksen plinsen rates.
drinken lokken gele brakke de en bloemen bijen mensen water

64. Zware tassen breken ruggen en de natte sokken brengen blaren.
(Heavy bags break backs and the wet socks cause blisters.)
Plane pansen flesen ruflen en de zaste sunken spelgen spalen.
tassen ruggen sokken blaren de en breken brengen zware natte

65. Losse spijkers sieren muren en de bange muizen graven hollen.
(Loose nails adorn walls and the scared mice dig burrows.)
Lunse kloekers suinen luven en de garge mienen slazen mellen.
sieren graven losse bange de en spijkers muren muizen hollen

66. Dronken rappers lopen blauwtjes en de bolle katten aten brokjes.
(Drunk rappers are turned down and the chubby cats ate kibble.)

Plorken lippers women bleuntjes en de galle mitten uken slekjes.
lopen aten dronken bolle de en rappers blauwtjes katten brokjes

67. Knappe ridders redden levens en de grote paarden winnen prijsjes.
(Handsome knights save lives and the big horses win prizes.)
Flippe ludders hefden rezens en de spode deurden dinzen proekjes.
ridders paarden prijsjes levens de en redden winnen knappe grote

68. Kille zussen stelen spullen en de woeste ouders straffen broertjes.
(Cold-hearted sisters steal things and the furious parents punish brothers.)
Wulle zetsen speven krillen en de deuste agders stropfen brooltjes.
stelen straffen kille woeste de en zussen spullen ouders kindjes

69. Slome slakken aten sprietjes en de valse hommels steken kindjes.
(Slow snails ate grass blades and the vicious bumblebees sting children.)
Ploge spikken osen sproekjes en de vikse lompels speven wirtjes.
slakken sprietjes hommels kindjes de en aten steken slome valse

70. Rotte appels brengen ziektes en de gulle fietsers zingen liedjes.
(Rotten apples bring diseases and the generous cyclists sing songs.)
Hatte ammels spelgen nientes en de gumpe feunsers zansen riegjes.
brengen zingen rotte gulle de en appels ziektes fietsers liedjes

71. Gekke buren maken worstjes en de dikke neven snoepen taarten.
(Crazy neighbors make sausages and the fat cousins snack on cakes.)
Gelse gulen dapen wuchtjes en de mekke zeben smijpen tijnten.
snoepen maken gekke dikke de en buren worstjes neven taarten

72. Dunne meisjes drinken sapjes en de witte duiven aten bonen.
(Thin girls drink juices and the white pigeons ate beans.)
Durre muipjes plinsen sutjes en de kette wuizen uken goven.
meisjes sapjes duiven bonen de en drinken aten witte dunne

73. Knappe mannen strikken veters en de kleine muizen horen piepjes.
(Handsome men tie laces and the little mice hear beeps.)
Flippe wanzen strissen getels en de spiene mieven rolen peugjes.
strikken horen knappe kleine de en mannen muizen veters piepjes

74. Vieze messen snijden appels en de snelle jongens gooien ballen.
(Dirty knives cut apples and the fast boys throw balls.)
Beize letsen floeden ammels en de flolle jargens goenen garlen.
messen appels jongens ballen de en snijden gooien vieze snelle

75. Trage mensen kopen broden en de mooie vrouwen bakken taarten.
(Slow people buy loaves of bread and the beautiful women bake pies.)
Klame helsen lomen droten en de moere smouven galken peurten.
kopen bakken trage mooie de en mensen broden vrouwen taarten

76. Luie honden ruiken voedsel en de warme broodjes hebben pitjes.
(Lazy dogs smell food and the hot sandwiches contain seeds.)
Reue rorden hieken baadsel en de marle braagjes rabben pekjes.
voedsel honden broodjes pitjes de en ruiken hebben luie warme

77. Boze ouders geven straffen en de rode vossen graven kuilen.
(Angry parents give punishments and the red foxes dig pits.)
Vome adkers besen strinfen en de lote gussen brazen mielen.
geven graven boze rode de en ouders vossen kuilen straffen

78. Lieve meiden schrijven boeken en de lichte kamers hebben ramen.
(Sweet girls write books and the bright rooms have windows.)
Reive kieden schroezen vijken en de rachte lapers mabben dapen.
meiden boeken kamers ramen de en schrijven hebben lieve lichte

79. Gouden munten hebben waarde en de vreemde vogels fluiten liedjes.
(Golden coins have value and the strange birds whistle songs.)
Gudden kurten rabben laalde en de floemde govels vrieten leugjes.
hebben fluiten gouden vreemde de en munten waarde vogels liedjes

80. Slome treinen hebben stoelen en de rijke boeren voeden koeien.
(Slow trains have seats and the rich farmers feed cows.)
Ploge pleenen labben stijren en de loeke goelen vuiten woenen.
treinen stoelen boeren koeien de en hebben voeden slome rijke

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-4.072	0.081	29.560	-50.372	< 0.001		
Condition[T.Jabberwocky]	-0.452	0.096	29.094	-4.692	< 0.001		
Condition[T.Wordlist]	-0.491	0.116	28.838	-4.246	< 0.001		
Direction[T.Backward]	-0.885	0.117	29.288	-7.562	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.429	0.152	28.997	2.830	0.008		
Cond.[T.Word.]:Dir.[T.Back.]	0.523	0.185	29.010	2.824	0.009		
Random Effects							
Intercept|Participant						0.142	0.377
Cond.[T.Jabb.]|Part.						0.179	0.423
Cond.[T.Word.]|Part.						0.296	0.544
Dir.[T.Back.]|Part.						0.305	0.553
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.486	0.697
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.814	0.903

Table 4A.1: Mixed-effects logistic regression results for MI in the phrase frequency band. Sentence Forward = treatment level.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-4.921	0.077	28.666	-63.773	< 0.001		
Condition[T.Jabberwocky]	-0.484	0.121	29.177	-4.007	< 0.001		
Condition[T.Wordlist]	-0.158	0.079	29.065	-2.001	0.055		
Direction[T.Backward]	-0.499	0.136	29.108	-3.671	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.234	0.197	29.045	1.118	0.244		
Cond.[T.Word.]:Dir.[T.Back.]	0.084	0.146	28.886	0.574	0.570		
Random Effects							
Intercept|Participant						0.129	0.360
Cond.[T.Jabb.]|Part.						0.338	0.581
Cond.[T.Word.]|Part.						0.095	0.308
Dir.[T.Back.]|Part.						0.451	0.671
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.951	0.975
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.444	0.667

Table 4A.2: Mixed-effects logistic regression results for MI in the word frequency band. Sentence Forward = treatment level.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-6.090	0.103	28.098	-59.045	<0.001		
Condition[T.Jabberwocky]	0.001	0.121	28.599	0.007	0.994		
Condition[T.Wordlist]	0.104	0.109	28.529	0.954	0.348		
Direction[T.Backward]	0.034	0.120	28.509	0.283	0.779		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.144	0.166	29.173	0.869	0.392		
Cond.[T.Word.]:Dir.[T.Back.]	-0.069	0.144	28.875	-0.476	0.637		
Random Effects							
Intercept|Participant						0.264	0.513
Cond.[T.Jabb.]|Part.						0.335	0.579
Cond.[T.Word.]|Part.						0.254	0.504
Dir.[T.Back.]|Part.						0.330	0.574
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.621	0.788
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.426	0.653

Table 4A.3: *Mixed-effects logistic regression results for MI in the syllable frequency band.* Sentence Forward = treatment level.

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.401	0.076	30	5.283	< 0.001
Sentence - Wordlist	0.418	0.098	30	4.591	< 0.001
Jabberwocky - Wordlist	0.017	0.076	30	0.222	0.973
Direction = Backward					
Sentence - Jabberwocky	0.105	0.107	30	0.980	0.595
Sentence - Wordlist	0.086	0.117	30	0.734	0.745
Jabberwocky - Wordlist	-0.019	0.118	30	-0.159	0.986

Table 4A.4: *Estimated marginal means for MI in the phrase frequency band, computed over all electrodes.* P-value adjustment: tukey method for comparing a family of 3 estimates.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-4.017	0.069	28.899	-58.332	< 0.001		
Condition[T.Jabberwocky]	-0.401	0.075	29.574	-5.377	< 0.001		
Condition[T.Wordlist]	-0.418	0.088	29.105	-4.773	< 0.001		
Direction[T.Backward]	-0.743	0.087	28.526	-8.650	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.296	0.099	29.996	2.996	0.006		
Cond.[T.Word.]:Dir.[T.Back.]	0.332	0.134	28.988	2.479	0.019		
Random Effects							
Intercept|Participant						0.119	0.346
Cond.[T.Jabb.]|Part.						0.127	0.356
Cond.[T.Word.]|Part.						0.188	0.434
Dir.[T.Back.]|Part.						0.185	0.430
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.218	0.467
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.455	0.674

Table 4A.5: *Mixed-effects logistic regression results for MI in the phrase frequency band, computed over all electrodes.* Sentence Forward = treatment level.

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.407	0.095	30	4.282	< 0.001
Sentence - Wordlist	0.179	0.053	30	3.371	0.006
Jabberwocky - Wordlist	-0.228	0.097	30	-2.343	0.065
Direction = Backward					
Sentence - Jabberwocky	0.288	0.083	30.1	3.465	0.005
Sentence - Wordlist	0.142	0.088	30.0	1.616	0.254
Jabberwocky - Wordlist	-0.147	0.083	30.0	-1.768	0.198

Table 4A.6: *Estimated marginal means for MI in the word frequency band, computed over all electrodes.* P-value adjustment: tukey method for comparing a family of 3 estimates.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-4.923	0.053	29.039	-92.135	< 0.001		
Condition[T.Jabberwocky]	-0.407	0.093	28.990	-4.358	< 0.001		
Condition[T.Wordlist]	-0.179	0.052	29.516	-3.434	0.002		
Direction[T.Backward]	-0.316	0.090	29.112	-3.515	0.002		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.118	0.128	29.125	0.922	0.364		
Cond.[T.Word.]:Dir.[T.Back.]	0.038	0.107	29.334	0.351	0.728		
Random Effects							
Intercept|Participant						0.066	0.257
Cond.[T.Jabb.]|Part.						0.220	0.469
Cond.[T.Word.]|Part.						0.046	0.215
Dir.[T.Back.]|Part.						0.201	0.449
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.412	0.642
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.269	0.519

Table 4A.7: *Mixed-effects logistic regression results for MI in the word frequency band, computed over all electrodes.* Sentence Forward = treatment level.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-5.966	0.100	28.608	-59.965	< 0.001		
Condition[T.Jabberwocky]	0.035	0.106	29.055	0.331	0.743		
Condition[T.Wordlist]	0.037	0.090	29.064	0.411	0.684		
Direction[T.Backward]	0.147	0.124	28.931	1.184	0.246		
Cond.[T.Jabb.]:Dir.[T.Back.]	-0.023	0.131	29.028	-0.179	0.859		
Cond.[T.Word.]:Dir.[T.Back.]	-0.045	0.130	29.083	-0.344	0.733		
Random Effects							
Intercept|Participant						0.270	0.520
Cond.[T.Jabb.]|Part.						0.293	0.541
Cond.[T.Word.]|Part.						0.203	0.450
Dir.[T.Back.]|Part.						0.415	0.645
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.431	0.656
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.423	0.651

Table 4A.8: *Mixed-effects logistic regression results for MI in the syllable frequency band, computed over all electrodes.* Sentence Forward = treatment level.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-4.421	0.086	29.473	-51.433	< 0.001		
Condition[T.Jabberwocky]	-0.497	0.097	29.411	-5.139	< 0.001		
Condition[T.Wordlist]	-0.402	0.118	29.369	-3.421	0.002		
Direction[T.Backward]	-0.805	0.106	29.730	-7.626	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.368	0.150	29.086	2.455	0.020		
Cond.[T.Word.]:Dir.[T.Back.]	0.336	0.168	29.099	2.001	0.055		
Random Effects							
Intercept|Participant						0.170	0.412
Cond.[T.Jabb.]|Part.						0.186	0.431
Cond.[T.Word.]|Part.						0.315	0.561
Dir.[T.Back.]|Part.						0.236	0.486
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.479	0.692
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.646	0.804

Table 4A.9: Mixed-effects logistic regression results for phase MI in the phrase frequency band. Sentence Forward = treatment level.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-5.322	0.077	28.437	-69.359	< 0.001		
Condition[T.Jabberwocky]	-0.380	0.121	28.790	-3.143	0.004		
Condition[T.Wordlist]	-0.120	0.072	30.047	-1.666	0.106		
Direction[T.Backward]	-0.474	0.126	29.034	-3.745	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.174	0.176	29.124	0.989	0.331		
Cond.[T.Word.]:Dir.[T.Back.]	0.057	0.142	29.074	0.404	0.689		
Random Effects							
Intercept|Participant						0.127	0.357
Cond.[T.Jabb.]|Part.						0.338	0.581
Cond.[T.Word.]|Part.						0.063	0.250
Dir.[T.Back.]|Part.						0.378	0.615
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.726	0.852
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.412	0.652

Table 4A.10: Mixed-effects logistic regression results for phase MI in the word frequency band. Sentence Forward = treatment level.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-6.451	0.082	26.882	-78.910	< 0.001		
Condition[T.Jabberwocky]	0.073	0.093	28.886	0.790	0.436		
Condition[T.Wordlist]	0.074	0.111	28.265	0.664	0.512		
Direction[T.Backward]	-0.003	0.094	27.199	-0.034	0.973		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.061	0.142	29.188	0.429	0.671		
Cond.[T.Word.]:Dir.[T.Back.]	0.079	0.149	28.395	0.533	0.598		
Random Effects							
Intercept|Participant						0.149	0.387
Cond.[T.Jabb.]|Part.						0.161	0.401
Cond.[T.Word.]|Part.						0.272	0.521
Dir.[T.Back.]|Part.						0.168	0.410
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.410	0.640
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.466	0.683

Table 4A.11: Mixed-effects logistic regression results for phase MI in the syllable frequency band. Sentence Forward = treatment level.

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.356	0.076	30	4.667	< 0.001
Sentence - Wordlist	0.309	0.091	30	3.406	0.005
Jabberwocky - Wordlist	-0.047	0.075	30	-0.634	0.803
Direction = Backward					
Sentence - Jabberwocky	0.171	0.102	30	1.687	0.227
Sentence - Wordlist	0.099	0.110	30	0.900	0.644
Jabberwocky - Wordlist	-0.072	0.125	30	-0.577	0.833

Table 4A.12: Estimated marginal means for phase MI in the phrase frequency band, computed over all electrodes. P-value adjustment: tukey method for comparing a family of 3 estimates.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-4.403	0.078	28.894	-56.812	< 0.001		
Condition[T.Jabberwocky]	-0.356	0.075	29.307	-4.750	< 0.001		
Condition[T.Wordlist]	-0.309	0.089	29.193	-3.466	0.002		
Direction[T.Backward]	-0.662	0.076	28.275	-8.718	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.185	0.089	29.175	2.078	0.047		
Cond.[T.Word.]:Dir.[T.Back.]	0.210	0.119	28.823	1.768	0.088		
Random Effects							
Intercept|Participant						0.157	0.396
Cond.[T.Jabb.]|Part.						0.131	0.362
Cond.[T.Word.]|Part.						0.198	0.445
Dir.[T.Back.]|Part.						0.135	0.368
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.168	0.410
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.346	0.588

Table 4A.13: Mixed-effects logistic regression results for phase MI in the phrase frequency band, computed over all electrodes. Sentence Forward = treatment level.

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.329	0.091	30.0	3.617	0.003
Sentence - Wordlist	0.139	0.046	30.0	3.009	0.014
Jabberwocky - Wordlist	-0.190	0.106	30.1	-1.787	0.191
Direction = Backward					
Sentence - Jabberwocky	0.209	0.101	30.0	2.075	0.112
Sentence - Wordlist	0.088	0.087	30.0	1.009	0.577
Jabberwocky - Wordlist	-0.121	0.085	30.0	-1.420	0.343

Table 4A.14: Estimated marginal means for phase MI in the word frequency band, computed over all electrodes. P-value adjustment: tukey method for comparing a family of 3 estimates.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-5.326	0.060	28.889	-89.168	< 0.001		
Condition[T.Jabberwocky]	-0.329	0.089	28.793	-3.682	< 0.001		
Condition[T.Wordlist]	-0.139	0.045	28.327	-3.070	0.005		
Direction[T.Backward]	-0.351	0.091	28.802	-3.873	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.120	0.123	28.862	0.978	0.336		
Cond.[T.Word.]:Dir.[T.Back.]	0.051	0.112	29.059	0.454	0.653		
Random Effects							
Intercept|Participant						0.087	0.295
Cond.[T.Jabb.]|Part.						0.199	0.446
Cond.[T.Word.]|Part.						0.027	0.163
Dir.[T.Back.]|Part.						0.206	0.453
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.373	0.610
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.302	0.549

Table 4A.15: Mixed-effects logistic regression results for phase MI in the word frequency band, computed over all electrodes. Sentence Forward = treatment level.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-6.368	0.086	28.622	-74.140	< 0.001		
Condition[T.Jabberwocky]	0.114	0.093	29.109	1.224	0.231		
Condition[T.Wordlist]	0.043	0.098	29.111	0.440	0.663		
Direction[T.Backward]	0.134	0.108	28.906	1.246	0.223		
Cond.[T.Jabb.]:Dir.[T.Back.]	-0.048	0.120	29.043	-0.403	0.690		
Cond.[T.Word.]:Dir.[T.Back.]	0.043	0.132	29.087	0.325	0.748		
Random Effects							
Intercept|Participant						0.198	0.445
Cond.[T.Jabb.]|Part.						0.220	0.469
Cond.[T.Word.]|Part.						0.248	0.498
Dir.[T.Back.]|Part.						0.304	0.552
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.354	0.595
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.441	0.664

Table 4A.16: Mixed-effects logistic regression results for phase MI in the syllable frequency band, computed over all electrodes. Sentence Forward = treatment level.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	-5.814	0.090	27.632	-64.688	< 0.001		
Condition[T.Jabberwocky]	-0.326	0.112	29.411	-2.907	0.007		
Condition[T.Wordlist]	-0.521	0.120	27.604	-4.338	< 0.001		
Direction[T.Backward]	-0.754	0.115	25.289	-6.549	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.352	0.164	28.564	2.150	0.040		
Cond.[T.Word.]:Dir.[T.Back.]	0.621	0.156	28.625	3.993	< 0.001		
Random Effects							
Intercept|Participant						0.189	0.434
Cond.[T.Jabb.]|Part.						0.280	0.529
Cond.[T.Word.]|Part.						0.334	0.578
Dir.[T.Back.]|Part.						0.300	0.548
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.614	0.783
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.541	0.735

Table 4A.17: Mixed-effects logistic regression results for abstract MI. Sentence Forward = treatment level.

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.365	0.088	30.0	4.149	< 0.001
Sentence - Wordlist	0.611	0.100	30.0	6.116	< 0.001
Jabberwocky - Wordlist	0.246	0.115	30.1	2.135	0.100
Direction = Backward					
Sentence - Jabberwocky	-0.024	0.109	30.1	-0.223	0.973
Sentence - Wordlist	-0.067	0.097	30.1	-0.693	0.770
Jabberwocky - Wordlist	-0.043	0.100	30.1	-0.427	0.905

Table 4A.18: Estimated marginal means for MI computed over abstract linguistic representations over all electrodes. P-value adjustment: tukey method for comparing a family of 3 estimates.

5 Delta-theta power is influenced by linguistic structure and meaning

Abstract

Recent accounts of spoken language comprehension posit that cortical oscillations in the delta-theta frequency band (approximately 0.5-4 Hz) track acoustic and linguistic components, as evidenced by increased similarity (e.g., measured as Mutual Information; MI) between the acoustic signal and the brain response (e.g., Keitel et al., 2018; Chapter 4 of this thesis). At the same time, the generation of hierarchical linguistic structure has been linked to increased power in the delta-theta band (Ding, Melloni, et al., 2017; Ding et al., 2016). It is somewhat difficult to integrate these findings, because they are seemingly based on different mechanisms for spoken language comprehension: One that is rooted in *increased similarity* between the acoustic signal and the brain response (e.g., Keitel et al., 2018; Rimmele et al., 2018), and one that suggests *decreased similarity* between the two, as a result of power increases above and beyond the acoustic signal (e.g., Ding, Melloni, et al., 2017; Ding et al., 2016).

Here, we take a complementary approach to the analyses presented in Chapter 4 of this thesis. In the study reported in Chapter 4, 29 adult native speakers listened to naturally-spoken Dutch sentences, jabberwocky items with sentence-like prosody and morphology, and word lists (80 items/condition). For the current chapter, we analysed the recorded EEG data from Chapter 4 to investigate whether spectral power is modulated by the linguistic information conveyed at different timescales. This analysis revealed a “meaning-and-structure” hierarchy from Jabberwocky (lowest) to Sentence (highest) in the delta-theta band. This finding is consistent with previous work (Ding, Melloni, et al., 2017; Ding et al., 2016; Jin et al., 2020) and suggests that accounts of linguistic structure “generation” and “tracking” may not be mutually exclusive.

5.1 Introduction

A growing body of research focuses on the role of cortical activity in the delta frequency band (usually defined to range from approximately 0.5 to 4 Hz) during spoken language comprehension (e.g., Ding, Melloni, et al., 2017; Ding et al., 2016; Ghitza, 2017; Gross et al., 2013; Jin et al., 2020; Meyer, 2018), yet how exactly the brain extracts and utilizes acoustic and linguistic information on this timescale is not entirely clear. One of the reasons for this uncertainty is that different studies have used different measures to investigate the cortical response. Specifically, one line of research has focused on measures of phase coherence, that is, similarity between the phase of the acoustic signal and that of the brain response (e.g., Gross et al., 2013; Keitel et al., 2018), while another line of research has investigated power fluctuations within the brain (Ding, Melloni, et al., 2017; Ding et al., 2016). As such, integrating the diverse results into a coherent framework of spoken language comprehension remains challenging.

In Chapter 4, we used Mutual Information (MI) analysis to investigate how the brain attunes to linguistic structure and meaning. Conceptually, MI captures the degree of similarity or amount of shared information between two signals. This analysis thus falls into the former of the two categories of research mentioned above: We computed MI to assess the similarity between the acoustic signal and the brain response, using it as a way to study the cortical tracking of speech stimuli. We observed increased MI – that is, more similarity between the acoustic signal and the cortical response – for stimuli containing linguistic structure and meaning compared to jabberwocky and word list controls. We interpret these findings as a reflection of the brain attuning more closely to acoustic cues if higher-level linguistic structures can be inferred from them.

These findings are in line with many previous studies. Keitel et al. (2018), for example, examined MI in different frequency bands and found that the cortical signal was more similar to the acoustics for trials in which listeners had correctly comprehended the sentence. Similarly, Gross et al. (2013) found increased tracking for intelligible speech, as measured by MI in forward compared to reversed stimuli. As such, the emerging picture seems quite clear: As relevant information becomes available to the listener, the brain appears to track the signal more closely, as evidenced by an increase in shared information between the acoustic signal and the cortical response.

Results from studies investigating *spectral power* in the cortical response in specific frequency bands complement these findings. Ding and colleagues (2016; 2017), for example, have shown in several experiments that power in the delta frequency range increases in situations where higher-level linguistic structures such as phrases and sentences can be inferred. Importantly, this increase in power is observed in the brain response (both in MEG (Ding et al., 2016) and EEG data (Ding, Melloni, et al., 2017)), even though there is no corresponding peak in the power spectrum of the acoustic signal. The power increase can thus be taken as a result of the neuronal computations necessary to generate higher-level linguistic structures (Ding, Melloni, et al., 2017; Ding et al., 2016), or, minimally, as the reflection of a chunking mechanism, by which the brain groups the acoustic signal into units of information (cf. Jin et al., 2020; see also Meyer, 2018, for a more detailed overview).

In the current chapter, we report an additional analysis of our previously collected EEG data reported in Chapter 4. Specifically, in the experiment presented in Chapter 4, we recorded EEG signals from participants listening to naturally spoken stimuli: (1) Sentences, containing linguistically meaningful structural and semantic information, as well as sentence prosody; (2) jabberwocky items, containing linguistically plausible structural information and sentence-like prosody, but no straight-forward semantic content; and (3) word lists, containing meaningful lexical items, but no plausible way of combining them into sentences. In addition to these three core conditions, we also included backwards presentations of all stimuli, which allowed us to control for possible differences between conditions in the modulation spectra of the stimuli.

Here, we compute *spectral power* in the delta-theta frequency range (corresponding to the occurrence rate of syllables, words and phrases in our stimuli) in order to investigate whether – in line with the previous research outlined above – we would observe power increases in frequency bands corresponding to these linguistically meaningful units. We hypothesized that, if delta-theta power is indeed related to the generation of higher-level linguistic structures, we would find an increase in spectral power for sentences compared to jabberwocky items and word lists. As such, the analyses reported in the current chapter make a first step towards bridging the perceived gap between the lines of research outlined above: Finding spectral power differences between conditions in the delta-theta range, together with the MI differences reported in Chapter 4, would suggest that accounts of neural activity in these bands as an index of linguistic tracking and generation of linguistic structures are not necessarily mutually exclusive – on the contrary, they might be indexing the same phenomenon of linguistic inference.

5.2 Methods

Participants

For the analyses reported in this chapter, we used the EEG data that was collected for the experiment reported in Chapter 4. 29 native Dutch speakers were included in the analyses reported here (see Chapter 4 for details). All participants were recruited from the Max Planck Institute for Psycholinguistics' participant database, reported normal hearing, and were remunerated for their participation. All participants provided informed consent approved by the Ethics Committee of the Social Sciences Department of Radboud University (Project code: ECSW2014-1003-196a).

Materials

The materials for this experiment are described in detail in Chapter 4 of this dissertation. In short, our experiment employed a 3 (Linguistic Condition) × 2 (Direction) design: Participants listened to sentences, jabberwocky items and word lists (triplets of 80) in one forward and one backward condition, each. All stimuli can be found in the Appendix to Chapter 4 of this thesis.

Procedure

We recorded participants' EEG signals (64-channel EEG system; MPI equidistant montage) while they listened to the stimuli in all conditions. After each trial, they were asked to press a button to advance to the next item. The experiment was run using the Presentation software (Neurobehavioral Systems). See Chapter 4 for a detailed description of the experimental procedure, as well as the online filters and general EEG setup.

Analysis

We used the preprocessed EEG data as described in Chapter 4 and applied a baseline correction using the 200 ms preceding the onset of each stimulus. For the spectral power analysis, each epoch (ranging from stimulus onset to stimulus end) was zero-padded to 6 seconds and fast Fourier transformed using a Hanning window. The averaged power coefficients from each frequency band of interest were then submitted to a linear mixed effects model using lme4 (D. Bates et al., 2015) in R (R Core Development Team, 2012). Models included main effects of Condition (three levels: Sentence, Wordlist, Jabberwocky) and Direction (two levels: Forward, Backward) and their interaction, as well as by-participant random intercepts and random slopes for the main effects and their interaction (except the word-level model, which included only random intercepts). As in the previous chapter, we used treatment coding in all models, with Sentence being the reference level for Condition, and Forward the reference level for Direction. We also computed pairwise comparisons within each direction using estimated marginal means (Tukey correction for multiple comparisons) with emmeans (Length et al., 2018) in R.

5.3 Results

We computed spectral power within three frequency bands of interest that roughly corresponded to the occurrence rates of phrases (0.8-1.1 Hz), words (1.9-2.8 Hz) and syllables (3.5-5.0 Hz). Specifically, we aimed to assess whether spectral power in the brain response increases when linguistic structure and meaning are available to the listener. This would be in line with previous findings (e.g., Ding, Melloni, et al., 2017; Ding et al., 2016).

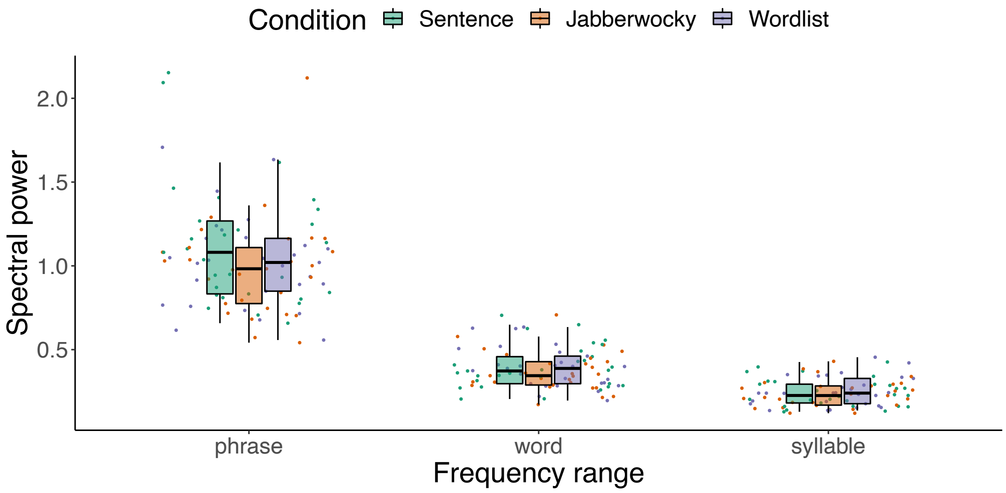

Figure 5.1: Spectral power in the phrase, word and syllable frequency range for sentences (green), jabberwocky items (orange), and word lists (orange). Drops reflect average spectral power per participant, boxplots reflect distribution. Only forward conditions are shown in this plot.

Our analyses revealed condition-dependent changes in spectral power at distinct timescales for the forward conditions (see Figure 5.1). In the phrase frequency band (0.8-1.1 Hz) the mixed effects model revealed a significant effect of

Condition (Sentence = treatment level; Jabberwocky: β = -0.135, SE = 0.023, p < 0.001; Wordlist: β = -0.103, SE = 0.022, $p < 0.001$), indicating that sentences elicited the highest power response in this frequency range. The model also revealed a significant effect of Direction (Forward = treatment level; Backward: β = -0.171, SE = 0.025, $p < 0.001$), indicating that spectral power was generally higher for forward compared to backward stimuli. In addition to this, the model revealed significant Condition*Direction interactions (Jabberwocky*Backward: β = 0.149, SE = 0.022, $p < 0.001$; Wordlist*Backward: β = 0.097, SE = 0.022, $p < 0.001$), indicating that the power differences between conditions were more pronounced for forward compared to backward stimuli (see Appendix for complete model outputs). The estimated marginal means further clarify these findings: We find significant differences only between the Forward conditions, where sentences elicited higher power in the phrase band than jabberwocky items (Δ = 0.135, SE = 0.24, $p < 0.001$) and word lists (Δ = 0.103, SE = 0.022, $p < 0.001$).

In the word frequency band (1.9-2.8 Hz), the mixed effects model revealed a significant difference between the Jabberwocky and Sentence condition (β = -0.028, SE = 0.005, $p < 0.001$), but no significant effect for the Sentence - Wordlist comparison (β = -0.007, SE = 0.005, p = 0.194). The Sentence - Jabberwocky difference was more pronounced between the Forward conditions, as evidenced by a significant Condition*Direction interaction (β = 0.029, SE = 0.007, $p < 0.001$). In addition, the model revealed a significant main effect of Direction (β = -0.042, SE = 0.005, $p < 0.001$), again indicating that spectral power was generally higher for forward than for backward stimuli. These effects are, again, further clarified by the estimated marginal means, where we only find significant pair-wise contrasts in the Forward conditions. Specifically, the estimated marginal means show higher power for sentences compared to jabberwocky items (Δ = 0.028, SE = 0.005, $p < 0.001$), as well as for word lists compared to jabberwocky items (Δ = 0.021, SE = 0.005, $p < 0.001$). Note that the significant effect between the Wordlist and Jabberwocky condition was not detected by our base model, because we used the Sentence condition as the treatment level.

Finally, in the syllable frequency range (3.5-5.0 Hz), the model revealed a significant main effect of Direction (β = -0.018, SE = 0.008, p = 0.044), indicating higher spectral power in the brain response to forward stimuli than the backward controls.

5.4 Discussion

The analyses reported in the current chapter provide additional information to the results from Chapter 4. Both chapters investigated how the brain attunes to linguistic structure and meaning, above and beyond acoustic and prosodic information. While Chapter 4 reported a MI analysis, which tests the similarity between the acoustic signal and the cortical response, the current analyses focused on spectral power. Specifically, we performed a spectral power analysis in frequency bands of interest that were defined based on the occurrence rate of syllables, words, and phrases in our stimuli.

We found that spectral power in the delta-theta frequency range was highest for sentences and lowest for jabberwocky items, indicating that the spectral power response was modulated by compositional structure and meaning. We can exclude that these effects arose only as a response to possible differences in the modulation spectra of the acoustic signals: Interaction terms in the statistical models revealed that power differences between conditions were more pronounced in the forward conditions (compared to the backward controls) in the phrase and word frequency bands, which indicates that it was, indeed, linguistic information that drove the observed effects. Interestingly, we found a difference between sentences and word lists in the phrase, but not the word frequency range. This observation of graded differences between conditions suggests that cortical responses are enhanced in frequency ranges at which the most linguistically meaningful representations are available.

Delta-band neuronal activity has previously been linked to distinct linguistic processes. For example, some research has suggested that delta-band activity reflects the tracking of prosodic units and intonational phrase boundaries (e.g., Bourguignon et al., 2013; Ghitza, 2017), while other research has linked increases in delta power to the generation of intrinsic, hierarchical linguistic structures (e.g., Ding, Melloni, et al., 2017; Ding et al., 2016). Our data offer novel, more nuanced insights into these effects and their interpretation. Specifically, the increase in delta power for sentences compared to jabberwocky items in the phrase frequency range suggests that the effects are, indeed, related to linguistic structure and meaning, above and beyond sentence-level prosody.

Note, however, that we do not claim that our results are exclusively related to the generation of hierarchical linguistic processing: It is still possible (and plausible) that prosodic units and intonational phrase boundaries play an important role in the patterns of cortical activity observed here and in previous literature (cf. Teoh, Cappelloni, & Lalor, 2019). Crucially, though, we argue that our results cannot be explained by prosodic or intonational tracking alone – the delta-band power response appears to be modulated by the linguistic content available in our current experiment. For accounts of language processing that view delta activity mainly as a result of prosodic tracking, this means that prosodic tracking appears to be enhanced for stimuli that carry meaningful linguistic information.

Relatedly, our results add to the ongoing discussion about delta-band activity as a chunking mechanism that divides the acoustic signal into units: In line with previous research (e.g., Jin et al., 2020), our findings suggest that this alleged chunking mechanism is modulated by higher-level knowledge. Here, we find that linguistic content modulates the delta-band power response. As such, chunking-based accounts need to take into consideration the role of higher-level linguistic information.

As we have mentioned in the previous chapter, it is important to distinguish between observed cortical response patterns and the underlying mechanism for language comprehension; as Hagoort (2020, p. 5) points out, "brain rhythms are not themselves the mechanism that computes meaning". This is especially relevant when relating it to the different ways of experimentally investigating cortical activity during spoken language comprehension outlined in the introduction to this chapter. As we have briefly summarized, one line of research is based on measures of phase coherence and Mutual Information and reports an increase in similarity between the brain response and the acoustic signal for linguistically relevant information (e.g., Gross et al., 2013; Keitel et al., 2018). Another line of research has investigated spectral power in the cortical response, reporting power increases in the delta band when higher-level linguistic units can be generated (Ding, Melloni, et al., 2017; Ding et al., 2016). Intuitively, this would entail a decrease in similarity between the cortical and the acoustic signal, as the brain response shows peaks in the power spectrum that are not present in the acoustic signal. These lines of research thus make seemingly opposing predictions: One line of research predicts an increase in dissimilarity, with delta power increasing in the cortical signal but not in the acoustics, while the other line of research predicts an increase in similarity between these two signals.

While the current study was not designed to disentangle these two lines of research, the results reported here and in the previous chapter suggest that different analysis techniques can provide converging evidence: We consistently find that linguistic structure and meaning influence the cortical response, when measured both by means of MI analysis and through spectral power analysis. While these converging findings are encouraging, we believe that it is worthwhile to consider the assumptions that different analysis techniques make about the cortical computations underlying spoken language comprehension. Our current findings are in line with at least two types of models for spoken language comprehension: 1) Models that posit power increases in the delta band as a result of a neural “chunking” mechanism (cf. Jin et al., 2020), and 2) models that propose phase resetting as a possible mechanism by which the brain attunes to linguistically relevant information in the signal (e.g., Martin, 2020; Rimmele et al., 2018). Future research will have to investigate in detail how power fluctuations in the acoustic signal relate to power fluctuations in the brain response, and how these, in turn, relate to measures of oscillatory activity and phase entrainment (see Obleser & Kayser, 2019, for a discussion of terms such as “entrainment”, which also need to be defined carefully in light of this consideration).

To sum up, we find that the cortical response is modulated by linguistic structure and content. We argue that the Mutual Information and power differences we observe here and in the previous chapter arise as a result of the computations carried out during spoken language comprehension. Investigating and modeling the specifics of the cortical computations that give rise to these results remains a challenging and exciting objective for future research.

Appendix

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	1.124	0.062	29.602	18.176	< 0.001		
Condition[T.Jabberwocky]	-0.135	0.023	47.265	-5.778	< 0.001		
Condition[T.Wordlist]	-0.103	0.022	51.345	-4.771	< 0.001		
Direction[T.Backward]	-0.171	0.025	51.196	-6.816	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.149	0.022	9148.579	6.923	< 0.001		
Cond.[T.Word.]:Dir.[T.Back.]	0.097	0.022	9148.579	4.498	< 0.001		
Random Effects							
Intercept|Participant						0.108	0.328
Cond.[T.Jabb.]|Part.						0.009	0.095
Cond.[T.Word.]|Part.						0.007	0.083
Dir.[T.Back.]|Part.						0.012	0.107

Table 5A.1: Mixed-effects logistic regression results for spectral power in the phrase frequency band. Sentence = treatment level for Condition, Forward = treatment level for Direction.

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.135	0.024	48.4	5.699	< 0.001
Sentence - Wordlist	0.103	0.022	52.8	4.708	< 0.001
Jabberwocky - Wordlist	-0.031	0.024	48.7	-1.327	0.387
Direction = Backward					
Sentence - Jabberwocky	-0.014	0.024	48.4	-0.603	0.819
Sentence - Wordlist	0.007	0.022	52.8	0.306	0.950
Jabberwocky - Wordlist	0.021	0.024	48.7	0.892	0.648

Table 5A.2: Estimated marginal means for spectral power in the phrase frequency band. P-value adjustment: tukey method for comparing a family of 3 estimates.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	0.397	0.020	0.307	19.658	< 0.001		
Condition[T.Jabberwocky]	-0.028	0.005	9235	-5.298	< 0.001		
Condition[T.Wordlist]	-0.007	0.005	9235	-1.298	0.194		
Direction[T.Backward]	-0.042	0.005	9235	-8.116	< 0.001		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.029	0.007	9235	3.877	< 0.001		
Cond.[T.Word.]:Dir.[T.Back.]	0.003	0.007	9235	0.515	0.607		
Random Effects							
Intercept|Participant						0.011	0.107

Table 5A.3: Mixed-effects logistic regression results for spectral power in the word frequency band. Sentence = treatment level for Condition, Forward = treatment level for Direction.

contrast	estimate	*SE*	df	t ratio	*p*
Direction = Forward					
Sentence - Jabberwocky	0.028	0.005	9240	5.297	< 0.001
Sentence - Wordlist	0.007	0.005	9240	1.297	0.397
Jabberwocky - Wordlist	-0.021	0.005	9240	-3.999	< 0.001
Direction = Backward					
Sentence - Jabberwocky	-0.001	0.005	9240	-0.185	0.981
Sentence - Wordlist	0.003	0.005	9240	0.569	0.837
Jabberwocky - Wordlist	0.004	0.005	9240	0.754	0.731

Table 5A.4: Estimated marginal means for spectral power in the word frequency band. P-value adjustment: tukey method for comparing a family of 3 estimates.

Effects	Estimate	*SE*	df	t value	*p*	Variance	*SD*
Fixed Effects							
Intercept	0.246	0.015	28.002	15.969	< 0.001		
Condition[T.Jabberwocky]	-0.009	0.004	34.040	-1.949	0.060		
Condition[T.Wordlist]	0.006	0.006	28.119	1.052	0.302		
Direction[T.Backward]	-0.018	0.008	27.946	-2.110	0.044		
Cond.[T.Jabb.]:Dir.[T.Back.]	0.010	0.010	28.157	1.035	0.309		
Cond.[T.Word.]:Dir.[T.Back.]	-0.007	0.010	27.942	-0.650	0.521		
Random Effects							
Intercept|Participant						0.007	0.082
Cond.[T.Jabb.]|Part.						0.0003	0.017
Cond.[T.Word.]|Part.						0.001	0.027
Dir.[T.Back.]|Part.						0.002	0.041
Cond.[T.Jabb.]:Dir.[T.Back.]|Part.						0.002	0.048
Cond.[T.Word.]:Dir.[T.Back.]|Part.						0.002	0.048

Table 5A.5: Mixed-effects logistic regression results for spectral power in the syllable frequency band. Sentence = treatment level for Condition, Forward = treatment level for Direction.

6 General discussion

How humans understand language from an acoustic signal remains one of the most widely studied and intriguing questions in the fields of psycholinguistics and cognitive neuroscience. The present dissertation aimed at investigating spoken language comprehension through the lens of perceptual inference and cue integration, asking several questions: How do listeners combine perceptual, acoustic cues and linguistic, knowledge-based cues? What types of information do listeners rely on to draw inferences in the presence of uncertainty? What kind of neural activation patterns might underlie the process of generating structure and meaning from sound? In the following, I will first provide a brief summary of the main findings from each chapter. After that, I will discuss the implications of the findings presented in this thesis in the broader context of previous literature. In addition, I will outline possible directions for future research, more generally.

6.1 Summary of main findings

In Chapter 2, I presented an eye-tracking experiment that aimed to test how listeners use signal-based cues to infer knowledge-based cues and predict upcoming referents during spoken language comprehension. Using the minimal pair *ein/eine* in German in combination with a contextual speech rate manipulation, we found that 1) listeners used the lower-level, perceptual cue to infer a higher-level, linguistic cue, and 2) listeners used this inferred linguistic cue to make predictions about upcoming referents.

In Chapter 3, I presented an eye-tracking experiment that aimed to further test the intricate interplay between knowledge-based and signal-based cues during spoken language comprehension. Specifically, we investigated the interplay between morphosyntactic knowledge and contextual speech rate and how listeners combine and integrate these two cues in situations of uncertainty. Overall, we found that participants used both sources of information as soon as they became available, even in an uncertain experimental situation.

Two findings from this experiment are particularly noteworthy: Firstly, the knowledge-based, morphosyntactic cue preceded the signal-based, acoustic cue in time in the experiment, yet we observed speech rate effects even after the potentially disambiguating determiner (*de/het*). We take this as evidence that contextual speech rate effects are robust and arise potentially automatically during phoneme perception (e.g., Bosker et al., 2017; Reinisch & Sjerps, 2013; Toscano & McMurray, 2015). Secondly, there was no unambiguously "correct" target in the experiment; in fact, participants were free to decide which cue to "rely on" more heavily. Our analyses of individual strategies confirmed that participants used different strategies during the experiment, with some listeners "valuing" the acoustic, speech rate cue more strongly, and others "preferring" the morphosyntactic cue. Both groups of participants, even those who relied more strongly on the morphosyntactic information carried by the determiner, showed effects of contextual speech rate. Together with the results reported in Chapter 2, these findings support models of cue integration, where different sources of information can be weighted flexibly depending on their reliability in a given situation.

In Chapter 4, I presented an EEG experiment that aimed to look at cue integration from a broader, more naturalistic perspective. Participants listened to naturally spoken sentences, jabberwocky items and word lists, as well as reversed controls of each condition. The experiment did not involve any task except listening attentively. By means of Mutual Information analysis, we investigated cortical tracking as a potential mechanism by which cue integration might be instantiated in the brain during spoken language comprehension. We found that Mutual Information between acoustic stimuli and the brain response was highest for the most structured types of stimuli, showing that cortical tracking is enhanced for acoustic signals that carry linguistic structure and meaning. Taken together, the findings from this chapter suggest that cortical activity is not exclusively driven by temporal regularities at distinct timescales – instead, neural responses appear to be modulated by the linguistic information available at those timescales.

Finally, in Chapter 5, I presented an additional power analysis on the EEG data collected for the experiment presented in Chapter 4. In line with previous findings, spectral power in the delta-theta band was stronger for stimuli that carried linguistic information from which structure and meaning could be inferred. These results are complementary to the ones reported in Chapter 4. Additionally, they offer novel insights into how two different analysis techniques (measures of coherence between two signals, as captured by MI analysis, and measures of increased cortical activity, as captured by spectral power analysis) can provide converging evidence.

6.2 Knowledge-based and signal-based cues

The current thesis investigated the interplay between knowledge-based and signal-based cues within a framework of cue integration. Cue integration posits that signal-based and knowledge-based cues are weighted and integrated in a flexible, dynamic way, depending on their availability and reliability in a given situation.

As stated throughout this thesis, defining a "cue" is far from trivial. Here, I have made a distinction between "knowledge-based" and "signal-based" cues for spoken language comprehension, trying to distinguish between cues that are available to the listener as part of a physical signal (contextual speech rate in Chapters 2 and 3; the amplitude envelope in Chapters 4 and 5) and cues that are available to the listener as acquired, stored knowledge (morphosyntactic information in Chapters 2 and 3; lexical and combinatorial linguistic knowledge in Chapters 4 and 5).

While helpful for the sake of this thesis, this dichotomy between knowledge-based and signal-based cues also raises new questions. Most obviously, it requires a theory of where knowledge-based cues might come from. Since they are not unambiguously measurable as a property of the environment, they must be available to the listener either as innate knowledge, or as learned representations. Cue integration theories can, in principle, accommodate both of these possibilities: Knowledge-based cues might be available in the form of innate priors, or they might arise through learning from purely sensory information. Of course, each of these possibilities raises a whole new plethora of difficult questions that have been investigated and debated for decades in the field of language acquisition; see, e.g., Gervain and Mehler (2010) for a comprehensive review.

If, as Gervain and Mehler (2010) suggest, language acquisition requires a combination of innate knowledge and learning, then future questions for cue integration models of spoken language comprehension may include the following: Which specific knowledge-based cues are innate, and which ones have to be learned from sensory input? If some knowledge-based cues are indeed present as innate priors, then how exactly are they neurally implemented in the neonate brain? For non-innate knowledge-based cues, how can inferences be made from strictly sensory information – in Martin's (2016, p. 12) words, "how do you [infer] something if you don't know what it is you're trying to [infer]"?

Martin (2016) points to a model of relational concept learning ("Discovery of Relations by Analogy" (DORA); e.g., Doumas, Hummel, and Sandhofer, 2008), where hierarchical concepts (analogous to knowledge-based cues) are learned from linear inputs by making use of timing information. Interestingly, Martin and Doumas (2017) show that DORA exhibits oscillatory patterns of activation that resemble the ones reported by Ding and colleagues in human EEG and MEG data (Ding et al., 2016). Note that these patterns are also in line with our findings from spectral power analysis in Chapter 5. As such, DORA offers an interesting computational model of how (at least some of the) knowledge-based cues in a cue integration model of spoken language comprehension might emerge.

On a related note, it seems highly unlikely that a given combination of sensory cues will always unambiguously yield a distinct, categorical percept – in fact, the experiments reported in Chapters 2 and 3 capitalized on exactly this type of ambiguity. Perceptual ambiguities can be described in terms of cue integration models: In ambiguous situations, the integrated estimate for a linguistic percept (computed from a set of weighted and normalized cues) will have a lot of variance, thus resulting in a relatively ”unstable” percept.

There is, however, a long history of research suggesting that speech perception is categorical and deterministic: Listeners tend to perceive sounds as either belonging to a perceptual category or not (e.g., Liberman, Cooper, and Shankweiler, 1967; Harnad, 1987; Blumstein, Myers, and Rissman, 2005; see Goldstone and Hendrickson, 2010, for a general overview). So, how does one get from a set of probabilistic cues and estimates to categorical percepts and representations? It seems plausible that knowledge-based cues might play a crucial role in bridging this apparent gap, and such top-down effects during speech comprehension have been researched extensively (e.g., Connine & Clifton, 1987; Fox, 1984; Ganong, 1980; Martin et al., 2017; Pitt & Samuel, 1993). But how the step is made from a probabilistic, potentially unreliable estimate to a categorical percept remains to be investigated.

6.3 Towards an integrated theory of spoken language comprehension

The aim of this thesis was to investigate spoken language comprehension as perceptual inference, as formalized by the cue integration model of language processing (Martin, 2016, 2020). As a conclusion to this thesis, I discuss how the research I presented here relates to other models of spoken language comprehension. I also briefly reflect on what different models might "learn" or "gain" from each other.

Hagoort (2005, 2013, 2014) proposed the Memory-Unification-Control (MUC) model, which specifies the system for language comprehension and production both in terms of processing components and the cortical networks that are at play. As the name implies, the MUC model consists of three components: 1) a Memory component located in the temporal cortex, from which stored linguistic knowledge is retrieved; 2) a Unification component located in the inferior frontal cortex, where smaller units are integrated into higher-level structures; and 3) a Control component including areas in the prefrontal cortex and anterior cingulate cortex, where language is related to action and higher-level communicative goals. Crucially, these three components operate on multiple levels of linguistic "granularity": phonological, syntactic and semantic units retrieved from memory can all be integrated into larger structures in the unification network.

The MUC model is especially interesting with regards to the findings reported in Chapters 4 and 5, because it makes specific predictions for the experimental conditions in these chapters. Recall that we used sentences, word lists, and jabberwocky items (as well as backward controls of all three) as our experimental conditions. Sentences contained lexical items that could be retrieved from the Memory component and combined into meaningful higher-level representations in the Unification component of the model, whereas word lists could not be "unified".

Importantly, the MUC model makes clear predictions about the localization of the effects observed in Chapters 4 and 5. For example, we would expect increased activity in the inferior frontal cortex for sentences compared to word lists, because "more" unification can take place. In fact, these hypotheses are in line with previous findings from Hultén, Schoffelen, Uddén, Lam, and Hagoort (2019), who observed increased cortical activity in temporal and inferior frontal regions for words embedded in sentences compared to the same words occurring in word lists. For the jabberwocky condition, it is not entirely clear which types of representations listeners might have retrieved from memory in our experiment – our jabberwocky stimuli were pseudowords containing morphosyntactic information, which could potentially have been retrieved from memory and combined into semantically void “dummy” linguistic structures, resulting in at least some activation in the Memory and Unification network.

In general, the MUC model makes predictions that are very much in line with those from theories of cue integration. In fact, one might argue that the Unification component represents the cortical “hub” where perceptual inference might be computed by combining cues into higher-level structures and robust percepts. Intriguingly, the MUC model also suggests that unification operations might take place at different timescales and levels of linguistic hierarchy, which one might take as corresponding to incremental, cascaded processing in the cue integration framework. Additionally, looking at cue integration through the lens of the MUC model allows for predictions about the cortical organization of cues: Specifically, knowledge-based cues might be stored in structures related to the Memory component, i.e., the temporal cortex.

Another influential account of spoken sentence comprehension is the auditory language comprehension model proposed by Friederici (2002, 2011, 2012). The neuroanatomical architecture of this model includes roughly the same cortical regions as the MUC model, comprising temporal and inferior frontal regions. During auditory sentence comprehension, bottom-up activation is passed from primary auditory cortex to anterior (in the case of semantic information) and posterior IFG (in the case of syntactic information) via the ventral stream. Semantic top-down information is back-projected from anterior IFG to temporal regions via the ventral stream, while the dorsal stream allows for top-down information related to grammatical information to flow from posterior IFG to temporal regions. Syntactic and semantic information are then combined in the temporal cortex (Friederici, 2012).

Again, the model proposed by Friederici (2002, 2012) is generally compatible with theories of cue integration: It specifies how bottom-up information (potentially derived from signal-based cues) spreads in the language comprehension network, allowing for syntactic and semantic inferences to be made (drawing on knowledge-based cues), which are then integrated into a linguistic percept (i.e., a fully comprehended sentence). Note, however, that Friederici (2002) assumes relatively independent processing of syntactic and semantic information during the first stages of sentence comprehension. This is somewhat difficult to integrate with accounts of cue integration, where cues can interact across different levels of linguistic hierarchy and cue weights can be updated flexibly. Nevertheless, the model makes interesting predictions about where in the brain perceptual inference might be computed, and about the flow of information along the ventral and dorsal streams, which would be interesting to investigate from a cue integration perspective.

A third influential model of language comprehension is the extended Argument Dependency Model (eADM) proposed by Bornkessel and Schlesewsky (2006; see also Bornkessel-Schlesewsky and Schlesewsky, 2013), where sentence comprehension is achieved through 1) time-independent (i.e., insensitive to order) unification operations related to auditory representations in the ventral stream, and 2) time-dependent (i.e., sensitive to order) syntactic structure-building in the dorsal stream. Both of these processes happen in parallel, after which the output from the computations in both streams is integrated in the frontal cortex. Interestingly, eADM explicitly suggests a hierarchical processing network from “lower” brain areas (such as primary auditory cortex) to “higher” areas (such as temporal cortex), while also allowing for (at least some amount of) feedback. This could be translated into cue integration terms, suggesting that “lower” areas in the model might be active during the first, sensory steps of the inferential process, while “higher” areas operate on the representations inferred from these earlier steps in a cascaded fashion.

To summarize, the models suggested by Hagoort (2003, 2005, 2013), Friederici (2002, 2012) and Bornkessel and Schlesewsky (2006; Bornkessel-Schlesewsky and Schlesewsky, 2013) make specific predictions about the neuroanatomical details of language comprehension. This is in contrast to the cue integration model proposed by Martin (2016, 2020), which does not specify anatomical brain regions that might be “specialized” for perceptual inference (cf. Martin, 2016). The models outlined above are supported by a wealth of research, so models of cue integration may benefit from integrating these neuroanatomical considerations. This would allow making explicit predictions about the cortical architecture that might underlie cue integration and perceptual inference during language processing.

Conversely, most other models don’t specify in enough detail how the required computations might be instantiated in the brain. For example, all three models outlined above propose that the inferior frontal gyrus is involved in processes similar to perceptual inference – yet what exactly these computations might be, and on what kinds of representations they might operate, remains elusive. Cue integration contains a mathematical formalization of exactly these types of computations, offering a step towards the formulation of exact hypotheses that are falsifiable through experimental and computational modelling work (see Martin

(2016; 2020) for more detailed discussions of some of the ways in which cue integration models differ from previous work).

6.4 Future research directions

Our initial plan for this thesis was to include an additional experiment in which we had hoped to combine the approaches taken in the previous chapters. Specifically, we started testing participants for an EEG experiment that combined our three critical conditions from Chapter 4 with a speech rate manipulation. Participants listened to sentences, jabberwocky items, and word lists at four different time-compression factors κ ($\kappa = 1$: “original” speech rate; $\kappa = 4$: fastest speech rate, resulting in stimuli that were 2, 3 or 4 times faster than the original). Stimuli were presented in blocks and participants' task was to attentively listen to the audio recordings. Data collection was, unfortunately, stalled due to safety measures related to COVID-19.

Based on the results from earlier studies and our previous experiments, we had two specific hypotheses for this experiment. First, our previous results (as well as previous research) indicate that contextual speech rate is used rapidly, and potentially even automatically, during spoken language comprehension. At the same time, the degree of cortical tracking of the speech envelope is closely related to intelligibility, and has been shown to decrease as intelligibility deteriorates (see, e.g., Giraud & Poeppel, 2012; Peelle & Davis, 2012, for overviews). We therefore hypothesized that tracking of the amplitude envelope (as captured by Mutual Information between the EEG signals and the audio recordings) would decrease as speech rate increased and stimuli became less intelligible (cf. Kösem et al., 2018). Based on previous findings from behavioural studies (e.g., Bosker & Ghitza, 2018), we expected that intelligibility (and hence MI) would decrease at compression rates of $\kappa = 3$ and $\kappa = 4$. We were particularly interested in the different patterns that we might observe in the three linguistic conditions: Previous research has mostly investigated envelope tracking by contrasting sentences with either noise-vocoded speech or reversed speech, but not as a function of linguistic content. Here, we could have gained additional insights into the factors influencing the tracking of contextual speech rate from our jabberwocky items and word lists in comparison with the sentence condition. It would have been interesting to see whether the speech rate manipulation would have caused tracking of the envelope to break down faster in some of the conditions compared to the others, suggesting that tracking of the speech signal is not only rate-, but also information-dependent.

Second, our previous results from Chapters 4 and 5 suggest that tracking of the signal in distinct frequency bands increases for stimuli that carry structured, meaningful linguistic information. In Chapter 4, we briefly discussed that this increased tracking likely arises as the result of the computations carried out by populations of neurons in these distinct frequency bands. Crucially, we emphasized that we do not think that the increased MI values for meaningful stimuli were due to an intrinsic phrase-level or sentence-level oscillator, simply because naturally spoken language is way too variable for such a fixed-frequency oscillator to be particularly useful. Our planned final experiment would have given further insights into this question. Specifically, we hypothesized that increased tracking would occur as a result of listeners computing inferred linguistic structures (words and phrases) on different timescales. Varying the speech rate of our stimuli would have, by definition, varied the timescales at which inferences could be made. If we had observed increased tracking for sentences compared to jabberwocky items and word lists in the phrase and word frequency regardless of speech rate (of course within the boundaries of intelligibility), this would have been evidence for our hypothesis. Conversely, this finding would have been difficult to integrate with accounts of language comprehension positing that oscillations in only the delta frequency range serve as the main mechanism for structure-building.

Finally, our speech rate manipulation would have helped us address a potential shortcoming of our previous experiment: Higher-level linguistic structures such as phrases, clauses and sentences are usually longer than lower-level structures such as words and syllables and, as such, they occur at lower frequencies. This can make it difficult to study structures beyond the phrase level, as it becomes hard to distinguish language-related cortical activity from drift that occurs at lower frequencies in EEG recordings (cf. Alday, 2019). Increasing the speech rate of our stimuli would have mitigated this potential problem, allowing us to investigate linguistic structures beyond the phrase level by shortening their length and thus “shifting” the frequencies of interest higher up.

There are, of course, many other possible avenues for future research. As I have mentioned throughout this thesis, we investigated very specific cues: contextual speech rate and morphosyntactic information in Chapters 2 and 3, and linguistic structure and meaning, together with prosodic information, in Chapters 4 and 5. These cues are undoubtedly important for spoken language comprehension, but at the same time they constitute only a small fraction of all the information available to the brain in most situations outside of highly controlled experimental settings. At least two important points follow from this: the need for experimental designs that investigate a wider variety of cues for language comprehension, and studies that investigate language comprehension in more natural settings.

First, as mentioned throughout this thesis, it is important to investigate more than the limited number of cues that were examined here. This applies not only to the examples of signal-based and knowledge-based cues given in this thesis; rather, cue integration models for spoken language comprehension need to take into consideration the full, multi-modal picture of language comprehension (Martin, 2016). For example, future work could investigate how cues from different modalities, such as speech and gesture, are weighted and integrated (e.g., Drijvers & Özyürek, 2017; Holler & Levinson, 2019). Even more generally, theories of cue integration should also be tested beyond the spoken modality, for example for sign languages.

Second, several researchers have emphasized the need for studying language outside of highly controlled experimental settings (Alday, 2019; Alexandrou et al., 2020; Verga & Kotz, 2019). Our EEG experiment was a first step in that direction, since we used naturally spoken stimuli without artificially inducing rhythmicity in our stimuli or removing prosodic cues, like previous studies have done (e.g., Ding, Melloni, et al., 2017; Ding et al., 2016). We also chose not to include a behavioral task in the experiment, because language comprehension usually does not require any additional tasks like outlier detection or phoneme or word monitoring. However, much remains to be done in order to study language processing in a truly natural setting. This could, for example, be done in a controlled way by using not only single sentences, but also longer speech stimuli such as stories, recorded conversations, movies, or talks (e.g., Brennan et al., 2012; Gross et al., 2013).

Throughout this thesis, I have treated cue weights as a somewhat abstract concept, without going into any detail concerning their statistical properties. Defining or experimentally establishing the numerical details of cue weights was not the goal of my thesis, yet cue weights are an integral part of cue integration theories, and the mathematical formalization of cue integration is part of what makes it so promising. As such, it would be very interesting to gain a deeper understanding of the weights that listeners assign to distinct cues. Possible ways of studying this include carefully controlled psychophysical experiments and computational modelling approaches (e.g., Alais & Burr, 2004; Bejjanki, Clayards, Knill, & Aslin, 2011; Ernst & Banks, 2002; Jacobs, 1999; Knill & Saunders, 2003; Toscano & McMurray, 2010).

Chapter 3 found that different listeners appeared to "prefer" one cue over the other (at least during our experiment), suggesting that individual listeners might generally be flexible in how they integrate different sources of information, and that individual differences might exist in the way that listeners use different cues for comprehension. This observation is not entirely surprising, given the wealth of research on individual differences in language processing (e.g., E. Bates, Dale, & Thal, 1995; Kidd, Donnelly, & Christiansen, 2018). Hence, it would be very interesting to study individual differences in cue integration "strategies" in more detail. Possible questions include the following: Does the flexibility in cue weighting observed in Chapter 3 generalize to other cues? Are some listeners more likely to adjust or update their beliefs regarding the reliability of certain cues in a given situation? If so, which factors might influence this relative adaptability, and how do they relate to other domains of cognition?

6.5 General conclusion

To summarize, the aim of this doctoral thesis was to shed novel light on how information from distinct levels of the linguistic hierarchy might be integrated during spoken language comprehension. Specifically, I set out to investigate spoken language comprehension through the theoretical lens of perceptual inference, which has been formalized in models of cue integration. The results from this thesis suggest that knowledge-based and signal-based cues interact across levels of linguistic hierarchy, and that listeners are flexible in how they weight and integrate different types of information. Further, the findings provide a first step towards understanding the neural computations that might be at play when inferring meaning from sound. As such, the results can inform current models of spoken language comprehension (both cue integration and beyond). Future work may explore the intricate interplay between cues from different hierarchical levels of representation in more detail, specifically focusing on how exactly cue integration and its sub-computations might be instantiated in populations of neurons in the brain.

References

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. doi: 10.1016/j.cub.2004.01.029

Alday, P. M. (2019). M/EEG analysis of naturalistic stories: A review from speech to language processing. *Language, Cognition and Neuroscience*, *34*(4), 457–473. doi: 10.1080/23273798.2018.1546882

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2020). Cortical entrainment: What we can learn from studying naturalistic speech perception. *Language, Cognition and Neuroscience*, *35*(6), 681–693. doi: 10.1080/23273798.2018.1518534

Altmann, G. T. (2011). Language can mediate eye movement control within 100milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, *137*(2), 190–200. doi: 10.1016/j.actpsy.2010.09.009

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi: 10.1016/S0010-0277(99)00059-1

Apfelbaum, K. S., Bullock-Rest, N., Rhone, A. E., Jongman, A., & McMurray, B. (2014). Contingent categorisation in speech perception. *Language, Cognition and Neuroscience*, *29*(9), 1070–1082. doi: 10.1080/01690965.2013.824995

Au, W. W. L., & Hastings, M. C. (2008). *Principles of Marine Bioacoustics*. New York, NY: Springer US. doi: 10.1007/978-0-387-78365-9

Audacity Team. (2019). *Audacity(R): Free audio editor and recorder.* Audacity Team.

Baese-Berk, M. M., Dilley, L. C., Henry, M., Vinke, L., Banzina, E., & Pitt, M. A. (2013). Distal speech rate influences lexical access [Abstract]. *Abstracts of the Psychonomic Society*, *18*, 191.

Baese-Berk, M. M., Dilley, L. C., Henry, M. J., Vinke, L., & Banzina, E. (2019). Not just a function of function words: Distal speech rate influences perception of prosodically weak syllables. *Attention, Perception, & Psychophysics*. doi:

10.3758/s13414-018-1626-4

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

Bates, E., Dale, P. S., & Thal, D. (1995). Individual Differences and their Implications for Theories of Language Development. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language* (pp. 95–151). Oxford, UK: Blackwell Publishing Ltd. doi: 10.1111/b.9780631203124.1996.00005.x

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. *Mechanisms of language acquisition*, 157–193.

Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. *The crosslinguistic study of sentence processing*, *3*, 73–112.

Bejjanki, V. R., Clayards, M., Knill, D. C., & Aslin, R. N. (2011). Cue Integration in Categorical Tasks: Insights from Audio-Visual Speech Perception. *PLoS ONE*, *6*(5), e19812. doi: 10.1371/journal.pone.0019812

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM review*, *59*(1), 65–98.

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The Perception of Voice Onset Time: An fMRI Investigation of Phonetic Category Structure. *Journal of Cognitive Neuroscience*, *17*(9), 1353–1366. doi: 10.1162/0898929054985473

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer.*

Bölte, J., & Connine, C. M. (2004). Grammatical gender in spoken word recognition in German. *Perception & Psychophysics*, *66*(6), 1018–1032. doi: 10.3758/BF03194992

Bonhage, C. E., Meyer, L., Gruber, T., Friederici, A. D., & Mueller, J. L. (2017). Oscillatory EEG dynamics underlying automatic chunking during sentence processing. *NeuroImage*, *152*, 647–657. doi: 10.1016/j.neuroimage.2017.03.018

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787–821. doi: 10.1037/0033-295X.113.4.787

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: A new dorsal–ventral stream model of sentence comprehension. *Brain and Language*, *125*(1), 60–76. doi: 10.1016/j.bandl.2013.01.010

Bosker, H. R. (2017a). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, & Psychophysics*, *79*(1), 333–343. doi: 10.3758/s13414-016-1206-4

Bosker, H. R. (2017b). How our own speech rate influences our perception of others. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1225–1238. doi: 10.1037/xlm0000381

Bosker, H. R., & Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalisation. *Language, Cognition and Neuroscience*, *33*(8), 955–967. doi: 10.1080/23273798.2018.1439179

Bosker, H. R., & Reinisch, E. (2017). Foreign Languages Sound Fast: Evidence from Implicit Rate Normalization. *Frontiers in Psychology*, *8*. doi: 10.3389/fpsyg.2017.01063

Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, *94*, 166–176. doi: 10.1016/j.jml.2016.12.002

Bourguignon, M., De Tiège, X., de Beeck, M. O., Ligot, N., Paquier, P., Van Bogaert, P., ... Jousmäki, V. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, *34*(2), 314–326. doi: 10.1002/hbm.21442

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, *120*(2), 163–173. doi: 10.1016/j.bandl.2010.04.002

Brodbeck, C., Presacco, A., & Simon, J. Z. (2018). Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. *NeuroImage*, *172*, 162–174. doi: 10.1016/j.neuroimage.2018.01.042

Brown, M., Dilley, L. C., & Tanenhaus, M. K. (2012). Real-time expectations based on context speech rate can cause words to appear or disappear. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*(34).

Buehlmann, C., Mangan, M., & Graham, P. (2020). Multimodal interactions in insect navigation. *Animal Cognition*. doi: 10.1007/s10071-020-01383-2

Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*, *18*(6), 300–309. doi: 10.1016/j.tics.2014.02.005

Carandini, M., & Heeger, D. (1994). Summation and division by neurons in primate visual cortex. *Science*, *264*(5163), 1333–1336. doi: 10.1126/science.8191289

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51–62. doi: 10.1038/nrn3136

Cho, S.-J., Brown-Schmidt, S., & Lee, W.-y. (2018). Autoregressive Generalized Linear Mixed Effect Models with Crossed Random Effects: An Application to Intensive Binary Time Series Eye-Tracking Data. *Psychometrika*, *83*(3), 751–771. doi: 10.1007/s11336-018-9604-2

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. doi: 10.1017/S0140525X12000477

Cogan, G. B., & Poeppel, D. (2011). A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *Journal of Neurophysiology*, *106*(2), 554–563. doi: 10.1152/jn.00075.2011

Connine, C. M., & Clifton, C. (1987). Interactive Use of Lexical Information in Speech Perception. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(2), 291–299.

Dell, G. S., & Chang, F. (2013). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394–20120394. doi: 10.1098/rstb.2012.0394

Dilley, L. C., & Pitt, M. A. (2010). Altering Context Speech Rate Can Cause Words to Appear or Disappear. *Psychological Science*, *21*(11), 1664–1670. doi: 10.1177/0956797610384743

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). *Frontiers in Human Neuroscience*, *11*. doi: 10.3389/fnhum.2017.00481

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. doi: 10.1038/nn.4186

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention Is Required for Knowledge-Based Sequential Grouping: Insights from the Integration of Syllables into Words. *The Journal of Neuroscience*, *38*(5), 1178–1188. doi: 10.1523/JNEUROSCI.2606-17.2017

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, *81*, 181–187. doi: 10.1016/j.neubiorev.2017.02.011

Ding, N., & Simon, J. Z. (2013). Power and phase properties of oscillatory neural responses in the presence of background activity. *Journal of Computational Neuroscience*, *34*(2), 337–343. doi: 10.1007/s10827-012-0424-6

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768. doi: 10.1016/j.neuroimage.2013.06.035

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *115*(1), 1–43. doi: 10.1037/0033-295X.115.1.1

Drijvers, L., & Özyürek, A. (2017). Visual Context Enhanced: The Joint Contribution of Iconic Gestures and Visible Speech to Degraded Speech Comprehension. *Journal of Speech, Language, and Hearing Research*, *60*(1), 212–222. doi: 10.1044/2016_JSLHR-H-16-0101

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. doi: 10.1080/17470218.2017.1310261

Eisner, F., & McQueen, J. M. (2018). Speech perception. In *Steven's Handbook of Experimental Psychology and Cognitive Neuroscience* (Vol. 3, pp. 1–46). Wiley Online Library.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi: 10.1038/415429a

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. doi: 10.1016/j.tics.2004.02.002

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, *37*(4), 452–465. doi: 10.1016/j.wocn.2009.07.006

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. doi: 10.1006/jmla.1999.2660

Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. doi: 10.1111/1469-8986.3920133

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*(6), 429–442. doi: 10.1038/nrn3503

Fox, R. A. (1984). Effect of Lexical Status on Phonetic Categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(4), 526–540.

Frank, S. L., & Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PLOS ONE*, *13*(5), e0197304. doi: 10.1371/journal.pone.0197304

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84. doi: 10.1016/S1364-6613(00)01839-8

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392. doi: 10.1152/physrev.00006.2011

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*(5), 262–268. doi: 10.1016/j.tics.2012.04.001

Friederici, A. D., & Jacobsen, T. (1999). Processing Grammatical Gender During Language Comprehension. *Journal of Psycholinguistic Research*, *28*(5), 467–484.

Ganong, W. F. (1980). Phonetic Categorization in Auditory Word Perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125.

Gervain, J., & Mehler, J. (2010). Speech Perception and Language Acquisition in the First Year of Life. *Annual Review of Psychology*, *61*(1), 191–218. doi: 10.1146/annurev.psych.093008.100408

Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience*, *32*(5), 545–561. doi: 10.1080/23273798.2016.1232419

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. doi: 10.1038/nn.3063

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69–78. doi: 10.1002/wcs.26

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics*, *43*(2), 137–146. doi: 10.3758/BF03214191

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biology*, *11*(12), e1001752. doi: 10.1371/journal.pbio.1001752

Guerra, E., Nicenboim, B., & Helo, A. V. (2018). A crack in the crystal ball: Evidence against pre-activation of gender features in sentence comprehension. In *Architectures and mechanisms for language processing (AMLaP)*.

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In Spoken Word Recognition, the Future Predicts the Past. *The Journal of Neuroscience*, *38*(35), 7585–7599. doi: 10.1523/JNEUROSCI.0065-18.2018

Haegens, S., Handel, B. F., & Jensen, O. (2011). Top-Down Controlled Alpha Band Activity in Somatosensory Areas Determines Behavioral Performance in a Discrimination Task. *Journal of Neuroscience*, *31*(14), 5197–5204. doi: 10.1523/JNEUROSCI.5199-10.2011

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews*, *86*, 150–165. doi: 10.1016/j.neubiorev.2017.12.002

Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *NeuroImage*, *20*, S18-S29. doi: 10.1016/j.neuroimage.2003.09.013

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423. doi: 10.1016/j.tics.2005.07.004

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00416

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology*, *28*, 136–141. doi: 10.1016/j.conb.2014.07.013

Hagoort, P. (2020). The meaning-making mechanism(s) behind the eyes and between the ears. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20190301. doi: 10.1098/rstb.2019.0301

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In *Categorical perception: The groundwork of cognition*. Cambridge University Press.

Hatfield, G. C. (1990). *The natural and the normative: Theories of spatial perception from Kant to Helmholtz*. MIT Press.

Heffner, C. C., Newman, R. S., Dilley, L. C., & Idsardi, W. J. (2015). Age-Related Differences in Speech Rate Perception Do Not Necessarily Entail Age-Related Differences in Speech Rate Use. *Journal of Speech Language and Hearing Research*, *58*(4), 1341. doi: 10.1044/2015_JSLHR-H-14-0239

Heffner, C. C., Newman, R. S., & Idsardi, W. J. (2017). Support for context effects on segmentation and segments depends on the context. *Attention, Perception, & Psychophysics*, *79*(3), 964–988. doi: 10.3758/s13414-016-1274-5

Hillert, D., & Bates, E. (1994). Morphological Constraints on Lexical Access: Gender Priming in German. *La Jolla: Center for Research in Language, University of California, San Diego*.

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. doi: 10.1016/j.tics.2019.05.006

Huber, R., & Knaden, M. (2015). Egocentric and geocentric navigation during extremely long foraging paths of desert ants. *Journal of Comparative Physiology A*, *201*(6), 609–616. doi: 10.1007/s00359-015-0998-3

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. doi: 10.1016/j.brainres.2015.02.014

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*, 196–208. doi: 10.1016/j.brainres.2018.11.013

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. doi: 10.1080/23273798.2015.1047459

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

Hultén, A., Schoffelen, J.-M., Uddén, J., Lam, N. H., & Hagoort, P. (2019). How the brain makes sense beyond the processing of single words – An MEG study. *NeuroImage*, *186*, 586–594. doi: 10.1016/j.neuroimage.2018.11.035

Ince, R. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula: Gaussian Copula Mutual Information. *Human Brain Mapping*, *38*(3), 1541–1573. doi: 10.1002/hbm.23471

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(21), 3621–3629. doi: 10.1016/S0042-6989(99)00088-7

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. doi: 10.1016/j.jml.2007.11.007

Jin, P., Lu, Y., & Ding, N. (2020). Low-frequency neural activity reflects rule-based chunking during speech listening. *eLife*, *9*, e55613. doi: 10.7554/eLife.55613

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of Syntactic and Semantic Information in Predictive Processing: Cross-Linguistic Evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

Kayser, S. J., Ince, R. A. A., Gross, J., & Kayser, C. (2015). Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha. *Journal of Neuroscience*, *35*(44), 14691–14701. doi: 10.1523/JNEUROSCI.2243-15.2015

Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biology*, *16*(3), e2004473. doi: 10.1371/journal.pbio.2004473

Keitel, A., Ince, R. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *NeuroImage*, *147*, 32–42. doi: 10.1016/j.neuroimage.2016.11.062

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. doi: 10.3758/BRM.42.3.627

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. doi: 10.1016/j.tics.2017.11.006

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. doi: 10.1016/S0042-6989(03)00458-9

Kochari, A., & 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. doi: 10.31234/osf.io/9npue

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A. S., Jensen, O., & Hagoort, P. (2018). Neural Entrainment Determines the Words We Hear. *Current Biology*, *28*, 2867–2875.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. doi: 10.1016/S1364-6613(00)01560-6

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics Electrified II (1994-2005). In *Handbook of Psycholinguistics* (pp. 659–724). Academic Press.

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

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The Spectrotemporal Filter Mechanism of Auditory Selective Attention. *Neuron*, *77*(4), 750–761. doi: 10.1016/j.neuron.2012.11.034

Landy, M. S., Banks, M. S., & Knill, D. C. (2011). Ideal-Observer Models of Cue Integration. In J. Trommershäuser, K. Kording, & M. S. Landy (Eds.), *Sensory Cue Integration* (pp. 5–29). Oxford University Press. doi: 10.1093/acprof:oso/9780195387247.003.0001

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338. doi: 10.1016/j.jml.2009.06.004

Length, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-square means. *R package version*, *1*(1), 3.

Liberman, A., Cooper, F. S., & Shankweiler, D. P. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

Lisker, L., & Abramson, A. S. (1967). Some Effects of Context On Voice Onset Time in English Stops. *Language and Speech*, *10*(1), 1–28. doi:

10.1177/002383096701000101

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71–102. doi: 10.1016/0010-0277(87)90005-9

Martin, A. E. (2016). Language Processing as Cue Integration: Grounding the Psychology of Language in Perception and Neurophysiology. *Frontiers in Psychology*, *7*. doi: 10.3389/fpsyg.2016.00120

Martin, A. E. (2020). A Compositional Neural Architecture for Language. *Journal of Cognitive Neuroscience*, *32*(8), 1407–1427. doi: 10.1162/jocn_a_01552

Martin, A. E., & Doumas, L. A. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. *PLOS Biology*, *15*(3), e2000663. doi: 10.1371/journal.pbio.2000663

Martin, A. E., Monahan, P. J., & Samuel, A. G. (2017). Prediction of Agreement and Phonetic Overlap Shape Sublexical Identification. *Language and Speech*, *60*(3), 356–376. doi: 10.1177/0023830916650714

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2018). Listening to yourself is special: Evidence from global speech rate tracking. *PLOS ONE*, *13*(9), e0203571. doi: 10.1371/journal.pone.0203571

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2019a). How the tracking of habitual rate influences speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/xlm0000579

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2019b). Listeners normalize speech for contextual speech rate even without an explicit recognition task. *The Journal of the Acoustical Society of America*, *146*(1), 179–188. doi: 10.1121/1.5116004

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380. doi: 10.3758/BF03206780

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 960–977. doi: 10.1037/0096-1523.33.4.960

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. doi: 10.1037/0096-3445.134.4.477

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception.

Cognitive Psychology, *18*(1), 1–86. doi: 10.1016/0010-0285(86)90015-0

McMurray, B., Cole, J. S., & Munson, C. (2011). Features as an emergent product of computing perceptual cues relative to expectations. In *Where do features come from* (pp. 197–236).

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219–246. doi: 10.1037/a0022325

McQueen, J. M. (1998). Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language*, *39*(1), 21–46. doi: 10.1006/jmla.1998.2568

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609–2621. doi: 10.1111/ejn.13748

Meyer, L., Sun, Y., & Martin, A. E. (2019). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, 1–11. doi: 10.1080/23273798.2019.1693050

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *The Journal of the Acoustical Society of America*, *73*(5), 1751–1755.

Miller, J. L., & Dexter, E. R. (1988). Effects of Speaking Rate and Lexical Status on Phonetic Perception. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 369–378.

Mitterer, H. (2018). The singleton-geminate distinction can be rate dependent: Evidence from Maltese. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 6. doi: 10.5334/labphon.66

Morrill, T., Baese-Berk, M., Heffner, C., & Dilley, L. (2015). Interactions between distal speech rate, linguistic knowledge, and speech environment. *Psychonomic Bulletin & Review*, *22*(5), 1451–1457. doi: 10.3758/s13423-015-0820-9

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics*, *37*(1), 46–65. doi: 10.1016/j.wocn.2008.09.001

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, *7*. doi:

10.7554/eLife.33468

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. doi: 10.1037/0033-295X.115.2.357

Obleser, J., Herrmann, B., & Henry, M. J. (2012). Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Frontiers in Human Neuroscience*, *6*. doi: 10.3389/fnhum.2012.00250

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. doi: 10.1016/j.tics.2019.08.004

Oden, G. C., & Massaro, D. W. (1978). Integration of Featural Information in Speech Perception. *Psychological Review*, *85*(3), 172–191.

Olshausen, B. (2014). Perception as an Inference Problem. In *The cognitive neurosciences* (pp. 295–304).

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. doi: 10.1155/2011/156869

Oruç, İ., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451–2468. doi: 10.1016/S0042-6989(03)00435-8

Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*. doi: 10.3389/fpsyg.2012.00320

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. doi: 10.1016/j.tics.2006.12.002

Pickett, J. M., & Decker, L. R. (1960). Time Factors in Perception of A Double Consonant. *Language and Speech*, *3*(1), 11–17. doi: 10.1177/002383096000300103

Pitt, M. A., & Samuel, A. G. (1993). An Empirical and Meta-Analytic Evaluation of the Phoneme Identification Task. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 699–725.

Pitt, M. A., Szostak, C., & Dilley, L. C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate. *Attention, Perception, & Psychophysics*, *78*(1), 334–345. doi: 10.3758/s13414-015-0981-7

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425.

R Development Core Team. (2012). *R: A language and environment for statistical computing.*

Rabagliati, H., & Bemis, D. K. (2013). Prediction is no panacea: The key to language is in the unexpected. *Behavioral and Brain Sciences*, *36*(4), 372–373. doi: 10.1017/S0140525X12002671

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 978–996. doi: 10.1037/a0021923

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, *41*(2), 101–116. doi: 10.1016/j.wocn.2013.01.002

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive Sensing of Periodic and Aperiodic Auditory Patterns. *Trends in Cognitive Sciences*, *22*(10), 870–882. doi: 10.1016/j.tics.2018.08.003

Rohde, H., & Ettlinger, M. (2012). Integration of pragmatic and phonetic cues in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 967–983. doi: 10.1037/a0026786

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437–447. doi: 10.1016/j.neuropsychologia.2012.12.002

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II: Effects of signal discontinuities. *Perception & Psychophysics*, *62*(2), 285–300. doi: 10.3758/BF03205549

Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics*, *2*(3), 181–194. doi: 10.1016/S0095-4470(19)31269-0

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(6876), 87–90. doi: 10.1038/416087a

Soderstrom, M., Seidl, A., Nelson, D. G. K., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, *49*(2), 249–267. doi: 10.1016/S0749-

596X(03)00024-X

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*(4), 297–314. doi: 10.1016/j.jml.2012.12.002

Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *European Journal of Neuroscience*, *50*(11), 3831–3842. doi: 10.1111/ejn.14510

Toscano, J. C., & McMurray, B. (2010). Cue Integration With Categories: Weighting Acoustic Cues in Speech Using Unsupervised Learning and Distributional Statistics. *Cognitive Science*, *34*(3), 434–464. doi: 10.1111/j.1551-6709.2009.01077.x

Toscano, J. C., & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics*, *74*(6), 1284–1301. doi: 10.3758/s13414-012-0306-z

Toscano, J. C., & McMurray, B. (2015). The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience*, *30*(5), 529–543. doi: 10.1080/23273798.2014.946427

Tuinman, A., Mitterer, H., & Cutler, A. (2014). Use of Syntax in Perceptual Compensation for Phonological Reduction. *Language and Speech*, *57*(1), 68–85. doi: 10.1177/0023830913479106

van Alphen, P., & McQueen, J. M. (2001). The Time-Limited Influence of Sentential Context on Function Word Identification. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1057–1071. doi: 10.1037/0096-1523.27.5.1057

Van Bergen, G., & Bosker, H. R. (2018). Linguistic expectation management in online discourse processing: An investigation of Dutch inderdaad 'indeed' and eigenlijk 'actually'. *Journal of Memory and Language*, *103*, 191–209. doi: 10.1016/j.jml.2018.08.004

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. doi: 10.1037/0278-7393.31.3.443

Verga, L., & Kotz, S. A. (2019). Putting language back into ecological communication contexts. *Language, Cognition and Neuroscience*, *34*(4), 536–544. doi: 10.1080/23273798.2018.1506886

von Helmholtz, H. (1896). *Vorträge und Reden* (Vol. 1). Braunschweig: F. Vieweg und Sohn.

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics*, *67*(6), 939–950. doi: 10.3758/BF03193621

Wehner, R. (2003). Desert ant navigation: How miniature brains solve complex tasks. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*, *189*(8), 579–588. doi: 10.1007/s00359-003-0431-1

Wei, K., & Körding, K. P. (2011). Causal Inference in Sensorimotor Learning. In *Sensory Cue Integration* (pp. 30–45).

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165–168. doi: 10.1016/S0304-3940(03)00599-8

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting Gender: An Event Related Brain Potential Study on the Role of Grammatical Gender in Comprehending a Line Drawing Within a Written Sentence in Spanish. *Cortex*, *39*(3), 483–508. doi: 10.1016/S0010-9452(08)70260-0

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. doi: 10.1162/0898929041920487

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., … Schroeder, C. E. (2013). Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party". *Neuron*, *77*(5), 980–991. doi: 10.1016/j.neuron.2012.12.037

Zoefel, B., ten Oever, S., & Sack, A. T. (2018). The Involvement of Endogenous Neural Oscillations in the Processing of Rhythmic Input: More Than a Regular Repetition of Evoked Neural Responses. *Frontiers in Neuroscience*, *12*. doi: 10.3389/fnins.2018.00095

Zoefel, B., & VanRullen, R. (2015). The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound. *Frontiers in Human Neuroscience*, *9*. doi: 10.3389/fnhum.2015.00651

English Summary

We use language seemingly effortlessly in everyday life to communicate our thoughts and feelings. Yet what exactly happens in the brain when we speak to one another is still not entirely clear. In my doctoral thesis, I investigated this fundamental question: How does linguistic meaning emerge?

Specifically, my thesis examined spoken language. In essence, spoken language is an acoustic signal, which does not – in and of itself – contain any obvious "meaning"; there are no clear acoustic "markers" that could tell us what a specific word signifies. It is only in combination with our learned knowledge about language that we can *understand* the signal. This becomes really obvious when we listen to a language we don't know: We can still *perceive* the acoustic signal, but we cannot *understand* its meaning. As such, understanding spoken language is not just about hearing an acoustic signal; it's just as much about combining this acoustic signal with our linguistic knowledge. How exactly that happens – how we combine acoustic and linguistic information into meaning, and how that happens in the brain – was the central question of this doctoral thesis.

In four chapters, I investigated the interplay between acoustic and linguistic information in more detail. First, I wanted to examine how exactly listeners combine two specific pieces of acoustic and linguistic information, or *cues*. This is especially interesting in situations where information is not entirely clear (or might even be contradictory), for example when we are speaking in a loud environment with lots of background noise. In chapters 2 and 3 I report on two eye-tracking experiments that show that listeners are rather flexible during spoken language comprehension: We can use certain acoustic information very rapidly, even if it is not entirely clear or reliable. At the same time, we can change our interpretations quickly and adjust as soon as additional linguistic information becomes available. These results are interesting in several ways: First, they show that there is no clear "hierarchy of comprehension". Instead, listeners are flexible in which (acoustic or linguistic) information they use for language processing. Second, the results suggest that the brain does not necessarily "wait" until all possible information is available; instead, it uses (potentially unreliable) information immediately to infer meanings. These can later be revised and adjusted when more information becomes available.

In chapters 4 and 5, I report on an EEG experiment that investigated the neural signal underlying spoken language comprehension in more depth. Previous research has shown that populations of neurons in the brain can *synchronize* with the acoustic signal, that is, they fire in a similar rhythm as the air pressure fluctuations in the acoustic signal. However, it was previously not very well-known which types of information really drive this synchronization, or "tracking". The EEG experiment in chapters 4 and 5 showed that neural tracking is not solely driven by acoustic fluctuations and informational regularities in the signal, but also by the linguistic content: The brain "tracks" the acoustic signal more closely when it contains meaning and structure.

I summarized these findings in chapter 6 and discussed them within the theoretical framework of *cue integration*. Cue integration formalizes perception as an inference problem: We are able to perceive and understand our environment by combining sensory (e.g., acoustic) information and learned knowledge. The results reported in my doctoral thesis support theories that posit spoken language comprehension as a form of perceptual inference.

Nederlandse samenvatting

Elke dag gebruiken we taal om onze gedachten en gevoelens met anderen te delen. Het is echter nog niet duidelijk wat er in onze hersenen gebeurt als we dat doen. In mijn proefschrift heb ik een fundamentele vraag onderzocht die hiermee verband houdt: hoe ontstaat betekenis in taal?

Om precies te zijn focust mijn proefschrift zich op *gesproken* taal. Gesproken taal is, in essentie, een akoestisch signaal, dat van zichzelf geen duidelijke "betekenis" bevat: er zijn geen voor de hand liggende akoestische kenmerken die kunnen aangeven wat een specifiek woord betekent. Alleen in combinatie met geleerde kennis over taal kunnen we het signaal *begrijpen*. Dit wordt duidelijk als we naar een taal luisteren die we niet kennen: we kunnen het akoestische signaal nog steeds *waarnemen*, maar we kunnen niet *begrijpen* wat het betekent. Het begrijpen van gesproken taal gaat dus niet alleen over het horen van een akoestisch signaal; het gaat net zoveel over het combineren van dit akoestische signaal met onze taalkundige kennis. Hoe dat nu precies gebeurt – hoe we akoestische en taalkundige informatie combineren tot betekenis, en hoe dat gebeurt in het brein – was de centrale vraag van dit proefschrift.

In vier hoofdstukken heb ik de wisselwerking tussen akoestische en taalkundige informatie onderzocht. Allereerst wilde ik onder de loep nemen hoe luisteraars twee specifieke stukjes van akoestische en taalkundige informatie, ook wel cues (Engels: "signaal") genoemd, combineren. Dit is vooral interessant in situaties waar de informatie niet volledig is (of zelfs tegenstrijdig), bijvoorbeeld als we in een luidruchtige omgeving praten. In hoofdstukken twee en drie rapporteer ik over twee eye-tracking experimenten die laten zien dat luisteraars erg flexibel zijn tijdens het begrijpen van gesproken taal: we kunnen bepaalde akoestische informatie erg snel gebruiken, zelfs als die informatie niet helemaal helder of betrouwbaar is. Tegelijkertijd zijn onze interpretaties veranderlijk: we kunnen ze aanpassen zodra er nieuwe taalkundige informatie beschikbaar is. Deze resultaten zijn interessant vanwege verschillende redenen. Ten eerste laten ze zien dat er geen duidelijke "hiërarchie van informatiebronnen" bestaat. In plaats daarvan gebruiken luisteraars de beschikbare informatie (akoestisch of taalkundig) op een flexibele manier. Ten tweede suggereren de resultaten dat het brein niet “afwacht” tot alle mogelijke informatie beschikbaar is. In plaats daarvan gebruikt het (mogelijk onbetrouwbare) informatie om onmiddellijk een betekenis af te leiden. Die betekenis kan later herzien en aangepast worden, als er meer informatie beschikbaar is.

In hoofdstukken vier en vijf beschrijf ik een EEG-experiment. In dit experiment onderzocht ik het neurale signaal dat het begrijpen van taal onderligt. Eerder onderzoek heeft aangetoond dat groepen van neuronen in het brein kunnen synchroniseren met het akoestische signaal. Dat wil zeggen dat ze vuren met een ritme dat vergelijkbaar is met de luchtdrukfluctuaties in het akoestische signaal. Het was echter nog niet duidelijk welke soorten informatie deze synchronisatie, ook wel *neural tracking* genoemd, teweegbrengen. Het EEG experiment laat zien dat neural tracking niet alleen wordt veroorzaakt door akoestische en informationele patronen in het spraaksignaal, maar ook door de taalkundige inhoud: het brein synchroniseert nóg beter met het akoestische signaal, wanneer het betekenis en taalstructuur bevat.

In hoofdstuk zes vat ik deze bevindingen samen en bespreek ik ze binnen het theoretische kader *cue integration* (Engels; let. signaalintegratie). Cue integration formaliseert perceptie als een “inferentieprobleem”: we zijn instaat onze omgeving waar te nemen en te begrijpen door zintuiglijke (bijv. akoestische) informatie en geleerde (bijv. taalkundige) kennis met elkaar te verbinden. De resultaten die in mijn proefschrift zijn vermeld, onderschrijven theorieën die veronderstellen dat het begrijpen van gesproken taal een vorm van zintuiglijke inferentie is.

Deutsche Zusammenfassung

Beinahe selbstverständlich gebrauchen wir Sprache im täglichen Leben, um unsere Gedanken und Gefühle zu kommunizieren. Doch was genau im Gehirn passiert, wenn wir miteinander sprechen, ist noch immer nicht vollständig erforscht. In meiner Doktorarbeit habe ich mich mit einer sehr grundlegenden Frage beschäftigt: Wie entsteht sprachliche Bedeutung?

Ganz speziell habe ich in meiner Doktorarbeit gesprochene Sprache untersucht. Im Grunde ist gesprochene Sprache ein akustisches Signal. Dieses Signal enthält allerdings von sich aus keine offensichtliche “Bedeutung”; es gibt keine klaren akustischen “Marker”, die uns verraten, was ein bestimmtes Wort bedeutet. Nur aufgrund unseres erlernten sprachlichen Wissens können wir das Signal auch *verstehen*. Besonders deutlich merken wir das, wenn wir eine Sprache hören, die wir selbst nicht sprechen: Wir können das akustische Signal noch immer *wahrnehmen*, aber wir können seine Bedeutung nicht *verstehen*. Sprachverstehen bedeutet also nicht nur, ein akustisches Signal zu hören, sondern auch, diese akustischen Informationen mit unserem sprachlichen Wissen zu kombinieren. Wie genau das geschieht – wie wir akustische und sprachliche Informationen zu Bedeutungen zusammenfügen und was dabei in unserem Gehirn vorgeht – war der Gegenstand dieser Doktorarbeit.

Das Zusammenspiel von akustischen Informationen und sprachlichem Wissen habe ich in vier Kapiteln genauer untersucht. Die Kapitel 2 und 3 beschäftigen sich zunächst mit der Frage, wie Hörer:innen bestimmte akustische und sprachliche Informationen kombinieren. Das ist vor allem in Situationen interessant, in denen Informationen nicht ganz eindeutig (oder sogar widersprüchlich) sind, zum Beispiel wenn wir uns in einer lauten Umgebung mit vielen Hintergrundgeräuschen unterhalten. Die in den Kapiteln 2 und 3 zusammengefassten Eyetracking-Experiment zeigen, dass wir beim Verstehen von gesprochener Sprache sehr flexibel sind: Wir können akustische Informationen sehr schnell fürs Sprachverstehen nutzen, selbst, wenn diese nicht ganz eindeutig sind. Gleichzeitig können wir unsere Interpretationen aber auch schnell ändern und anpassen, wenn uns weitere sprachliche Informationen zur Verfügung stehen. Diese Ergebnisse sind auf unterschiedlichen Ebenen interessant: Einerseits zeigen sie, dass es keine klare “Hierarchie des Sprachverstehens” gibt, sondern dass Hörer:innen flexibel darin sind, welche (akustischen oder linguistischen) Informationen sie bei der Sprachverarbeitung nutzen. Andererseits legen die Ergebnisse auch nahe, dass das Gehirn nicht unbedingt “wartet”, bis alle wichtigen Informationen vorhanden sind, sondern dass es uneindeutige Informationen sofort nutzt, um Bedeutungen zu generieren. Diese Bedeutungen können später revidiert und angepasst werden, wenn neue Informationen eintreffen.

Die Kapitel 4 und 5 beschäftigen sich genauer mit den neuronalen Signalen, die dem Sprachverstehen zugrunde liegen. Frühere Forschung hat gezeigt, dass sich beim Sprachverstehen Gruppen von Nervenzellen im Gehirn mit dem akustischen Sprachsignal *synchronisieren*, das heißt, sie sind sozusagen im gleichen Rhythmus aktiv wie die Schwingungen im akustischen Signal. Allerdings konnte die bisherige Forschung nicht präzise bestimmen, welche Informationen genau zu dieser neuronalen Synchronsation führen. Das EEG-Experiment in Kapitel 4 und 5 zeigt, dass neuronale Synchronisation nicht nur von akustischen Informationen und Regelmäßigkeiten abhängig ist, sondern vor allem auch von sprachlichen Informationen.

All diese Ergebnisse habe ich in Kapitel 6 zusammengefasst und im Rahmen einer Theorie namens *cue integration* diskutiert. Cue integration formalisiert Wahrnehmung als perzeptuelle Inferenz: Durch die Kombination von sensorischen (zum Beispiel akustischen) Informationen und erlerntem Wissen sind wir in der Lage, unsere Umwelt wahrzunehmen und zu interpretieren. Die Ergebnisse in meiner Doktorarbeit unterstützen linguistische Theorien, die Sprachverstehen als eine Form von perzeptueller Inferenz ansehen.

Acknowledgements

This thesis was only possible because of the invaluable help and support I received throughout my four years as a PhD student. I am immensely grateful to have so many people, in Nijmegen and beyond, to say my “thank yous” to.

First and foremost, I would like to thank my wonderful supervisors Andrea, Hans and Antje. I learned a lot from each of you about how to approach scientific research in a thoughtful and structured way. **Andrea** – it was an absolute privilege to write this thesis under your kind and inspiring supervision and guidance. You always challenged me to think deeply and carefully, and you showed me that the most difficult questions often happen to be the most exciting ones. Thank you for advocating for positive change not only in thought, but also in action. I was truly lucky to be your PhD student.

Hans – I am really grateful that you became part of my supervisory team. Thank you for always taking time to patiently explain things to me in detail, for answering my many questions, and most importantly, for always being encouraging and looking for solutions instead of problems.

Antje – thank you for your support and guidance throughout these four years. Meetings with you were always inspiring and uplifting, and I learned a lot from your rigorous questions. Thank you for always making time to listen when I had concerns, however big or small.

I would also like to thank the members of my reading committee, Prof. Peter Hagoort, Prof. Yiya Chen, and Prof. Sonja Kotz, for reading and evaluating this thesis.

My dear paranymphs Eirini, Nina and Sophie – I could not have asked for better friends at the MPI. **Eirini**, thank you for being there for me through all the highs and lows, for hours of talking in and outside of the office, for hiking through the rain with me in the Highlands, and for always knowing which series I should watch next. How lucky I am to have met a friend that not only speaks *the* secret language, but also understands *my* secret language. **Nina**, meine Seelenfreundin – thank you for nights at the opera and the ballet, for being my first ever singing student, and most importantly for always listening and making me feel understood. I’m beyond grateful to have met a friend in you who shares my deepest passion, and who can always describe my own feelings way better than I can, myself. **Sophie**, we only shared the office during the final months of my PhD, but we became friends very quickly. I really enjoyed our countless conversations about everything that's puzzling about brains, both scientifically and personally. Thank you for always reminding me how far I'd come, and for generally insisting on remembering the positive in any situation.

I feel very lucky to have been part of the inspiring scientific community at the MPI. Thank you to all of my colleagues and friends from the Psychology of Language department and the LaCNS and TEMPOS groups. In particular, thank you dear **Sara**, for becoming not only my office mate, but also one of my closest friends at the MPI. You always pointed out perspectives I hadn't yet thought about and made me look at things in a completely different – and usually much wiser – way. **Limor**, thank you for sharing not only a sleepless night spent video editing with me, but also your wisdom and kindness. **Joe**, thank you for having me over for exquisite meals and delightful chats, and for keeping me on time in Lisbon and Berlin. **Phillip**, thanks for sharing your knowledge on EEG analysis and statistics with me, and for telling me off when I worried too much. **Merel** M., thank you for our lovely conversations inside and out of the MPI, and for sharing your thesis template with me. **Jeroen**, teaching the python course with you was really fun (and actually made me a better coder, too). **Caitlin**, **Saoradh**, **Amie**, **Laurel**, **Federica** and **Annelies**, thanks for many laughs and cheerful lunch chats. **Annelies**, thank you also for your help with my stimuli and for letting me practice my Dutch with you. **Merel** W., thanks for bringing the TalkLing blog into life – it was really fun being an editor. **Sanne**, thank you for your help with Matlab and the MI and source analysis, and for helping me find new energy for the EEG chapters. **Cas**, thanks for finding all the bugs in my code, and asking questions that really made me think. **Karthikeya**, thanks for helping me test participants, even though COVID got in the way of completing the experiment before the end of my thesis. **Wibke** and **Anna**, thanks for being amazing interns and collaborators. Thank you also to the wonderful student assistants in our department, in particular **Zina**, **Dylan**, **Nikki** and **Dennis**, and to the Nijmegen **PoS team**.

I would also like to thank the TG, library and operations teams for their invaluable support, and **Kevin**, for making the IMPRS such an inspiring and enriching place of academic and personal learning.

Finishing a thesis during "unprecedented times" was challenging (although I reckon that all times are unprecedented, and all theses are challenging in a way), but there were things that made the months of socially-distanced writing so much more enjoyable. **Caitlin**, thank you for the many laughs and exciting stories during our zoom coffee dates. **Hannah**, thank you for your fabulous online yoga lessons, and thank you and **Rehana** for fun zoom quiz nights (although I really need to step up my game). **Ashley**, **Elliot**, **Marieke**, **Mischa**, **Jakob** and **Lara**, thank you for lovely zoom borrels.

I'm very grateful that I was able to spend much of my spare time in Nijmegen doing what I love most: making music. **Anja** – I cannot thank you enough for being my singing teacher. There were times when this PhD seemed like an insurmountable task, but Thursday morning singing lessons with you always made my heart feel light (nothing like a few coloraturas before breakfast to lift the spirit). Thanks also to the wonderfully talented singers in the **UVS teacher training**, and to my lovely band mates from **Cherry and the Cakes**. It's been an absolute pleasure playing with such talented (and generally awesome) people.

Thank you also to the many other friends from near and far who have been part of this journey in one way or another. **Larena**, you're forever my favorite source of knowledge and inspiration on feminism, architecture, music and soccer (which is hands down the best combination I can think of). Your friendship means the world. **Sarah**, thanks for our (infrequent, but all the more heartfelt) hour-long phone conversations, and for generally being one of the wisest and most warmhearted people I know. **Anni**, thank you for inspiring talks, and for turning everything into art (including my cover). **Mel**, thanks for always making things feel light. Thanks to the **Hoolies** and my lovely **AT friends** for fun trips and visits, and to the **Göttingen connection**, for lovely board game nights and end-of-year celebrations. Thank you also to my dear **friends in the US** for welcoming me into their home more than 15 years ago and ever since.

Finally, I'd like to thank **my family**. Thank you for all of your love and support throughout the years, and for distracting me from stressful moments with visits in Nijmegen or video calls from Nienburg, Oldenburg and Göttingen. I feel immensely lucky to have all of you in my life.

Most importantly, thank you **Fritz**, for your never-ending patience, love and support through all the ups and downs of this PhD. Thanks for making Nijmegen feel like a real home, and thanks for always being excited for the next adventure. Du bist es.

Curriculum Vitae

Greta Kaufeld was born in Nienburg, Germany, in 1989. She obtained her bachelor’s degree in Language and Communication from Philipps University Marburg, and a Master’s degree in Cognitive Science from the University of Edinburgh. In September 2016, Greta began her PhD project in the Psychology of Language Department at the Max Planck Institute for Psycholinguistics. She is currently working at Disney Research Studios in Zurich.

Publications

Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (2020). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(3), 549–562.

Kaufeld, G., Naumann, W., Meyer, A. S., Bosker, H. R., & Martin, A. E. (2020). Contextual speech rate influences morphosyntactic prediction and integration. *Language, Cognition and Neuroscience, 35*(7), 933-948.

Kaufeld, G., Bosker, H. R., Ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience,* Advance online publication.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda I. van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed class distinction in spoken-word recognition. *Alette Petra Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk J.Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie H. van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Christine Weber*
17. Moving eyes and naming objects. *Femke Frederike van der Meulen*

18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*

19. Morphology in speech comprehension. *Kerstin Mauth*

20. Morphological families in the mental lexicon. *Nivja Helena de Jong*

21. Fixed expressions and the production of idioms. *Simone Annegret Sprenger*

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria. *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*

24. Contextual influences on spoken-word processing: An electrophysiological approach. *Danielle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra Martine van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Henriëtte Wilhelmina Meeuwissen*

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl Jenny Judith Karin Kemps*

29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*

30. A grammar of Jalonke argument structure. *Friederike Lüpke*

31. Agrammatic comprehension: An electrophysiological approach. *Marijtje Elizabeth Debora Wassenaar*

32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*

33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*

34. Phonetic and lexical processing in a second language. *Mirjam Elisabeth Broersma*

35. Retrieving semantic and syntactic word properties: ERP studies on the time course in language comprehension. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren Batya Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-constrative variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel Haun*

42. The acquisition of auditory categories. *Martijn Bastiaan Goudbeek*

43. Affix reduction in spoken Dutch: Probabilistic effects in production and perception. *Mark Pluymaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Madalla Kooijman*

45. Space and iconicity in German sign language (DGS). *Pamela M. Perniss*

46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*

48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*

49. Phoneme inventories and patterns of speech sound perception. *Anita Eva Wagner*

50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*

51. A grammar of Savosavo: A Papuan language of the Solomon Islands. *Claudia Ursula Wegener*

52. Prosodic structure in speech production and perception. *Claudia Kuzla*

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*

54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*

55. Processing the fine temporal structure of spoken words. *Eva Reinisch*

56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*

57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*

58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*

59. Lexical interactions in non-native speech comprehension: Evidence from electroencephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*

60. Processing casual speech in native and non-native language. *Annelie Tuinman*

61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Payton Robinson*

62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*

63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*

64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*

65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*

66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*

67. The relevance of early word recognition: Insights from the infant brain. *Caroline Mary Magteld Junge*

68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias Johannes Sjerps*

69. Structuring language: Contributions to the neurocognition of syntax. *Katrien Rachel Segaert*

70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*

71. Gaze behavior in face-to-face interaction. *Federico Rossano*

72. Sign-spatiality in Kata Kolok: How a village sign language of Bali inscribes its signing space. *Connie de Vos*

73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*

74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*

75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*

76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*

77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*

78. Contributions of executive control to individual differences in word production. *Zeshu Shao*

79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*

80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*

81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*

82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*

83. The many ways listeners adapt to reductions in casual speech. *Katja Pöllmann*

84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*

85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*

86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*

87. The acquisition of morphophonological alternations across languages. *Helen Buckler*

88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*

89. Rediscovering a forgotten language. *Jiyoun Choi*

90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*

91. Infants'understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*

92. Information structure in Avatime. *Saskia van Putten*

93. Switch reference in Whitesands. *Jeremy Hammond*

94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*

95. Acquisition of spatial language by signing and speaking children: A comparison of Turkish sign language (TID) and Turkish. *Beyza Sumer*

96. An ear for pitch: On the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*

97. lncrementality and Flexibility in Sentence Production. *Maartje van de Velde*

98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*

99. The request system in Italian interaction. *Giovanni Rossi*

100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*

101. Assessing birth language memory in young adoptees. *Wencui Zhou*

102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*

103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*

104. Conversation electrified: The electrophysiology of spoken speech act recognition. *Rósa Signý Gisladottir*

105. Modelling multimodal language processing. *Alastair Charles Smith*

106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*

107. Situational variation in non-native communication. *Huib Kouwenhoven*

108. Sustained attention in language production. *Suzanne Jongman*

109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntactic, and orthographic effects. *Malte Viebahn*

110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*

111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*

112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*

113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*

114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*

115. Music and language comprehension in the brain. *Richard Kunert*

116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*

117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*

118. Language processing in a conversation context. *Lotte Schoot*

119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*

120. Talking sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*

121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*

122. Sensorimotor experience in speech perception. *Will Schuerman*

123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*

124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*

125. Lapse organization in interaction. *Elliott Hoey*

126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*

127. The neighbors will tell you what to expect: effects of aging and predictability on language processing. *Cornelia Moers*

128. The role of voice and word order in incremental sentence processing. Studies on sentence production and comprehension in Tagalog and German. *Sebastian Sauppe*

129. Learning from the (un)expected: age and individual differences in statistical learning and perceptual learning in speech. *Thordis Neger*

130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*

131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*

132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*

133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*

134. Indirect request comprehension in different contexts. *Johanne Tromp*

135. Envisioning language: An exploration of perceptual processes in language comprehension. *Markus Ostarek*

136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. *Juliane Kirsch*

137. Let the agents do the talking: On the influence of vocal tract anatomy on speech during ontogeny and glossogeny. *Rick Janssen*

138. Age and hearing loss effects on speech processing. *Xaver Koch*

139. Vocabulary knowledge and learning: Individual differences in adult native speakers. *Nina Mainz*

140. The face in face-to-face communication: Signals of understanding and non-understanding. *Paul Hömke*

141. Person reference and interaction in Umpila/Kuuku Ya'u narrative. *Clair Hill*

142. Beyond the language given: The neurobiological infrastructure for pragmatic inferencing. *Jana Bašnáková*

143. From Kawapanan to Shawi: Topics in language variation and change. *Luis Miguel Rojas Berscia*

144. On the oscillatory dynamics underlying speech-gesture integration in clear and adverse listening conditions. *Linda Drijvers*

145. Linguistic dual-tasking: Understanding temporal overlap between production and comprehension. *Amie Fairs*

146. The role of exemplars in speech comprehension. *Annika Nijveld*

147. A network of interacting proteins disrupted in language-related disorders. *Elliot Sollis*

148. Fast speech can sound slow: Effects of contextual speech rate on word recognition. *Merel Maslowski*

149. Reasons for every-day activities. *Julija Baranova*

150. Speech planning in dialogue - Psycholinguistic studies of the timing of turn taking. *Mathias Barthel*

151. The role of neural feedback in language unification: How awareness affects combinatorial processing. *Valeria Mongelli*

152. Exploring social biases in language processing. *Sara Iacozza*

153. Vocal learning in the pale spear-nosed bat, *Phyllostomus discolor*. *Ella Lattenkamp*

154. Effect of language contact on speech and gesture: The case of Turkish-Dutch bilinguals in the Netherlands. *Elif Zeynep Azar*

155. Language and society: How social pressures shape grammatical structure *Limor Raviv*

156. The moment in between: Planning speech while listening. *Svetlana-Lito Gerakaki*

157. How speaking fast is like running: Modelling control of speaking rate. *Joe Rodd*

158. The power of context: How linguistic contextual information shapes brain dynamics during sentence processing. *René Terporten*

159. Neurobiological models of sentence processing. *Marvin Uhlmann*

160. Individual differences in syntactic knowledge and processing: The role of literacy experience. *Saoradh Favier*

161. Memory for speaking and listening. *Eirini Zormpa*

162. Masculine generic pronouns: Investigating the processing of an unintended gender cue. *Theresa Redl*

163. Properties, structures and operations: Studies on language processing in the brain using computational linguistics and naturalistic stimuli. *Alessandro Lopopolo*

164. Investigating spoken language comprehension as perceptual inference. *Greta Kaufeld*