Sentence comprehension with a competing talker

The elusive nature of informational interference

Huarda Yareri Valdes-Laribi

PhD

University of York

Psychology

July 2016

Abstract

In everyday environments, we often have to attend to one person's speech (target speech) while ignoring another (competing speech). A competing talker can impair speech processing through both energetic masking (acoustic degradation at the periphery) and informational, cognitively-demanding aspects of the mask. We refer to the latter as informational interference. We hypothesized that informational interference depletes processing resources that could otherwise be allocated to recognizing and understanding target speech. Consequently, informational interference should be more pronounced when the task is more resource-demanding (more or less complex syntax) or when the participants' own processing demands are elevated (non-native listeners). Finally, modulating the semantic content of the competing talker's utterances should influence the degree of informational interference.

Using a speeded picture-selection task, we assessed native and non-native listeners' understanding of spoken sentences varying in syntactic complexity, played with a competing talker or a matched energetic mask, at various signal-to-noise ratios (SNRs). In a follow-up experiment, the semantic content of the competing talker sentences was manipulated to be congruent, incongruent or unrelated to the target sentence. Participants' performance was measured with accuracy and reaction times from button presses, as well as eye-tracking. Selective attention, short-term and working memory were assessed to determine the contribution of these cognitive factors to informational interference.

Although syntactic complexity affected participants' performance, the competing talker was not more detrimental than the energetic mask controls, contrary to our hypothesis. This pattern was comparable for native and non-native listeners, and across SNRs. In the follow-up experiment there was no difference between semantically incongruent and neutral competing sentences, but semantically congruent sentences led to faster sentence processing, indicating facilitation or priming. This indicates that the content of the competing talker is not indiscriminately inhibited. Moreover, individual differences in memory and selective attention were not related to differences in the speeded-selection task, regardless of the mask. These results provide little support for the existence of a uniquely informational source of speech masking.

Contents

List of tables

List of figures

Acknowledgments

Without the opportunities and funding provided by the European Union, this PhD thesis would not have been possible. I am very grateful to have been part of a vibrant and inspiring community of researchers within the Marie Curie ITN INSPIRE network.

I would like to express my profound gratitude to my supervisor Sven Mattys for his expert guidance and support. Thank you for your generosity in time and your enthusiasm, which never failed to encourage me throughout the PhD process. You have helped me to navigate the sometimes murky waters of research, with patience and gentle steering.

I would also like to thank Dorothea Wendt for never hesitating to offer her help and time, regardless of her own work commitments. Thank you for your patience and the endless hours of technical and moral support that you have given me without counting.

Thank you to Shirley-Ann Rueschemeyer and Quentin Summerfield for encouraging me to critically evaluate my arguments and for the stimulating discussions we have had during my Thesis Advisory Panel meetings.

Thank you to Torsten Dau, Ewen MacDonald and all the team at the Centre for Applied Hearing Research at the Technical University of Denmark for their warm welcome and precious brainstorming sessions.

Many thanks to Garreth Prendergast and Daniel Baker for all your help in my Matlab battles, and to Daniel Baker and Adele Goman for accepting to sit through several hours of recording sentences.

I am grateful to my friends at the University of York and each of the INSPIRE fellows for being a source of inspiration and support, for the serious and less serious conversations we have had, for the fun and for the knowledge you have shared with me. To my dear friends Dorina Strori and Victoria Brattan, this experience would not have been the same without you. Thank you for your treasured friendship, which I will keep close to my heart wherever our paths may lead us.

My friends and family have encouraged and supported me from my very first steps on this journey. I am forever indebted to you for keeping my spirits high and reminding me what is important in life. Melanie, Emanuele, Hélène, Jessica, Amandine, and Milena, thank you for your friendship, your words of wisdom and encouragement, and your humour.

To Alice, Charlotte and Juliet, little bubbles of joy, thank you for your laughter and excitement, thank you for listening to my stories and for keeping me grounded when I needed it the most.

Thank you to Sid-Ahmed and Christiane for always offering a peaceful haven and open arms. To Ouahiba, Brian, Kalen and Elouan, you have helped me to carry on, with your enthusiasm and perspective from across the pond.

My parents' unfaltering love, support, guidance, reassuring words and expert advice have allowed me to manoeuvre through the PhD journey without losing my balance.

Finally, I could not have embarked on this journey, let alone finish it, without the unwavering support, optimism, and selfless love from my best friend and life companion, Nathalie.

Author's declaration

This thesis comprises the candidate's own original work and has not, whether in the same or different form, been submitted to this or any other university for a degree. All sources are acknowledged as References. All experiments were designed by the candidate under the supervision of Prof Sven Mattys. All data collection and analyses were conducted by the candidate. Dr Dorothea Wendt (Technical University of Denmark) provided advice on the eye-tracking methodology, data collection and analyses. I am grateful to Prof Martin Cooke (University of the Basque Country) for assistance in creating the speech-modulated noise stimuli for all experiments. The pictures for this thesis were hand-drawn by Ling Ge (University of York), and modified by the candidate.

This research was conducted within the FP7 Marie Curie Initial Training Network *Investigating Speech Processing In Realistic Environments* (INSPIRE), funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-PEOPLE-2011-290000.

Selected aspects of this thesis have been presented at the following conferences:

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (January 2016). *Sentence comprehension with a competing talker: informational interference or semantic facilitation?* Poster presented at the Speech Processing in Realistic Environments (SPIRE) workshop, Groningen, NL.

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (September 2015). *Competing talker interference in sentence comprehension*. Paper presented at the 19th Conference of the European Society for Cognitive Psychology (ESCOP), Paphos, CY.

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (June 2015). *Informational interference from a competing talker: a thought-provoking but elusive construct.* Poster presented at the Third International Conference on Cognitive Hearing Science for Communication (CHSCOM), Linköping, SE.

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (May 2015). *Informational interference from a competing talker: a thought-provoking but elusive construct*. Paper presented at the Psycholinguistics in Flanders conference (PiF), Marche-en-Famenne, BE.

Mattys, S., Valdes-Laribi, H. (May 2015). *When a competing talker is easy to ignore: The elusive nature of informational interference.* Paper presented at the 169th meeting of the Acoustical Society of America (ASA), Pittsburgh, PA, USA.

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (January 2015). *Native and non-native sentence comprehension in the presence of a competing talker.* Poster presented at the INSPIRE workshop on talker-listener interactions, University College London, London, UK.

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (January 2015). *Native and non-native sentence comprehension in the presence of a competing talker.* Paper presented at the Experimental Psychology Society meeting (EPS), University College London, London, UK.

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (November 2014). *Native and non-native sentence comprehension in the presence of a competing talker.* Poster presented at the 55th annual meeting of the Psychonomic Society, Long Beach, CA, USA.

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (September 2014). *Native and non-native sentence comprehension in the presence of a competing talker.* Paper presented at the Postgraduate and Academic Researchers in Linguistics at York conference (PARLAY), University of York, York, UK.

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (September 2014). *Native and non-native sentence comprehension in the presence of a competing talker.* Poster presented at the annual conference of the British Society of Audiology (BSA), Keele University, Keele, UK.

Valdes-Laribi, H., Wendt, D., MacDonald, E., Cooke, M., Mattys, S. (May 2014). *Understanding sentences in the presence of a competing talker.* Paper presented at the Psycholinguistics in Flanders conference (PiF), Ostend, BE.

Chapter 1: Introduction and literature review

Our everyday oral interactions often take place in suboptimal listening conditions. Whether it is the whirring of a fan, the traffic in the street, the chatter of a classroom, an open-plan office or a busy party, we are exposed to adverse listening conditions in most communicative situations. Although psycholinguists have been studying speech perception and comprehension and untangling their different components for decades, many studies have typically used optimal, quiet listening conditions. Those studies that have considered sub-optimal listening conditions have usually focused on the lower levels of language, specifically sound perception and identification. However, our task in everyday conversations is not only to *perceive* speech sounds, but also to *understand* the words and sentences that we perceive. Despite sentence comprehension and syntactic processing being at the heart of real-world interactions, few studies have focused on the interplay between syntactic processing and adverse conditions, and even fewer have investigated syntactic processing with a competing talker.

A competing talker poses an interesting challenge because it produces two potential sources of interference, known as energetic masking (EM) and informational masking (IM). EM focuses on the spectro-temporal overlap between a target and a mask. IM is broadly construed as the detrimental effect of a mask once EM has been accounted for. The hypothesis that will guide this thesis is that dealing with informational masking requires greater processing resources than EM, and hence, its effect should be particularly detrimental to speech tasks that require a substantial amount of processing resources. In this chapter, after a brief introduction to adverse conditions and energetic vs. informational masking, I will review studies that have investigated the effect of a competing talker on speech perception and comprehension, and those that have investigated the effect of EM on syntactic processing. I will then review the cognitive processes thought to be involved in language processing in adverse conditions.

1.1 Adverse conditions

Speech perception in adverse conditions has received considerable attention in the past decades. In a review of studies investigating speech perception in adverse conditions, Assmann & Summerfield (2004) defined the term 'adverse conditions' as " any perturbation of the communication process resulting from either an error in production by the speaker, channel distortion or masking in transmission, or a distortion in the auditory system of the listener" (p.232).

A complementary definition of adverse conditions has been advanced by Mattys, Davis, Bradlow, and Scott (2012), who considered an adverse condition to be "any factor leading to a decrease in speech intelligibility on a given task relative to the level of intelligibility when the same task is performed in optimal listening situations". Mattys et al. have proposed a classification of adverse conditions according to their origin, their effect, and their approximate frequency of occurrence. According to these authors, the origin of adverse conditions can be separated into three main categories: (1) a degradation at the source; (2) a degradation in the environment, or during the transmission of the signal; (3) difficulties attributed to the listener ('receiver limitations'). The 'masking in transmission' category mentioned by Assmann and Summerfield (2004) broadly corresponds to the environmental or transmission degradation mentioned by Mattys et al. (2012). Masking, or degradation in the environment, is the type of adverse condition that will be the main focus of the next section.

1.2 Masking as an adverse condition

Masking occurs whenever at least one competing sound source interferes with the perception and/or processing of a target sound source. Broadly speaking, there are two types of masking: energetic and informational (Kidd, Mason, Richards, Gallun, & Durlach, 2007; Shinn-Cunningham, 2008). The earliest mention of these terms seems to have been in an abstract by Pollack (1975), although masking had been studied in much earlier experiments (e.g. French & Steinberg, 1947; Miller, 1947). A competing talker can lead to both energetic and informational masking, as I shall outline below.

1.2.1 Controlling for energetic masking

Energetic masking (EM) takes place when the competing signal overlaps spectro-temporally with the target, thereby degrading the acoustic signal at the auditory periphery, or cochlear level (e.g., Brungart, 2001; Watson, Kelly, & Wroton, 1976). The consequence of EM is that portions (or all) of the target signal become less audible at the auditory periphery.

More recently, Stone, Füllgrabe and Moore (2012) and Stone and Moore (2014) have argued that EM should in fact be termed 'modulation masking'. Indeed, EM implies that the energy of the competing signal interferes with the energy of the target signal, whereas these authors point out that it is more often the amplitude modulations of the competing signal that interfere with the perception of the amplitude modulations in the target signal. What is usually thought of as EM is composed both of EM and modulation masking. Although this distinction is important to the study of low-level perceptual masking, these types of masking are not the focus of this thesis, and it is thus not crucial to determine whether the masking in my experiments is due to energetic or modulation masking (or indeed both). To simplify terminology, I will only use the term ‘energetic masking’ (EM).

Competing speech can overlap with target speech in both the time and the frequency domains, thereby creating EM. However, it also creates non-energetic masking, or informational masking, which I will describe in the next section. Non-speech or unintelligible speech maskers have been used to isolate the EM component of competing speech. This allows EM to be maximised while reducing higher-level (informational) components of masking, such as the linguistic content of the mask. Non-speech maskers can vary in their spectral characteristics and whether they are continuous or modulated/fluctuating. These characteristics determine the amount of glimpsing opportunities they provide to the listener. The spectro-temporal ‘dips’ in modulated masks (including speech) allow portions of the target energy to be ‘glimpsed’ through the mask, also known as ‘dip listening’ (Cooke, 2006; Howard-Jones & Rosen, 1993a). Glimpses can be thought of as spectro-temporal areas where the target speech is least degraded by the masker. When investigating the unique contribution of informational masking from a competing talker, an ideal energetic mask control should provide the same glimpsing opportunities as the competing talker.

Commonly used energetic maskers varying in their spectral characteristics include white noise, pink noise, checkerboard noise (Howard-Jones & Rosen, 1993), speech-shaped noise, spectrally-rotated speech, and time-reversed speech. All of these maskers have been compared to a competing talker to control for EM, however they do have different acoustic characteristics. Stationary white noise is an effective energetic masker, as it has a flat long-term spectral density across all frequencies, which reduces glimpsing. It is thus not as useful if the goal is to match the EM of competing speech. Another example of a stationary masker which has been widely used in intelligibility experiments is speech-shaped stationary noise. This type of masker has a long-term frequency spectrum matching that of the speech from which it was made. Speech-shaped stationary noise can be matched to the spectrum of the target speech, thus creating maximal energetic masking. However, stationary noise by definition does not mimic the amplitude modulations of speech, which help listeners to ‘listen in the dips' when the target speech is louder than the competing speech. Thus, although speech-shaped stationary noise shares the long-term frequency spectrum of speech, its EM is greater than that of competing speech since its amplitude is not modulated.

White noise, pink noise and speech-shaped noise can all be amplitude-modulated to match the amplitude contour of a competing talker, in which case the mask will be described as fluctuating. Spectrally-rotated speech and time-reversed speech are by definition fluctuating maskers, since they preserve the fluctuations of the original speech signal (albeit at different points in time for time-reversed speech). Energetic maskers can thus be designed to match (with various degrees of precision) the spectro-temporal amplitude structure of a competing talker, thereby providing similar glimpsing opportunities. One way of mimicking amplitude modulations of competing speech is to create speech-modulated noise, also known as speech-shaped fluctuating noise, using the intensity envelope of a speech signal. Speech-modulated noise is similar to speech-shaped noise, but in addition to sharing the average spectral characteristics of speech, it follows the temporal amplitude modulations of the speech from which it was created (Brungart, 2001; Moore, 2013). In this way, conditions can be created in which the average spectral overlap of the noise masker is similar to that of a speech masker, also allowing for listening in the dips when the intensity of the target signal is greater than that of the competing talker. However, although speech-modulated noise preserves the long-term average spectral characteristics and the intensity envelope of the original speech signal, on average it provides fewer opportunities to glimpse the target signal than competing speech. This is due to the fact that the spectrum of speech varies across time, with certain spectral regions containing more or less energy depending on the phonemes uttered. In contrast, the spectrum of speech-modulated noise represents the average speech spectrum but with relatively constant energy across all spectral regions. Speech-modulated noise has been used extensively as an energetic mask control for competing speech, but its main disadvantage is the additional EM created by its different spectral profile at any given instant in time.

To circumvent the issue of the additional EM from speech-modulated noise, time-reversed speech can also be used to control for EM. This masker simply flips the signal in the time domain, rendering a speech-like stimulus with the same long-term average spectrum as the original speech, containing no identifiable semantic information. However, unlike speech-modulated noise, time-reversed speech does not preserve the same amplitude envelope as the original speech signal. Time-reversed speech produces similar opportunities for glimpsing to forward competing speech, since the spectral characteristics of speech are preserved. However, since the temporal envelope is reversed, the glimpses do not occur at the same time, and the typical shape of the temporal envelope of speech is distorted. Furthermore, forward speech is characterised by quick onsets and slow decays (Rosen, 1992), reflecting the large number of plosives due to the biomechanical constraints of the vocal tract, whereas reversed speech has abrupt offsets, which leads to more forward masking.

Another energetic masker based on the competing speech signal is spectrally rotated speech. This masker can be created by inverting the speech signal (usually low-pass filtered) around a given frequency, which creates an unintelligible mask to the untrained ear (Green, Rosen, Faulkner, & Paterson, 2013 trained participants to recognise spectrally rotated speech). This mask still contains the same amplitude modulations and certain characteristics of speech such as formant and quasi-harmonic structure, as well as intonation and rhythm. However the spectral characteristics are by definition different at any given point in time.

Finally, competing speech itself has also been used as an energetic mask control (in addition to its IM properties). By using speech in an unknown or made-up language, the semantic content of the competing speech is inaccessible to the listener. However, it is not possible to match the EM of a competing talker in a native language with that of the competing talker in an unknown language, if only because of the different acoustic characteristics of different languages. Multi-talker babble is also a predominantly energetic masker, with a reduced amount of IM. Indeed, although a single competing talker leads to a relatively high proportion of IM, as the number of talkers increases, this proportion decreases in favour of EM. This is due to the decrease in spectro-temporal dips and an increase in overall energy in the competing signal. Multi-talker babble can be constructed from speech in the same language as the target speech, in a different language, or indeed in a made-up language, for example from the International Speech Test Signal (Holube, Fredelake, Vlaming, & Kollmeier, 2010), although this masker has been found to be more distracting than other non-intelligible masks (Francart, van Wieringen, & Wouters, 2011). The advantage of using multi-talker babble is that it sounds speech-like and is a modulated masker. In addition, when the language is unknown to the listener, this should reduce higher-level IM. However, this kind of mask by definition does not mimic the energy in a single competing talker, and in fact leads to greater EM than the competing talker itself.

Despite the fact that none of the energetic maskers described above is a perfect acoustic match for competing speech – indeed the only perfect acoustic match would be the speech itself - they have been used extensively to control for the EM properties of competing speech. Ultimately, the choice of masker depends on the research question the study tries to address (Francart et al., 2011).

1.2.2 Informational masking

Informational masking (IM) is broadly construed as the detrimental effect of a mask once EM has been accounted for (also sometimes referred to as "non-energetic masking", Durlach, 2006). IM is thought to involve more central processes than EM (Kidd et al., 2007), although the exact mechanisms involved are still under investigation. A visual analogy of the difference between EM and IM is illustrated in Figure 1.1. The intact target speech is represented by the picture of the butterfly on the left. The noise (EM) is represented by a grid of grey lines, whereas the competing talker (EM+IM) is represented by a grid with the picture of a flower. When the mask is superimposed on the butterfly, the EM+IM condition results in a picture with information from both pictures that can be perceived as the flower or the butterfly, whereas the EM-only condition results in a picture that is perceived only as the butterfly, albeit degraded.

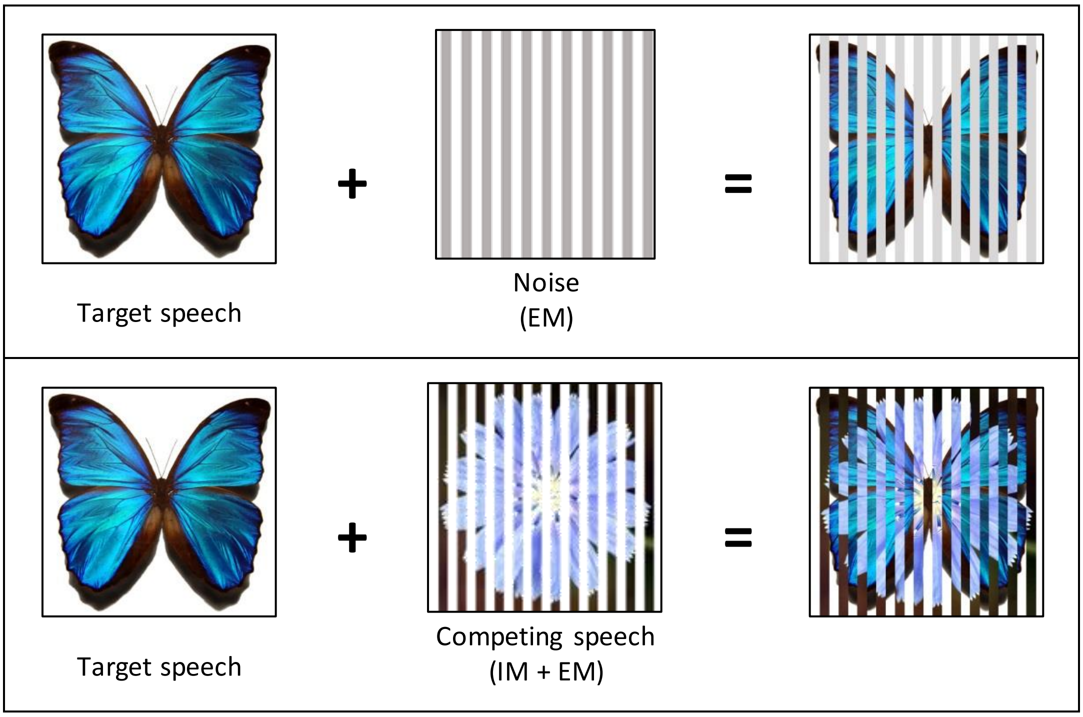


Figure 1.1. Visual analogy of energetic masking (top panel) and informational masking (bottom panel).

The definition of IM as anything that is not EM gives rise to the possibility of IM actually encompassing several components. Broadly speaking, these components can be categorised into low-level IM and high-level cognitive IM. The low-level components of IM have received the most attention. For example, the difficulty in segregating two acoustically similar voices has been attributed to IM (e.g. Brungart, 2001; Moore, 2013). Higher-level components such as voice familiarity (Brungart, Simpson, Ericson, & Scott, 2001) or the semantic content of the utterance are also components of IM. Kidd et al. (2007) have emphasised that IM cannot be pinned down to one specific phenomenon, and that it can refer to any stage of processing taking place after the auditory periphery. These authors also include low-level components such as perceptual grouping and source segregation as well as high-level cognitive functions such as attention and memory within IM.

In an attempt to further clarify the components of IM, Cooke et al (Cooke, Garcia Lecumberri, & Barker, 2008) identified four sources of IM: (1) misallocation of masker signal components to the target signal (e.g., migration of a fricative patch from the masker to the target, resulting in the listener reporting a different percept), (2) competing attention of the masker (especially if the semantic content is relevant to the listener), (3) higher cognitive load (often due to competing attention and interference), (4) interference from a language known to the listener (more interference when the language is known). Although it could be argued that some of these components share some characteristics, this nomenclature is useful for studying IM from a psycholinguistic perspective, since it distinguishes low-level factors (such as point 1) from high-level factors (such as points 2, 3, and 4). The current study will focus on aspects relating to points (2) and (3), namely the allocation of processing resources to both the target and the masker signal, that in turn leads to increased cognitive load, or a depletion of domain-general cognitive resources.

A complementary view to the components of IM has been theorised by Shinn-Cunningham (2008), who suggests that IM can be primarily explained by failures of auditory attention. According to this view, IM can arise when a listener fails to separate a target auditory 'object' from a competing source (failure of object formation), or when the listener fails to maintain attention on the target, due to the competing source grabbing their attention (failure of object selection). Failure of object formation corresponds to the low-level aspects of IM previously described (e.g. segregation) whereas failure of object selection corresponds to both low-level and high-level aspects of IM. Failures of object formation are typical in cases when the target and competing voices are acoustically similar (e.g. same gender), which leads to a difficulty in segregating the two streams. Failures of object selection can arise even when target and competitor have been successfully segregated, in particular when the competing speech is louder or more salient, when the target and competitor are similar, or when there is uncertainty about the target. These situations lead to involuntary attention on the competing signal due to its bottom-up salience, which overrides top-down attention. In this context, bottom-up salience can be due to the loudness of the signal, but also other features such as the semantic content of the competing signal, for example if one hears one's own name across the room (Shinn-Cunningham, 2008). In the context of this thesis, I will be focusing on the high-level object selection components of IM due to a competing talker, which I have termed 'informational interference'.

1.2.3 Informational interference due to a competing talker

Many studies have investigated the effect of a single competing talker mask on speech intelligibility. These studies have mostly compared single-talker masks with multi-talker babble, time-reversed speech, stationary speech-shaped noise or speech-modulated noise. There have been contradictory findings with regard to the detrimental effect of a competing talker compared to matched energetic masks. Some studies suggest that a competing talker is more detrimental to intelligibility than an energetic mask, whereas others suggest that a competing talker has as detrimental an effect as an energetic mask alone, and in some cases it is even less detrimental than EM.

Table A.1 (Appendix A) compiles several studies investigating the effect of a competing talker on speech perception and comprehension, classified by whether they found a detrimental effect of the competing talker or not. There is no one factor that explains the difference between the studies in this table that have found a detrimental effect of mask and those that haven't. A range of masks has been used for both types of studies, as well as a range of target and competitor materials, signal to noise ratios (SNRs), and voice genders.

For example, Brungart and colleagues (Brungart, 2001; Brungart et al., 2001) compared performance in an intelligibility task where a target sentence was masked by one, two or three competing talkers as a function of SNR and type of masker (competing talker(s) or speech-modulated noise). Intelligibility was lower in the single competing talker condition than in the speech-modulated noise condition, which was attributed to the informational mask of the competing talker. Similarly, Trammell and Speaks (1970) compared forward competing speech to reversed competing speech in an intelligibility task, and found that the 50% speech reception threshold (SRT) was significantly higher (worse performance) for the forward competing speech. However, when comparing speech maskers with time-reversed speech maskers, Dirks and Bower (1969) and Hygge, Rönnberg, Larsby, Arlinger, and Rönnberg (1992) found no difference in performance between the mask with semantic content and the reversed speech. Dirks and Bower (1969) concluded that the semantic content of a speech mask did not add to the difficulty of identifying sentences. Likewise, Hygge et al. (1992) found evidence that a linguistic mask with or without content (forward vs. reversed speech) does not affect listeners differently.

A direct comparison between a competing talker and noise is further complicated by signal-to-noise ratio (SNR) considerations. For EM from noise maskers, as SNR decreases (i.e. as the target speech intensity decreases in relation to the masker), intelligibility reduces monotonically. In contrast, performance with a single competing talker (EM + IM) seems unchanged when the SNR is between 0dB and -10dB (Brungart, 2001; Dirks & Bower, 1969).

It would seem therefore that under certain circumstances a competing talker can be more detrimental to intelligibility than its energetic mask alone, under others the effect of both mask types is comparable, and in yet others a competing talker is in fact less detrimental than energetic mask controls.

One of the difficulties of studying the unique contribution of the high-level components of IM from a competing talker is partialling out EM as well as low-level IM. As mentioned earlier, low-level components of IM include sound source separation, and migration of phonetic information from mask to target. High-level components of IM involve language-specific characteristics, such as lexical selection and semantic content.

The use of maskers in a native or non-native language is one way the low-level components can be teased apart from the high-level components of IM. Presumably, if the semantic content of speech does not have a major effect, then there should be little to no difference between a mask presented in a known versus unknown language. In contrast, if the semantic content of speech does have an effect, then a known language should be more detrimental than an unknown language. In accordance with this latter prediction, most studies investigating English speech recognition with a different language masker have shown that an unknown language does indeed lead to a release from masking. The masker languages have included Spanish (Lecumberri & Cooke, 2006), 2-talker Dutch (Freyman, Balakrishnan, & Helfer, 2001), and 2-talker Mandarin (Van Engen and Bradlow, 2007). It is not clear from these studies whether the release from masking with an unknown/non-native language is due to linguistic factors (access to semantic content in the mask) or lower-level spectro-temporal characteristics of the mask.

In a study aiming to tease apart linguistic factors from spectro-temporal factors in release from masking in an unknown language, Calandruccio, Dhar and Bradlow (2010) measured IEEE sentence recognition with 2-talker speech masks in a known language (English) or an unknown language (Mandarin). To separate the effect of higher-level semantic content (termed linguistic interference in this study) as opposed to lower-level acoustic and phonetic content of the speech mask, the authors also included Mandarin-accented English conditions that varied in intelligibility and, in one experiment, speech-modulated noise and stationary speech-shaped noise. Note that all targets and maskers were recorded by male voices, increasing EM and low-level IM. Participants' performance was found to depend on the SNR. At -3 dB, there was no difference between the Mandarin mask and the native English mask, presumably because the SNR was not challenging enough to bring out the effect of linguistic interference. However, at the more challenging -5 dB SNR, performance was better for the Mandarin mask compared to the native English mask, indicating that linguistic interference can play a role in release from IM. Across both SNRs, the least intelligible accented English (i.e. the one with the least readily accessible lexical information) led to the highest performance compared to the two other accented English conditions. This is in accordance with the hypothesis that release from masking can be influenced by the linguistic content of the mask. These results were confirmed by a follow-up experiment comparing sentence repetition masked by speech-modulated noise created from each of the 2-talker masks, at -5 dB SNR. Performance with the SMN created from the Mandarin and the native English masks did not differ, confirming that the lack of difference in the -3 dB SNR in the first experiment was likely driven by the EM of the two masks, and that the lower SNR brought out the effect of linguistic interference from the native English mask. Furthermore, when comparing performance between the SMN conditions and the 2-talker mask conditions, the only difference was for the native English condition which showed release from masking in the SMN compared to the 2-talker mask. The authors concluded that although spectral differences between the target and maskers in their experiments did explain part of the release from masking observed, linguistic differences between the target and maskers also played a role in release from masking, in particular when the task was challenging for the auditory and cognitive systems (in their case with the lower SNR).

Another experiment that points to the role of SNR in release from IM with different languages was conducted by Gautreau, Hoen, and Meunier (2012). These authors presented French target words with Italian, Irish Gaelic, or French masks, with an additional speech-modulated noise mask to control for EM effects. The task was a speeded lexical decision to the target French stimuli. The authors found that at 0dB SNR, there was no difference in reaction times between the speech-modulated noise (SMN) and language masks, but did find a difference at -5dB. At this SNR, participants' lexical decisions were slower when the mask was a known language (French) than when it was SMN. Furthermore, they found that participants were slower with French and Italian than with Irish Gaelic. This was attributed to the segmental and prosodic similarities between French and Italian. However, they found no difference in reaction times when the mask was an unknown language (Italian or Irish Gaelic) compared to the corresponding speech-modulated noise. On the basis of the absence of difference in reaction times between SMN and a competing talker in an unknown language, Gautreau et al. (2012) concluded that the unknown language masks (Italian and Irish) and the SMN masked the target speech on an acoustic level, and not on a linguistic level. In contrast, when the masker was the same language as the target, masking was both acoustic (energetic) and linguistic (informational). The results of this study and the previous one are in accordance with the hypothesis that the higher level, linguistic content of a speech mask acts as a specific type of interference, and that the lower level, acoustic or energetic content of the mask does not interfere as much.

The studies mentioned above mostly used measures of intelligibility such as speech reception thresholds (SRTs) with performance levels as low as 50% , and signal-to-noise ratios as low as -30dB (Lew & Jerger, 1991). However, at this level of intelligibility, the acoustic signal is highly degraded and might not reflect typical acoustic environments. The cost of processing target speech masked by competing speech when intelligibility is high may not be detectable when using SRTs at 50% performance accuracy.

A handful of studies have supported the idea that the cost of informational masking from a competing talker may be visible only in certain circumstances, in particular when processing resources or effort are measured. Brungart et al. (2013) undertook a series of studies investigating different types of speech or energetic maskers and their effects on performance in a variety of increasingly complex listening tasks. In the following paragraphs I will describe these studies in detail, given that the conclusions are important to the motivation behind my own experiments. Across three different experiments, listeners were asked to perform tasks of increasing complexity using the coordinate response measure or CRM corpus (Bolia, Nelson, Ericson, & Simpson, 2000) or modified sentences from the Revised Speech Perception in Noise Test or R-SPIN (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984) in different listening conditions. The R-SPIN sentences are composed of low predictability sentences, e.g. “I want to know about the crop”, and high predictability sentences e.g. “The farmer harvested his crop”.

The masks in Brungart et al. (2013) consisted of a competing talker (CT), time-reversed competing talker (RCT), four-talker babble, speech-modulated noise (SMN), and stationary speech-shaped noise (SSN). These experiments also manipulated task complexity, with the hypothesis that the informational masking from competing speech taps into central cognitive resources which are not affected by energetic masking alone. The measure of interest was therefore the relative cost of performing a more complex task under the different masking conditions. Complexity was operationalised in three ways across three different experiments. In the first experiment, participants completed a 2-alternative forced choice task. There were three tasks increasing in complexity: a detection task, a discrimination task and an identification task. All of these tasks presented the stimuli with one of two masks: a different CRM sentence spoken by a speaker of the same gender as the target or speech-shaped noise. In the detection task, participants had to identify the target CRM sentence from a sequence of two stimuli containing a CRM target sentence and a masker only. In the discrimination task, participants had to identify the target CRM sentence from a sequence of two stimuli containing a CRM target sentence and a reversed CRM sentence. Finally, in the identification task, participants were presented with a display containing a colour-number combination prior to hearing a sequence of two CRM sentences, and they had to indicate which of the sentences corresponded to the display. Speech reception thresholds at 75% correct were calculated using SNRs ranging from -56 dB to 8 dB in 4 dB steps. The authors report that the SRT_{75} for the identification task (the most complex task) was -18 dB for the competing speech compared to -8 dB for the noise masker. In other words, the competing speech led to release from masking compared to the noise. However, the authors also interpolated and reported performance across the three tasks when the SNR was fixed at the level required to obtain 75% correct in the detection task. This method of presenting the data is useful to compare the relative decrement or improvement for each mask condition between tasks varying in complexity. When the results were analysed in this way, it became apparent that although there was no decrement in performance for either the speech or the noise masks between the detection and the discrimination tasks (the two least complex), there was a decrease in performance for the speech mask but not for the noise mask in the identification task. This first experiment highlights the importance of going beyond simple SRTs when studying the effect of a competing talker.

The second experiment by Brungart et al. (2013) required participants to select the correct colour-number combination, based on the information in the target sentence. Target sentences were either presented in isolation or masked by one of five masks: speech-shaped noise, speech-modulated noise, competing speech (CRM sentence), reversed competing speech, or babble. Complexity was manipulated across four conditions: (1) monaural, (2) target in known ear, (3) target in unknown ear, and (4) respond to both ears (in order of least to most complex). In the monaural condition, the target and masker were both presented to the participant in one ear only. In the “target in known ear” condition, the masked target was presented in one ear, while another competing CRM sentence (also masked) was presented in the other ear. Participants were told in advance which ear the target sentence would be presented in. The “target in unknown ear” condition was identical to the previous condition except that participants did not know in advance which ear the target would be presented in. Finally, the “respond to both ears” condition was also identical to conditions (2) and (3), except that participants had to select the colour-number combination described by both the left and the right ears. SRTs at 80% correct were calculated for 19 SNRs ranging from -27 dB to +21 dB. SRTs were reported for the monaural condition and showed that the lowest SRT (i.e. better performance) was in the competing talker and reversed competing talker conditions, followed by the speech-modulated noise, then the speech-shaped noise and finally the babble. If the authors had limited their analysis to SRTs only, the conclusion would have been that competing speech is not more detrimental than energetic mask controls. The authors also reported performance for each task when the SNR was set at SRT_{80} for each mask based on the easiest (monaural) task. They found that although performance decreased as task complexity increased for all mask types, the detrimental effect of task complexity was greatest for the competing speech and the reversed competing speech. Once again, the conclusions were very different depending on how the results were analysed and whether task complexity was taken into account or not.

The third experiment described in Brungart et al. (2013) manipulated task complexity by introducing a working memory task within the speech intelligibility task. Target sentences were taken from the high probability set of R-SPIN sentences and were modified to produce a new set of anomalous sentences in addition to the high probability set. For example when the high probability sentence was “His plans meant taking a big risk”, the new anomalous sentence was “His doctor drank a lost risk”. The maskers consisted of two-talker competing speech in the same voice as the target sentences (passages from fairy tales), or two-talker speech-shaped noise. Participants were asked to repeat each masked sentence (0-back task), or to repeat the masked sentence that was presented directly before the most recent sentence heard (1-back task). The manipulations in this task were thus in the sentence type (high probability vs anomalous), the type of mask (speech vs noise), and the type of memory load (0-back or low memory load vs 1-back or high memory load). The SRT yielding 80% correct responses was calculated for the 0-back task, and showed that the speech mask led to a higher SRT than the noise mask, i.e. the noise mask was less detrimental than the speech mask. The SRTs were higher for the high probability sentences compared to the low probability sentences, but it did not depend on mask type. When the SNR was set to the value corresponding to the SRT_{80} in the 0-back task, it became apparent that there was a greater decrement in performance for the 1-back task when the mask was speech compared to noise.

In summary, across three different experiments using equivalent speech reception thresholds, Brungart et al. (2013) found that increasing the complexity of the listening task had a more detrimental effect for speech maskers than for noise maskers. This was visible in the larger drop in performance for the CT and RCT conditions compared to the babble, SSN, and SMN conditions. The authors conclude that speech maskers require additional cognitive resources or effort to be allocated, and that this additional effort was only visible because they used tasks increasing in complexity. The authors attribute this increased effort to informational masking, in particular the difficulty of extracting “the acoustic and phonetic elements of a speech signal from those of a potentially confusable speech masker”. As mentioned in previous sections, difficulties in segregating target speech from competing speech correspond to lower levels of informational masking. Indeed, the experiments reported by Brungart et al. (2013) used voices of the same gender for the target and masker, thus increasing difficulty due to segregation. Because of this, it is not possible from these experiments to determine whether the higher levels of informational masking (e.g. linguistic interference from the content of the mask) also played a role in increasing the cognitive resources or effort involved in dealing with competing speech.

Taken together, the studies reviewed in this section have shown mixed results with regard to the effect of a competing talker on sentence recognition and intelligibility. Some authors claim that a competing talker has a detrimental effect on sentence recognition beyond its energetic component, whereas other authors claim that a competing talker is just as detrimental as energetic mask controls. Brungart et al. (2013) showed that the type of task and the measure used are paramount to observing the added cognitive load induced by competing speech. However, they investigated low-level informational masking. Furthermore, these studies measured speech recognition and intelligibility, but did not include measures of comprehension. Although recognition is the first step to successful listening, it does not stop there, as the message then has to be interpreted and understood. Presumably, this step requires additional processing resources or effort, further increasing the possible detrimental effect of a competing talker.

In an attempt to provide a more complete picture of listening in adverse conditions, the experiments in this thesis focus on measures of sentence comprehension. Furthermore, these measures were chosen to capture the effort or cognitive resources required to process sentences in the presence of a competing talker, rather than intelligibility alone. In the present work, I will more specifically attempt to measure the cost involved in understanding sentences, rather than simply hearing and repeating them. The term that will be the focus of this thesis is “informational interference”. As previously mentioned, this term refers to the higher-level aspects of IM (without low-level IM such as segregation). Informational interference should lead to greater listening effort, which may not be directly measurable in intelligibility tasks. I hypothesise that the cost of informational interference can be quantified through online measures of processing load, such as reaction times. Although the term “listening effort” is not well defined, McGarrigle et al. (2014) suggest that it is “the mental exertion required to attend to, and understand, an auditory message”. I propose that informational interference arises even when a target speech signal is masked by a competing speech signal at a signal-to-noise ratio allowing intelligibility to remain high, through an increased reliance on cognitive processes (and therefore increased listening effort). I will adopt the view that a speech masker competes with the target signal through its potential linguistic relevance to the listener, thus prompting the listener to automatically process the mask, or elements of it. This, in turn, would involve additional cognitive processes (or “mental exertion”) that would otherwise have been allocated to the target signal. The effect of informational interference would be particularly notable if the target signal also involves additional cognitive processes (e.g., syntactically complex sentences) and if we adopt the view that there is a general pool of limited processing resources (e.g. Kahneman, 1973; Rudner et al., 2011). In the next section, I will explore the notion of cognitive resources and the role that cognition plays in listening to speech in adverse conditions.

1.3 The role of cognition in masking and sentence comprehension

In the previous section, I mentioned that informational interference should increase the involvement of cognitive resources, and I hypothesised that informational interference from a competing talker involves the higher-order, cognitive aspects of IM. What do these cognitive resources consist of, and what are these cognitive aspects of IM? I will explore the role that cognition plays in sentence comprehension in adverse conditions before focusing on the specificities of sentence comprehension with informational interference. Multiple components of cognition are involved in the complex task of understanding masked or degraded speech, including working memory (WM) and executive processes such as inhibition, attention and cognitive control. A listener must select the relevant speech stream, inhibit irrelevant information, maintain attention to the relevant stream, piece together an imperfect target signal, identify the phonemes and words of the utterance, keep words in WM and reassemble the input to form a meaningful sentence following syntactic properties stored in long-term memory, and finally make a decision based on the understanding of the sentence. At any stage in this complex chain, if any of the processes breaks down, the entire task could be compromised. The development of the field of 'cognitive hearing science' (Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009) in recent years highlights the increasing interest in combining research from cognitive psychology and hearing sciences, emphasising the interactions between hearing and cognition. Unsurprisingly, many studies have investigated the cognitive factors involved in listening to speech in adverse listening conditions, in particular WM and selective attention. Most of these studies have looked at individual differences in hearing impaired and/or older listeners, as these groups often exhibit greater difficulty both in speech perception in adverse conditions and in certain cognitive functions.

Akeroyd (2008) reviewed the link between individual differences in speech processing in adverse conditions and cognitive ability, spanning twenty studies from 1989 to 2008. These studies cover a range of linguistic levels (phoneme level, word level and sentence level – though not syntactic processing per se), adverse conditions (modulated and unmodulated noise maskers, speech maskers, and time-compressed speech), and tests tapping into different cognitive resources (e.g. working memory, attention, visual analogues of speech in noise tasks, general IQ and academic ability). Nineteen out of the twenty studies reviewed by Akeroyd (2008) found relationships between the listening tasks and the cognitive tasks, albeit not for all cognitive measures. These results confirm the relevance of studying cognition in relation to speech in adverse conditions.

Note however that the studies mentioned in Akeroyd (2008) often used relatively low speech reception thresholds as dependent measures (e.g., 50%), which implies that the target speech was highly degraded. When the target speech is degraded, listeners must conduct more ‘guesswork’ than with highly intelligible speech, because parts of the signal will not have been heard. When portions of the speech are not available, cognitive resources (in particular WM) should be taxed because segments of the signal have to be kept in WM and pieced together to make inferences about the meaning of the speech. It is thus not surprising that WM would be involved with relatively unintelligible speech. However, what these studies do not address is how WM and other cognitive processes are involved when the SNR is high enough for the signal to be fully intelligible. This is the situation that most listeners are faced with in everyday environments. The involvement of cognitive resources in listening to and understanding speech in adverse conditions should be different for highly intelligible speech compared to unintelligible speech, because the listener does not have to reconstruct a degraded signal. In other words, cognitive resources are probably taxed in different ways under EM compared to IM, in particular informational interference from a competing talker.

Recent studies have started making reference to the concept of cognitive spare capacity as a way of determining the cognitive resources that are available during successful listening (Rudner et al., 2011). Cognitive spare capacity is a generic term that encompasses different cognitive components, including WM, attention, and executive functions such as inhibition and updating. Several tests have been proposed to measure cognitive spare capacity, usually measuring different constructs within the same test. For example, the Cognitive Spare Capacity Test (CSCT) claims to measure working memory storage, multimodal binding, and executive resources (Mishra, Lunner, Stenfelt, Rönnberg, & Rudner, 2013). The fact that tests like the CSCT measure so many different constructs at once is useful for agglomerating a range of processes in a single index, but it does not allow the unique contribution of each process to be pinpointed. In the following sections, I will explore how two specific cognitive processes, attention and WM, are related to listening. I hypothesise that the depletion of processing resources due to informational interference will exercise its effect by reducing working memory capacity and attentional resources beyond what would be expected with EM.

1.3.1 Listening and attention

Since the task of listening to speech in the presence of a competing talker involves selectively attending to the target while ignoring the irrelevant stream, in this section I will briefly review the link between listening and attention, and in particular the role of attention in recognising and understanding masked speech. Intuitively, the notion that attention is involved in successful listening is evident. Children are often told to "pay attention" to what the adult is saying, and most of us have had the experience of not following part of a conversation simply because our attention was not focused on the speaker. In everyday language, it is common to make a distinction between hearing and listening, where listening is usually thought of as hearing while paying attention. Attention is essential in listening even at low levels such as pure tone detection. For example, Baldwin and Galinsky (1999) measured audiometric pure-tone thresholds and found that participants' pure-tone thresholds were elevated by a secondary visual task requiring them to divide their attention between the two tasks. Thus even at a low level of acoustic identification, attention can have an influence on performance.

Attention also plays an important role at higher processing levels, for example in the recognition and comprehension of masked speech or when listening to one of two competing messages. The role of attention in listening to one of two competing messages was famously studied by Cherry (1953), who was interested in the "cocktail party problem", or how listeners "recognize what one person is saying while others are speaking at the same time" (p.976). Cherry devised a series of dichotic listening experiments where participants were asked to shadow the message in one ear while another message was presented in the other ear. Although capable of repeating the message in the attended ear, listeners were unaware of a change in language in the unattended message, were unable to report the content of the speech, and a majority did not notice when reversed speech was played instead of normal speech. However they did report low-level characteristics of the unattended message such as a change in voice gender or the use of a pure tone instead of speech. Thus, from these experiments it is possible that listeners selectively attend to the target channel based on low-level acoustic characteristics while filtering out the higher-level characteristics of the unattended message very early on.

Following on from Cherry's set of experiments, Moray (1959) further explored the finding that the unattended channel may be completely ignored, and what factors might capture listeners' attention. Participants were asked to shadow passages presented in one ear with a competing message in the other ear. The prose was interspersed with instructions such as “you may stop now” or “change to the other ear”, presented in the target channel or the competing channel. Participants were not informed of the presence of these instructions. Crucially, the instructions were preceded either by “All right” or by the participants’ full name. Most listeners were still impervious to the competing message regardless of its content, but about 33% were aware of the instructions when they were preceded by their own name. Moray termed this the “identification paradox”: although the content of the competing message was “blocked below the level of conscious perception”, listeners’ attention was captured when their own name was presented. Moray hypothesised that the focus of attention on the target message does not block out low-level features such as “simple sounds”, and that words in the competing channel are treated as meaningless sounds. He further hypothesised that the extraction of meaning from both messages partly takes place “below the level of conscious perception”, and that it is only when the competing channel contains “important” information that attention is no longer focused exclusively on the target channel.

Moray’s results were replicated by Wood and Cowan (1995) in a more controlled manner (e.g. larger sample size, acoustic similarities controlled across conditions). Similarly to Moray (1959) they found that 34.6% of participants heard their own name in the competing channel. In addition, for those participants, an attention switch to the competing channel occurred for a short time after hearing their name, evidenced in increased errors and greater response lags for the two target words immediately following their name. However, these participants did not show an increase in errors or in response lag for the target word that was presented at the same time as they heard their name, indicating that they did not just happen to change their focus of attention to the competing stream at that point, but rather that their name captured their attention, leading to a temporary attention switch to the competing stream.

These studies demonstrate the importance of considering attention when studying the issue of speech recognition and understanding in adverse conditions, particularly how listeners follow one of two competing messages. Models of attention can provide frameworks for such experimental findings. Broadly speaking, these models can be separated into early and late-selection models. One of the earliest models of attention was developed by Broadbent (1958), partly based on listening experiments similar to Cherry (1953) and Moray (1959) where participants were asked to attend to one of two messages presented dichotically (e.g.

Broadbent, 1952a, 1952b). Broadbent suggested that sensory information is filtered and selected at a very early stage of processing. Once the relevant streams of information are selected, semantic processing occurs only for those selected streams. However, this model does not fully account for the "identification paradox" described by Moray (1959), since listeners' attention can be captured at a later stage when the content is their name. According to Broadbent's early filter model, these results could be explained by occasional attentional drifting to the other channel, which would not allow for the target message to be processed during that time. However, the fact that participants in Wood and Cowan (1995) did not show delayed or less accurate responses for the target word presented at the same time as their name indicates that their attention was not focused solely on the competing stream at that particular point in time. Thus, although Broadbent's model does account for the very early filtering of information based on acoustic characteristics of the unattended stream such as results reported in Cherry (1953), it does not account for the capture of attention and semantic processing at such a late stage, such as in Moray (1959) and Wood and Cowan (1995).

In contrast to Broadbent's early filter model, Deutsch and Deutsch (1963) proposed that the filtering of information takes place at a later processing stage. According to this view, irrelevant information is only filtered out by attention once objects have been fully perceived. Thus, even seemingly unattended-to stimuli are processed, which accounts for the "identification paradox". However, the model posits that although there is semantic processing of the unattended stream, it does not involve working memory or awareness as long as the information in the unattended stream is not relevant or important to the listener.

A few years later, Treisman (1969) developed and modified Broadbent's early filter model, to account for results such as Moray's. The resulting "attenuation model" was also a model of early selection, but instead of assuming that irrelevant information is filtered out completely after the first stage of low-level acoustic analysis, the model proposes that the irrelevant information is attenuated, or weakened. A series of hierarchical analysers then process the information from the target stream. As long as the system still has the required capacity, some or all of the irrelevant information is also analysed. Crucially, some words have lower thresholds than others, leading to a greater probability of their being processed (they are more attention-grabbing). For example, one's own name will always have a low threshold of activation, whereas irrelevant or nonsensical information has a high threshold of activation. Whether or not the competing words are attended voluntarily therefore depends on their threshold of activation but also on the general capacity of the system. If the capacity of the system is reduced by a task requiring high processing resources, the threshold of activation of a competing word will be higher.

In an attempt to reconcile the early vs. late filter debate, Lavie developed "load theory" (e.g. Lavie & Dalton, 2014; Lavie & Tsal, 1994; Lavie, 2005) based mainly on the findings of studies in visual attention. The authors argue that the data supporting early selection arise from experiments with high perceptual load, whereas the data supporting late selection arise from experiments with low perceptual load. According to this theory, perceptual capacity is limited, so when a task exceeds the capacity of the perceptual system (i.e. by imposing high perceptual load), the irrelevant stimuli cannot be processed. Perceptual load has been manipulated by changing the number of items displayed, the perceptual similarity of the items, or indeed the "processing requirements" of the task (Lavie & Dalton, 2014). In other words, when perceptual load is high, early selection occurs, whereas late selection can occur when perceptual load is low. Several experiments testing the perceptual load theory of attention in vision support the idea that when the perceptual load is high (e.g., several different distractors sharing similar visual properties with the target), interference from an irrelevant stimulus is lower. In contrast, when the perceptual load is low (several repeated distractors that are very different from the target), interference from an irrelevant stimulus is higher. One aspect of perceptual load theory that is relevant to this section is that, in addition to perceptual load determining whether early or late selection takes place, higher-level cognitive mechanisms (e.g. working memory) also play an important role in regulating distraction. These higher-level cognitive mechanisms include cognitive or executive control. In Lavie's load theory, cognitive load has the opposite effect to perceptual load: when the load on cognitive control is high, a competing stimulus increases interference.

Subsequent studies have extended the perceptual load theory to hearing, based on a series of multimodal tasks. Both visual and auditory tasks varying in perceptual load or cognitive control load were presented as dual-task paradigms and yielded similar results to the visual-only experiments. For example, Francis (2010) conducted an experiment to determine whether spatial release from masking in perceiving speech with a competing talker stemmed from a release from perceptual load or cognitive load. Participants had to identify words while in the presence of competing speech (the two voices were different genders). Cognitive load was manipulated by introducing a secondary visual memory task. Perceptual load was manipulated by asking participants to respond to cues for either pitch or modulation (low load) or both pitch and modulation (high load) of a non-speech amplitude-modulated tone. The authors found that increasing the perceptual load decreased the interference from the competing talker, whereas increasing the cognitive load (visual memory) increased interference from the competing talker, in accordance with load theory.

In an experiment investigating perceptual load across vision and hearing, Macdonald and Lavie (2011) showed that increasing visual perceptual load decreased participants' ability to notice the presence of a pure tone presented simultaneously ("inattentional deafness"), again in accordance with the load theory of attention. Raveh and Lavie (2015) showed similar cross-modal effects, again using tone detection while performing a high or low-load visual search task. However, unlike the previous study they did not include cognitive load, and the main task did not involve processing sounds, let alone speech.

Although the above tasks support Lavie's load theory of attention and point to a possible extension in the auditory domain, Murphy, Fraenkel, and Dalton (2013) suggest that load theory does not always hold true in the purely auditory domain. In a first experiment, participants were asked to determine when a target phoneme was presented via the loudspeaker directly in front of them, within a string of flanker phonemes. While participants heard the string of phonemes from the front speaker, a flanker (distractor) phoneme was presented via a loudspeaker to their left or right. Perceptual load was manipulated by increasing the similarity between target and flanker phonemes[1]. In the second experiment, participants had to respond to pure-tone targets presented in one ear while white noise was presented in the other ear. Perceptual load was manipulated by asking participants to attend to either stimulus duration alone (low load) or stimulus duration and frequency (high load). The distractor in this experiment was an unrelated word presented during the last trial in the unattended ear (in addition to the white noise). The third experiment used the same procedure and stimuli as the second one, except that the target pure tones were replaced with words. For all of these experiments, the authors found no evidence of a greater effect of interference under low perceptual load than under high perceptual load. They suggest that load theory does not apply to the auditory domain, and that the auditory system has "surplus capacity" to deal with the information in competing streams as well as the target stream, independently of the perceptual load induced by the target stream. Although the

[1] : For example if the target phoneme was /ti:/, it could be presented as part of a string of six phonemes composed of five times /ɛks/ (low load) and a /ti:/ or as part of a sequence of six different phonemes, e.g. /si:/ /ɛm/ /di:/ /ti:/ /ɛs/ /ʒi:/ (high load)

manipulations in these experiments were designed to manipulate perceptual load, it is still unclear how perceptual load might be instantiated in the auditory domain in other tasks.

Finally, as previously mentioned, Shinn-Cunningham (2008) has suggested an adaptation of theories of visual attention to auditory attention, in particular related to informational masking. In this model, attention operates at both the lower level of spectro-temporal grouping (object formation) and the higher level of object selection. Shinn-Cunningham takes care to specify that her model is not a hierarchical model but a heterarchical relationship between the different components. In other words, there is a constant interaction between the different levels of the model, and in particular attention is involved at both low and high levels of processing. It is therefore neither an exclusively late selection model nor an exclusively early selection model, but an intermediate account of attentional processing.

The debate about what conditions give rise to attention being allocated at an early stage or a late stage in the auditory domain, and whether the early and late accounts are mutually exclusive is still ongoing. In the case of interference from a competing talker, which is the focus of this thesis, there are several possibilities of what the mechanisms of attention could be. If attention is allocated at an early stage (such as Broadbent's early filter model), listeners could apply a perceptual filter based on the acoustic characteristics of the voices (e.g. female vs. male) to select the target and block out the competitor. In this case, the semantic content of the competitor should not interfere with the main task of following the target. However, if the selection takes place at a later stage, and listeners' attention is shared across both streams despite knowing which stream is irrelevant, then we could expect that the semantic content of the competing utterance would be available to the listener. In this case, we might expect relevant semantic content to interfere with processing of the target sentence. The middle-ground view posited by Lavie leads to the prediction that the competing talker would interfere with target speech processing only in situations with low perceptual load and/or high cognitive load. One way of increasing cognitive load is to increase working memory demands. Following Shinn-Cunningham's model, attention operates both during object formation and object selection, depending on factors such as the listener's goals and previous knowledge. Thus the content of a competing talker could lead to a failure in object selection, for example if it is relevant to the listener or the task.

1.3.2 Listening and working memory

Another cognitive component that has been found to correlate with speech intelligibility in adverse conditions is working memory (WM). Before looking at some studies that have found links between adverse listening conditions and working memory, I will give a brief overview of the components of WM and how WM capacity can be measured. While definitions and models of working memory vary (e.g., Cowan, 1999 or Nairne, 1990), probably the most widely accepted model of WM is Baddeley's multi-component model (Baddeley, 1992, 2000, 2012), of which the latest version is represented in Figure 1.2. This model has the advantage of being able to accommodate many of the aspects of other WM models, namely Cowan's embedded processes model (for a discussion, see Baddeley, 2012).

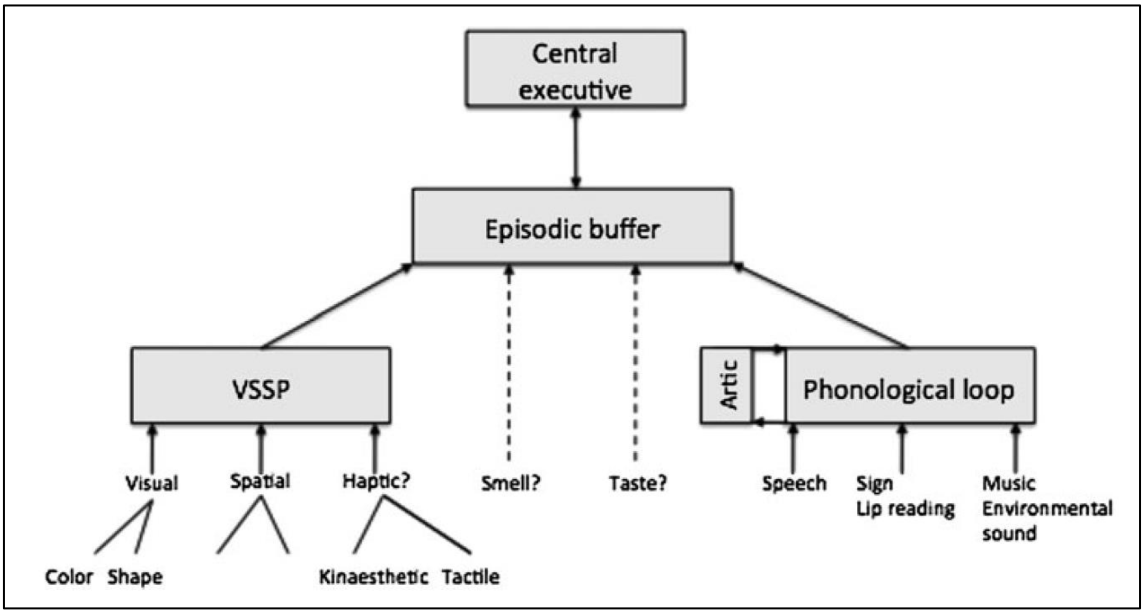


Figure 1.2. Baddeley's (2012) multi-component working memory model, with speculative flows of information from perception (e.g. speech, sign, lip-reading, environmental sound) to WM. Of particular interest to the questions in this thesis are the central executive, the episodic buffer and the phonological loop (with its articulatory loop).The visuo-spatial sketchpad (VSSP) is not of immediate relevance to this thesis as it deals with visuo-spatial information.

According to this framework, WM refers to "a limited capacity system allowing the temporary storage and manipulation of information necessary for such complex tasks as comprehension, learning and reasoning" (Baddeley, 2000). Central to this definition are the concepts of *storage* and *manipulation*. Broadly speaking, the storage component of verbal WM involves the phonological loop and the episodic buffer, whereas manipulation of verbal information is carried out by the central executive in addition to the phonological loop and episodic buffer. The central executive can be thought of as "virtually a homunculus" (Baddeley, 2012), since it involves many processes such as selective and divided attention, inhibition, storage, and decision making. Related to the previous section in this chapter exploring attention, it is important to note that attention is actually included in the executive processes of working memory. Other models also include attention within the construct of working memory (Cowan, 1999), but as yet the exact mechanisms and interactions between WM and attention have not been agreed upon (Shah & Miyake, 1999).

When the central executive is not engaged, this is referred to as short-term memory (STM). Tasks measuring verbal short-term memory capacity (or storage only) include non-word repetition and forward digit spans. Verbal STM is paramount in language development, with several studies having shown a correlation between vocabulary and children's phonological short-term memory as measured by the non-word repetition span (e.g. Gathercole & Baddeley, 1990; Gathercole, Willis, Emslie, & Baddeley, 1992). Some authors suggest that phonological short-term memory plays a fundamental role in vocabulary development (Baddeley, Gathercole, & Papagno, 1998) and foreign vocabulary learning (Papagno, Valentine, & Baddeley, 1991). Due to the importance of verbal STM in language development and processing, and the fact that verbal STM is involved by definition in WM, I decided to include a test of non-word repetition in my experiments in an attempt to pinpoint the level at which memory may influence performance with a competing talker.

Verbal WM capacity (WMC) or span is typically assessed using tasks where participants have to process, store, and manipulate verbal information. Such tasks include reading span measures (e.g. Daneman & Carpenter, 1980) and backward digit spans, and can be in the auditory modality or the visual modality, or both. In typical reading span tests, individuals judge the plausibility of a sequence of written sentences, while remembering the first or last word of each sentence for later recall. The number of sentences in each set increases until the participant can no longer recall a predetermined number of words. The maximum number of items within the set is the individual's verbal WM span or WMC. An equivalent test in the auditory modality is the listening span test. Arguably, tests such as the listening or reading span tap into other cognitive functions in addition to working memory, and may in fact be indicators of more than just working memory capacity. Indeed, participants must carry out a semantic analysis of the content of the sentence. Although working memory might be involved in efficient semantic processing, semantic processing is not a core component of working memory. A high span in these tests may therefore be an indication of semantic processing in addition to working memory capacity. A classic working memory test that involves minimal semantic processing is the backward digit span test, where individuals have to retain a series of digits and repeat them in reverse order.

1.3.2.1 Working memory and adverse listening conditions

Previous research has indicated that STM and/or WM could be involved in listening to speech in adverse conditions, and it could be particularly involved in IM. For example, discussing the causes of IM, Kidd et al. (2007) highlight the possibility that IM can arise from "limitations on the short-term storage and retrieval of sounds in memory, or interruptions in the processing of stored sounds". As mentioned above, short-term storage and retrieval correspond to STM, a component of WM. Likewise, in a study investigating the link between WM capacity and the ability to ignore irrelevant speech, Conway, Cowan, and Bunting (2001) found that those people whose performance decreased most when hearing their own name in the "unattended" ear were also those whose WMC was lowest. This suggests that there may be a link between the ability to process speech in IM (competing speech) and WM.

In relation to the specific components of WM that are involved with dealing with a competing talker beyond its energetic masking, Sörqvist and Rönnberg (2012) investigated normal-hearing listeners' ability to understand and remember facts from stories that were masked by either speech or spectrally-rotated speech, at +5 dB SNR. Although the focus was on long-term episodic memory, this study is of particular interest to my own research questions due to the use of a comprehension measure in addition to the intelligibility measure. In addition to testing long-term episodic memory (via comprehension questions), intelligibility was tested beforehand by playing isolated sentences from the stories followed by a 4-Alternative Forced Choice (AFC) recognition task. To determine the components of WM involved, two WM tests were administered: a reading span and the size-comparison span or SICSPAN (Sörqvist, Ljungberg, & Ljung, 2010). The SICSPAN requires participants to answer a series of questions about the size of two objects (e.g. "Is ELEPHANT smaller than MOUSE?"). Each question is followed by the presentation of a semantically related word (e.g. LION) that participants are asked to remember for later recall. Similarly to other span tasks, the sequence of questions and words to remember increases until the participant can no longer remember all the words in the sequence. According to the authors, the SICSPAN is a measure of WM that requires "cognitive control of semantic confusion" and involves inhibition of irrelevant information more than the reading span does. As such, it was expected to be a better predictor than the reading span of listeners' performance in the comprehension task with a competing talker compared to spectrally-rotated speech, and this is in effect what the authors found. Two conclusions can be drawn from this study in relation to the current thesis. The first is the detrimental effect of a competing talker above its energetic masking, both for the intelligibility task and the comprehension task. This effect was found despite the relatively high SNR of +5 dB. The second is the relationship between the SICSPAN and performance in the competing talker condition, which was greater than with the spectrally-rotated speech. This relationship indicates that competing speech taps into different cognitive processes than energetic masking alone, and that the specific cognitive processes may involve inhibition of irrelevant information in particular. Although the authors of this study included inhibition as a sub-process of WM, it could arguably be measured independently from WM (e.g. Stroop task or Flanker task).

On the assumption that computational resources are limited (e.g. Kahneman, 1973) and domain-general (e.g. Camos, Lagner, & Barrouillet, 2009; Vergauwe, Barrouillet, & Camos, 2010), a concurrent task that increases demands in cognitive resources should also reduce memory resources. Using a speeded word recognition task presented with a concurrent WM task, Francis and Nusbaum (2009) found that one competing talker was more detrimental to performance on a WM task than several competing talkers (babble). Presumably, the increased load induced by WM demands reduced listeners' capacity to deal with the interference from a single competing talker compared to babble. Although babble leads to greater overall EM than a single competing talker, a competing talker involves a greater proportion of IM, which may be competing with the WM demands for more central processing resources. However, WM also seems to be involved in dealing with EM, because the authors also found that when demands on WM were high, recognition of degraded synthetic speech decreased.

Another example of the sharing of limited processing resources in STM while dealing with competing speech was provided by Salamé and Baddeley (1987). They asked participants to perform a digit recall task (a standard phonological loop or STM task) with either a noise mask or a speech mask in an unknown language. Their results showed greater impairment in the speech than the noise condition, which led the authors to propose that a component of the phonological loop, the articulatory loop, is disrupted when presented with irrelevant speech, even though the speech is meaningless to the listener.

The relationship between WM, STM and the ability to deal with speech in adverse listening conditions can be summarised as follows. Adverse listening conditions (including EM and IM) increase general reliance on central processing resources, of which WM is part.

Individuals with greater WMC have a larger pool of resources to tap into, which leads them to perform better in adverse conditions than individuals with lower WMC. In the case of a competing talker mask, the articulatory loop (part of the phonological loop) may be disrupted by the mere presence of speech. Individuals with greater STM spans may be less affected by the presence of competing speech than those with lower STM spans. In addition, competing speech may compete for attention within the central executive, as well as increasing the need to inhibit the irrelevant stimulus, once again within the central executive.

Of particular interest to the notion of an interaction between WM and the ability to deal with speech in adverse conditions is the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008), which formally introduced WM as a key factor in speech processing in adverse conditions. The ELU model was developed to detail the role of WM and long-term memory (LTM) in language understanding for a wide range of conditions and all language modalities, including adverse listening conditions, for hearing impaired listeners and sign-language users. In the ELU, WM is a key predictor of intelligibility, since it allows short-term maintenance of the energetically or informationally impaired signal for delayed integration. Rönnberg et al. (2013) define WM as “a limited capacity system for temporarily storing and processing the information required to carry out complex cognitive tasks such as comprehension, learning, and reasoning” (p. 2). Within the ELU framework, working memory capacity (WMC) or working memory span measures an individual’s ability both to store information and process it. Presumably, the processing aspect of WM in this case corresponds to the manipulation aspect of WM in Baddeley’s model. The preferred test of WMC used by proponents of the ELU has been the reading span, since in many studies this test has predicted hearing-impaired and normal hearing listeners’ performance in speech-in-noise tasks. However, as previously mentioned, this test involves semantic processing in addition to storage and manipulation of verbal information. It is possible that the relationship found between WM and listening in adverse conditions holds mainly when the WM task involves these additional processes that are not necessary components of working memory. Figure 1.3 shows the updated ELU model (Rönnberg et al., 2013).

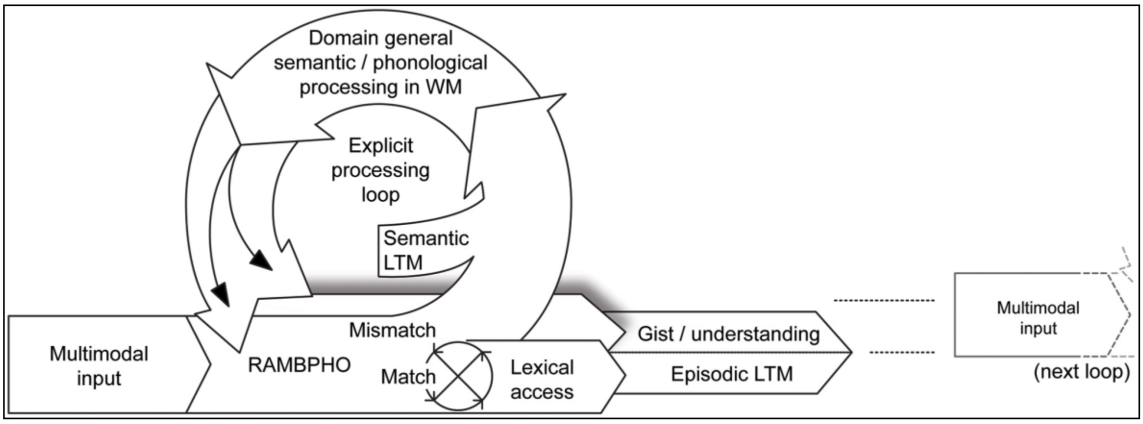


Figure 1.3. The Ease of Language Understanding (ELU) model (Rönnberg et al., 2013). Multimodal input is dealt with via an implicit bottom-up episodic buffer rapidly and automatically (RAMBPHO), leading to lexical access which feeds into episodic long-term memory. This implicit loop can be influenced by prior expectations and context-specific knowledge (e.g. accent). When phonological representations from the RAMBPHO do not match with the listeners' lexical representations, the explicit top-down WM component kicks in to resolve the mismatch and generate meaning. The explicit processing loop includes inhibitory control, attention and storage, and allows inferences to be generated about the gist of the message.

In the ELU model, two parallel processing loops interact to lead to successful listening: the faster implicit (or bottom-up) processing loop and the slower explicit (or top-down) processing loop which kicks in when there is a mismatch between the input and the long-term phonological representations, which can be due to a wide variety of adverse conditions (listener-related, environmental, or transmission-related). The implicit processing loop contains an episodic buffer called RAMBPHO, because this is where "multimodal speech information is Rapidly, Automatically, and Multimodally Bound into a Phonological representation" (Rönnberg et al., 2013). As long as the phonological representations in RAMBPHO correspond to lexical representations in long-term memory, processing is fast and implicit, with little involvement of the explicit processing loop. Explicit WMC is recruited when there is mismatch or uncertainty in the matching between phonological representations in RAMBPHO and lexical representations in LTM, allowing the listener to use phonological and semantic LTM information to infer the gist of what is being said or signed. In this model, explicit WMC includes a range of processes, for example inference-making, semantic integration, attention switching, storage of information, and inhibition of irrelevant information.

The ELU model is based on evidence that WMC (mostly, but not exclusively, measured with the reading span) is related to the ability to recognise and understand speech in adverse conditions in general. These adverse conditions include EM, but also IM due to a competing talker. In fact, the authors have speculated that WMC is particularly involved in tasks with a competing talker, in particular inhibition of semantic information from the competing talker.

The studies above suggest that WM is involved in speech perception in adverse conditions. Some of the evidence points to a specific involvement of WM when the masker is speech, in particular the ability to inhibit irrelevant information. One of the aims of this thesis is to further explore the link between WM and the ability to deal with informational interference from a competing talker. I hypothesise that a competing talker depletes processing resources by increasing the reliance on working memory, in particular the executive components of selective attention and inhibition. To determine the component of working memory that is involved (i.e. phonological loop only or central executive), I will administer a phonological short-term memory task in addition to working memory tasks. Individual differences in WMC and STM span should be related to individual differences in the ability to understand sentences in the presence of a competing talker.

Furthermore, any aspect of the task that increases reliance on WM should provide valuable information about the role WM plays in understanding sentences with a competing talker. In the experiments presented in this thesis, WM demands were manipulated by including target sentence structures increasing in syntactic complexity. This is based on the assumption that increased syntactic complexity leads to greater WM demands. The next section explores the notion of syntactic complexity with relation to adverse conditions and WM.

1.3.2.2 Working memory, syntactic complexity and adverse conditions

A listener's task in a communicative environment is to extract meaning from the speech they hear. Meaning can usually only be fully extracted once the listener has processed the speech at the acoustic, phonetic, morphological, lexical, syntactic and pragmatic levels. In this thesis, I will be concentrating on the syntactic level, as this has often been overlooked by speech-in-noise studies, which mostly present sentences that do not require particularly challenging syntactic processing. Many authors have shown that WM plays an important role in syntactic processing (Gibson, 1998; P. C. Gordon, Hendrick, & Johnson, 2004; P. Gordon, Hendrick, & Johnson, 2001; Grodner & Gibson, 2005; R. L. Lewis, Vasishth, & Van Dyke, 2006).

Syntactic complexity has typically been carried out by contrasting subject and object relative clauses. A relative clause is a subordinate of a main clause, which is introduced by a relative pronoun such as "who" or "that". An example of a subject relative (SR) clause is "Show the boy who is following the girl", where "the boy" is the subject and "the girl" is the object of the verb. The corresponding object relative (OR) clause is "Show the boy who the girl is following", where "the boy" is now the object and "the girl" is the subject of the verb. A vast body of research has shown that OR clauses are more difficult to process than SR clauses (Baird & Koslick, 1974; Caplan & Waters, 1999; Ford, 1983; Gennari & Macdonald, 2008; Gordon et al., 2004; Holmes & O'Regan, 1981; Just & Carpenter, 1992; King & Just, 1991; Mak, Vonk, & Schriefers, 2002; Traxler, Morris, & Seely, 2002; Waters & Caplan, 1996; Cooke et al., 2002; however see Carreiras, Duñabeitia, Vergara, de la Cruz-Pavía, & Laka, 2010 for an exception to this in Basque and Hsiao & Gibson, 2003 in Chinese).

Using the SR/OR contrast, Just and Carpenter (1992) have proposed that syntactic processing taps into the same WM resources as tasks typically used to measure verbal WMC, in particular the Reading Span task (Daneman & Carpenter, 1980). Using a self-paced reading task of subject and object relative clause sentences, Just and Carpenter (1992) found that participants with smaller reading spans (worse WMC) had longer reading times at the verb of both the main and the embedded clauses, compared to participants with larger reading spans (better WMC). The authors interpreted their results within the framework of the theory they call "capacity constrained comprehension". According to this theory, when confronted with a high-demand task, the available resources are depleted, and there is a degradation of the capacity to store and compute information. Following the above results, one would expect that the detrimental effect of a competing talker should be exacerbated when processing sentences of increasing syntactic complexity.

Previous studies have investigated the link between syntactic complexity, WM and sentence intelligibility in adverse conditions. For example, Miller and Isard (1963) asked participants to shadow different types of sentences heard with a speech-shaped noise mask, at varying signal-to-noise ratios (-5dB to +15dB in steps of 5dB). Sentences were of three types:

(1) Grammatical and semantically plausible
"The academic lecture attracted a limited audience"
(2) Grammatical and semantically implausible
"The odourless lecture became a filthy audience"
(3) Ungrammatical and semantically implausible
"From hunters house motorists the carry"

Participants' ability to repeat the sentences in both quiet and speech-shaped noise was best in grammatical and meaningful sentences (1), followed by grammatical and implausible sentences (2) and finally ungrammatical and implausible sentences (3). Furthermore, the grammatical sentences were more resistant to noise than the implausible and the ungrammatical sentences. Indeed, accuracy for the grammatical sentences decreased less abruptly as the SNR decreased. The authors concluded that both syntactic and semantic "rules" influence sentence perception in noise and in quiet. The greater difficulty encountered by participants in shadowing the less grammatical sentences in speech-shaped noise could reflect a higher cognitive load induced by these ungrammatical sentences. Note, however, that these authors only used an energetic mask and not a competing talker, and they did not use a measure of comprehension but rather of intelligibility.

A recent study by Kidd, Mason, and Best (2014) investigated listeners' ability to rely on syntactic structure to preserve the integrity of a target stream of speech masked by either noise bursts or speech. The authors hypothesised that syntactic structure aids binding of the words into a coherent stream compared to a random string of words with no syntactic structure. Indeed, the syntax of the target sentence should allow attention to be focused and maintained on the relevant stream through its predictability. Target sentences were taken from the 'BU corpus' (Kidd, Best, & Mason, 2008) which consists of sentences with the following structure "<name> <verb> <number> <adjective> and <object>", for example "Sue found six red hats". Each of the five (monosyllabic) word categories has eight possibilities, making this a closed set corpus. The target sentences either followed the structure in the example (syntactic condition) or were composed of words from the corpus mixed in a pseudo-random order (random condition). All sentences were masked by two maskers composed of either five random words or five speech-shaped noise bursts. All target and competing talker sentences were spoken by female talkers. The sentences could be presented from one of three speaker locations, and both location and voice were used to let participants know which sentence was the target. Participants were asked to select the correct answer for each word within each target sentence by choosing from eight alternatives presented on the screen. The authors found that participants benefited from the syntax of the sentence more when the masker was the two-talker competing speech than when it was composed of speech-shaped noise bursts. In other words, there was a greater difference between the 'syntactic' and the 'random' conditions when the maskers were speech than when they were noise. The authors interpret these results as suggesting that predictability in the 'syntactic' condition allowed participants to form coherent streams, and that this helped them to overcome the detrimental effect of informational masking. However, it is important to note that the level of informational masking at play in this study was probably at lower levels of segregation, since the voices used were very similar. Indeed the authors mention stream formation, which is at a lower level of informational masking than the levels I will be investigating in this thesis.

In the two studies described above (Kidd et al., 2014; Miller & Isard, 1963), although the authors' goal was to investigate the role of syntax, they actually compared syntactically correct with syntactically incorrect sentences. This is altogether a different question to that of how listeners deal with sentences varying in syntactic complexity.

More recently, Wingfield, McCoy, Peelle, Tun, and Cox (2006) assessed participants' comprehension of sentences played at a normal or speeded speech rate by asking young and older adults with or without hearing loss to indicate the gender of the agent of the sentences. Speeded speech rate is assumed to increase processing load and can be considered an adverse listening condition. Sentence structures were either subject relative clauses (less complex) or object relative clauses (more complex), as below.

(1) Subject-relative clause, male agent: "Men that assist women are helpful."
(2) Object-relative clause, male agent: "Women that men assist are helpful."
(3) Subject-relative clause, female agent: "Women that assist men are helpful."
(4) Object-relative clause, female agent: "Men that women assist are helpful."

The authors found main effects of speech rate and hearing acuity for both subject and object relatives. The increased speech rate (increased adverse condition) did not decrease overall performance for the young normal-hearing adults, however the object relative clauses led to slightly lower performance in the fastest speech rate for this group. Furthermore, there was a greater detrimental effect of speech rate for the older adults compared to the younger adults in the object relative condition but not in the subject relative condition, suggesting that it is only when the load is high that some individual differences can emerge. Note that, while a sentence comprehension task was used in this study (albeit an indirect one), the adverse condition was not instantiated through a mask, let alone a competing talker.

Additional evidence that syntactic complexity may magnify the detrimental effect of adverse conditions was provided by Tun, Benichov, and Wingfield (2010) These authors investigated the effect of syntactic complexity on reaction times to a comprehension task that was presented at three different intensity levels. Although once again there was no mask involved in this experiment, reducing the intensity level of the target signal increases the perceptual difficulty, so the lower intensity levels could be considered adverse conditions. The authors asked young and older adults with normal hearing and mild-to-moderate hearing loss to perform a speech comprehension task presented at three different intensities. Sound levels were adjusted individually depending on each participant's SRTs: either 15dB, 20 dB or 25 dB above the individual SRT. There was a main effect of syntactic complexity across all groups. The young normal-hearing adults did not show an effect of intensity. However, similarly to Wingfield et al. (2006), syntactic complexity (SR-OR contrast) magnified the differences between hearing-impaired and normal-hearing groups. The hearing impaired groups had longer response latencies in the comprehension task where the sentences were presented at lower amplitudes. Thus, the increased cognitive load induced by syntactic complexity magnified the effect of perceptual load in the groups that were most susceptible to both cognitive load and perceptual load. Although a reduced amplitude signal can be considered an adverse condition since it requires greater perceptual effort, once again it does not allow the effect of a competing talker to be assessed.

The previous studies have demonstrated that syntactic complexity can increase processing load, thus magnifying the effect of perceptual load due to adverse conditions. However, some studies have found that syntactic complexity does not always lead to an increased detrimental effect of adverse conditions. Carroll and Ruigendijk (2013) conducted a study in German with normal-hearing native participants. Using a word-monitoring paradigm followed by a comprehension task ('whodunit'), sentences were presented unmasked and masked by speech-shaped noise (SNR -3dB). Participants were briefly shown a target word on the screen, followed by a sentence presented over the headphones, which could contain the target word. Participants were asked to press a button whenever they heard the target word. Following sentence presentation, participants answered a comprehension question relating to the sentence, which ensured they had processed the syntax and meaning of the sentence. Two sets of contrasting syntactic structures were presented: less complex SR vs. more complex OR, and less complex Subject-Verb-Object (SVO) structure vs. more complex Object-Verb-Subject (OVS) structure. These sentences were taken from the Oldenburg Linguistically and Audiologically Controlled Sentences or OLACS (Uslar et al., 2013), which contains seven syntactic structures: SVO, OVS and ambiguous OVS sentences, and SR, OR, ambiguous SR and ambiguous OR sentences. Although they did find a main effect of syntactic complexity and a main effect of noise, Carroll and Ruigendijk (2013) found no interaction in reaction times between the presence or absence of noise and syntactic complexity when the sentences were relative clauses. However, they found an interaction between noise type and sentence structure when analysing the error rates for the comprehension task. Although they do not report the details of this interaction, the plotted results indicate a much smaller difference between noise and silence in the SR condition (6.6% error rate for silence vs 7.1% for noise, i.e. 0.5% difference) than in the OR condition (15.5% error rate for silence vs 22.5% for noise, i.e. 7% difference). Thus, the conclusions for relative clauses are not clear. When reaction times were considered, object relative clauses were not more detrimentally affected by adverse conditions in this population, contrary to what Tun et al. (2010) and Wingfield et al. (2006) found with hearing-impaired listeners. However, when error rates were considered, object relative clauses seemed more detrimentally affected by the presence of noise than subject relative clauses. Furthermore, Carroll and Ruigendijk (2013) found an interaction between noise and syntax when the sentences presented were SVO vs. OVS structures.

Similar results were found by Wendt, Kollmeier, and Brand (2015) in a study investigating the relationship between syntactic complexity and comprehension of speech in noise for normal-hearing and hearing-impaired native German adults. These authors used an eye-tracking paradigm (Wendt, Brand, & Kollmeier, 2014) to determine whether hearing impairment influences the duration of sentence processing, in particular for more cognitively demanding syntactic structures. Target sentences were presented in quiet or masked either by speech-shaped noise or by speech-modulated noise. The signal-to-noise ratio was adjusted to correspond to participants' individual SRT at 80% correct. The SNR averages for the normal-hearing group with speech-modulated noise ranged between -7.8 dB and -9.8 dB depending on the structure, and between -3.6 dB and -4.4 dB for the stationary noise. The SNR averages for the hearing-impaired group varied between 0.1 dB and 2.3 dB for the modulated noise, and between -1.5 dB and -0.5 dB for the speech-shaped noise. Target sentences were also taken from the OLACS corpus, and consisted of the subject-verb-object (SVO, least difficult)), object-verb-subject (OVS, more difficult) and ambiguous object-verb-subject (ambOVS, most difficult) sentence structures. In this experiment, participants were shown visual stimuli depicting two scenes. For example, if the target sentence was "Die nasse Ente tadelt der treue Hund" (The wet duck [accusative] reprimands the loyal dog [nominative] - ambOVS), they would be shown a scene with a dog reprimanding a wet duck on one side of the screen, and on the other side of the screen they would be shown a scene with a wet duck reprimanding a dog. Participants were asked to indicate which picture corresponded to the target sentence by pressing a button on the keyboard. Participants' gaze was tracked in order to determine at what point in the sentence their eye fixations were reliably more towards the correct target picture than the incorrect picture. In addition to the sentence comprehension task, a series of cognitive tests was administered: the digit span, word span, and a Stroop task (susceptibility to interference). It was hypothesised that participants with hearing impairment would show longer processing times than their normal-hearing counterparts, and in particular for the more complex syntactic structures. The authors found a main effect of noise: sentence processing time increased in the noise conditions compared to the quiet condition for all participants. They also found a main effect of sentence complexity across both groups. There was no interaction between syntactic complexity and noise, i.e. the presence of background noise did not exacerbate the delay caused by processing more complex syntax. However, the hearing-impaired group was more affected in noise by the more complex syntactic structures than the normal-hearing group. Furthermore, they found that the normal-hearing group's processing times for the more complex sentences in noise were correlated to susceptibility to interference as measured by reaction times in the Stroop task. However, normal-hearing participants' short-term memory (word span) and working memory (digit span) did not correlate with sentence processing time. Only the hearing-impaired listeners' processing times were correlated with short-term memory (word span) and working memory (digit span). The authors contrasted noise maskers and did not include competing speech, which might help to explain the lack of interaction between noise and syntactic complexity. Indeed, a competing talker may tap into the same processing resources as dealing with complex syntax, whereas noise maskers may not involve central processing resources as much. Finally, this study highlighted the benefit of using an online measure of sentence processing in adverse conditions (eye-tracking).

The results of the experiments reviewed thus far indicate that syntactic complexity (for example contrasting subject and object relative clauses) may be an effective way of manipulating processing load of the target sentence, thereby increasing the difficulty of speech processing in adverse conditions.

The next two studies investigated the involvement of WM in understanding syntactically complex sentences played against energetic maskers. Although the authors did not compare energetic maskers to a competing talker, the results could indicate the extent to which WM is already involved, if at all, when energetic maskers are used.

The first of these studies was an offline "Object Manipulation Task" where participants used figurines to represent the actions described in nine types of English sentences varying in syntactic complexity presented in 8-talker babble (Dillon, 1995). The results showed that, as the signal-to-noise ratio decreased (ranging from 0 dB to -6 dB), syntactic errors and response latencies increased. However, Dillon did not find any relationship between syntactic complexity and WM (reading span). This conclusion could support the hypothesis that the type of WM involved in syntactic processing is not particularly implicated in dealing with EM. It is interesting to note, however, that the measure used in this study was not online, and therefore might not be sensitive enough to capture the differences at play.

The second study investigated online processing of English subject and object relative clauses played in speech-shaped noise at -3dB SNR, -4.5dB SNR, and in quiet in relation to WM (Yampolsky, Waters, Caplan, Matthies, & Chiu, 2002). Normal-hearing participants were asked to perform a self-paced listening task, and to answer acceptability judgment questions at the end of each sentence. The self-paced listening task used the Auditory Moving Windows paradigm, where participants are presented with a phrase at a time (see examples below) and must press a button when they are ready to move on to the next phrase. Sentences were either subject or object relatives, semantically acceptable or unacceptable, as in the examples below (phrase boundaries are denoted by the slash /):

i) Acceptable, SR: /It was/ /the fire/ /that/ /injured/ /the policeman/ /on the highway/.
ii) Acceptable, OR: /It was/ /the policeman/ /that/ /the fire/ /injured/ /on the highway/.
iii) Unacceptable, SR: /It was/ /the man/ /that/ /delighted/ /the camera/ /in the film/.
iv) Unacceptable, OR: /It was/ /the camera/ /that/ /the man/ /delighted/ /in the film/.

Participants took longer to process the object relative clauses, as measured by the self-paced listening task (or auditory window paradigm). Listening times were not proportionally longer during the more resource-demanding portions of the more complex syntactic structures compared to the least complex structures. However, noise magnified the difficulty of the complex sentences in the offline sentence comprehension task, since the difference in performance between the quiet and the noise conditions was larger for the most complex sentences compared to the least complex sentences. Similarly to Carroll and Ruigendijk (2013), the interaction between syntactic complexity and noise depended on the type of measure: interestingly, the online measures did not reveal an interaction whereas the offline comprehension measures did. Furthermore, the authors found no relationship between the (online) self-paced listening task measures in noise and measures of WM (as measured by an alphabet span, a subtract-2 span, and a listening span). The authors concluded that the type of WM involved in syntactic processing is probably not involved in dealing with EM, in this case speech-shaped noise.

Studies investigating comprehension of complex syntactic structures in adverse conditions have mainly used speech-shaped stationary noise (Carroll & Ruigendijk, 2013; Yampolsky et al., 2002), speech-modulated noise (Carroll, 2012; Wendt et al., 2015), multi-talker babble (Dillon, 1995), increased speech rate (Wingfield et al., 2006), or decreased sound levels (Tun et al., 2010), and only one has used speech maskers (Kidd et al., 2014). As reported in the previous paragraphs, an interaction between the presence or absence of noise and syntactic complexity has not been systematically found, and, when found, this interaction was modulated by the type of syntactic structure as well as the language used. In English (Dillon, 1995; Yampolsky et al., 2002) and in German (Carroll & Ruigendijk, 2013; Wendt et al., 2015), the few studies investigating subject and object relative clause comprehension did not show an interaction between WM (as measured by reading span tasks, alphabet span, listening span and 'subtract-2' span), noise, and syntactic complexity.

The studies reviewed above mostly used energetic maskers or low-level informational masking. Investigating sentence comprehension using a speech masker would allow us to tease apart the factors linked to EM and low-level IM from those linked to higher levels of IM, i.e. informational interference. By increasing syntactic complexity (and thus cognitive load), the effect of a competing talker might be magnified for more complex sentences compared to simpler sentences. This is the hypothesis that guided the experiments reported in Chapters 2 to 5 of this thesis.

1.3.3 Long-term memory: language proficiency

In addition to working memory, long-term memory is an essential component to successfully understanding speech. Listeners rely on their prior knowledge of native phonemes, how these fit together to form words which are stored in long-term memory, as well as knowledge of the syntactic rules (and exceptions) governing the language and how these may change the meaning of a sentence. Studying populations with an incomplete or impaired language model allows the effect of language-independent processes (such as lower-level acoustic components of masking) to be disentangled from those that require a fully-formed language model. Listeners with an incomplete or impaired language model (such as non-native listeners) are likely to require more processing resources to understand target speech. On the assumption that a competing talker involves additional processing resources beyond energetic masking alone, the effect of a competing talker should be particularly magnified for non-native listeners.

1.3.3.1 Non-native listening in adverse conditions

Speech perception and comprehension in adverse listening conditions are difficult for native listeners, and even more so for non-native listeners, who have to contend with limited knowledge of the language in addition to dealing with suboptimal listening conditions. Lecumberri et al. (2010) reviewed the "dual challenges of *imperfect signal* and *imperfect knowledge*" faced by non-native listeners. Although the term "non-native" refers to a very heterogeneous group, a number of general conclusions can be drawn from the studies investigating non-native speech perception in adverse listening conditions. In their review, Lecumberri et al. (2010) showed that the effect of decreasing the signal-to-noise ratio is greater for non-native listeners than for native listeners, but only when listeners were asked to process words or sentences, and not for tasks where low-level acoustic properties or phonemes were involved. This points to a difficulty in the higher-level aspects of speech processing in adverse listening conditions, which is echoed by Cutler, Weber, Smits, and Cooper (2004) and Mattys et al. (2010) who found that, at the phoneme level, native and non-native listeners' overall performance was equally affected by a masker. The increased difficulty of speech tasks in adverse conditions for non-native listeners might therefore not be due to a difficulty in low-level phonetic aspects of speech perception such as phoneme misidentifications, but it could be situated at a higher word or sentence level, only appearing when non-native listeners' processing resources are taxed due to more complex linguistic demands.

On average, native listeners perform better than non-native listeners in measures of sentence comprehension, in particular with complex syntactic structures. Adult non-native listeners seem to have less detailed and shallower syntactic representations (in their second language) than native listeners, and knowledge of the syntax of their native language may in some cases be interfering with the correct syntactic processing in the non-native language (Clahsen & Felser, 2006). This in turn could lead to a higher level of processing resources (such as WM) being used, with less cognitive spare capacity available, leading to more effortful listening and lower performance in the listening task.

Non-native speakers are however at an advantage over native listeners in certain conditions. When speech maskers are in a second language, in most cases non-native listeners seem to benefit from the lack of interference from their first language (Lecumberri & Cooke, 2006; Rhebergen, Versfeld, & Dreschler, 2005; Van Engen & Bradlow, 2007).

1.4 Interim conclusion for Chapter 1

After a brief overview of the concept of adverse conditions, this chapter explored the notion of masking by a competing talker. Energetic and informational masking were introduced as two different types of masking involving either low-level peripheral (EM) or high-level central (IM) processes. Informational interference was defined as the linguistic and cognitive aspects of IM which should arise in the presence of a competing talker, due to the linguistic interference and attentional capture of the competing talker, and which could be compounded by additional WM demands of the target sentences. Attention and WM were further explored as possible mechanisms modulating informational interference. In the next section, I will provide an overview of the aims for each of the experiments reported in this thesis.

1.5 Aims of the current thesis

The overarching goal of this thesis was to investigate the nature of the interfering effect of a competing talker on speech comprehension. In particular, I focused on the higher-order cognitive and linguistic aspects of IM, which I refer to as “informational interference”. Informational interference does not include lower-level aspects of IM, such as source segregation (object formation). In this thesis, informational interference designates a specific type of adverse condition induced by the presence of a competing talker that increases cognitive load and listening effort. I hypothesised that informational interference from a competing talker depletes processing resources that could otherwise be allocated to recognising and understanding the target speech. Consequently, informational interference should be more pronounced when any characteristics of the task or the listener involve increased processing demands. In all experiments the competing talker was compared to energetic mask controls, to isolate the contribution of informational masking beyond energetic masking. Furthermore, the target voice and competing voice were different genders, to minimise low-level IM (segregation difficulties).

Using a speeded picture-selection task specifically developed for this thesis, the online comprehension of varying syntactic structures thought to require different degrees of processing resources was measured using reaction times and accuracy. In some experiments, eye-tracking was added as a more sensitive online measure of sentence processing. Previous studies have investigated syntactic processing in noise (Carroll & Ruigendijk, 2013; Wendt et al., 2015; Yampolsky et al., 2002), but have not included masking by a competing talker. In this thesis, I compared the effect of a competing talker with that of energetic mask controls, to distinguish results due to EM or low-level IM from those due to informational interference.

Studies investigating the effect of speech masks have traditionally focused on word or sentence transcription, but rarely on comprehension and syntactic processing. However, everyday listening primarily involves comprehension, and syntactic processing forms part of successful understanding. Furthermore, syntax can be used to manipulate the load required to understand sentences. Thus, syntactic complexity was introduced as a variable in all experiments, to manipulate processing load and introduce the additional 'real-world' element of comprehension and syntactic processing.

This thesis consists of four empirical chapters describing six experiments: two intelligibility experiments designed to select signal-to-noise ratios and four sentence comprehension experiments. Each experiment brings complementary evidence towards understanding the mechanisms that give rise to informational interference. Together, these experiments shed light on previous contradictory findings suggesting that a competing talker is either more detrimental than an energetic mask control (e.g. Brungart, 2001; Koelewijn, Zekveld, Festen, & Kramer, 2012; Trammell & Speaks, 1970) or is just as detrimental or even less detrimental than an energetic mask control (e.g. Dirks & Bower, 1969; Festen & Plomp, 1990; Hygge, Rönnberg, Larsby, Arlinger, & Rönnberg, 1992; Qin & Oxenham, 2003). The underlying question in this thesis was whether a competing talker is indeed more detrimental to sentence comprehension than EM alone. A series of more specific questions (outlined below) motivated each experiment, in an attempt to identify the underlying factors behind the emergence of informational interference from a competing talker.

In addition to these questions, measures of WM and attention were administered across all studies, to identify the nature of the processing resources involved, from an individual-difference standpoint. Although previous research from other groups has evidenced links between WM and speech recognition in adverse conditions in general, I hypothesised that individuals with greater working memory capacity and attentional resources would be less affected by the competing talker in particular, beyond the effect of EM alone. This could be because the inhibitory aspects of WM and attention are more specifically taxed in the presence of a competing talker due to the potentially relevant linguistic information competing with the target information.

In the following paragraphs, I describe the specific underlying questions for each of the four studies. All studies shared the same method, however the mask types and signal-to-noise ratios differed depending on the specific research question. In all cases, participants were presented with a line drawing depicting three characters, and were asked to press a button as quickly and accurately as possible in response to a spoken injunction (target sentence). The target sentences varied in their syntactic complexity: simple sentences such as “Show the girl with the red shoes”, more complex subject relatives such as “Show the girl who is holding the boy”, and even more complex object relatives such as “Show the girl who the boy is holding”. Informational interference was quantified by the difference between a competing talker condition and energetic mask controls. Some of the studies were supplemented with eye-fixation data.

1.5.1 Is informational interference influenced by the syntactic complexity of the target utterance?

Chapter 2 describes Experiments 1 and 2, which aimed to determine whether informational interference would be influenced by the syntactic complexity of a target utterance. Under the assumption that processing resources are limited, any aspect of the target sentences that increases processing resources should also increase informational interference. The main hypothesis of these first two experiments was that the syntax of the target sentence (simple, subject relative, object relative) and the type of mask type (no mask, competing talker, speech-modulated noise) would show an interaction. As mentioned in the previous sections, prior research has shown that object relatives are more difficult to process than subject relatives. Therefore, the greater processing load induced by object relative sentences was expected to magnify the effect of a competing talker. Speech-modulated noise created from each of the competing talker utterances was used as a control for the EM of the competing speech signal. Informational interference was estimated as the decrement of the competing talker compared to the speech-modulated noise. The masked sentences were presented at a signal-to-noise ratio of -5dB, chosen based on a transcription task (Experiment 1). The prediction was that the more complex the syntactic structure, the greater the effect of a competing talker, as compared to EM alone.

1.5.2 Is informational interference influenced by the language proficiency of the listeners?

Chapter 3 describes Experiment 3, which explored non-native listeners' susceptibility to informational interference from a competing talker. As mentioned in the previous sections, non-native listeners expend greater processing resources to understand complex syntax, and to process speech in adverse conditions. The hypothesis behind Experiment 3 was therefore that informational interference would be increased for non-native listeners. To explore the online processing of sentences with a more time-sensitive measure, eye-tracking was added for this experiment. Furthermore, because speech-modulated noise may in fact have more energetic masking than the competing talker, time-reversed speech was used as an additional energetic mask. Time-reversed speech and speech-modulated noise have both been used to investigate the effect of a competing talker on speech recognition, but as mentioned above they each have different acoustic characteristics that lead to different EM patterns. By using both types of energetic mask controls, any difference between the competing talker mask and both energetic masks can be attributed to EM in general, and not specifically to that type of energetic mask.

1.5.3 Is informational interference influenced by the intelligibility of the target utterances?

In Experiments 4 and 5 (Chapter 4), the signal-to-noise ratio was drastically reduced to explore the effect of intelligibility on the emergence of informational interference. Experiment 4 was designed to select the SNR for Experiment 5. By decreasing the signal-to-noise ratio to a level where participants could still report most of the words (82-83% at -22dB and -25dB SNR) but could no longer reach near-perfect intelligibility, Experiment 5 investigated whether low intelligibility of the target utterance gives rise to a greater detrimental effect of the competing talker. Accuracy, reaction times and eye-tracking were once again used to measure performance. The hypothesis for Experiment 5 was that reducing the intelligibility of target sentences would increase the overall difficulty of the task, thus revealing the possible detrimental effect of the competing talker. An alternative hypothesis was that reducing intelligibility would in fact have no effect on informational interference from a competing talker. Indeed, it is possible that EM and high-level IM (informational interference) do not tap into the same pool of resources. EM taps into low-level perceptual processes that may not overlap with the high-level processes involved in high-level IM. Low SNR leads to sensory degradation, which is not hypothesised to be part of high-level IM. Although EM increases as

SNR decreases, the effect of informational interference may in fact be independent from the effect of EM.

1.5.4 Is informational interference influenced by the semantic content of the competing talker utterances?

In the final chapter (Chapter 5), Experiment 6 explored the influence of the semantic content of the competing sentences on target sentence comprehension. The hypothesis for Experiment 6 was that informational interference would be greater when the content of the competing speech captured the listeners' attention, much like the own-name effect reported by Moray (1959). In Experiment 6, the content of the competing talker utterances was manipulated to be semantically congruent, incongruent, or unrelated to the target utterance. Once again, performance was measured using accuracy, reaction times and eye-tracking. It was expected that when the competing talker utterance was completely unrelated to the target utterance, informational interference would be less pronounced than when the competing talker utterance was related (either congruent or incongruent). In this experiment, although there was no energetic mask control, informational interference was measured by contrasting performance in each of the neutral conditions to the congruent and incongruent conditions.

Chapter 2: Effect of syntactic complexity on informational interference from a competing talker

This chapter describes the stimulus creation, an intelligibility experiment (Experiment 1) that formed the basis for selecting the signal-to-noise ratio (SNR) in Experiment 2, and the sentence comprehension with speeded picture-selection experiment (Experiment 2).

The target sentences that were created for this study followed one of three syntactic structures: simple, subject relative (SR), and object relative (OR). The high imageability of the semantic content ensured that they could be illustrated with pictures. The competing talker (CT) sentences in Experiments 1 and 2 followed a simple syntactic structure and were semantically unrelated to the target sentences. The energetic mask control for Experiments 1 and 2 consisted of speech-modulated noise (SMN) created from the competing talker mask. Since the difficulty due to acoustic segregation of target and masker (low-level IM) is not a component of informational interference, by choosing a male target and a female competitor segregation difficulty was reduced, thus focusing on the effects of informational interference.

The aim of Experiment 1 was to select the SNR for Experiment 2. Participants' task in Experiment 1 was to transcribe masked target sentences, presented at three different SNRs. We aimed to select the SNR that yielded comparable transcription scores across conditions while still presenting a challenge to intelligibility (i.e. performance not at ceiling).

In Experiment 2, reaction times to a speeded-picture selection task were used as a measure of processing cost. The picture-selection task was preferred over a more conventional transcription task, to ensure that participants would process the syntactic structure of the target sentences, and not merely have to repeat elements of the sentence or only process simple syntactic structures. On the assumption that participants' reaction times should be slower in conditions requiring more processing resources, I expected reaction times in the CT condition to be slower than the energetic mask control (speech-modulated noise). Furthermore, assuming that object relative sentences are more resource-demanding than subject-relative sentences, we assessed the processing resources required to understand these sentence structures as a function of the type of mask. In line with previous research, we expected slower reaction times for the OR sentences than for the SR sentences. Critically, following the hypothesis that more processing resources are involved in dealing with the interference of a competing talker, we expected an interaction between mask type and sentence type, where informational interference (as measured by the difference between the CT condition and the energetic mask condition) would be particularly pronounced for the most resource-demanding syntactic structure, namely the OR sentences.

To assess how much of the expected pattern of results could be explained by individual differences in cognitive processes known to correlate with masked speech recognition (e.g., Akeroyd, 2008; Zekveld, Festen, & Kramer, 2013), we also collected measures of short-term memory, working memory and selective attention/ inhibition for each participant.

The main goal of these cognitive measures was to determine whether individual differences in the sentence comprehension task may be associated with performance in short-term memory, working memory and selective attention tests. Indeed, previous studies have shown associations between working memory in particular and performance in speech intelligibility tasks (Rönnberg et al., 2013). We thus expected participants with higher scores in the working memory task (listening recall) to be less affected by the presence of a mask (CT or SMN). By implication, we also expected participants with higher working memory scores to be less affected by the CT than the SMN mask.

Short-term memory performance, assessed by a non-word repetition test, could also explain some of the variability in participants' susceptibility to masking, since it is a component of working memory and has been shown to be related to language development and processing (Baddeley et al., 1998). However it was expected to predict less variability than working memory performance. Indeed, short-term memory should not be as heavily involved in the task of understanding sentences in the presence of competing speech as working memory if it is the executive and/or inhibitory aspect of WM that influences the ability to deal with informational interference.

Finally, since the picture-selection task requires selectively attending to the target sentence while inhibiting an irrelevant mask, performance on the flanker task was also expected to predict participants' susceptibility to interference from a mask, and in particular susceptibility to interference from a competing talker.

2.1 Method for Experiments 1 and 2

2.1.1 Participants

Participants for both Experiments 1 and 2 were monolingual native speakers of British English, who reported never having experienced hearing difficulties (including tinnitus) or speech-language impairments (including dyslexia). Participants were students from the University of York and received payment or course credit for their time. There were 12 participants (5 females) in Experiment 1 (transcription task), whose mean age was 22;10 years (*SD* = 2;9). In Experiment 2 (picture selection task), there were a further 36 participants (31 females), whose mean age was 20;5 years (SD = 1;10). Each participant only took part in one experiment.

2.1.2 Materials

Experiment 1 was designed to select the SNR for Experiment 2. Stimuli for Experiment 1 were auditory-only and consisted of spoken sentences masked by either competing speech or speech-modulated noise presented at a range of SNRs (-10 dB, -5 dB, 0 dB). In Experiment 2, the auditory stimuli were identical to Experiment 1 at a SNR of -5 dB, with an additional unmasked condition. Since this experiment consisted of a speeded picture-selection task, each target sentence was accompanied by a line drawing depicting the content of the sentence. The details of all stimuli are described in the following paragraphs.

2.1.2.1 Auditory stimuli: target sentences and masks

2.1.2.1.1 Target sentences

Two hundred ninety-one target sentences were created specifically for this thesis, and followed one of three syntactic structures: simple sentences (N=165), SR sentences (N = 63) and OR sentences (N=63). All sentences can be found in Appendix B.

Simple sentences. These were 165 simple noun phrase sentences, of which 120 were used as experimental trials and 45 as filler trials. The syntactic structure contained no embedding, unlike the SR and OR sentences. They were constructed based on the syntactic structure below:

Show the *noun1* with the *adjective* *noun2*.

e.g., Show the elephant with the orange hat.

Thirteen different adjectives, 32 animate nouns (*noun1*) and 11 inanimate nouns (*noun2*) were used to construct these sentences. Each simple sentence contained 7 words, ranging from 7 to 11 syllables per sentence, with 8.12 syllables on average (*SD* = 0.95). The range of sentence lengths was 1990 to 2923 ms, with an average length of 2378 ms (*SD* = 171).

Relative clause sentences. There were 63 SR sentences and 63 OR sentences. Sixty of each sentence type were used as experimental trials, and 3 of each were used as familiarisation trials prior to the start of the experiment. The SR sentences were constructed based on the syntactic structure below, where the head noun (*noun1*) is modified by a subject relative clause, denoted within the square brackets.

Show the *noun1* [that $verb_{aux+gerund}$ the *noun2*].
e.g., Show the elephant [that is following the crocodile].

The OR sentences were constructed based on the syntactic structure below, where the head noun (*noun1*) is modified by an object relative clause, denoted within the square brackets.

Show the *noun1* [that the *noun2* $verb_{aux+gerund}$].
e.g., Show the elephant [that the crocodile is following].

Crucially, each SR sentence had a corresponding OR sentence using the same nouns (*noun1* and *noun2*) and the same verb. Sentences were controlled for syllable length and frequency in English. Specifically, within each sentence, noun1 and noun2 were matched for length in syllables and frequency, using the CELEX database (Baayen, Piepenbrock, & Rijn, 1993). Across all relative clause sentences there were 34 different verbs and 32 different animate nouns, chosen based on their high degree of imageability. Sentence length in syllables ranged from 9 to 14, with an average of 10.19 syllables (*SD* = 1.49). SR sentence lengths in ms ranged from 2212 to 3015 ms, with an average length of 2616 ms (*SD* = 163). OR sentence lengths ranged from 2280 to 2995 ms, with an average length of 2687 ms (*SD* = 167).

2.1.2.1.2 Competing talker sentences.

The competing sentences were taken from the Hearing in Noise Test, or HINT (Nilsson, Soli, & Sullivan, 1994). All sentences reported in Nilsson et al. (1994) were used, including the three practice lists (35 sentences) and the 25 experimental lists (250 sentences). Eleven sentences were modified to replace American words with British words (Appendix C). In order to have at least as many CT sentences as target sentences, nine new sentences were created based on the original sentences, yielding a total of 294 sentences.

A target sentence masked by two HINT sentences will be referred to as a 'target-masker pair'. Because the HINT sentences were shorter than the target sentences, two HINT sentences were concatenated for each target-masker pair, with a 50ms silence inserted between the two sentences to prevent the perception of an abrupt transition. I will refer to a pair of HINT sentences that masks one target sentence as a 'competing talker utterance'. Two hundred and ninety-one competing talker utterances were created to mask the 291 target sentences. Each of these 291 utterances was created using a different combination of sentences, so no two utterances were identical.

Each competing talker utterance was assigned to one of the 291 target sentences, ensuring that the semantic content of the target sentences and corresponding HINT sentences was as contrasted as possible. For example, if the target sentence referred to a boy and a girl, the competing talker sentence did not contain reference to a boy or a girl, and where possible did not refer to female or male protagonists, nor did it include the pronouns 'he' or 'she'. For a full list of the competing talker utterances assigned to target sentences, please refer to Appendix B.

2.1.2.1.3 Energetic mask control

To isolate the effect of masking from a competing talker beyond its energetic component, speech-modulated noise (SMN) was used as an energetic mask control. The SMN maskers were created using Matlab (Release 2010a), following the technique described in Brungart (2001). A 30-second fragment of speech-shaped noise (SSN) was first generated from the concatenated competing talker (HINT) sentences. The resulting SSN had the same average frequency spectrum as the concatenated HINT sentences. A random sample of the SSN was then cross-multiplied with the intensity envelope of each competing talker utterance (HINT sentence pair). Each resulting SMN sound file simulated the acoustic energy of its competing talker utterance counterpart, while removing the informational/linguistic content. Each SMN sound file root mean square level was then matched to that of the competing talker sentences.

2.1.2.1.4 Target-to-masker pairing procedure.

The competing talker was female and the target talker was male. We chose to contrast the gender of the talkers to facilitate acoustic segregation of target and masker, thus capturing the unique contribution of informational interference independent of lower-level aspects of informational masking or energetic masking. Both target and competing talker were monolingual native speakers of Standard Southern British English who recorded the target sentences and the HINT sentences in a sound-insulated booth using a TASCAM DR-100 Portable Digital Recorder. All but six of the target sentences were recorded in one sitting. Each of the HINT sentences and the target sentences was extracted from the stream using Cool Edit Pro (Version 2.0, 2002). One hundred ms of silence were manually inserted at the beginning and end of each target sentence, and at the beginning and end of each competing talker utterance (pair of HINT sentences). Using Praat (Boersma & Weenink, 2012), the root mean square level of all auditory stimuli (target, competing talker, SMN) was manipulated so that it was normalised across sound files. The masker sound files (competing talker and SMN) were normalised to an intensity of 68 dB, in the arbitrary units used by Praat. The intensity of the target sentences was normalised depending on the desired SNR: 58 dB (-10 dB SNR), 63 dB (-5 dB SNR), 68 dB (0 dB SNR).

Using Matlab (Release 2010a), each target sentence sound file was combined with one competing talker utterance and separately with the corresponding speech-modulated noise. The alignment of target sentences with their corresponding mask was carried out from the end of the sound files, such that the competing talker utterance ended 100ms after the offset of the target sentence. The mask always started before the target sentence, with varying lead times, as shown in **Table 2.1**. We chose to vary the lead times between mask and target sentences to reduce the predictability of the onset of the target sentence.

Target sentence structure	Range of lead time (ms)	Average length of lead time (SD)
Simple	25 -1234	560 (272)
Subject relative	118 -926	545 (186)
Object relative	84 - 1313	517 (197)

Table 2.1. Range and average lead times (with standard deviations) in ms between target sentence and mask for each sentence type (Simple, SR, OR).

2.1.2.2 Pictures for picture-selection task

In Experiment 2, participants were asked to show one of three characters depicted on a picture on the screen, as illustrated in **Figure 2.1**, to assess their comprehension of the target sentences. Each picture had three characters corresponding to those mentioned in the target sentences. For example, if the target sentence was "*Show the girl who is holding the boy*" (SR) or "*Show the girl who the boy is holding*" (OR), the picture depicted a girl holding elephant boy holding another girl. Half of the pictures showed characters facing left and the other half had characters facing right. All pictures can be found in Appendix B, along with the corresponding target sentences.

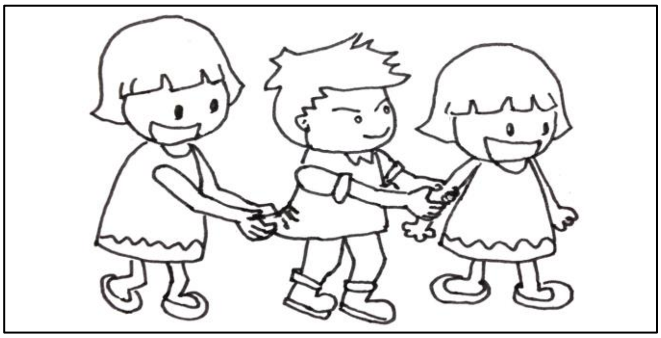


Figure 2.1. Example of a picture for a subject relative sentence ("Show the girl who is holding the boy") and the corresponding object relative sentence ("Show the girl who the boy is holding"). The correct answer is the character on the left and the character on the right, respectively.

The 60 pictures illustrating the relative clause sentences followed the character layout in **Figure 2.2** (YXY), where character X (in this case the boy) was always flanked by two similar characters Y (in this case the two girls). One of the two Y characters was always the agent of an action on character X, whereas the other Y character was always the patient of the action by character X. The same picture could thus be used for the SR and the OR condition. Half of the pictures for the relative clause sentences were exactly the same for the SR and OR conditions, and the other half had opposite orientations, with characters in the SR conditions facing one direction and characters in the OR conditions facing the opposite direction. The correct answer for these pictures was one of the external characters (Y character). A full breakdown of the expected answers can be found in Table 2.2.

For the 120 pictures accompanying the simple sentences, there were two possible character layouts. The first layout was similar to the SR and OR pictures, i.e. character X flanked by two characters Y (Figure 2.2, YXY).

Figure 2.2. Example of a picture for a simple sentence, same YXY layout as SR/OR ("Show the girl with the black shirt")

The second configuration was either YYX (e.g., Figure 2.3) or XYY. These configurations were used to ensure that participants would also pay attention to the central character, which was never the correct character in the YXY configuration. The correct answer for this configuration was either of the two Y characters.

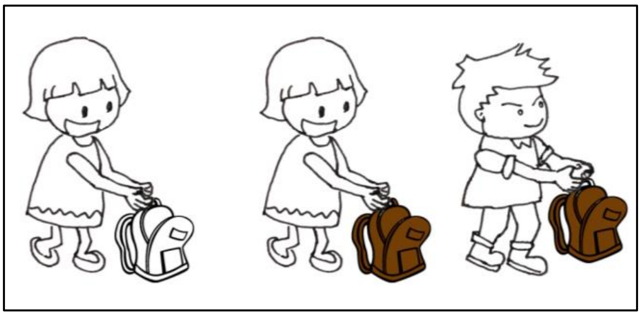


Figure 2.3. Example of a picture for a simple sentence, YYX layout.
("Show the girl with the brown bag")

Finally, the characters in the 45 filler sentence pictures could fall in any of the above configurations (YXY, YYX, XYY). Figure 2.4 shows a YXY example.

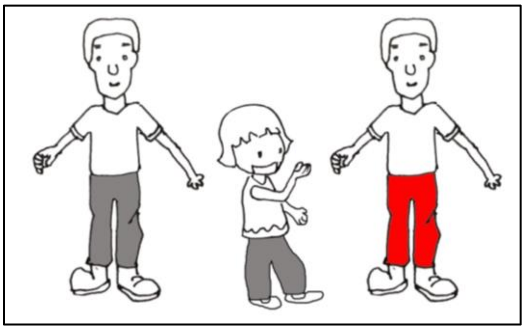


Figure 2.4. Example of a picture for a filler sentence, YXY layout.
("Show the girl with the grey trousers")

Table 2.2 shows a summary of all possible configurations and number of occurrences with each configuration and correct character (familiarisation items are not included in this table). The number of fillers and the counterbalancing of character positions within the pictures discouraged participants from employing the same strategy throughout the experiment, and encouraged them to maintain their attention on all three characters as much as possible.

<table>
<tr><th rowspan="3">Position of target character</th><th colspan="9">Number of occurrences per configuration and sentence type</th></tr>
<tr><th colspan="4">YXY</th><th colspan="2">YYX</th><th colspan="2">XYY</th><th rowspan="2">Total</th></tr>
<tr><th>Filler</th><th>Simple</th><th>SR</th><th>OR</th><th>Filler</th><th>Simple</th><th>Filler</th><th>Simple</th></tr>
<tr><td>Left</td><td>-</td><td>15</td><td>30</td><td>30</td><td>-</td><td>15</td><td>15</td><td>-</td><td>105</td></tr>
<tr><td>Middle</td><td>15</td><td>-</td><td>-</td><td>-</td><td>-</td><td>30</td><td>-</td><td>30</td><td>75</td></tr>
<tr><td>Right</td><td>-</td><td>15</td><td>30</td><td>30</td><td>15</td><td>-</td><td>-</td><td>15</td><td>105</td></tr>
<tr><td rowspan="2">Total</td><td>15</td><td>30</td><td>60</td><td>60</td><td>15</td><td>45</td><td>15</td><td>45</td><td rowspan="2">285</td></tr>
<tr><td colspan="4">165</td><td colspan="2">60</td><td colspan="2">60</td></tr>
</table>

Table 2.2. Number of occurrences per picture configuration, sentence type, and position of the target character.

Sixty-three black and white pictures were hand-drawn to illustrate each of the SR and OR sentence pairs. These pictures were scanned at high resolution and modified using Adobe Photoshop. To create the simple and filler sentence pictures, characters from the SR and OR pictures were digitally extracted and pasted into new digital image files. The colours, accessories, garments and different positions of the characters in the simple and filler sentence illustrations were all added and modified or drawn using Adobe Photoshop.

2.1.2.3 Cognitive measures

In addition to the picture-selection sentence comprehension task, Experiment 2 investigated individual differences in short-term memory, working memory and visual selective attention.

2.1.2.3.1 Short-term memory and working memory.

Two subtests of the Automated Working Memory Assessment, or AWMA (Alloway, 2007) were administered, to assess verbal short-term memory and verbal working memory. The AWMA is a computer-based standardised test that has been validated with a range of ages including a UK population of undergraduate university students aged 19 to 22. The non-word recall task assesses verbal short-term memory, and the listening recall task assesses verbal working memory.

2.1.2.3.2 Selective attention: flanker task.

The selective visual attention “flanker task” was based on the original letter version (Eriksen & Eriksen, 1974), but like Bunge, Dudukovic, Thomason, Vaidya, and Gabrieli (2002), we used arrows to reduce the linguistic content. Participants indicated the direction of a target arrow flanked by distracters. Examples of the stimuli are shown in **Figure 2.5**. There were three different conditions that differed according to whether the central arrow was the same or different to the distracting arrows. In the consistent condition, all five arrows pointed either left or right. In the inconsistent condition, the central arrow pointed in the opposite direction to the distracters.

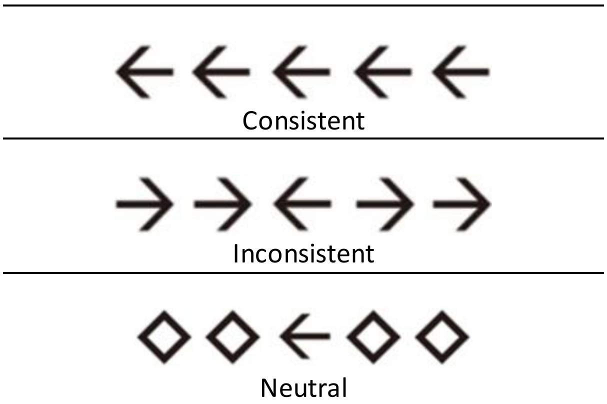


Figure 2.5. Examples of the three conditions for the flanker task (Consistent, Inconsistent and Neutral)

Previous studies using flanker tasks have shown that participants involuntarily process the irrelevant stimuli surrounding the central stimuli (Eriksen & Eriksen, 1974; Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988; Hazeltine, Poldrack, & Gabrieli, 2000). Thus, when the irrelevant (flanker) stimuli are different to the target stimulus, reaction times are slowed down, whereas when the target stimulus is the same as the irrelevant stimulus, there is a facilitation effect and reaction times are faster. For the present experiment, the measure used was the reaction time difference between the inconsistent condition and the consistent condition.

2.2 Design and Procedure

All participants were tested individually in a sound-insulated booth. Sentences were presented at the same intensity for all participants, via Sony MDR V700 headphones, using the DMDX software (Forster & Forster, 2003) on a Dell PC with a Creative Sound Blaster X-Fi Xtreme Audio sound card and an ART HeadAmp4 headphone amplifier, which allowed the experimenter to monitor progress and listen to the stimuli from outside the booth.

2.2.1 Experiment 1: signal-to-noise ratio selection

In Experiment 1, participants' task was to type the target sentences as accurately as possible, excluding the lead phrase "*Show the*". They were told that they had to focus on the man talking, although there would also be either a woman talking or noise. No time limit was given, and participants could start typing before or after the end of the sentence. Participants' responses appeared on the screen as they typed, allowing them to correct their typing as needed. The intertrial interval was 1 second. Participants' typed responses were collected with DMDX.

Only the 120 simple sentences, 60 SR sentences and 60 OR sentences were presented. Fillers and familiarisation sentences were excluded because they were not analysed in Experiment 2. Sentence type (simple, SR, OR) was a within-subject variable, with each participant hearing all three types of sentences. Mask type (HINT competing talker, speech-modulated noise) and SNR (-10 dB, -5 dB, 0 dB) were within-subject and within-item variables. Participants heard each sentence only once. Sentence type and SNR were randomised, and sentences were blocked into two blocks of 120 sentences, by mask type. Order of mask presentation was counterbalanced across participants. The experiment included a break halfway through at the end of the first block of 120 sentences.

The dependent variable was the percentage of correct keywords per sentence. Each sentence contained three keywords. Keywords were defined as the content words within each sentence (see Table 2.3 for examples). Keywords were counted as correct if spelled correctly, misspelled but phonologically identical, or if obvious typographical errors were made (e.g., adjacent letters on the keyboard), as long as these did not result in another lexical item.

Sentence type	Target sentences with keywords underlined
Simple	Show the <u>sheep</u> with the <u>grey</u> <u>ball</u>.
SR	Show the <u>sheep</u> that is <u>pulling</u> the <u>cow</u>.
OR	Show the <u>sheep</u> that the <u>cow</u> is <u>pulling</u>.

Table 2.3. Example of keywords (underlined) for each sentence type.

2.2.2 Experiment 2: sentence comprehension and speeded picture-selection task

Each individual session lasted one hour. Participants first completed the cognitive tests (visual attention task followed by the two memory tasks), and then the sentence comprehension task. The sentences were presented with the same Sony MDR V700 headphones as in Experiment 1, and the pictures were presented on a 22-inch Dell monitor, with a resolution of 1920 x 1080 pixels.

2.2.2.1 Sentence comprehension and speeded picture-selection task

In the picture-selection task, participants were told that they would hear a man talking, and that they would have to indicate as quickly and accurately as possible which of three characters on the screen the man was referring to, using 'g' 'h' and 'j' on the keyboard (stickers had been placed on these keys). The experimenter explained that in some cases there would be a female speaker talking at the same time, and at other times there would be noise, but that their task was to focus on what the man said. After a series of six familiarisation items (two per mask type), participants continued to the test items if they had no questions.

Sentence type (simple, SR, OR) was a within-subject variable, with each participant hearing all three types of sentences. Mask type (unmasked, HINT competing talker, speech-modulated noise) was a within-subject and within-item variable. Mask type was blocked, with the order counterbalanced across participants through a full three-way permutation. There were three blocks of 95 sentences, with a short break between blocks. Each participant was exposed to 285 sentences (in addition to the six familiarisation items), each sentence presented with either no mask, speech-modulated noise, or the competing talker. A given target sentence was never heard more than once.

For the SR and OR sentence conditions, correct responses corresponded to either the character on the left or the character on the right, as described in Table 2.2. The expected response for the subject and object relative conditions was counterbalanced such that half of the correct responses were the character on the left, and the other half the character on the right. Correct responses for the simple and filler pictures could be any of the three characters. The filler items were included so that participants' attention was kept on all three characters throughout the experiment.

2.2.2.2 Visual flanker task

In the visual flanker task, the experimenter first showed a print-out of all of the stimuli to participants, explaining that they would have to indicate whether the middle arrow was pointing to the right or to the left, using the right and left shift keys of the keyboard. It was emphasised that participants had to answer as quickly and accurately as they could. There were 13 practice items and 120 experimental items, with 40 items in each condition (inconsistent, consistent, neutral). Half of the expected answers were left button-presses and half were right button-presses. Condition type and response side were randomised.

2.2.2.3 Short-term and working memory tests

During the memory tests, the experimenter stayed in the booth with the participants, who were asked to follow the audio instructions presented to them via the computer. For both tests, three familiarisation items of increasing complexity familiarised the participant with the task. In each test there were six blocks with six items each. As the task progressed, the sequences of non-words or number of sentences became longer, from 1 non-word or sentence for block 1 to six non-words or sentences for block 6. The experimenter scored the answers as correct or incorrect on the computer. When the first four items of a block were correctly recalled, the programme moved on to the next block and attributed the maximum score of 6 for that block. Testing was automatically interrupted when a participant gave 3 or more incorrect responses within one block, and the score included the number of correct responses until the point at which the test was interrupted. If a participant were to correctly answer all items of all blocks, the program would attribute a maximum raw score of 36.

In the non-word recall test, the participant heard a sequence of nonsense words and was asked to repeat the words in the correct sequence. In the listening recall test, participants heard a sentence and had to first indicate whether it was true or false, and then recall the last word of the sentence. Similarly to the non-word recall test, the number of sentences to process and recall increased as the task progressed, from one to six.

During the familiarisation items for both tests, participants' questions were answered, but not during the test phase. The experimenter gave no indication of whether answers were correct or incorrect during the test phase.

2.3 Results

2.3.1 Experiment 1: signal-to-noise ratio selection

A summary of the results for Experiment 1 is shown in Figure 2.6. Breakdowns by SNR levels are shown in Figure 2.7. Accuracy across conditions was generally high, with condition averages ranging from 88% to 99%.

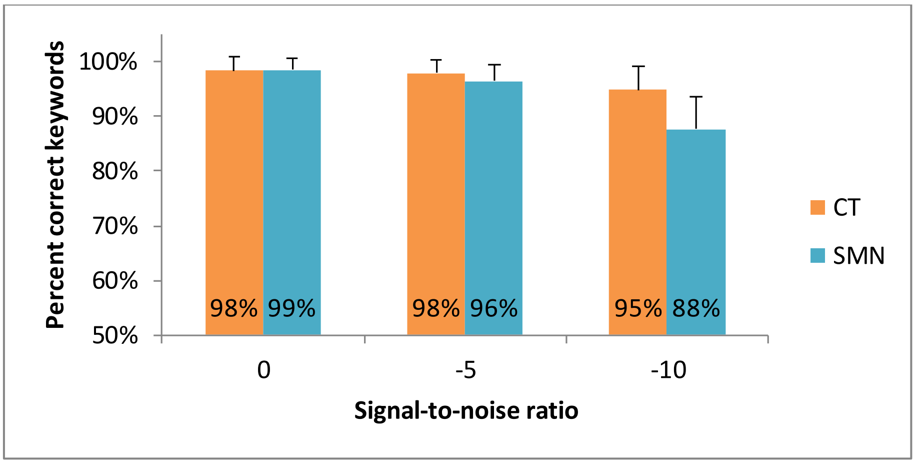


Figure 2.6. Experiment 1. Accuracy by mask (CT, SMN) and SNR (0 dB, -5 dB, -10 dB) for all three sentence types. Error bars indicate one standard error of the mean, by participants.

Three-way repeated measures Analyses of Variance (ANOVAs) by participants and by items were performed with percentage of accurate keywords per sentence as a dependent variable, and mask type (CT, SMN), SNR (0 dB, -5 dB, -10 dB) and sentence type (simple, SR, OR) as independent variables[2]. A main effect of SNR, F_1 (1.17, 12.82) = 40.10, p < .001, η_p^2 = .79; F_2 (1.63, 386.33) =, p < .001, η_p^2 = .24, showed that as SNR decreased, accuracy decreased. The main effect of SNR reflected a significant difference between 0dB and -5dB (p = .01 by participants, p = .02 by items), 0dB and -10dB (p < .001), -5dB and -10dB (p < .001), with a Bonferroni adjustment for multiple comparisons. A main effect of mask, F_1 (1, 11) = 23.21, p = .001, η_p^2 = .68; F_2 (1, 237) = 21.16, p < .001, η_p^2 = .08, indicated that accuracy was lower with SMN than with CT. An interaction between SNR and mask, F_1 (2, 22) = 23.95, p < .001, η_p^2 = .69; F_2 (1.65, 391.39) = 19.60, p < .001, η_p^2 = .08, showed that the mask effect was only significant in the -10 dB SNR condition (p < .001). However, this interaction cannot be interpreted given the ceiling effect. There was no significant main effect of sentence, F_1 (2, 22) = 1.31, p = .29, η_p^2 = .11 ; F_2 (2, 237) = 2..01, p = .137, η_p^2 = .02, no interaction between mask and sentence, F_1 (2, 22) = 2.15, p = .14, η_p^2 = .16; F_2 (2, 237) = 1.70, p= .185 , η_p^2 = .01, and no interaction between mask, SNR and sentence, F_1 (4, 44) = .39, p = .82, η_p^2 = .03; F_2 (3.30, 391.39) = 0.35, p= .811, η_p^2 = .00. Figure 2.7 illustrates these findings.

[2] When investigating the main effect of SNR, Mauchly's test indicated that the assumption of sphericity had been violated by participants, $X^2(2)$= 12.59, p = .002 and by items, $X^2(2)$= 60.74, p < .001. The assumption of sphericity was also violated in the by-items analysis for the SNR by mask interaction, $X^2(2)$= 55.95, p < .001. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

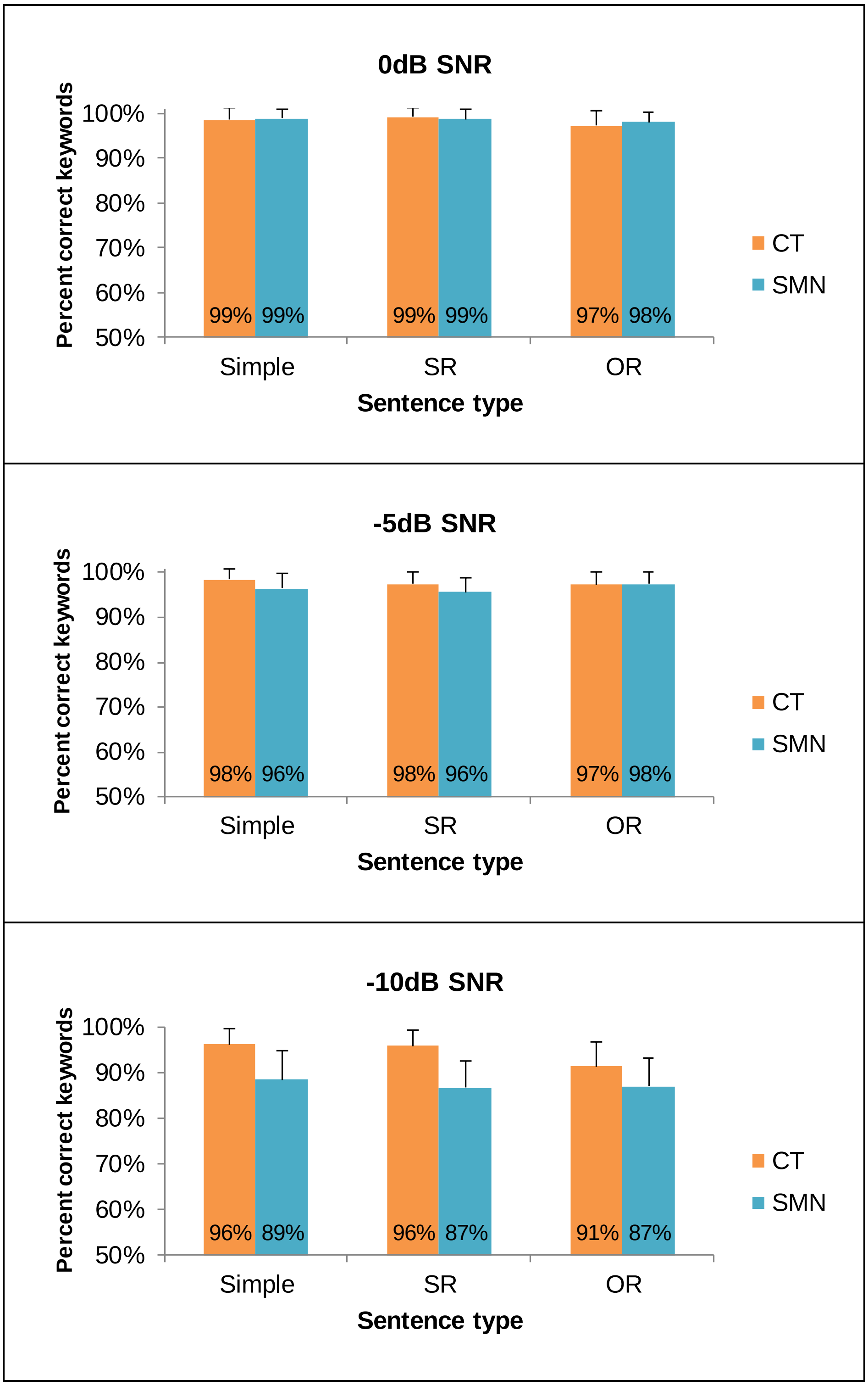


Figure 2.7. Experiment 1. Accuracy by mask (CT, SMN) and sentence type (simple, SR, OR) for each SNR. Error bars indicate one standard error, by participants.

To summarise, the goal of Experiment 1 was to select a SNR leading to comparably high intelligibility levels across conditions. Transcription accuracy was generally high across conditions. This reflected good intelligibility level, which is a prerequisite for investigating the processing resources involved in sentence comprehension in the presence of a competing talker, above and beyond its energetic masking. The lack of significant effect of sentence type indicated that this transcription task did not allow the effect of syntactic complexity to appear, most probably because transcription does not necessarily require the processing of the syntactic structure, since participants can succeed in this task simply by recognising each of the words separately. Based on these transcription data, the SNR was set at -5 dB SNR for Experiment 2, since both mask conditions showed comparable intelligibility (96-98%) but ceiling was not reached.

2.3.2 Experiment 2: sentence comprehension and speeded picture-selection task

2.3.2.1 Accuracy

Both accuracy and reaction times were recorded. A response was deemed accurate when the key pressed corresponded to the target character location. Across the three mask conditions and the three sentence types, accuracy remained constant and high (96-97%), as shown in Figure 2.8, which suggests that, consistent with the data in Experiment 1, the sentences were intelligible despite the masks.

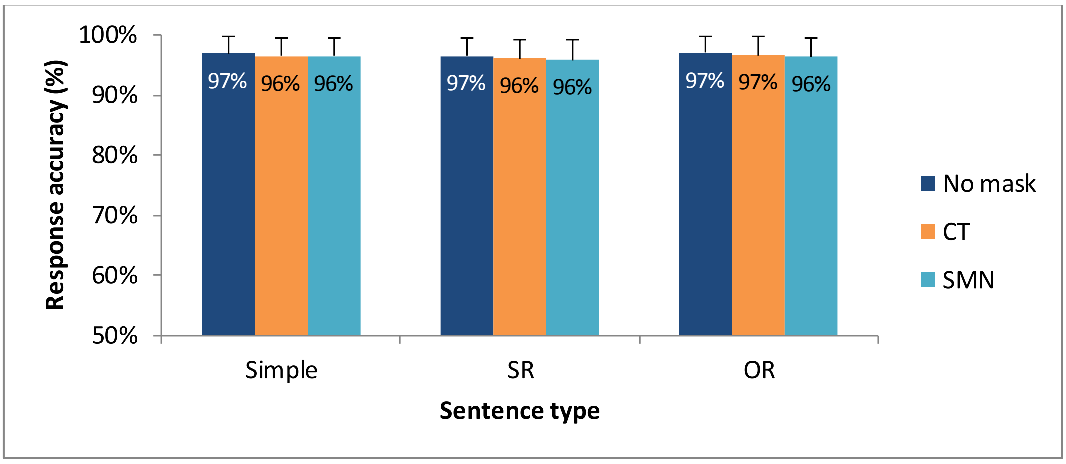


Figure 2.8. Experiment 2. Accuracy as a function of mask conditions and sentence types. Error bars indicate one standard error.

2.3.2.2 Reaction times

Reaction times were the main focus of these analyses (Figure 2.9), since we were mainly interested in the processing cost incurred by the various masking and syntactic conditions. Reaction times were measured from the onset of the target sentence. The reaction times used in the analyses included only correct responses and excluded outliers. Outliers were defined as reaction times greater than two standard deviations above the mean across all three masks and all sentence types (excluding familiarisation and filler items) on a subject-by-subject basis.

Two-way repeated measures Analyses of Variance (ANOVAs) by participants and by items were performed with reaction times (RT) per sentence as a dependent variable, and mask type (no mask, CT, SMN), and sentence type (simple, SR, OR) as independent variables.[3] There was a main effect of sentence type, F_1 (1.59, 55.72) = 128.15, $p < .001$, $\eta_p^2 = .79$; F_2 (2, 237) = 103.30, $p < .001$, $\eta_p^2 = .47$, showing that the OR sentences were slower to process than the SR sentences, which were in turn slower than the simple sentences (both $p < .001$, with a Bonferroni adjustment for multiple comparisons).

Although there was no main effect of mask, F_1 (2, 70) = .51, $p = .61$, $\eta_p^2 = .01$; F_2 (1.86, 441.08) = .97, $p = .37$, $\eta_p^2 = .00$, there was a significant two-way interaction between Mask and Sentence, F_1 (4, 140) = 6.18, $p < .001$, $\eta_p^2 = .15$, F_2 (4, 474) = 6.41, $p < .001$, $\eta_p^2 = .05$, indicating that the pattern of responses for the three masks was different depending on the sentence type. Numerically, for the simple sentences, participants were slowest in the SMN condition, followed by the CT condition, then the unmasked condition. For the SR sentences, there seems to be no difference across masks, and finally responses for the OR sentences were slower in the unmasked condition than in the CT, followed by the SMN, which was the fastest. However, none of these differences were statistically significant. It is thus most likely that the interaction between sentence and mask types was due to small differences within sentence conditions that only appeared within the interaction but were not otherwise meaningful.

[3] When investigating the main effect of sentence type, Mauchly's test indicated that the assumption of sphericity had been violated, $X^2(2)= 10.07$, $p = .007$ for the by-participants analysis, and $X^2(2)= 18.31$, $p < .001$ for the by-items analysis. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

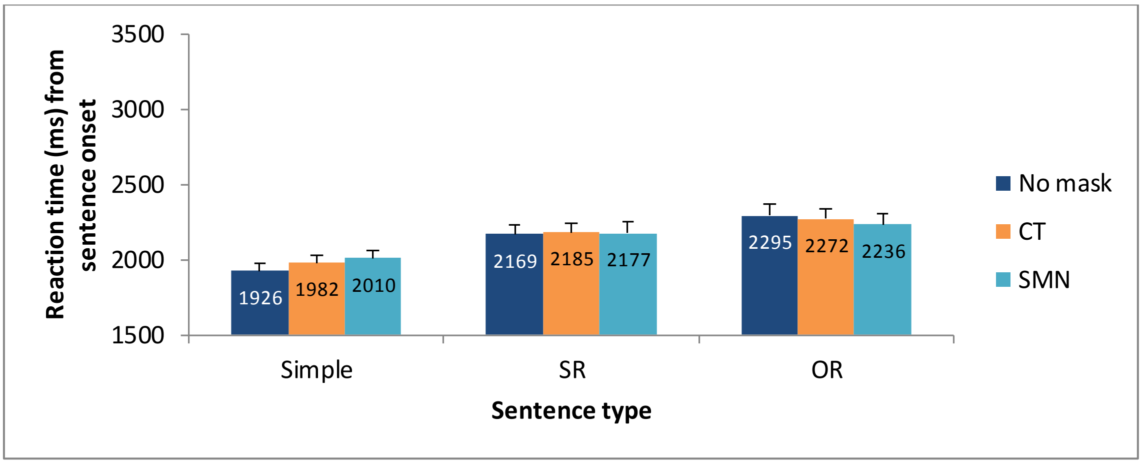
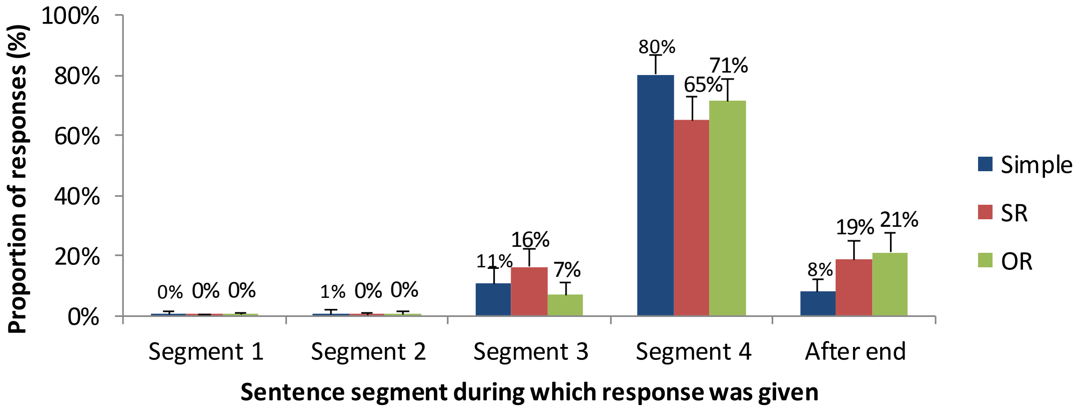


Figure 2.9. Experiment 2. Reaction time (ms) from sentence onset by sentence type (simple, SR, OR) and mask type (no mask, SMN, CT). Error bars indicate one standard error, by participants.

To further focus on the contrast between the two masks and the contrast between the SR and OR sentences, the same ANOVA as above was run, but omitting the no mask condition and the simple sentence condition. A main effect of sentence type, F_1 (1, 35) = 16.72, $p < .001$, $\eta_p^2 = .32$, confirmed the comparatively slower reaction times to OR sentences. However, there still was no main effect of mask, F_1 (1, 35) = .631, $p = .432$, $\eta_p^2 = .02$; F_2 (1, 118) = 1.31, $p = .255$, $\eta_p^2 = .01$, and no Mask by Sentence interaction, F_1 (1, 35) = .745, $p = .394$, $\eta_p^2 = .02$; F_2 (1, 118) = .80, $p = .373$, $\eta_p^2 = .01$. It would therefore seem that for the SR and OR sentences, there was no detrimental effect of either type of mask (SMN, CT), regardless of the sentence type.

2.3.2.3 Button presses for each sentence segment

In addition to the reaction time data, I analysed the time course of participants' responses in relation to sentence segments. Figure 2.10 shows the breakdown of the proportion of responses per sentence segment concatenated across masks. Only correct responses were included in the analysis.



Examples	**Segment 1**	**Segment 2**	**Segment 3**	**Segment 4**
Simple	Show the girl	with the	red	trousers
SR	Show the girl	who	is holding	the boy
OR	Show the girl	who	the boy	is holding

Figure 2.10. Experiment 2. Proportion of responses (%) per segment by sentence type (simple, SR, OR), averaged across all mask conditions (no mask, CT, SMN). Error bars indicate one standard error (by participants). The bottom part of the figure shows examples of segments for each sentence type.

The breakdown shows that most responses were given before the end of the sentence, during the last segment (80% for simple, 65% for SR and 71% for OR). This indicates that participants did not wait until after the end of the sentence to give their response. If they had waited until after the end of the sentence to answer, their online sentence processing would not have been reflected by the reaction times, and thus the processing cost might not have been accurately measured. This could have been a possible explanation for the lack of difference between masks. However, it is safe to assume that the reaction time measures did reflect participants' online sentence processing in this experiment.

2.3.2.4 Cognitive measures

In the visual flanker task, the average difference between the inconsistent and the consistent conditions was 43 ms (SD = 19), ranging from 3 ms to 81 ms.

Descriptive statistics for the standard scores for each of the memory tests are reported in Table 2.4. Individual scores can be found in Appendix D.

AWMA sub-test	Range	Mean	Standard deviation
Non-word repetition	73 - 118	98.04	10.62
Listening recall	70 - 122	95.72	14.75

Table 2.4. Experiment 2. Range, mean, and standard deviation of standard scores for each of the memory tests.

As mentioned in the introduction to this chapter, we expected each of the cognitive measures to be related to performance in the sentence comprehension task, to varying degrees. We predicted that participants who were less affected by masked speech (difference in reaction times between masked and unmasked conditions) would do better in the cognitive tasks. Furthermore, participants who were less affected by the competing speech compared to an energetic mask control (difference in reaction times between the competing talker and the speech-modulated noise condition) should exhibit better performance in the cognitive tasks. Finally, participants who are least affected by syntactic complexity (difference between SR and OR) were also expected to have better memory performance.

To investigate the link between the three cognitive tests and participants' performance in the sentence comprehension task, a series of correlations is reported in Table 2.5. Bivariate correlations were calculated between each of the cognitive tests (non-word repetition, listening recall, flanker task difference between consistent and consistent) and reaction time differences between the masked and unmasked conditions, the CT and SMN conditions, and the OR and SR conditions.

	Non-word repetition	Listening recall	Flanker task difference
Difference masked - unmasked	-.227	-.065	-.025
Difference CT –SMN	-.060	-.135	.168
Difference OR-SR	-.094	-.044	-.243

Table 2.5. Experiment 2. Bivariate correlations between each of the 3 cognitive measures and the difference in reaction times for the sentence comprehension task between masked and unmasked conditions, between the CT condition and EM control (SMN), and between OR and SR.

None of the cognitive measures were significantly correlated with the difference in reaction times between masked and unmasked conditions, between the CT and the SMN condition or between the OR and the SR sentences, using a Bonferroni-corrected $\alpha = .0056$. Previous studies have shown associations between similar cognitive measures and performance in speech in adverse conditions, and the lack of an association in this experiment could be due to the relative homogeneity of the profiles of these highly proficient, normal-hearing undergraduate students.

2.4 Discussion

In Experiment 2, our goal was to investigate informational interference, as evidenced by the additional difficulty involved in sentence comprehension with a competing talker compared to energetic masking alone. There was no evidence of a competing talker being more detrimental to sentence comprehension than speech-modulated noise. Indeed, neither of the masked conditions was more detrimental to sentence comprehension than the no mask condition. Although we found a main effect of syntactic structure, sentences believed to require more processing resources (OR) were not more affected by a competing talker than those requiring fewer resources (SR). There are several possible explanations for these findings. The first one is that there is genuinely no additional cost in ignoring a speech masker with linguistic content compared to an equivalent energetic masker. Although this finding may seem counterintuitive, there have been studies pointing in this direction, such as Dirks and Bower (1969) and Hygge, Rönnberg, Larsby, Arlinger, and Rönnberg (1992), who found no difference in intelligibility performance between a competing talker and time-reversed speech at various SNRs[4]. However, other intelligibility studies contrasting speech-modulated noise with a competing talker (Brungart, 2001; Brungart et al., 2001) have shown that a competing talker is more detrimental than speech-modulated noise.

The second explanation for these findings is that the competing talker does not tap into the same pool of processing resources as that needed for sentence processing. Waters and Caplan (1996) argue that the cognitive resources involved in syntactic processing are specific to syntax. If this is true, then the additional processing resources involved in understanding OR sentences compared to SR sentences do not come from the same pool of general processing resources as the additional resources that may be required to deal with a competing talker compared to an energetic mask control. However, although this could account for the lack of interaction between syntax and mask type, it does not explain the lack of main effect of mask.

A further possibility lies in the choice of population for this task. It is possible that the listening situation created in this experiment with native listeners was not challenging enough for a competing talker effect to emerge. These considerations motivated Experiment 3, in

[4] SNRs used by Dirks & Bower (1969) were -30dB, -24dB, -20dB, -18dB, -12dB, -10dB, 0dB, 10dB. SNRs adjusted by the participants in Hygge et al. (1992) ranged from -12.5dB to 7.7dB.

which the stimuli were unchanged but the listeners (non-native English speakers) were expected to show greater sensitivity to processing load.

Chapter 3: Effect of language proficiency and syntactic complexity on informational interference from a competing talker

This chapter presents Experiment 3, which explored a series of modifications of Experiment 2 (Chapter 2) that might elicit informational interference from a competing talker, and possibly give rise to an interaction with syntactic complexity. The first modification involved testing non-native listeners. Indeed, speech perception and comprehension in a second language (L2) is more demanding than in a first language (L1), in particular when listeners are confronted with L2 speech in adverse conditions (e.g. Lecumberri, Cooke, and Cutler, 2010). For example, in a study comparing native and non-native listeners' performance on an intelligibility task, Cooke, Lecumberri, & Barker (2008) presented stationary noise with target sentences at SNRs of +6, 0 and -6dB, and competing talker utterances with target sentences at +6, +3, 0, -3, -6 and -9dB. In all of these SNRs the native listeners were better than the non-native listeners at identifying sentences in noise, and the authors conclude that the non-native listeners are more affected by the competing talker than by the stationary noise.

In addition to this detrimental effect of masking, syntactic processing is expected to be less efficient for non-native listeners than for native listeners. Clahsen and Felser (2006) reviewed a series of studies showing differences between L1 and L2 syntactic processing and assessed four possible explanations for these differences: reduced knowledge of the grammar in L2, influence of the L1 on the L2, limited cognitive resources and changes in maturation during adolescence. All of these explanations have some degree of evidence backing them, but particularly relevant to the issue of informational interference is the possibility that the differences in syntactic processing observed between L1 and L2 learners could be due to a greater toll on working memory or other cognitive resources due to the added difficulty of processing speech in a non-native language. If the participants in our study find the sentence comprehension task more demanding because of depleted cognitive resources (due to the added load of non-native speech processing), this could in turn lead to informational interference, simply due to the lack of available resources to deal with the competing talker. Clahsen and Felser (2006) do however conclude that the difference between L1 and L2 syntactic processing is mainly in the processing of complex hierarchical structures such as *wh*-dependencies. These authors do not address relative clause processing. However, in a self-paced reading study investigating the processing of relative clause ambiguities with German learners of Dutch (Havik, Roberts, van Hout, Schreuder, & Haverkort, 2009), L2 participants were slower than L1 participants, and the authors concluded that the L2 learners were not as good at using the syntactic information in the sentences. This was despite the fact that the parsing preferences for these syntactic structures are the same in German and Dutch. Based on the conclusions of these studies, our hypothesis was that the participants in Experiment 3 would be slower to process the more complex syntactic structures (simple faster than SR faster than OR), and this effect would be enhanced in the presence of a competing talker.

The second modification consisted of measuring online sentence processing with eye-tracking in addition to reaction times and accuracy. Eye-tracking is used in a variety of ways to study language processing. One of these has been referred to as the 'visual world paradigm' (VWP). In this paradigm, participants are presented with a visual display, and their eye movements (fixations and/or saccades) are monitored online while they listen to a speech stimulus. The first study to show that participants' eye gaze is associated with the content of what they hear was conducted by Cooper (1974). In this study, participants were shown nine line drawings on a grid and were asked to listen to a short story at the same time. Crucially, participants were not told that their eye-gaze would be monitored. The story contained words that were semantically related to the contents of the drawings. For example, participants saw a zebra, a dog, a snake, a camera, a lion, a tree, a peacock, a pipe and grapes, and in the text of the story they heard the words 'snake', 'slithering' (related to the snake), 'zebra', 'grazing' (related to the zebra), and 'Africa' (related to the lion and the zebra). Cooper found that the proportion of eye-fixations to a particular drawing increased when it was semantically related to the words in the spoken story, and that this happened while the word was being heard or within 200 ms after word offset. Since this seminal study, the VWP has been used extensively in psycholinguistic research at a variety of linguistic levels, ranging from phonemic to syntactic (for a review, see Huettig, Rommers, and Meyer, 2011). In eye-tracking studies using the VWP, participants are either given a task, such as pointing or picking up an object, or are simply asked to "look and listen", with no specific task. In both cases, participants' gaze follows the objects or actions spoken or implied in the sentence (Huettig et al., 2011). More recently, the VWP paradigm has been used in the context of speech in noise tasks. Wendt, Brand, and Kollmeier (2014) used the VWP to determine whether processing of different syntactic structures differed when presented in noise or in quiet. This measure allowed them to determine the cost of processing sentences in noise with greater precision. In a subsequent study, Wendt, Kollmeier, and Brand (2015) used the same paradigm to study the effect of syntactic complexity and hearing loss on the online comprehension of sentences in different types of noise (modulated and stationary noise, but not with a competing talker). These authors found an effect of syntactic complexity and hearing loss, which was exacerbated by noise compared to quiet. The VWP thus seems an ideal technique to study the effect of syntactic complexity and the effect of a competing talker, which is the main focus of this thesis.

The third modification consisted of adding a second energetic mask control in addition to speech-modulated noise. Indeed, the masking properties of SMN may have led to increased EM. Time-reversed speech was introduced as a second EM control. Time-reversed speech preserves partial phonetic information such as vowels and fricatives, and still sounds speech-like. Although the intensity modulation of the time-reversed speech does not align with that of the original speech, its spectral masking fluctuates in time, unlike SMN. Indeed, SMN preserves the modulation contour of the original speech but it creates potential additional EM due to the decreased spectral dynamics. At a given point in time, there may in fact be less glimpsing opportunities in the SMN than with a competing talker. With time-reversed speech, the spectral dynamics are conserved, although they do not align with the original speech. The average glimpsing opportunities are therefore more likely to be similar between competing speech and time-reversed speech than between competing speech and speech-modulated noise. Given the different masking properties of SMN and time-reversed speech, and the fact that both of these masks have been widely used as energetic mask controls, we decided to include them both in Experiment 3.

The fourth and final modification was to randomly present the mask conditions from item to item. In Experiment 2 (native listeners), the mask conditions were blocked. Participants could therefore anticipate the type of mask that would be presented, which may have led them to employ an attentional strategy to block out the mask. Randomisation reduces the probability of habituation to a specific mask, thus decreasing the likelihood of a strategy being used.

In addition to the sentence comprehension task detailed above, a series of cognitive tasks was once again administered to establish whether there would be an association between performance in the sentence comprehension task and participants' performance in short-term memory (forward digit span), working memory (backward digit span and reading span), and selective attention (flanker task). The hypotheses were identical to Chapter 2, with the added non-native perspective.

If masked speech is more resource-demanding than unmasked speech, then we expected that participants with better working memory, short-term memory and/or selective attention would show smaller differences between their performance in masked speech and their performance in unmasked speech. Furthermore, participants with lower English proficiency should also show smaller differences between their performance in masked and unmasked speech.

The second hypothesis regarded the detrimental effect of the competing talker compared to the two energetic mask controls. Participants who have higher proficiency and/or better working memory, short-term memory, and selective attention should be less affected by the possible informational interference from a competing talker.

Finally, those participants who were most affected by the OR sentences compared to the SR sentences may also have been those who had higher working memory scores, if the resources involved in processing these sentences include working memory.

3.1 Method

3.1.1 Participants

Nineteen Danish-speaking participants were recruited for this experiment (8 female, 11 male). Eighteen were students in one of the main Higher Education institutions in Copenhagen, and one was a healthcare professional in Copenhagen. All reported that Danish was their main language, having studied in Danish and lived in Denmark for most of their life. The mean age was 24;9 years, (SD = 5;10, range = 19;6 to 40;6 years). Participants either reported having normal vision, or wore corrective glasses or contact lenses when necessary. Two participants reported having seen a speech and language therapist for articulation therapy when they were children. None had been diagnosed with dyslexia. Audiometric thresholds were obtained for all but one participant. The individual results of the audiometric tests can be found in Appendix G. Of the 18 participants who were tested, all but two had hearing thresholds of 20dB HL or better for each ear for all of the following frequencies: 125 Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 6000 Hz. Participant 6 had hearing thresholds at 30 dB HL in her right ear for the lower frequencies (125, 250, 500 Hz), as well as for 3000Hz. However this was most likely due to the incorrect placement of the headphones, as the participant reported never having experienced any hearing difficulties. Participant 17 had thresholds of 30 dB HL for 2000 Hz, and 35 dB HL for 6000 Hz in the left ear. This participant reported suffering from mild hearing loss at certain higher frequencies in his left ear due to recurring otitis media in infancy/childhood. However, because the hearing loss was unilateral and only affected one of the critical pure-tone frequencies for speech (2000 Hz), the participant's results were included in the sample.

3.1.2 Materials

3.1.2.1 Sentence comprehension and speeded picture-selection task

The sentence comprehension task was the same as described in Experiment 2, but items were divided equally among four masks rather than three (71 items per mask: 11 fillers, 30 Simple, 15 SR, 15 OR), and presentation was randomised across mask types and sentence types. The target sentences were the same as in Experiment 2. In addition to the three mask conditions in Experiment 2 (competing talker, speech-modulated noise and no mask), a time-reversed speech mask was created from the competing talker (reversed competing talker). Each of the competing talker sentences was flipped in the time domain, rendering them unintelligible but speech-sounding. The pictures were the same as those in Experiment 2.

3.1.2.2 English language proficiency

The rationale behind choosing to work with a non-native population was to investigate whether less proficient individuals would show increased interference from a competing talker and less efficient syntactic processing than native speakers. Participants' English language proficiency was estimated using the LexTALE (Lemhöfer & Broersma, 2012), which is a lexical decision task designed to assess vocabulary knowledge. This measure has been found to correlate with more general measures of English language proficiency, such as the Quick Placement Test and word translations (Lemhöfer & Broersma, 2012).

Participants also answered a self-report proficiency questionnaire in Danish (Appendix E), loosely based on the questions described by MacIntyre, Noels, and Clément (1997), aimed at assessing participants' everyday use of English, years of experience, and how comfortable they feel using English in different situations. The questionnaire was first written in English, and translated into Danish by a native speaker living and working in Copenhagen.

3.1.2.3 Short-term and working memory capacity

3.1.2.3.1 Short-term and working memory: forward and backward digit spans

A forward digit span test was used as a measure of phonological short-term memory, and a backward digit span test was used as a measure of verbal working memory, as the test includes an executive component. For both tests, a native Danish speaker (male) recorded several tokens of the digits one to nine, and chose the clearest tokens. The tests were implemented with Matlab. Participants heard a sequence of digits over the headphones, which they were asked to repeat either in the same order or starting with the last digit heard. The task started with a string of two digits, and gradually increased to a string of eight digits. There were 14 strings in total, so 2 trials for each length. A string of digits was scored as correct if all of the digits in the string were repeated in the expected order. The final score was the number of correct strings.

3.1.2.3.2 Working memory: reading span

In this task, participants were asked to read sentences in Danish and make a truth-value judgment about each sentence by indicating whether the sentence was true or false (button press). After each sentence, participants were presented with a letter that they had to keep in memory for later recall. The number of sentences and corresponding letters to recall varied from two to ten, randomly presented via the Psych Toolbox in Matlab. At the end of each sentence and letter sequence, participants were asked to recall and type the letters they had seen. The reading span scores were calculated based on the maximum number of letters in a correctly recalled sequence, and the number of correct truth-value judgments.

3.1.2.4 Visual flanker task

The same flanker task as described in Experiment 2 was administered, designed to assess visual attention. Since this task does not require linguistic processing, it can be used for native and non-native participants. The difference between the incongruent and the congruent conditions was once again the focus of the analysis.

3.1.2.5 Audiometry

Pure-tone thresholds for both ears were obtained following the British Society of Audiology's (2011) *Recommended procedure for pure-tone air-conduction threshold*. For logistical reasons, nine participants were tested using an Interacoustics AS216 Screening

Audiometer and nine were tested using an Interacoustics A222 Audio Traveller Audiometer, both with Sennheiser HDA200 headphones.

3.2 Design and Procedure

3.2.1 General procedure

Participants were tested individually in a sound-insulated booth. Each session lasted two and a half to three hours. A short vocabulary task was first administered, to ensure that all the words in the sentence comprehension task were known to participants by the start of the main part of the experiment, thus reducing the possibility of incorrect responses due to lack of lexical knowledge. Participants were given the list of nouns, adjectives and verbs from the sentence comprehension task, and were asked to write translations or definitions in Danish. When they were unsure or did not know the word, the experimenter explained it to them and asked them to repeat. They were then asked to read these words out loud, to check for major pronunciation differences.

After this vocabulary check, participants carried out the sentence comprehension experiment. The sentences were presented via Sennheiser HDA 200 headphones. In addition to reaction times and accuracy measures from the button presses, participants' eye-movements were monitored using an SR Research Eyelink 1000 Plus desk-mounted camera at a sampling rate of 1000Hz, with a chin rest to minimise head movements. Only the dominant eye was tracked by the camera. The pictures were presented on a 22" monitor with a resolution of 1680 x 1050 pixels. Participants were seated approximately 60cm from the screen, and the lighting was kept constant across participants. Participants were told that the camera would monitor their eye movements but that they should just look at the computer screen as they would normally do. The chin-rest was adjusted to a comfortable height for each participant, and participants were instructed to keep their head still and keep their hand on the keyboard ready to press one of the three corresponding buttons, without looking down at the keyboard. This ensured that the eye-tracking data were not affected by occasional glances away from the screen. A 9-point calibration was carried out at the beginning of the practice trials, after which participants were able to ask for clarification and readjust their position, to ensure they were comfortable and to reduce head movements during the task. Another 9-point calibration was carried out at the beginning of the main task, and after any head movements or breaks. Participants were encouraged to take breaks whenever they needed to, ensuring that their attention was held throughout the task. Like in Experiment 2, participants were first shown the picture on the screen for 1 second with no auditory stimulus, which allowed them to familiarise themselves with the visual display and form representations of the characters in the picture. The picture stayed on the screen until after the end of the sentence, which is typical in visual world paradigm experiments (Huettig et al., 2011).

After the sentence comprehension task, the experimenter administered the reading span task in Danish, the flanker task, the Lextale test (Lemhöfer & Broersma, 2012), the forward and backward digit span tasks in Danish, a pure-tone threshold audiometry test, and finally the proficiency questionnaire. Due to time constraints and/or technical glitches, not all participants completed each of the additional measures, but they all completed the sentence comprehension task.

3.2.2 Procedure for eye-tracking analysis

Each picture was divided into three regions of interest (ROI1, ROI2, ROI3, from left to right) corresponding to each of the three characters, manually defined using the SR Research Experiment Builder software. The delimitations of these regions were not visible to the participants. Figure 3.1 shows an example picture with the delimitations of the three ROIs shown for a simple sentence (top panel) and subject/object relative sentences (bottom panel). As shown in these examples, the ROI for the middle character usually overlapped with the left and right characters in the SR/OR pictures. This was due to the depicted action, which often involved contact between the middle character and the other two characters (e.g. holding, squeezing, biting).

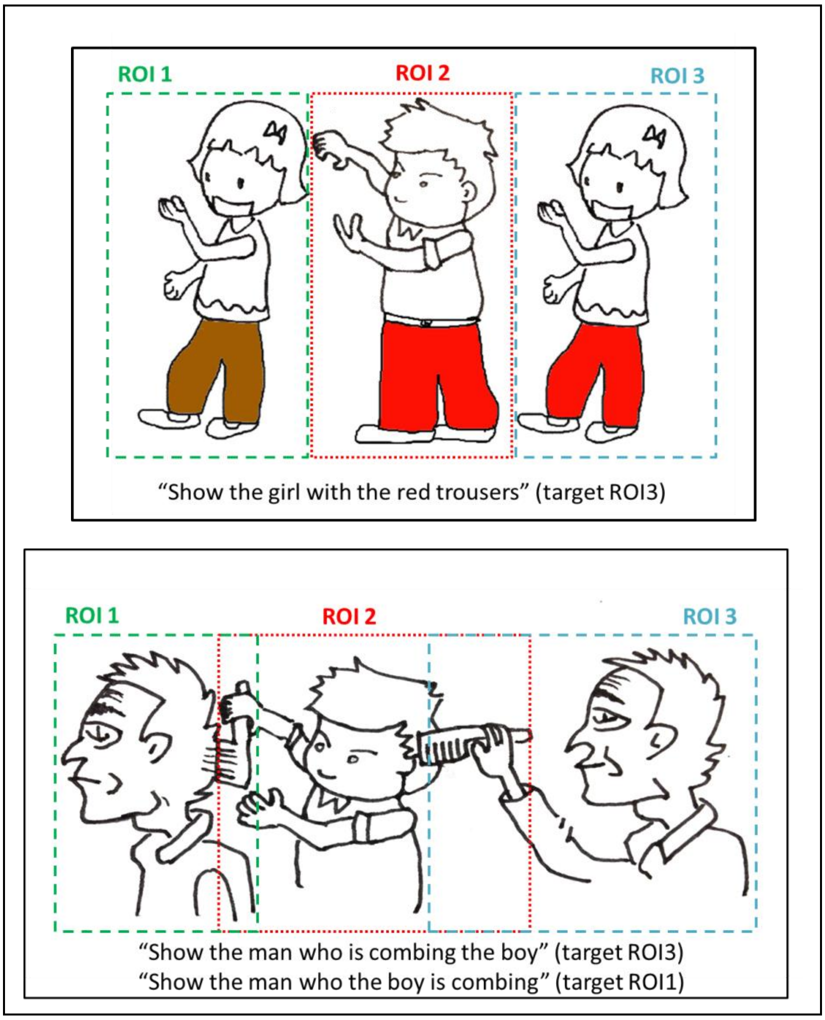


Figure 3.1. Example regions of interest (ROI) for a simple sentence (top panel) and relative clause sentences (bottom panel). The dashed lines delimit each region of interest, and were not visible to the participants.

In the SR/OR example in Figure 3.1 (bottom panel), ROI2 overlaps with ROI1 and ROI3. It includes the boy, both combs, and the hand of the man in ROI3. Given that the middle character (ROI2) was never a target character in the relative clause sentences, fixations that fell in the overlapping areas of the ROIs were attributed to ROI1 or ROI3 for the fixation rate calculations. The position of the target character was counterbalanced across sentence types to control for any left-to-right eye gaze bias.

In addition to defining ROIs for the pictures, time boundaries were defined for each segment within each sentence. This ensured that the time-course of eye-fixations could be analysed across sentences. Since the duration for each sentence was different, a re-scaling was carried out to enable comparisons across sentences.. Note that the rescaling did not alter the signal itself, rather it stretched or compressed the unit of time for each segment within each sentence at the data processing stage. Each sentence was divided into four segments, based on the syntactic structure of the sentence, as shown in Table 3.1 (relative clause sentences) and Table 3.2 (simple sentences).

		Segment 1	Segment 2	Segment 3	Segment 4
Subject Relative	Example segment	Show the man	who	is combing	the boy
	Average duration in ms	977	195	640	673
Object Relative	Example segment	Show the boy	who	the man	is combing
	Average duration in ms	1023	194	558	798
Average SR/OR	Average duration in ms	1000	195	599	736
	Duration in samples	100	20	60	70

Table 3.1. Average segment durations in milliseconds for each of the relative clause sentence types. The bottom row indicates the number of samples per segment used for the eye-tracking analysis, for both relative clause types.

		Segment 1	Segment 2	Segment 3	Segment 4
Simple Sentence	Example segment	Show the girl	with the	red	trousers
	Average duration in ms	992	339	389	657
	Duration in samples	100	30	40	70

Table 3.2. Average segment durations in milliseconds for the simple sentences, and corresponding length in samples used for the eye-tracking analysis.

For each sentence, the segments were re-scaled to correspond to the overall average duration in milliseconds, rounded to the nearest 100 milliseconds. This was then divided by ten to obtain the number of samples, so that one sample corresponded to roughly 10 milliseconds. For instance, segment 2 across subject and object relative sentences had an average duration of 195 milliseconds, which was rounded to 200, and divided by 10 to obtain the duration of 20 samples.

The following examples for two different SR sentences illustrate the process of re-scaling. In Table 3.3, the sentence "Show the crocodile that is following the elephant" is longer than the sentence "Show the horse that is watching the dog" (2696 ms and 2266 ms, respectively). Accordingly, the segment lengths (in ms) of each sentence are different.

	Segment 1	Segment 2	Segment 3	Segment 4
Sentence 1 content	Show the crocodile	that	is following	the elephant
Sentence 1 duration (ms)	1063	162	686	785
Sentence 2 content	Show the horse	that	is watching	the dog
Sentence 2 duration (ms)	894	226	619	527
Re-scaled duration in samples	100	20	60	70

Table 3.3. Examples of segment re-scaling for two SR sentences. Durations in ms are shown for each sentence, and the bottom row reports the final re-scaled duration in samples for all SR sentences.

Given that we needed to compare eye-fixations across all sentences, each segment was normalised or re-scaled to correspond to the average length across SR and OR sentences. For example, segment 4 in the shorter sentence ("the dog") was re-scaled to a longer value in samples than the original ms: from 527 ms to 70 samples (corresponding to 700 ms). Segment 4 in the longer sentence ("the elephant") was re-scaled to a shorter value in samples than the original ms: from 785 ms to 70 samples (corresponding to 700 ms). As a result of the re-scaling, eye-fixations could be compared across sentences that originally varied in duration.

3.3 Results

3.3.1 Sentence comprehension and speeded picture-selection task

3.3.1.1 Accuracy

The percent of button presses to the correct (target) character is reported in Figure 3.2. Accuracy was high for all conditions, indicating that the sentences were intelligible and understood correctly for the most part. However, contrary to the findings with the native participants, accuracy varied across sentence conditions in the non-native participants.

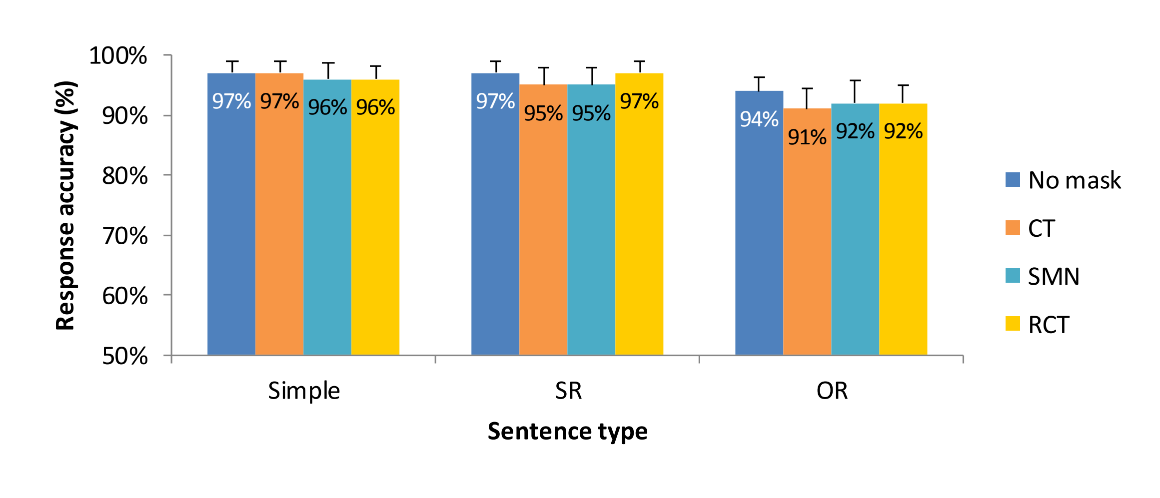


Figure 3.2. Experiment 3. Percent accurate button presses by sentence type (simple, SR, OR) and mask type (no mask, CT, SMN, RCT) in the sentence comprehension task. Error bars represent one standard error (by participants).

Two-way repeated measures ANOVAs by participants and by items were performed with percent accurate responses as a dependent variable, and mask type (no mask, competing talker, reversed competing talker, speech-modulated noise) and sentence type (simple, subject relative, object relative) as independent variables.[5] There was a main effect of sentence type, F_1 (1.26, 22.62) = 4.69, p = .034, η_p^2 = .21; F_2 (2, 236) = 12.08, p < .001, η_p^2 = .09. Although none of the pairwise comparisons in the by-participants analysis were significant, the corresponding by-items pairwise analyses revealed that the simple sentences (M = .97, SD = .09) were more accurate than the OR sentences (M = .92, SD = .14), p < .001, and that the SR sentences (M = .96, SD = .10) were more accurate than the OR sentences, p = .003, all with Bonferroni adjustment for multiple comparisons.

There was no main effect of mask, F_1 (2.28, 41.07) = 1.12, p = .341, η_p^2 = .06; F_2 (2.85, 673.68) = 1.04, p = .370, η_p^2 = .00; and no Mask by Sentence interaction, F_1 (3.04, 54.73) = .203, p = .892, F_2 (5.71, 673.68) = .38, p = .881, η_p^2 = .00. These analyses indicate that although accuracy was affected by syntactic complexity, there was no detrimental effect of a mask, despite the added non-native component.

One of the more specific hypotheses was that accuracy in the subject and object relative sentences would be modulated by mask type. In particular, we expected an interaction

[5] In the by-participants analysis, Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of mask, $X^2(5)$= 11.59, p = .041, for the main effect of sentence, $X^2(2)$= 15.21, p < .001, and for the interaction, $X^2(20)$= 50.92, p < .001. In the by-items analysis, Mauchly's test revealed that the assumption of sphericity had been violated for the main effect of mask, $X^2(5)$= 18.34, p = .003. Degrees of freedom were therefore corrected using Greenhouse-Geisser estimates of sphericity.

between mask type and sentence type, whereby the CT condition would be more affected by the OR sentences than the SR sentences, compared to the EM controls. In order to test this hypothesis, the following analyses focused on the relative clause sentences (SR and OR) and the masked conditions only (CT, SMN, RCT). Two-way repeated measures ANOVAs by participants and items were conducted, with accuracy as a dependent variable, and mask type (CT, SMN, RCT), and sentence type (SR, OR) as independent variables. A main effect of sentence type was apparent between the SR and OR sentences, F_1 (1, 18) = 6.21, p = .023, η_p^2 = .26; F_2 (1, 118) = 6.18, p = .014, η_p^2 = .05, reflecting the lower accuracy in the OR sentences than in the SR sentences. There was no main effect of mask, F_1 (2, 36) = .04, p = .956, η_p^2 = .00; F_2 (2, 236) = .14, p = .871, η_p^2 = .00. There was no mask by sentence interaction, F_1 (2, 36) = .23, p = .796, η_p^2 = .01; F_2 (2, 236) = .29, p = .752, η_p^2 = .00. Thus, we found no support for the hypothesis that the competing talker would affect accuracy in sentence comprehension differentially depending on the type of relative clause.

3.3.1.2 Reaction times

The same criteria as in Experiment 2 were used for the reaction times in Experiment 3, whereby the reaction times included only correct responses and excluded outliers. Outliers were defined as reaction times greater than two standard deviations above the mean across all four masks and all sentence types (excluding familiarisation and filler items) on a subject by subject basis. Figure 3.3 shows the average reaction times for each sentence type and mask type.

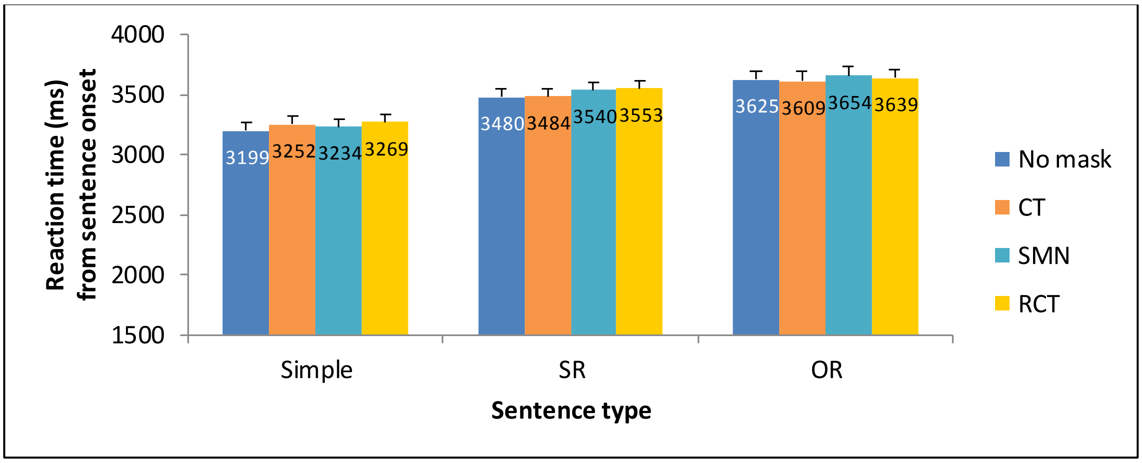


Figure 3.3. Experiment 3. Average reaction times (ms) from sentence onset, by sentence type (simple, SR, OR) and mask type (no mask, CT, SMN, RCT). Error bars indicate one standard error, by participants.

Two-way repeated measures ANOVAs by participants and items were performed with reaction times (RTs) per sentence as a dependent variable, and mask type (no mask, CT, SMN, RCT), and sentence type (simple, SR, OR) as independent variables.[6] There was a main effect of sentence type, F_1 (2, 36) = 117.10, $p < .001$, $\eta_p^2 = .87$; F_2 (2, 232) = 145.38, $p < .001$, $\eta_p^2 = .56$. All pairwise comparisons were significant at $p < .01$ by subjects and by items, with Bonferroni adjustment for multiple comparisons. This reflected the faster reaction times in the simple sentences (*M* = 3238, *SD* = 293), followed by the subject relative sentences (*M* = 3514, *SD* = 292), followed by the object relative sentences (*M* = 3632, *SD* = 333).

There was a main effect of mask, F_1 (3, 54) = 4.41, $p = .008$, $\eta_p^2 = .20$; F_2 (2.85, 660.16) = 4.21, $p = .007$, $\eta_p^2 = .02$, however none of the pairwise comparisons were significant at $\alpha = .05$. There was no mask by sentence interaction, F_1 (4, 72) = 1.17, $p = .329$, $\eta_p^2 = .06$; F_2 (5.69, 660.16) = .96, $p = .448$, $\eta_p^2 = .01$.

As in the accuracy analysis, we then focused on the relative clause sentences (SR and OR) and the masked conditions (CT, SMN, RCT). Two-way repeated measures ANOVAs by participants and items were conducted, with reaction times as a dependent variable, and mask type (CT, SMN, RCT), and sentence type (SR, OR) as independent variables.[7] There was a significant main effect of sentence, F_1 (1, 18) = 18.09, $p < .001$, $\eta_p^2 = .50$; F_2 (1, 116) = 6.96, $p = .009$, $\eta_p^2 = .06$, confirming that the OR sentences were slower than the SR sentences. The main effect of mask was significant in the by-participants analysis and showed a trend towards significance in the by-items analysis, F_1 (2, 36) = 3.53, $p = .040$, $\eta_p^2 = .16$; F_2 (1.89, 219.51) = 2.59, $p = .080$, $\eta_p^2 = .02$, with pairwise comparisons revealing a difference between the CT condition and the SMN condition ($p = .042$) by participants, with a Bonferroni adjustment for multiple comparisons. This difference was in the opposite direction to our hypothesis, since SMN was slower than CT. However the effect did not reliably generalise across participants and items, suggesting that this result should be taken with caution. There was no Mask by Sentence type interaction, F_1 (2, 36) = .38, $p = .686$, $\eta_p^2 = .02$, F_2 (1.89, 219.51) = .47, $p = .616$, $\eta_p^2 = .00$.

To summarise the reaction time analyses, we found a main effect of syntactic complexity, where the simple sentences were answered more quickly than the SR sentences, followed by the OR sentences. Although a main effect of mask was found when comparing all

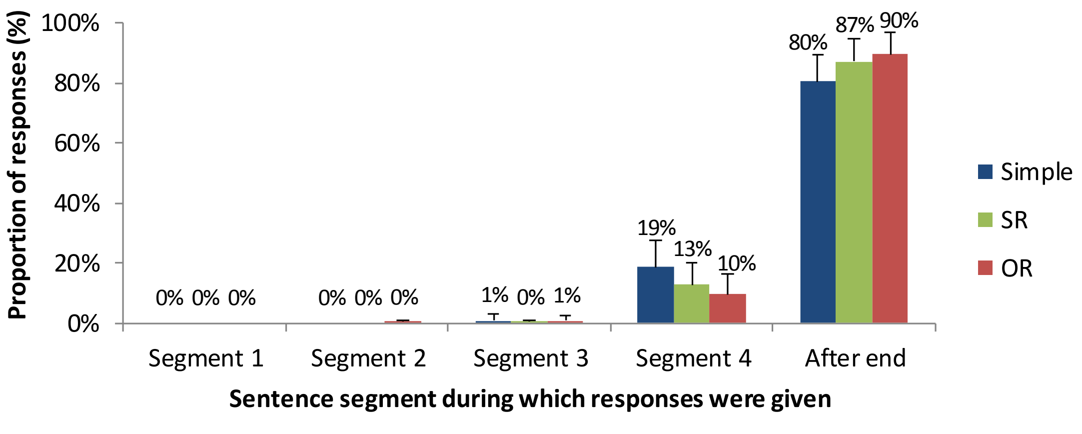
---

[6] Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of mask in the by-items analysis, $X^2(5) = 19.19$, $p = .002$. Greenhouse-Geisser corrections were therefore applied.

[7] In the by-items analysis, Mauchly's test revealed that the assumption of sphericity had been violated, $X^2(2) = 6.74$, $p = .03$. Degrees of freedom were corrected accordingly with Greenhouse-Geisser values.

four mask conditions, none of the pairwise comparisons were significant. When only the three masker conditions were compared (CT, RCT and SMN), a mask effect was only noted in the by-participants analysis. Given the very small numerical differences in reaction times between the masks (at most 52 ms, between CT and no mask), and the lack of a robust and consistent difference between the masks, it appears that the mask types had little effect on reaction times for the non-native listeners. Both the effect of syntactic complexity and the lack of a main effect of mask type or mask by sentence interaction were also found for the native listeners in Experiment 2. In this respect, native and non-native participants showed similar patterns of responses in their reaction times, except that the non-native listeners seem to have delayed responses compared to the native listeners, and they already showed an effect of syntactic complexity in their accuracy whereas the natives were at ceiling.

As in Experiment 2, responses were broken down by segments. The proportion of responses for each segment and sentence type across all masks is shown in Figure 3.4 (only correct responses were included). Most responses were given after the end of the sentence (80% simple, 87% SR and 90% OR), which was not the case for the native listeners in Experiment 2 (8% simple, 19% SR, 21% OR, see Figure 2.10). Furthermore, a negligible proportion of responses was given during segment 3 (1% simple, 0% SR, 1% OR), whereas most responses (80%, simple, 65% SR, 71% OR) had already been made during that time for the native participants.



Examples	**Segment 1**	**Segment 2**	**Segment 3**	**Segment 4**
Simple	Show the girl	with the	red	trousers
SR	Show the girl	who	is holding	the boy
OR	Show the girl	who	the boy	is holding

Figure 3.4. Experiment 3. Proportion of responses (%) per segment by sentence type (simple, SR, OR), averaged across all mask conditions (no mask, CT, SMN, RCT). Error bars indicate one standard error (by participants). The bottom part of the figure shows examples of segments for each sentence type.

The fact that the non-native participants did not respond until after the end of the sentence indicates that either they genuinely needed additional time to process the sentences before responding, or they were cautious in reporting their responses, even though they might have made their decision earlier in the sentences. The following eye-tracking analysis allowed us to distinguish between those two explanations.

3.3.1.3 Eye-tracking

The proportion of eye fixations per sample falling within each of the three regions of interest (see Figure 3.1) was calculated. Only ROI1 and ROI3 were kept, since they corresponded to the target and competitor characters for the subject and object relative sentences. The simple sentences where the target character was in ROI2 were not analysed, to facilitate the comparison between sentences. For each participant and each sentence, a proportion of fixations to the target character and a proportion of fixations to the competitor character were calculated over the duration of the sentence. The first step in the analysis was to determine the point at which participants reached their decision, i.e. when the proportion of fixations to the target character was significantly higher than the competitor. The decision point was defined as the point when the target and competitor fixations were significantly different, as long as the target fixations were significantly greater than the competitor fixations for at least 20 samples (corresponding to approximately 200 ms) after this point. This duration was based on a conservative estimate of the oculomotor planning delay, which is estimated at 200 milliseconds (e.g. Huettig & Altmann, 2005; McMurray, Clayards, Tanenhaus, & Aslin, 2008). A paired-samples permutation test based on a *t*-statistic was used[8] to compare the target and competitor fixation rates at each time sample, following the methodology described in Blair & Karniski, 1993. This test allows multiple comparisons across all samples of the trial to be carried out, while controlling for the familywise error rate, as well as being more powerful than a Bonferroni correction, in particular given that we expected each time sample to be correlated with the previous one. The test determines whether the difference between target and competitor is significantly different to 0. This Matlab permutation test function outputs a value of the *t*-statistic for the difference between target and competitor at each time sample, as well as the *p*-value for each time sample, and the critical value of *t* at which $p = .05$. Figure 3.5 shows an example for SR sentences in the CT condition. Equivalent figures for all other conditions can be found in Appendix H, section H.1 (Figure H.1 to Figure H.12).In the upper

---

[8] Using the mult_comp_perm_t1 function in Matlab, written by David Groppe, 2010.

panel of Figure 3.5, the fixation rates for target and competitor have been plotted from the onset of the sentence until the end of the trial (totalling 430 time samples). The bottom panel of Figure 3.5 shows the values of the *t*-statistics at each time sample (blue line), with the critical *t*-values (p = .05) plotted as red dotted lines above and below zero.

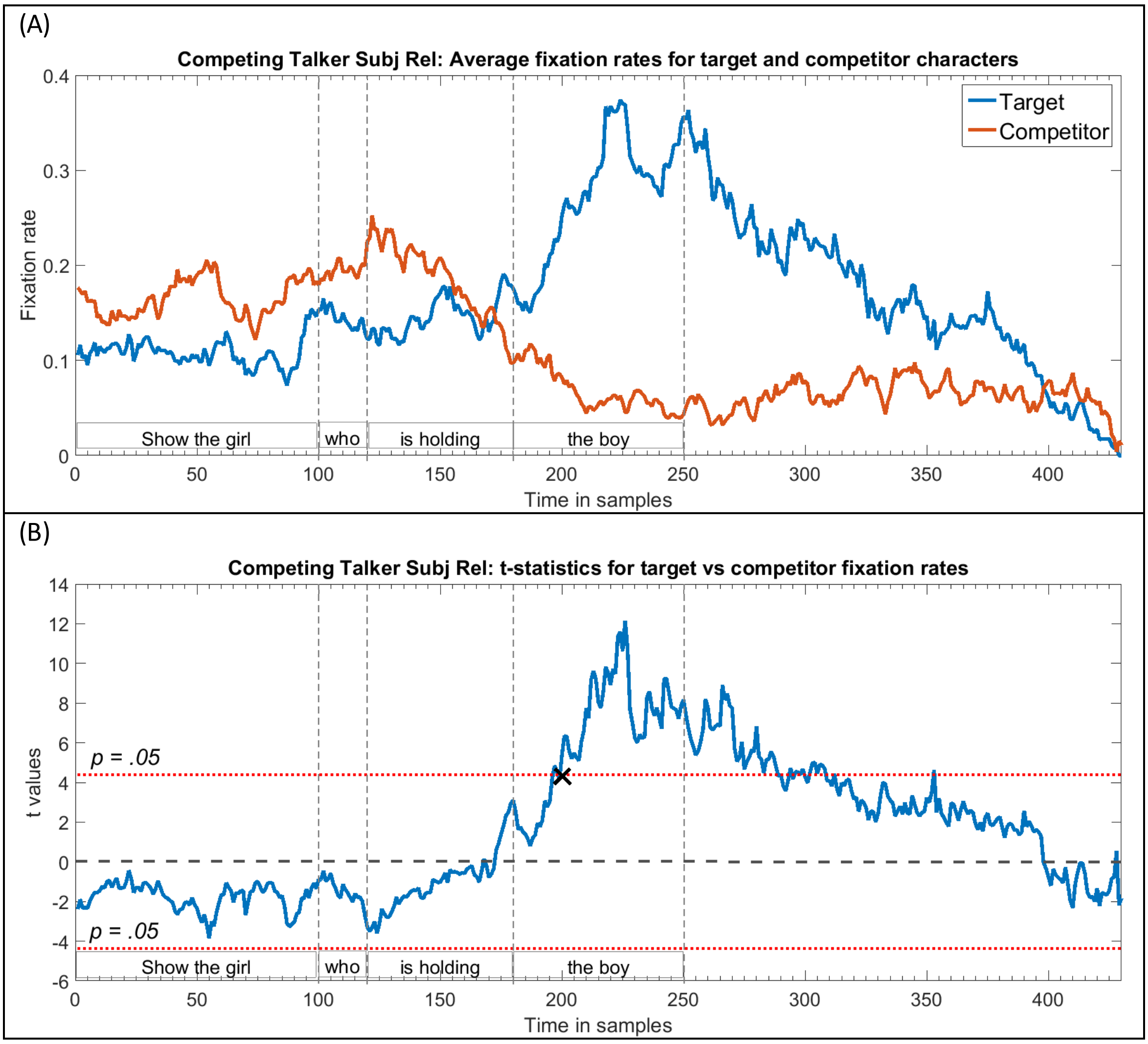


Figure 3.5. Experiment 3. Average fixation rates to the target and the competitor (A), and values of *t*-statistics for the difference between target and competitor (B), for the CT condition with SR sentences. In both panels, the horizontal dashed lines indicate the borders of the sentence segments, with an example sentence above the X-axis. In panel (B), the red dotted lines above and below 0 are plotted at the critical value of t where p = .05, and the black cross indicates the average point at which participants fixated the target more than the competitor for 20 samples or more.

I then determined the point in the sentence at which the proportion of fixations to the target was significantly higher than to the competitor, for a minimum of 20 samples, roughly equivalent to 200 milliseconds[9] to take into account the oculomotor delay. This was deemed to

[9] In segment 1, 200 ms = 20 samples ; in segment 2, 200 ms = 20.6 samples ; in segment 3,200 ms = 20.03 samples; in segment 4, 200 ms = 19.03 samples.

be the point at which participants' decision was reached, i.e., when they had understood the sentence. Table 3.4 shows the values of these decision points for each of the sentence types and mask types. These same decision points are indicated with black crosses in Figure 3.5 and in each of the figures in Appendix H, section H.1.

	No mask	**Competing talker**	**Reversed competing talker**	**Speech-modulated noise**	**Average**
Simple	156	168	162	160	**162**
Subj Rel	211	200	188	208	**202**
Obj Rel	192	210	199	226	**207**
Average	**186**	**193**	**183**	**198**	**190**

Table 3.4. Experiment 3. Point in time (expressed in samples) at which the target character was fixated significantly more than the competitor for at least 20 samples (corresponding to 200 ms).

It is important to note that the decision points were calculated based on averages, and using this methodology it is not possible to calculate them for each participant, thus precluding the use of inferential statistics for the decision points. The decision points for the SR and OR sentences fell within the last segment of the sentence. The decision points for the simple sentences all fell within the second-to-last segment of the sentence. From these descriptive data, it appears that the least demanding condition was the simple sentence with no mask (156 samples), and the most difficult was the OR with speech-modulated noise (226 samples). The simple sentences were all resolved before the end of the third segment, which is not surprising given that the information at that point is sufficient to make an unambiguous decision. For instance, in the example shown in Figure 3.1, once the participant had heard "Show the girl with the red…", there was only one possible answer, since the other girl does not have red elements. The surprisingly late decision moment in the SR with no mask condition could be due to the relative ease of the task. Indeed, it is possible that when a participant did not find an item challenging, their gaze wandered around the screen more than when an item was more demanding, leading to a later decision point. [10] It is relatively safe to conclude this, given that the accuracy and reaction time data clearly showed that the subject relative sentences were less demanding than the object relative sentences. In addition, the decision points for the subject relative sentences always fell before the decision points in the object relative sentences for all masked conditions (CT, RCT, SMN). Focusing only on the masked conditions across all sentences, these data indicate that the least demanding mask was the

[10] One could however argue that the simple sentences should also have led to more random gazes since these sentences are even less demanding, yet this was not the case.

reversed competing talker (average decision point of 183 samples), followed by the competing talker (average decision point of 193 samples), and finally the speech-modulated noise (average decision point of 198 samples). This does not follow the same pattern as the reaction time data. However the decision point values do not allow us to conclude anything about the statistical significance of these differences.

Figure 3.6 shows the average fixation rate difference between the target and competitor for each mask type (separate lines) and sentence type (separate graphs). These figures provide information about participants' certainty, as well as the time-course of sentence processing. The peak of the curves can be interpreted as representing certainty: the higher the peak the greater the difference between eye-fixations to the target and eye-fixations to the competitor, the greater the certainty. Furthermore, these curves provide information about when participants start to reach their decision, i.e. when the curves start to rise, complementing the decision points calculated above.

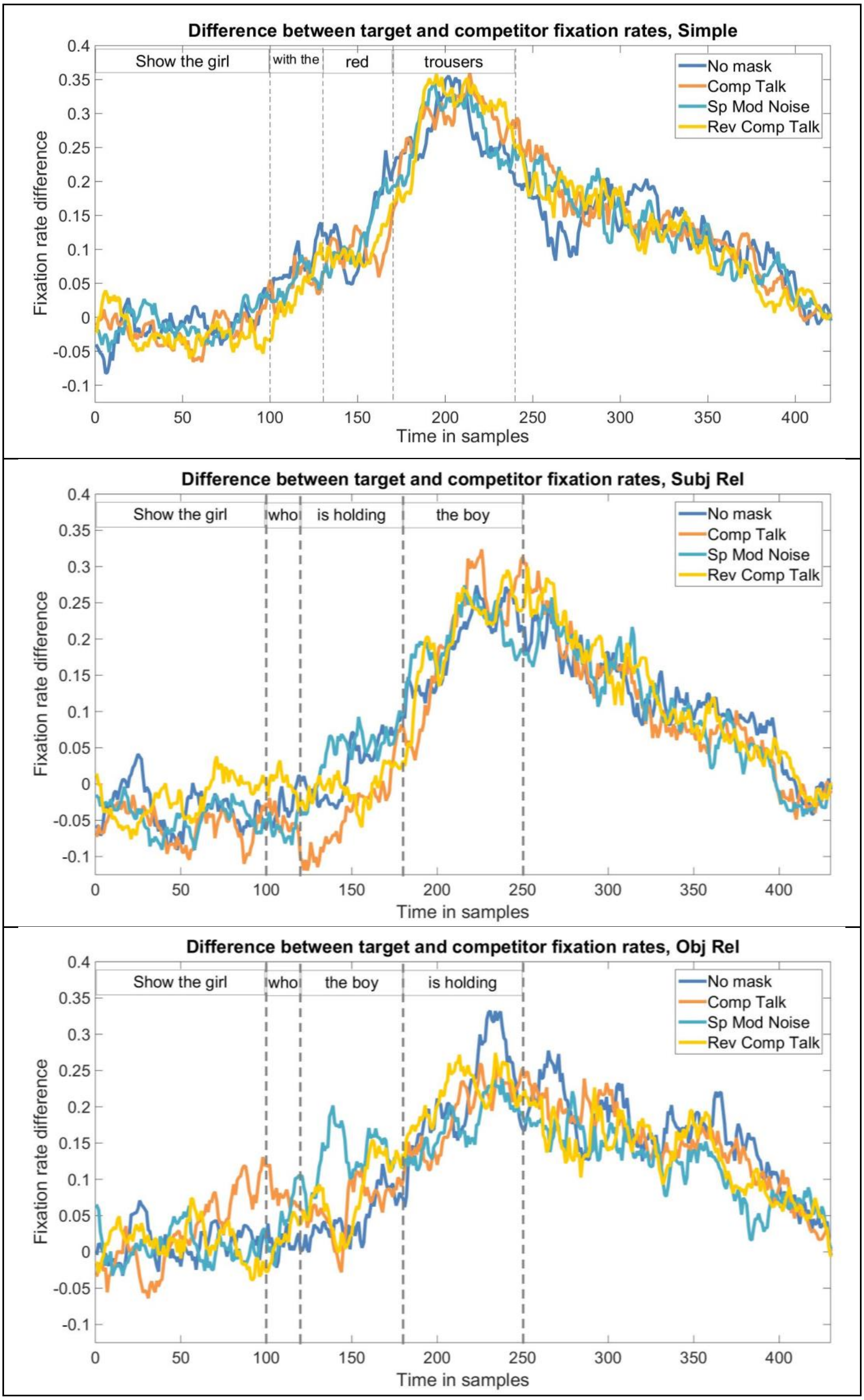


Figure 3.6. Experiment 3. Fixation rate differences between target and competitor characters for each mask condition (no mask, CT, RCT, SMN). The top panel shows the simple sentences, the middle panel the SR sentences, and the bottom panel the OR sentences.

To determine whether there were statistically significant differences between the masks, pairwise comparisons were calculated for each sentence type, with 99.17% bootstrapped (10'000 resamples) confidence intervals to correct for the 6 multiple comparisons. The difference in fixation rate difference for SR sentences between the CT condition (dark orange curve in the middle panel of Figure 3.6) compared to the no mask condition (dark blue curve in the middle panel of Figure 3.6) for SR sentences is plotted in Figure 3.7. All other pairwise comparisons are reported in Appendix H, section H.2 (Figure H.13 for the simple sentences, Figure H.14 for SR sentences, and in Figure H.15 for OR sentences).

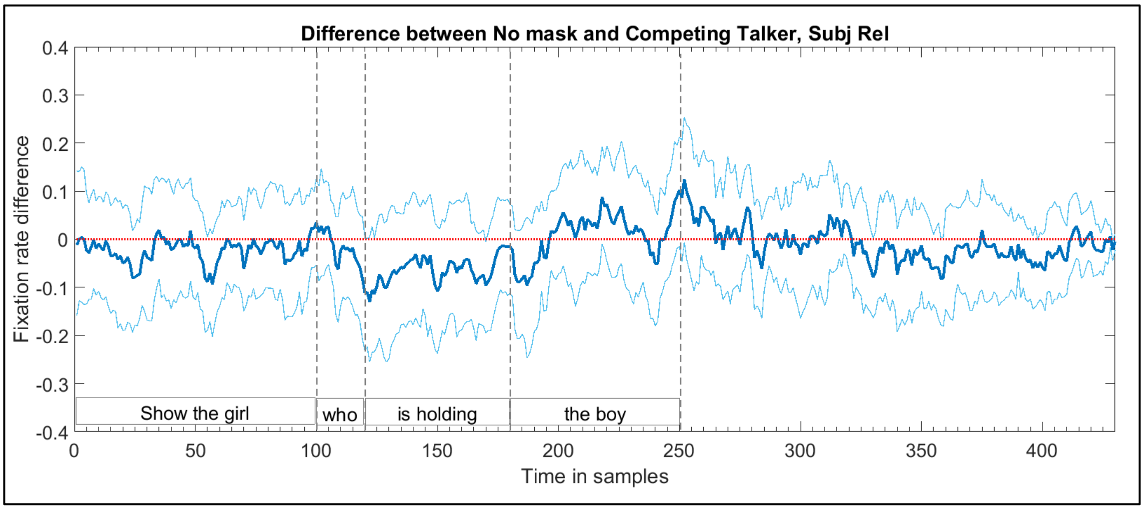


Figure 3.7. Experiment 3. Fixation rate difference (dark blue line) between the no mask condition and the competing talker condition, with 99.17% confidence intervals (light blue lines) to correct for multiple comparisons.

In Figure 3.7, the dark blue line represents the average difference between the fixation rate difference for the no mask condition and the fixation rate difference for the CT condition and the light blue lines represent the upper and lower 99.17% confidence intervals. If there was a significant difference during sentence presentation, the confidence intervals would depart significantly from the 0 line (horizontal red line). Once again I adopted the 20 sample threshold, whereby the difference between the masks had to take place for at least 20 samples to be considered reliable. Anything less than this was probably due to temporary differences that did not affect sentence processing. Although the confidence intervals departed from 0 in the SMN condition vs the CT condition and the SMN condition vs the RCT condition, it was for less than 15 samples, and did not take place during the time when a decision could have been made with regard to the sentence processing. We can therefore conclude that based on the eye-tracking data, there was no difference between mask types[11].

3.3.2 Cognitive tests and English proficiency

A number of additional measures of cognitive functions (attention and memory) and language proficiency were collected, with the same hypothesis as in Experiment 2. We expected individual differences in language proficiency, working memory, short-term memory and visual attention to modulate participants' performance in the main sentence comprehension task. In the next paragraphs, I briefly outline the results for each of the tests separately, followed by their relationship with the sentence comprehension task.

3.3.2.1 English language proficiency

LexTALE. All 19 participants completed this test. Individual results can be found in Appendix F section F.1 (Table F.1). Participants' results in the LexTALE evidenced a range of scores, from 34% to 96% correct, with an average of 72.1% (SD = 17.2).

Self-report questionnaire. Eighteen participants completed the self-report questionnaire (Appendix F, section F.2). Fourteen participants spoke other languages in addition to Danish and English, including German, Norwegian, Swedish, Spanish, French, Afrikaans and Vietnamese. Three participants lived outside of Denmark for longer than two years, but continued to speak Danish with their family. On average, participants had studied English at school for 9 years (SD = 1.71). Only 5 had taken a standardised test of English (TOEFL, Cambridge Language Assessment, IELTS), which was insufficient to make any comparisons based on their standardised test scores. An overall score was calculated based on average scores for questions in sections 8 and 9, which were both based on a scale from 1 to 10. All questions in these two sections were included except for the two relating to work since a few participants were not working. Section 8 required participants to indicate how often they used English in a variety of contexts, with separate questions for reading/writing and for spoken language use. The higher the score, the more often participants used English. Section 9 reflected participants' degree of comfort using English in different contexts. The higher the score, the higher the participant's degree of comfort in a variety of situations. Individual results for sections 8 and 9 can be found in Appendix F section F.2.

[11] This was also evidenced in the results of permutation tests between each of the two masks, where none of the differences ever reached significance (note that this method does not control for familywise error rate between the six pairwise comparisons, but given the lack of effect it should not be an issue).

The average score for section 8, 'English usage', was 5.3 (SD=1.6), indicating that participants used English moderately across a variety of contexts. Note, however, that there were large individual differences (3.3 to 7.8) and that participants mainly used English when watching television or movies (*M* = 8.9), and when reading and writing for their studies (*M* = 7.9), but infrequently with their family (*M* = 3.0) and friends (*M* = 4.5).

The average score for section 9, 'degree of comfort', was 8.7 (SD = 0.9), indicating that despite not using English very often in their everyday life, participants felt very comfortable using English in a variety of situations. Scores showed less variation across participants (ranging from 6.5 to 9.9) and across contexts (ranging from 7.7 to 9.8 averaged across participants) compared to the 'English usage' composite.

The internal reliability of the questionnaire (all items in sections 8 and 9 together) was good, $\alpha = .89$. Unsurprisingly, the two self-rated proficiency sections (usage and comfort) were significantly positively correlated, $r(16) = .615$, $p = .007$. These two sections were therefore treated as one, yielding an overall average score of 7.27 (SD = 2.86).

A Pearson correlation showed that the LexTALE scores were significantly positively correlated with the average proficiency questionnaire scores, $r(16) = .562$, $p = .015$. This correlation indicates that the two proficiency measures probably do tap into the same construct, although the proficiency questionnaire provides additional qualitative information. The LexTALE and proficiency questionnaire scores were converted to Z-scores and then combined to create one composite proficiency score per participant.

3.3.2.2 Memory

3.3.2.2.1 Digit span

Seventeen participants completed the forward and backward digit span tasks. The digit span was defined as the length of the longest correctly recalled list of digits. Table 3.5 shows the average forward and backward digit spans. Individual results can be found in Appendix F section F.3, Table F.3.

Digit span	Range	Mean	Standard deviation
Forward	5-8	6.47	1.12
Backward	3-8	5.06	1.43

Table 3.5. Experiment 3. Range, mean and standard deviation for each of the digit span tasks.

The individual results were used to investigate possible relationships between verbal short-term and working memory, and performance in the sentence comprehension task. This will be covered in section 3.3.2.4.

3.3.2.2.2 Reading span

Seventeen participants completed the reading span task. Table 3.6 shows the results averaged across participants. Individual results can be found in Appendix F section F.3.

Reading span	Range	Mean	Standard deviation
Letter	2-6	4.59	1.12
Meaning	5-10	7.29	1.61

Table 3.6. Experiment 3. Range, mean and standard deviation for the letter and meaning spans in the reading span task.

Two measures were derived from this task: the letter span and the meaning span. The letter span corresponded to the length of the longest string of letters correctly recalled after the sentence presentation, and is the measure we were interested in. The meaning span corresponded to the number of correct judgments when reading the sentence. This latter measure is an indication of how attentive the participants were on the task, and ensures that they were actually processing the sentences and not merely focusing on the letters presented at the end of each sentence. The average letter span was 4.59 (*SD* = 1.12) and ranged from 2 to 6. The average meaning span was 7.29 (SD = 1.61), and ranged from 5 to 10. The meaning span results indicate that participants were genuinely engaging in both parts of the task.

Bivariate correlations between the three memory tasks were calculated. None of the correlations were significant at $\alpha = .05$. The backward and forward digit spans were positively correlated, $r(15) = .447$, $p = .072$, as were the backward digit span and the reading span, $r(14) = .371$, $p = .157$ and the backward digit span and the reading span, $r(14) = .201$, $p = .455$. As previously mentioned, the reading span and backward digit span tasks are widely accepted as measures of working memory (Conway et al., 2005). As such, there is a theoretical reason for considering these two tests together in subsequent analyses. Furthermore the correlation between the forward digit span and the two other memory tasks was deemed sufficient to group these tests together, yielding a composite memory score based on individual Z scores.

3.3.2.3 Visual flanker task

Sixteen participants completed the flanker task. Individual results can be found in Appendix F, Table F.5.The difference between the inconsistent condition and the consistent condition was calculated for each participant as a measure of the cost of inhibiting the visual distractor. On average, this difference was 47.5ms (*SD* = 24.1).

3.3.2.4 Relationship between cognitive measures and sentence comprehension task

In Experiment 3, as in Experiment 2, we were interested in the relationship between the cognitive measures and performance in the sentence comprehension task. Table 3.7 summarises the Pearson's correlations for the composite proficiency score, composite memory score, and flanker task reaction time difference, with the differences between conditions in the sentence comprehension task. Accuracy and reaction time differences were calculated between the masked conditions (average of CT, RCT, and SMN) and the unmasked condition, and between the competing talker condition and the EM controls (average of RCT and SMN), as well as between the OR and SR sentences.

	Proficiency composite	Memory composite	Flanker task difference
Accuracy difference masked – unmasked	- .110	.000	- .032
Accuracy difference CT – (RCT + SMN)	- .351	- .398	.464
Accuracy difference OR - SR	- .267	- .211	- .244
Reaction time difference masked-unmasked	- .103	.139	.039
Reaction time difference CT – (RCT + SMN)	.232	- .324	- .032
Reaction time difference OR - SR	.422	.118	- .202

Table 3.7. Experiment 3. Bivariate correlations between the composite proficiency scores, composite memory scores and the flanker task difference in reaction times with the accuracy and reaction time differences between masked and unmasked conditions, the accuracy and reaction time differences between the CT condition and the energetic mask controls (RCT and SMN), and the accuracy and reaction time differences between OR and SR. No correlations were significant at $\alpha = .0056$ (Bonferroni-corrected).

Although previous analyses had shown that there was no main effect of mask, there may have been individual differences that could have shown up in these correlation analyses. However, none of the correlations were significant, for either the accuracy or the reaction time differences. In conclusion, the proficiency and cognitive measures did not shed light on possible individual differences in the sentence comprehension task reaction times.

3.4 Discussion

In keeping with previous studies contrasting subject and object relative clauses, we found that object relative clauses delayed reaction times compared to subject relative clauses for our group of non-native listeners. In contrast to Experiment 2, the effect of syntactic complexity was already apparent in the accuracy data, indicating that the non-native participants were more sensitive to syntactic complexity than the native participants. Although the participants in this experiment were non-native, their language proficiency was high, which accounts in great part for the very high accuracy. However, the fact that their reaction times were substantially slower than the native listeners' (although a direct comparison is not possible given several procedural changes between the experiments) and that their responses were given in large part after the end of the sentence suggests that, despite their high proficiency, the task was slightly more demanding for them. The eye-tracking data confirmed that observation, since the moment at which participants looked at the correct character happened on average during the final segment of the sentence and not before. In Experiment 2, participants’ button presses were already taking place during this last segment, which means that, had we measured eye movements on the native participants, we would probably have seen their eye fixations shifted slightly in time as well

However, contrary to what we had predicted, there was no difference between the types of mask, let alone an interaction between the mask type and the sentence type. Similarly to the findings of Experiment 2 (Chapter 2), no effect of informational interference was found in Experiment 3.

One possible explanation for the lack of difference between masks is the purported “bilingual advantage” in executive functions. Indeed, most participants had started learning English as children, and their proficiency was high enough to be considered bilingual by some definitions of the term. Several studies have reported that bilingual individuals are better at non-verbal executive control tasks (e.g. attentional control and inhibition measured by the Simon task or the Stroop task) than their monolingual counterparts (e.g. Bialystok, Craik, & Luk, 2012; Yang, Yang, & Lust, 2011). The bilingual advantage has been attributed to the additional inhibition and attentional control required to actively suppress the other language. Participants in Experiment 3 may have compensated for the difficulty of the speech-in-noise task by tapping into particularly developed attentional control. Additional evidence for this possibility was reported in an experiment investigating sentence comprehension with a competing talker by Italian-English late bilinguals and Italian or English monolinguals (Filippi, Leech, Thomas, Green, & Dick, 2012). Participants were presented with a sentence comprehension task using Italian sentences varying in syntactic complexity (canonical SVO vs non-canonical OVS). All sentences were masked by a competing talker of the opposite gender to the target talker, either in Italian or English. When both target and competitor were presented in Italian, bilingual participants' accuracy was higher than the monolinguals in the more difficult syntactic condition (OVS). However, this bilingual advantage was not observed in reaction times, or when the English monolinguals were compared to the bilinguals. The authors conclude that bilinguals are more able to inhibit the interference from a competing talker than monolinguals. They also report that bilingual participants whose second-language proficiency was higher were less affected by the competing talker. This was explained by the fact that more proficient bilinguals had more experience in attentional control.

If the bilingual advantage explanation were true in Experiment 3 of this thesis, one might have expected English proficiency to be correlated with the ability to deal with the competing talker. However, none of the proficiency measures correlated with the mask differences in the sentence comprehension task. Despite this, it is still possible that the so-called bilingual advantage may at least partly explain the lack of difference between mask conditions.

In addition to this bilingual advantage, non-native listeners in Experiment 3 may have been aided by the fact that subject and object relative clauses follow the same structure in Danish and English (e.g. Jensen De López, Sundahl Olsen, & Chondrogianni, 2014). The similarity between relative clauses in Danish and English may have allowed the Danish listeners in Experiment 3 to rely on their native knowledge of word order and syntactic structure to parse the sentences and respond to the comprehension task faster than if their native language had followed a different sentence structure to English. Studying a group of listeners whose native language is structured differently to English, e.g. German or Japanese, would allow to disentangle these issues.

Another explanation could lie in the SNR at which masked sentences were presented. Surprisingly, participants seemed to deal with the masked conditions just as well as with the unmasked condition. It is possible that the SNR was too high even for the non-native listeners, despite the fact that similar SNRs seem to be detrimental in other experiments with non-native listeners. Indeed, in their review of non-native speech perception studies in adverse conditions, Lecumberri et al. (2010) mentioned that 0dB SNR is the middle of the range for non-native studies.

One crucial difference between most studies of speech perception in adverse conditions and the present thesis is the inclusion of visual stimuli corresponding to the content of the target sentences. Most speech perception studies require the participants to rely solely on the acoustic input to resolve the task. In the case of a VWP, the visual information reduces the possible candidates, which could lead to a decreased reliance on the acoustic input. In the next chapter, I will investigate this idea by decreasing the signal-to-noise ratios, thus increasing the difficulty of the task

Chapter 4: effect of low intelligibility and syntactic complexity on informational interference from a competing talker

In this chapter I will investigate the influence of low intelligibility of the target signal on informational interference and sentence comprehension. One possible explanation for the absence of informational masking in Experiments 2 and 3 is the relative perceptual ease of the task. Indeed, the lack of a mask effect could lead to the conclusion that the signal-to-noise ratio may have been too favourable for any differences between masks to arise. This may be the case despite the fact that SNRs around -5 dB are not uncommon in research on masked speech (e.g. Brungart, 2001; Iyer, Brungart, & Simpson, 2010; Koelewijn, Zekveld, Festen, & Kramer, 2012). Indeed, the picture-selection task used in this thesis presents a highly restricted visual world which may reduce the lexical candidates and decrease task difficulty.

The SNR for Experiment 5 (sentence comprehension task) was thus decreased to -22 dB SNR for the SMN condition, and -25 dB SNR for the CT and RCT conditions. These SNRs were based on the results of Experiment 4, a transcription task similar to Experiment 1 (Chapter 2). The goal of these latter experiments was to select signal-to-noise ratios leading to a predetermined level of transcription accuracy. In Experiment 4 the picture was presented on the screen before and during the sentence presentation. This was done to follow the conditions in the sentence comprehension task more closely. Indeed, in the main sentence comprehension task (Experiments 2 and 3), the pictures allowed participants to disambiguate potentially unclear or unintelligible words by limiting the possible number of lexical candidates to those appearing in the pictures. In effect, Experiment 2 and 3 used a closed set of candidates, all visible on the screen, which should be easier than an open set. Experiment 1 was a measure of intelligibility of the masked sentences *without* the disambiguating information provided by the pictures. However, this may have led to underestimating the intelligibility in the sentence comprehension task with pictures. Indeed, participants' transcription accuracy would presumably increase with the disambiguating information from the pictures. The second difference between Experiments 1 and 4 was the choice of lower SNRs, which was a direct consequence of presenting the pictures at the same time. In Experiment 1, the final SNR was deliberately chosen to yield very high transcription accuracy, since we were interested in the effect of informational interference in conditions of high intelligibility. With lower ambiguity of the signal, lower SNRs were necessary to achieve the same level of performance in the transcription task. The goal in Experiment 4 was to identify the SNR that would lead to approximately 80% transcription accuracy with the picture, which was lower than the 96-98% accuracy of the chosen SNR in Experiment 1. This value was chosen instead of the more conventional cut-off of 50% because at 50% accuracy in the intelligibility task, the sentence comprehension task would have been near-impossible to perform. Indeed, more than 50% of the keywords need to have been heard to correctly interpret the sentence at all. Furthermore, we expected accuracy to decrease between the intelligibility task and the sentence comprehension task, since the latter involves processing the sentence in addition to identifying the words.

Although the main goal of Experiment 4 was to select the SNRs for Experiment 5, we did have a number of hypotheses. We did not expect to see an effect of sentence type, because the task did not require participants to process the syntax, they had no time limit, and were all native listeners. Based on the results of Experiment 1 at -10 dB SNR, we expected the SMN condition to lead to lower transcription accuracy than the CT or RCT conditions, for a given SNR. This would be expected if SMN is a more effective energetic masker than RCT. There was no specific prediction for the difference between the CT and the RCT conditions, although we expected informational interference to arise only in the sentence comprehension task (Experiment 5), since the intelligibility task is not an optimal way of measuring processing load. In other words, the effect of SMN might arise in the transcription task because of its energetic masking properties and increased perceptual load, whereas the effect of a competing talker was hypothesised to be due to an increased cognitive load, which should not affect transcription (at least not for a group of young normal-hearing native listeners).

The first hypothesis for Experiment 5 was that an effect of mask vs no mask would be evidenced across all measures, due to the low intelligibility of the masked conditions. The second hypothesis was identical to Experiments 2 and 3, predicting that the competing talker condition would lead to slower reaction times and lower accuracy compared to the energetic mask controls, as well as delayed sentence resolution and decreased certainty as evidenced by eye-movements. Furthermore, in addition to a main effect of sentence type, we expected the effect of the competing talker to be exacerbated by syntactic complexity.

Finally, we had the same predictions as in previous chapters regarding the relationship between susceptibility to masking and informational interference, and measures of memory and attention. We expected that participants with higher short-term memory, working memory, and/or attention should show less susceptibility to masking, and in particular that they would be less affected by interference from the competing talker compared to the energetic mask controls.

4.1 Method for Experiments 4a, 4b and 5

4.1.1 Participants

Participants for Experiments 4a, 4b and 5 were monolingual native speakers of English, who reported never having experienced hearing difficulties or speech-language impairments and had normal or corrected vision. All participants were students attending the University of York and received payment or course credit for their time. In Experiment 4a, there were 9 participants, 4 female and 5 male, with a mean age of 22;8 years (*SD* = 2;1). A further 9 participants took part in Experiment 4b, 4 female and 6 male, with a mean age of 21;8 years (*SD* = 2;2). In Experiment 5 there were an additional 36 participants, 34 female and 2 male, with a mean age of 20;5 years (*SD* = 1;6). Each participant took part in only one experiment.

4.1.2 Materials

4.1.2.1 Experiment 4: intelligibility task with pictures at low SNRs

Experiment 4 was designed to select the SNR for Experiment 5. Stimuli for this experiment consisted of the same 291 target sentences as in Experiments 1, 2 and 3, masked by the same competing talker (CT), reversed competing talker (RCT) and speech-modulated noise (SMN) as in Experiment 3. The accompanying pictures were the same as in Experiments 2 and 3. The SNRs were different to Experiments 1, 2, and 3, but were reached using the same procedure as in Experiment 1: the masker sound files (CT, RCT, SMN) were normalised to an intensity of 68 dB and the intensity of the target sentences was normalised depending on the desired SNR: 55 dB (-13 dB SNR), 52 dB (-16 dB SNR), 49 dB (-19 dB SNR), 46 dB (-22 dB SNR), 43 dB (-25 dB SNR) and 40 dB (-28 dB SNR). The time alignment of the target sentences with each masker file was identical to the previous experiments.

The SNRs in Experiment 4 thus started at -13 dB (3 dB under the lowest SNR in Experiment 1), and decreased in 3 dB steps down to -28 dB SNR. The step-size of 3 dB was smaller than in Experiment 1 (5 dB) for a more precise fine-tuning. The SNRs were separated across two experiments due to the exploratory nature of this experiment. Indeed, as will be shown in the following sections, the transcription accuracy in Experiment 4a was still too high, leading to the addition of Experiment 4b. Experiment 4a included the three highest SNRs (-13 dB, -16 dB, -19 dB) and Experiment 4b included the three lowest SNRs (-22 dB, -25 dB, -28 dB).

4.1.2.2 Experiment 5: sentence comprehension and speeded picture-selection task at low SNRs

The sentence comprehension task was the same as Experiment 3, but presented at different SNRs. As will be further detailed in the following sections, two SNRs were chosen based on the results of Experiment 4: -22 dB for the SMN condition, and -25 dB for the CT and RCT conditions.

4.1.2.2.1 Cognitive measures

In Experiment 5, the same cognitive measures were used as in Experiment 2 (Chapter 2): the non-word recall task from the AWMA (Alloway, 2007) to assess phonological short-term memory, the listening recall task from the AWMA to assess verbal working memory, and the 'flanker task' to assess visual selective attention.

4.2 Design and Procedure

Participants were tested individually in a sound-insulated booth. All stimuli were presented using the same headphones and monitor as in Experiments 1 and 2: Sony MDR v700 headphones and a 22-inch Dell monitor, with a resolution of 1920 x 1080 pixels.

4.2.1 Experiment 4: intelligibility with pictures at low SNRs

Similarly to Experiment 1, Experiment 4 was delivered and responses collected using the DMDX software (Forster & Forster, 2003). Participants' task was to type the target sentences as accurately as possible, excluding the lead phrase "*Show the*". There was no time limit, and participants could start typing as soon as they wanted to, before or after the end of the sentence. Participants' responses appeared on the screen as they typed, and they could correct their typing as needed. Whereas in Experiment 1 the filler trials and familiarisation sentences were excluded, in Experiment 4 all sentences were included, to render the experiment as similar to the main sentence comprehension task as possible. Only the masked sentences were presented (CT, RCT, SMN), and the presentation of items was fully randomised across sentences, mask types and SNRs. The experiment was divided into three blocks, enabling participants to take a break after each block as needed.

As in Experiment 1, the dependent variable was the percentage of correct keywords per sentence. Each sentence had three keywords, defined as the content words within each sentence (see Table 2.3 in Chapter 2 for examples). Keywords were counted as correct if spelled correctly, misspelled but phonologically identical, or if obvious typographical errors were made, as long as they did not result in a different lexical item.

4.2.2 Experiment 5: sentence comprehension and speeded picture-selection task at low SNRs

Each session lasted an hour and a half, including the cognitive testing. Participants in Experiment 5 first completed the sentence comprehension task (one hour), followed by the cognitive tests (flanker task and two memory tasks). The sentence comprehension task and procedure were the same as in Experiment 3, where participants were asked to respond as quickly and accurately as possible to the target sentence, while placing their head on the chin-rest for their eye-movements to be tracked. Participants' eye-movements were monitored using the same camera as in Experiment 3, a SR Research Eyelink 1000 Plus desk-mounted camera at a sampling rate of 1000 Hz, with a chin-rest to minimise head movements. Participants were seated approximately 60 cm from the screen and the lighting was kept constant. Participants were encouraged to take breaks when necessary. A 9-point calibration was carried out before the beginning of the experiment and each time the participant moved their head from the chin-rest. After seven familiarisation trials (two simple, two SR, two OR and one filler sentence, with all mask combinations across participants), the experimenter checked that participants were still seated comfortably and that they had understood the task. The picture was first shown for 1 second on its own, and stayed on the screen until after the end of the sentence. The eye-tracking analysis followed the same procedure as in Experiment 3 (see 3.2.2 Procedure for eye-tracking analysis), where sentence lengths were normalised to the same number of samples per segment, and the proportion of fixations to the regions of interest was calculated for each time sample as the sentence unfolded.

After the sentence comprehension task, participants completed the flanker task, the non-word repetition test, and the listening recall test.

4.3 Results

4.3.1 Experiment 4: intelligibility with pictures at low SNRs

4.3.1.1 Experiment 4a: -13 dB, -16 dB, -19 dB SNR

The proportion of correct keywords (maximum three) was calculated for each sentence. Figure 4.1 summarises the response accuracy by SNR and mask type, and Figure 4.2 shows accuracy by sentence type separated by SNR. Accuracy across conditions was high, with averages across masks ranging from 86% (SMN in -19 dB SNR) to 98% (CT and RCT in -13 dB SNR).

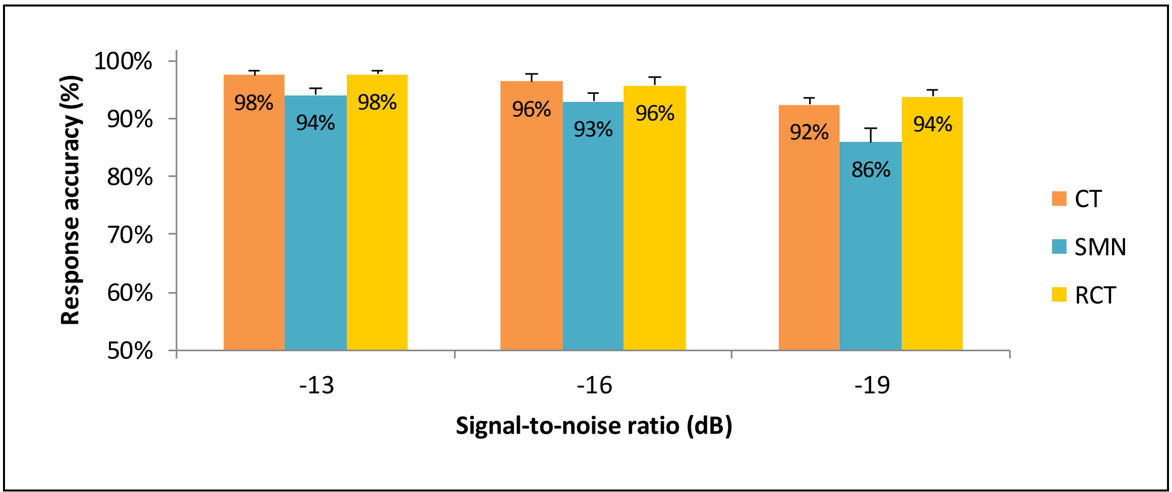


Figure 4.1. Experiment 4a. Response accuracy in percent of accurate keywords per sentence for each SNR (-13, -16, -19 dB) and mask type (CT, SMN, RCT), collapsed across sentence types. Error bars indicate one standard error (by participants).

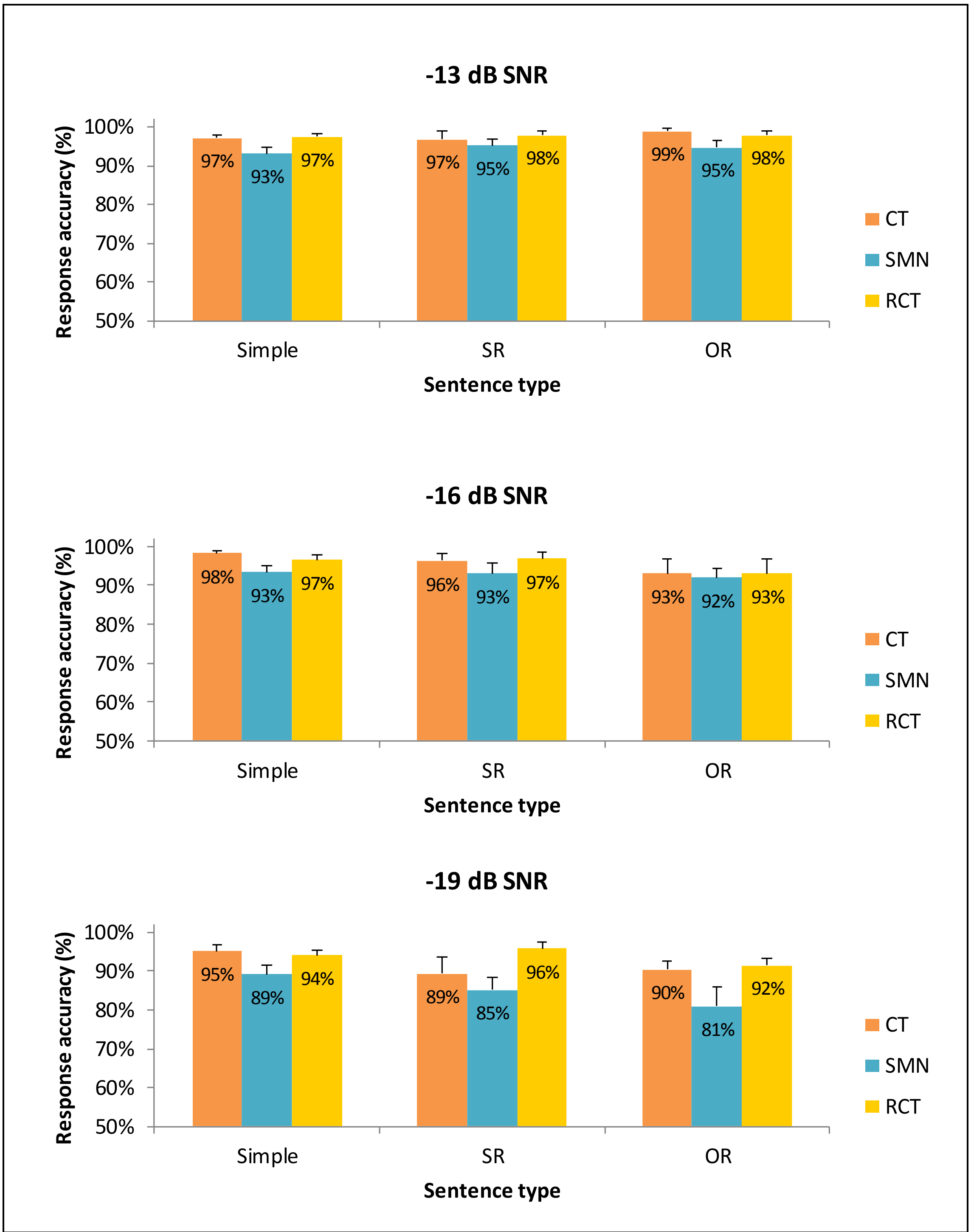


Figure 4.2. Experiment 4a. Response accuracy in percent of correct keywords per sentence for each sentence type (simple, SR, OR) and mask (CT, SMN, RCT), separated by SNR (-13, -16, 19 dB). Error bars indicate one standard error (by participants).

All descriptive statistics reported were calculated based on the by-participants analysis. Three-way repeated measures ANOVAs by participants and items were conducted with percent of correctly typed keywords as a dependent variable, SNR (-13 dB, -16 dB, -19 dB), mask (CT, RCT, SMN), and sentence (simple, SR, OR) as independent variables.[12]

There was a main effect of SNR, $F_1(2, 16) = 19.82$, $p < .001$, $\eta_p^2 = .71$, $F_2(1.71, 409.8) = 36.08$, $p < .001$, $\eta_p^2 = .13$. Pairwise comparisons showed that there was a significant difference between -19 dB SNR (*M* = 91%, *SD* = 5%) and -16 dB SNR (*M* = 95%, *SD* = 4%), with *p* = .007 by participants, and *p* < .001 by items. There was also a significant difference between -19dB SNR and -13dB SNR (*M* = 96%, *SD* = 3%), with *p* = .003 by participants, and *p* < .001 by items. The difference between -16dB and -13dB SNR was not significant by participants (*p* = .10) but was significant by items (*p* < .001). The lowest SNR (-19 dB) was the least accurate, followed by -16 dB, and -13 dB SNR led to the highest accuracy.

There was a main effect of mask, $F_1(2, 16) = 9.41$, $p = .002$, $\eta_p^2 = .54$, $F_2(1.72, 413.92) = 16.27$, $p < .001$, $\eta_p^2 = .06$. Pairwise comparisons with Bonferroni corrections revealed that there was a significant difference between the SMN condition (*M* = 91%, *SD* = 5%) and the RCT condition (*M* = 96%, *SD* = 3%), *p* = .003 by participants, *p* < .001 by items. There was a significant difference between the SMN condition and the CT condition (*M* = 95%, *SD* = 3%) in the by-items analysis (*p* < .001) but not in the by-participants analysis (*p* = .072). The RCT and CT conditions did not differ significantly, *p* = 1 by participants and by items. Numerically, the RCT condition was the most accurate, followed by the CT condition and finally the SMN condition.

There was no main effect of sentence type, $F_1(2, 16) = 2.33$, $p = .13$, $\eta_p^2 = .23$, $F_2(2, 240) = 2.92$, $p = .10$, $\eta_p^2 = .02$. However, although none of the interactions were significant at $\alpha = .05$ in the by-participants analysis, in the by-items analysis the SNR by sentence interaction was significant, $F_1(4, 32) = 2.53$, $p = .06$, $\eta_p^2 = .24$, $F_2(3.42, 409.80) = 3.49$, $p = .012$, $\eta_p^2 = .03$, as well as the SNR by mask interaction, $F_1(4, 32) = 2.16$, $p = .096$, $\eta_p^2 = .21$, $F_2(3.58, 858.76) = 3.58$, $p = .025$, $\eta_p^2 = .01$.

The SNR by sentence interaction was possibly due to the different pattern of results in the -13 dB SNR condition compared to the -16 dB SNR and -19 dB SNR conditions. At -13 dB SNR, the simple condition (*M* = 96%, *SD* = 3%) was just one percent lower than the SR (*M* = 97%, *SD* = 5%) and the OR (*M* = 97%, *SD* = 4%) conditions, whereas at -16 dB SNR and -19 dB

[12] Mauchly's test indicated that the assumption of sphericity had been violated in the by-items analysis for the SNR by mask interaction, $X^2(9) = 56.21$, $p < .001$, as well as for the main effect of SNR, $X^2(2) = 44.91$, $p < .001$, and the main effect of mask, $X^2(2) = 41.57$, $p < .001$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

SNR the difference between the sentence types was greater, with decreasing accuracy as complexity increased. However, because the interaction failed to generalise across participants and items this is to be interpreted with caution. The SNR by mask interaction was probably due to the greater difference between the SMN condition and the two other masks in -19 dB SNR compared to -13 dB and -16 dB SNR. The lowest SNR seems to have exacerbated the masking effect of the SMN compared to the other masks. However, as previously mentioned, both the SNR by sentence interaction and the SNR by mask interaction failed to generalise across participants and items, perhaps because of the small number of participants. Furthermore, the greater difference noted in the two lower SNRs may have been due to a ceiling effect in the -13 dB SNR.

Although the goal of this experiment was to select a SNR that would yield approximately 80% accurate responses across mask conditions, none of the SNRs fulfilled this criterion, since all mask conditions were above 86% accuracy on average. This led to further decreasing the SNRs in steps of -3 dB in Experiment 4b.

4.3.1.2 Experiment 4b: -22, -25, -28 dB SNR

Accuracy ranged from 65% (SMN at -28 dB) to 92% (CT at -22 dB). A summary of the results for Experiment 4b is shown in Figure 4.3, collapsed across sentences. Figure 4.4 shows the detailed breakdown by sentence and mask, for each SNR.

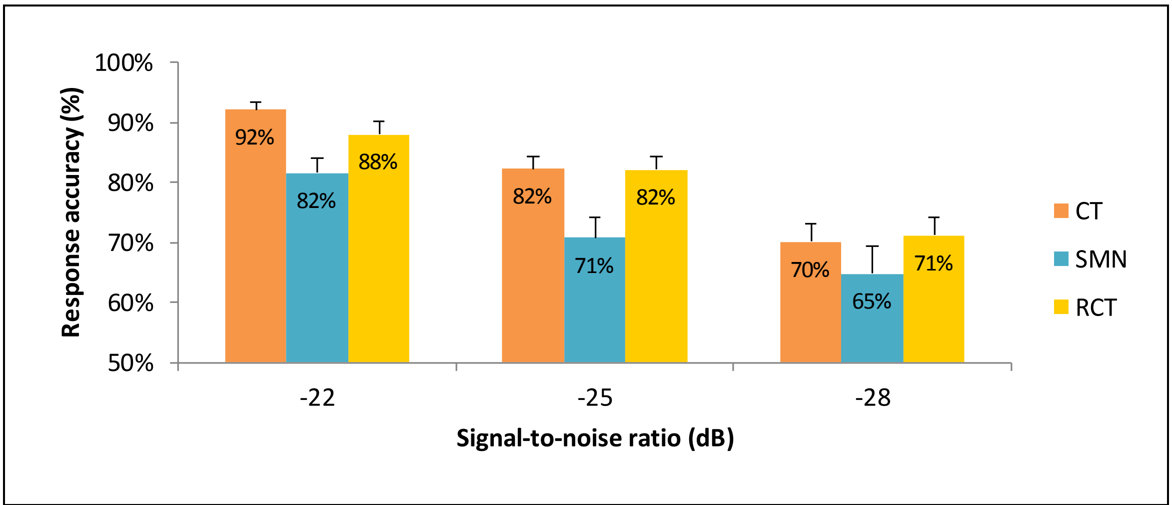


Figure 4.3. Experiment 4b. Response accuracy in percent of correct keywords per sentence for each SNR (-22, -25, -28 dB) and each mask type (CT, SMN, RCT), collapsed across sentence types. Error bars indicate one standard error (across participants).

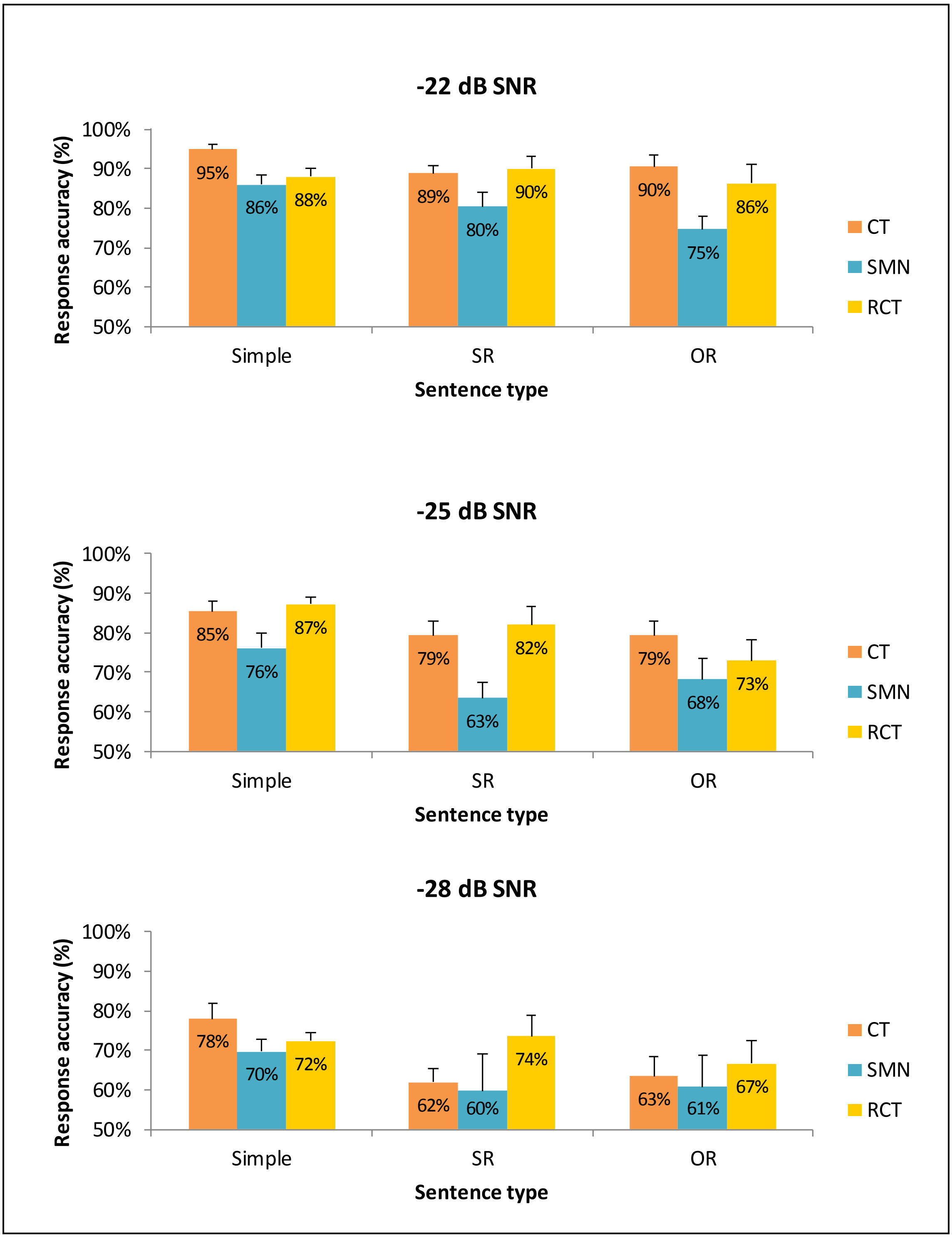


Figure 4.4. Experiment 4b. Response accuracy in percent of correct keywords per sentence for each sentence type (simple, SR, OR) and mask type (CT, SMN, RCT), separated by SNR. Error bars indicate one standard error (by participants).

All descriptive statistics stem from the by-participants analysis. Three-way repeated measures ANOVAs by participants and items were conducted with percent of correctly typed keywords per sentence as a dependent variable, SNR (-22 dB, -25 dB, -28 dB), mask (CT, RCT, SMN), and sentence (simple, SR, OR) as independent variables[13].

There was a main effect of SNR, $F_1(2, 16) = 39.48$, $p < .001$, $\eta_p^2 = .83$, $F_2(2, 480) = 99.90$, $p < .001$, $\eta_p^2 = .29$. All pairwise comparisons were significant at $p < .01$ by items and by participants, with Bonferroni corrections for multiple comparisons. The lowest SNR of -28 dB led to the lowest accuracy (*M* = 69%, *SD* = 11%), followed by -25 dB SNR (*M* = 78%, *SD* = 8%), and the highest SNR of -22 dB led to the highest accuracy (*M* = 87%, *SD* = 6%).

There was a main effect of mask, $F_1(2, 16) = 18.24$, $p < .001$, $\eta_p^2 = .70$, $F_2(1.91, 459.30) = 18.86$, $p < .001$, $\eta_p^2 = .07$. Pairwise comparisons showed that the SMN condition (*M* = 72%, *SD* = 10%) was significantly lower than the CT condition (*M* = 82%, *SD* = 6%), $p = .009$ by participants, $p < .001$ by items. The SMN condition was also significantly lower than the RCT condition (*M* = 80%, *SD* = 10%) $p = .002$ by participants, $p < .001$ by items. The CT and RCT conditions did not differ significantly ($p = 1$ by participants and by items).

There was a main effect of sentence, $F_1(2, 16) = 5.88$, $p = .012$, $\eta_p^2 = .42$, $F_2(2, 240) = 6.58$, $p = .002$, $\eta_p^2 = .05$. Pairwise comparisons with Bonferroni corrections showed that there was a significant difference between the simple sentences (*M* = 82%, *SD* = 5) and the SR sentences (*M* = 75%, *SD* = 8%), $p = .004$ by participants, $p = .032$ by items. The difference between the simple sentences and the OR sentences (*M* = 74%, *SD* = 9%) was significant in the by-items analysis ($p = .003$) but not in the by-participants analysis ($p = .078$). In both cases the simple sentences led to the highest accuracy. The SR sentences were not significantly different to the OR sentences ($p = 1$ by participants and by items). None of the interactions were significant at $\alpha = .05$.

The goal of Experiment 4 was to select a SNR that led to approximately 80% correctly typed keywords. This experiment highlighted the different masking properties of the SMN condition compared to the CT and the RCT conditions. Indeed, the main effect of mask was driven by the difference between the SMN condition and each of the two other masks. Speech-modulated noise appears to be a more effective masker than a competing talker or reversed speech, probably because of its broader frequency spectrum at a given point in time. Since one of the aims of using SMN is to create equivalent energetic masking as the competing

[13] Mauchly's test indicated that the assumption of sphericity had been violated in the by-items analysis for the main effect of mask, $X^2(2) = 11.02$, $p = .004$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

talker, rather than choosing the same SNR for all masks, we chose the SNR that led to the same transcription accuracy across all masks. The SNRs at which accuracy was 82% was chosen: this corresponded to -22 dB for the SMN condition and -25 dB for the CT and RCT conditions. To further ascertain that there was no difference in intelligibility between the masks at these two different SNRs, two-way repeated measures ANOVAs by participants and items were conducted with percent of correctly typed keywords per sentence as a dependent variable, mask (CT, RCT, SMN) and sentence (simple, SR, OR) as independent variables.

There was a main effect of sentence, $F_1(2, 16) = 5.70$, $p = .014$, $\eta_p^2 = .42$, $F_2(2, 240) = 6.15$, $p = .002$, $\eta_p^2 = .05$. Pairwise comparisons revealed no significant differences in the by-participants analysis, and a difference between the simple condition and the OR condition in the by-items analysis ($p = .002$), with the simple sentences leading to higher accuracy than the OR sentences There was no main effect of mask, $F_1(2, 16) = .08$, $p = .92$, $\eta_p^2 = .01$, $F_2(2, 480) = .73$, $p = .90$, $\eta_p^2 = .00$. There was no mask by sentence interaction, $F_1(4, 32) = .91$, $p = .47$, $\eta_p^2 = .10$, $F_2(4, 480) = .73$, $p = .57$, $\eta_p^2 = .01$.

The absence of mask by sentence interaction or main effect of mask justified the use of these two SNRs for Experiment 5, as it indicated that intelligibility was equivalent across SNRs and masks, allowing sentence comprehension to be the focus of the experiment rather than intelligibility alone.

4.3.2 Experiment 5: sentence comprehension and speeded picture-selection task at low SNRs

4.3.2.1 Accuracy

In Experiment 5, sentences were presented at the SNRs chosen in Experiment 4. A striking contrast between the accuracy results in Experiment 5 and the accuracy results in Experiments 2 and 3 (native and non-native listeners at -5 dB SNR) was the difference between the no mask condition and the masked conditions evidenced in Experiment 5. Figure 4.5 summarises the percent of accurate responses by sentence type and mask type for the sentence comprehension task.

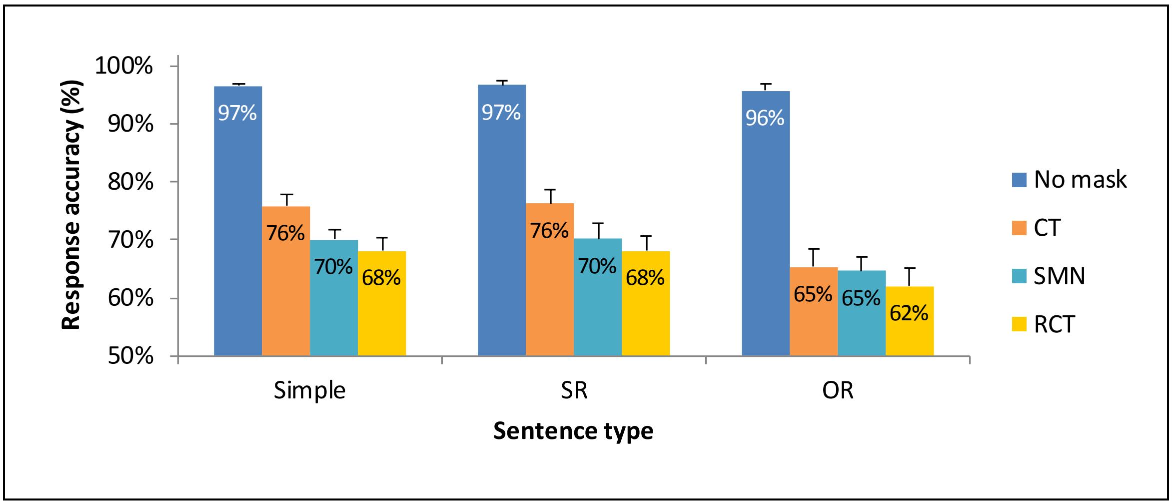


Figure 4.5. Experiment 5. Percent accurate button presses by sentence type (simple, SR, OR) and mask type (no mask, CT, SMN, RCT) in the sentence comprehension task. Error bars represent one standard error (by participants).

Two-way repeated-measures ANOVAs by participants and items were conducted with percent correct responses as a dependent variable, and mask type (no mask, CT, RCT, SMN) and sentence type (simple, SR, OR) as independent variables[14].

There was a significant main effect of mask type, $F_1(2.18, 76.19) = 150.58$, $p < .001$, $\eta_p^2 = .81$, $F_2(2.76, 646.14) = 141.97$, $p < .001$, $\eta_p^2 = .38$. Pairwise comparisons with Bonferroni correction for multiple comparisons revealed that the no mask condition was significantly different to each of the masked conditions at $p < .001$ in both the by-participants and by-items analysis. This reflects higher accuracy in the no mask condition (*M* = 96%, *SD* = 5%) compared to each of the masked conditions, which is not surprising given the low SNRs. Numerically, the CT condition (*M* = 73%, *SD* = 15%) led to higher accuracy than the RCT condition (*M* = 67%, *SD* = 15%) and the SMN (*M* = 69%, *SD* = 14%) condition, indicating that it was actually the least demanding condition. The CT condition was significantly different to the RCT condition, $p < .001$ by participants, $p = .006$ by items. The RCT and SMN conditions did not differ significantly ($p = .947$ by participants and $p = 1$ by items). A significant difference between the CT and the SMN conditions was found in the by-participants analysis ($p = .019$), but did not generalise by items ($p = .118$).

There was a significant main effect of sentence type, $F_1(2, 70) = 9.68$, $p < .001$, $\eta_p^2 = .22$, $F_2(2, 234) = 5.45$, $p = .005$, $\eta_p^2 = .04$. Pairwise comparisons showed that there was a

[14] Mauchly's test indicated that the assumption of sphericity had been violated for the by-participants analysis for the mask by sentence interaction, $X^2(20) = 32.13$, $p = .043$, as well as for the main effect of mask in the by-participants analysis, $X^2(5) = 19.56$, $p = .002$, and in the by-items analysis, $X^2(5) = 30.37$, $p < .001$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

significant difference between the OR sentences and the SR sentences, $p = .005$ by participants, $p = .021$ by items. There was also a difference between the OR sentences and the simple sentences ($p = .002$ by participants, $p = .007$ by items), but not between the simple and the SR sentences ($p = 1$ by participants, $p = 1$ by items). This reflected lower accuracy in the OR sentences ($M = 72\%$, $SD = 15\%$) compared to the SR ($M = 78\%$, $SD = 12\%$) and simple sentences ($M = 78\%$, $SD = 9\%$).

There was no significant mask by sentence interaction, $F_1(4.84, 169.38) = 2.09$, $p = .072$, $F_2(5.52, 646.14) = 1.00$, $p = .416$, $\eta_p^2 = .06$.

The next analysis focused only on the three masker conditions (without the no mask condition) and all sentence types, to further investigate the differences between the types of mask. Two-way repeated-measures ANOVAs by participants and items were conducted with percent correct responses as a dependent variable, and mask type (CT, RCT, SMN) and sentence type (simple, SR, OR) as independent variables[15].

The main effect of mask persisted after taking out the no mask condition from the analysis, $F_1(2, 70) = 12.23$, $p < .001$, $\eta_p^2 = .26$, $F_2(1.93, 453.39) = 6.08$, $p = .003$, $\eta_p^2 = .03$. Pairwise comparisons showed that the CT condition was significantly different to the RCT condition ($p < .001$ by participants, $p = .003$ by items) and to the SMN condition in the by-participants analysis only ($p = .010$ by participants, $p = .055$ by items). However this difference was in the opposite direction to our hypothesis, since the CT condition led to higher accuracy ($M = .72$, $SD = .15$) than the RCT condition ($M = .66$, $SD = .15$) and the SMN condition ($M = .68$, $SD = .14$). There was no significant difference between the RCT and the SMN condition ($p = .473$ by participants, $p = .643$ by items).

There was a main effect of sentence, $F_1(1.17, 60.10) = 9.66$, $p < .001$, $\eta_p^2 = .22$, $F_2(2, 235) = 5.12$, $p = .007$, $\eta_p^2 = .04$. Pairwise comparisons with Bonferroni corrections for multiple comparisons showed the same pattern as in the ANOVA with all masks included. The OR condition was significantly different to the SR condition ($p = .007$ by participants, $p = .025$ by items) and to the simple condition ($p = .002$ by participants, $p = .009$ by items), but the SR and the simple conditions did not differ significantly ($p = 1$ by participants and by items).

[15] Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of sentence in the by-participants analysis, $X^2(2) = 6.12$, $p = .047$. The assumption of sphericity was violated for the main effect of mask in the by-items analysis, $X^2(2) = 8.73$, $p = .013$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

There was no statistically significant mask by sentence interaction, $F_1(4, 140) = .88$, $p = .478$, $\eta_p^2 = .02$, $F_2(4, 470) = .427$, $p = .789$, $\eta_p^2 = .00$.

In conclusion, the accuracy data revealed an effect of sentence type with the OR sentences posing the greatest challenge. Although we had already observed this in the non-native listeners' accuracy data in Experiment 3, this effect had not been observed in the accuracy of Experiment 2 with native listeners, indicating that the decreased SNR may have affected overall processing load. This may have led to an increased cost of processing the more complex sentences. In addition to the burden induced by the decreased SNR, the randomisation of the mask conditions may have increased the difficulty of the task by decreasing predictability and forcing participants to adapt to a different mask more often.

The effect of mask had not been apparent in Experiments 2 (native listeners) and 3 (non-native listeners), however we did find a main effect of mask in this experiment. The biggest difference was between the no mask and the masked conditions. Given that we had chosen the SNRs of the masked conditions corresponding to 82% correctly reported keywords in the intelligibility task of Experiment 4, we expected accuracy to be lower than 82% in the sentence comprehension task for the masked conditions. Indeed, accuracy reached only 70% when averaged across masks, indicating that the task of transcribing a sentence is less demanding than that of processing the sentence, and further justifying the use of a sentence comprehension task as a different measure to a transcription task.

Although there was a substantial difference between the no mask and the masked conditions, the main effect of mask held even when the no mask condition was taken out of the analysis, albeit with relatively small effect sizes. However, this effect was in the opposite direction to our hypothesis, since the competing talker condition was less challenging (higher accuracy) than the reversed competing talker and the speech-modulated noise.

4.3.2.2 Reaction times

The same criteria were used for the reaction times in this experiment as in Experiments 2 and 3 (Chapters 2 and 3). Reaction times included only correct responses and excluded outliers. Outliers were defined as reaction times greater than two standard deviations above the mean across all four masks and all sentence types (excluding familiarisation and filler items) on a subject by subject basis. Figure 4.6 shows the mean reaction times by sentence type and mask type.

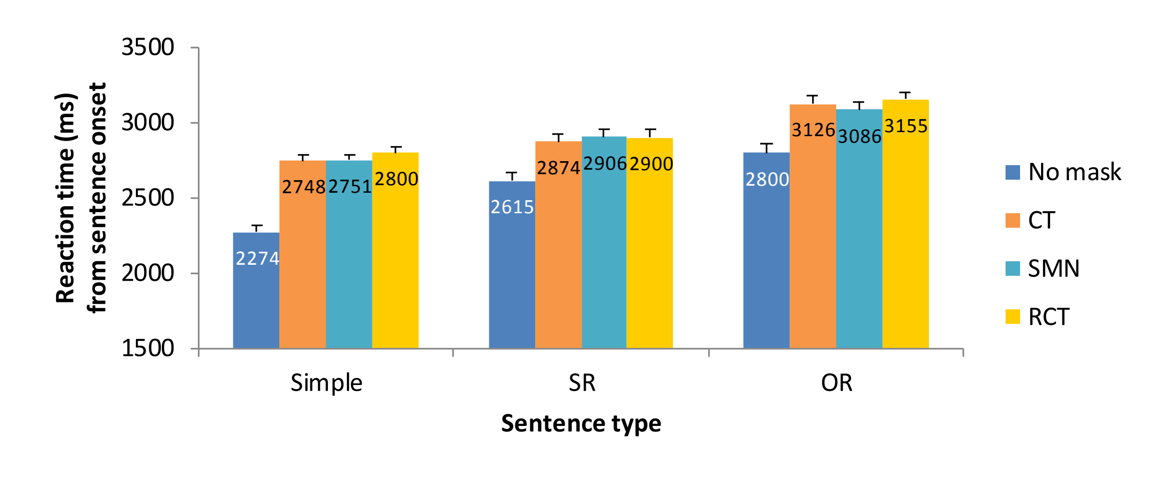


Figure 4.6. Experiment 5. Reaction time (ms) from sentence onset by sentence type (simple, SR, OR) and mask type (no mask, CT, SMN, RCT). Error bars indicate one standard error, by participants.

The same set of analyses was conducted for the RTs as for the accuracy data, to investigate the effects of sentence type and mask on reaction times. Two-way repeated-measures ANOVAs by participants and items were conducted with reaction time as a dependent variable, and mask type (no mask, competing talker, time-reversed speech, speech-modulated noise) and sentence type (simple, subject relative, object relative) as independent variables[16].

There was a main effect of mask, $F_1(2.12, 74.16) = 138.64$, $p < .001$, $\eta_p^2 = .80$, $F_2(2.78, 644.29) = 129.70$, $p < .001$, $\eta_p^2 = .36$. Numerically, the no mask condition was the fastest (M = 2563, SD = 314), followed by the SMN condition (M = 2914, SD = 275), the CT condition (M = 2916, SD = 286) and finally the RCT condition (M = 2952, SD = 304). Pairwise comparisons showed that the no mask condition was significantly different to each of the masked conditions ($p < .001$ for all comparisons by participants and by items), but none of the other comparisons were significant at $\alpha = .05$.

There was a main effect of sentence, $F_1(2, 70) = 222.56$, $p < .001$, $\eta_p^2 = .86$, $F_2(2, 232) = 80.35$, $p < .001$, $\eta_p^2 = .41$. All pairwise comparisons were significant between the sentence types ($p < .001$ by participants and by items). The simple condition (M = 2643, SD = 258) was the fastest, followed by the SR condition (M = 2824, SD = 308) and the OR condition (M = 3042, SD = 318).

[16] Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of mask in the by-participants analysis, $X^2(5) = 19.25$, $p = .002$, and in the by-items analysis, $X^2(5) = 28.89$, $p < .001$. The by-participants analysis also showed that the assumption of sphericity had been violated for the mask by sentence interaction, $X^2(20) = 60.94$, $p < .001$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

There was a statistically significant interaction between the effect of sentence and the effect of mask on reaction times, $F_1(3.6, 126.01) = 6.98$, $p < .001$, $\eta_p^2 = .17$, $F_2(6, 696) = 4.68$, $p < .001$, $\eta_p^2 = .04$. The pattern of results between sentence types was slightly different in the no mask condition compared to the masked conditions. Indeed, the difference was greater between the reaction times in the simple sentences and the reaction times in the SR sentences for the no mask condition compared to the masked conditions, where this difference was not as large. Furthermore, the difference between no mask and the masked conditions was greater in the simple sentences than in the SR and OR sentences. However, the overall pattern remained the same, as is apparent in Figure 4.6.

Focusing only on the masked conditions, two-way repeated measures ANOVAs by participants and by items were conducted with reaction time as a dependent variable, and mask type (CT, RCT, SMN) and sentence type (simple, SR, OR) as independent variables[17].

There was a main effect of mask, $F_1(2, 70) = 3.39$, $p = .039$, $\eta_p^2 = .09$, $F_2(1.92, 447.01) = 3.44$, $p = .035$, $\eta_p^2 = .01$, however none of the pairwise comparisons were statistically significant at $\alpha = .05$.

There was a main effect of sentence, $F_1(2, 70) = 148.01$, $p < .001$, $\eta_p^2 = .81$, $F_2(2, 233) = 40.57$, $p < .001$, $\eta_p^2 = .26$. All pairwise comparisons were significant by items and participants ($p = .017$ between simple and SR, and $p < .001$ for all other comparisons).

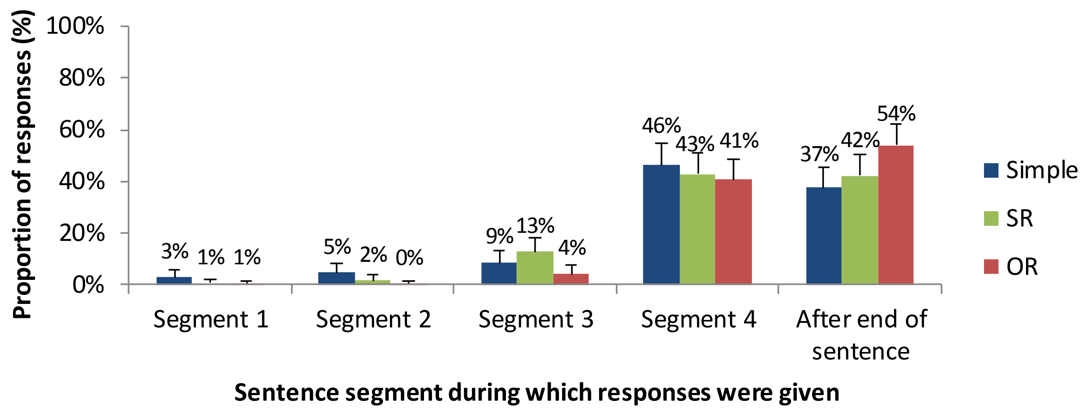
No statistically significant interaction was found, $F_1(2.52, 88.13) = .86$, $p = .448$, $\eta_p^2 = .02$, $F_2(4, 466) = .58$, $p = .676$, $\eta_p^2 = .00$, confirming that the interaction found in the previous analysis was mainly due to the different response pattern in the no mask vs the masked conditions.

In summary, the reaction time analyses showed that the no mask condition was faster than each of the masked conditions, and that reaction times increased as syntactic complexity increased. However, there was still no difference between the masks themselves.

[17] Mauchly's test indicated that the assumption of sphericity had been violated for the mask by sentence interaction in the by participants analysis, $X^2(9) = 38.60$, $p < .001$. The assumption of sphericity was violated for the main effect of mask in the by-items analysis, $X^2(2) = 10.07$, $p = .007$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

4.3.2.3 Button presses for each sentence segment

To gain a better understanding of the time-course of participants' button press responses, Figure 4.7 shows the breakdown of the proportion of responses per sentence segment, like in Experiment 2 (Chapter 2) and Experiment 3 (Chapter 3). Only correct responses were included.



Examples	**Segment 1**	**Segment 2**	**Segment 3**	**Segment 4**
Simple	Show the girl	with the	red	trousers
SR	Show the girl	who	is holding	the boy
OR	Show the girl	who	the boy	is holding

Figure 4.7. Experiment 5. Proportion of responses (%) per segment by sentence type (simple, SR, OR), averaged across all mask conditions (no mask, CT, SMN, RCT). Error bars indicate one standard error (by participants). The bottom part of the figure shows examples of segments for each sentence type.

Responses for all sentence conditions were distributed roughly equally between segment 4 and after the end of the sentence. Most responses for the simple and SR conditions occurred during segment 4 (55% and 56% respectively), whereas most responses for the OR condition occurred after the end of the sentence (54%).

To summarise the accuracy and reaction time results of Experiment 5, the no mask condition led to higher accuracy and faster reaction times than the masked conditions. However, contrary to predictions, the CT condition was not more detrimental than the EM controls. Indeed, accuracy measures revealed that the CT condition was more accurate than the EM controls, although reaction times did not follow this pattern.

When comparing the results of Experiments 2, 3, and 5, it appears that the non-native participants (Experiment 3) were either the most cautious or they needed the most time to process the sentences, waiting until the end of the sentence to give their responses. Participants in Experiment 5 gave a lower proportion of responses after the end of the sentence (44% averaged across sentences) than participants in Experiment 3 (86% averaged across sentences), but a higher proportion than participants in Experiment 2 (16% averaged across sentences). Although in Experiment 5 a large proportion of responses was also given during segment 4 (52%), in Experiment 3 this proportion was much smaller (14%), and in Experiment 2 it was substantially greater (72%). The native participants in Experiment 2 and Experiment 5 showed different patterns of responses, which could reflect several differences between the experiments. The first of these differences could be the uncertainty and difficulty arising in Experiment 5 due to decreased intelligibility compared to Experiment 2. In addition to this, the randomisation of the mask types in Experiment 5 but not in Experiment 2 could also have led to greater uncertainty, prompting participants to delay their responses or to need more time to process the sentences. Finally, the physical set-up of the eye-tracking may have led to a delay in button presses, since participants were asked to keep their heads immobile, leading to a generalised inhibition or delay of movement. This latter explanation seems unlikely, but may have played a part. The eye-tracking data for Experiment 5 should be able to shed light on the online processing of the sentences before the button presses, especially given that around half of participants' responses were after the end of the sentence.

4.3.2.4 Eye-tracking

The same procedure as described in Experiment 3 (Chapter 3) was used in Experiment 5. The proportion of eye fixations per sample falling within each of the three regions of interest was calculated. Like for Experiment 3, only ROI1 and ROI3 were considered. The simple sentences where the target character was in ROI2 were excluded. Only the sentences with correct button presses were included in the analysis. Only 35 participants were included in the eye-tracking analysis, due to a corrupt data file for participant 36.

Table 4.1 shows the point in time at which participants reached their decision (decision point) for each sentence type and mask type. The decision point was defined as the point in the sentence at which the target character was fixated significantly more than the competitor, for at least 20 samples (roughly equivalent to 200 milliseconds). The decision points are also indicated with black crosses in Figure I.1 to Figure I.12 in Appendix I section I.1. The hypothesis driving the inclusion of decision points as a measure of sentence processing was that more demanding conditions would lead to later decision points. This could reflect the higher cognitive load required to process sentences, delaying the point at which a sentence was understood. The average decision point should precede the average reaction time, and follow the same general pattern, if reaction times and eye fixations measure the same aspects of sentence processing.

	No mask	Competing talker	Reversed competing talker	Speech-modulated noise	Average
Simple	150	193	183	199	**181**
Subj Rel	182	170	256	229	**209**
Obj Rel	194	239	216	237	**222**
Average	**175**	**201**	**218**	**222**	

Table 4.1. Experiment 5. Point in time (expressed in samples) at which the target character was fixated significantly more than the competitor for at least 20 samples (corresponding to 200 ms).

Once again, it is not possible to derive inferential statistics from these values given that there are no individual decision points. Based purely on the mask averages shown in Table 4.1, the decision point was earliest for the no mask condition (175), followed by the competing talker condition (201), the reversed competing talker (218) and the speech-modulated noise (222). Looking at the conditions individually, in the simple and object relative sentences the decision points were earliest for the no mask condition (150 and 184, respectively), followed by the reversed competing talker (183 and 216). The competing talker and speech-modulated noise conditions for the simple and object relative sentences had very similar decision points (maximum difference of 6 samples). These data are consistent with the main effect of mask vs no mask observed in the reaction time and accuracy data. However, the earliest decision point in the subject relative sentences was for the competing talker (170), which is somewhat surprising considering that all other measures (accuracy, reaction times, and the overall pattern of eye-fixations reported below) indicate that the no mask condition was significantly less demanding than each of the masked conditions. The second-earliest decision point was for the no mask condition (182), followed by the speech-modulated noise (229), and finally the reversed competing talker.

The sentence condition averages followed the previously reported pattern of syntactic complexity, with the earliest decision moment in the simple sentence condition (181), followed by the subject relative sentence condition (209) and finally the object relative condition (222). In the simple sentence condition, the decision point was during the third segment of the sentence for the no mask condition, and during the first half of the fourth (last) segment of the sentence for the masked conditions. In the subject relative sentence condition, the decision point was at the end of the third segment for the competing talker condition, at the beginning of the fourth (last) segment for the no mask condition, at the end of the fourth (last) segment for the speech-modulated noise condition, and just after the end of the sentence for the reversed competing talker condition. In the object relative sentence condition, the decision point was during the last segment for all masks: at the beginning for the no mask condition, in the middle for the RCT condition, and at the end for the SMN and CT conditions. The decision points are consistent with the general trend in Figure 4.7 which showed that most button presses were made during the last segment of the sentence or after the end of the sentence, for all sentence types.

Figure 4.8 shows the average fixation rate difference between the target and competitor for each mask type (separate lines) and sentence type (separate graphs).

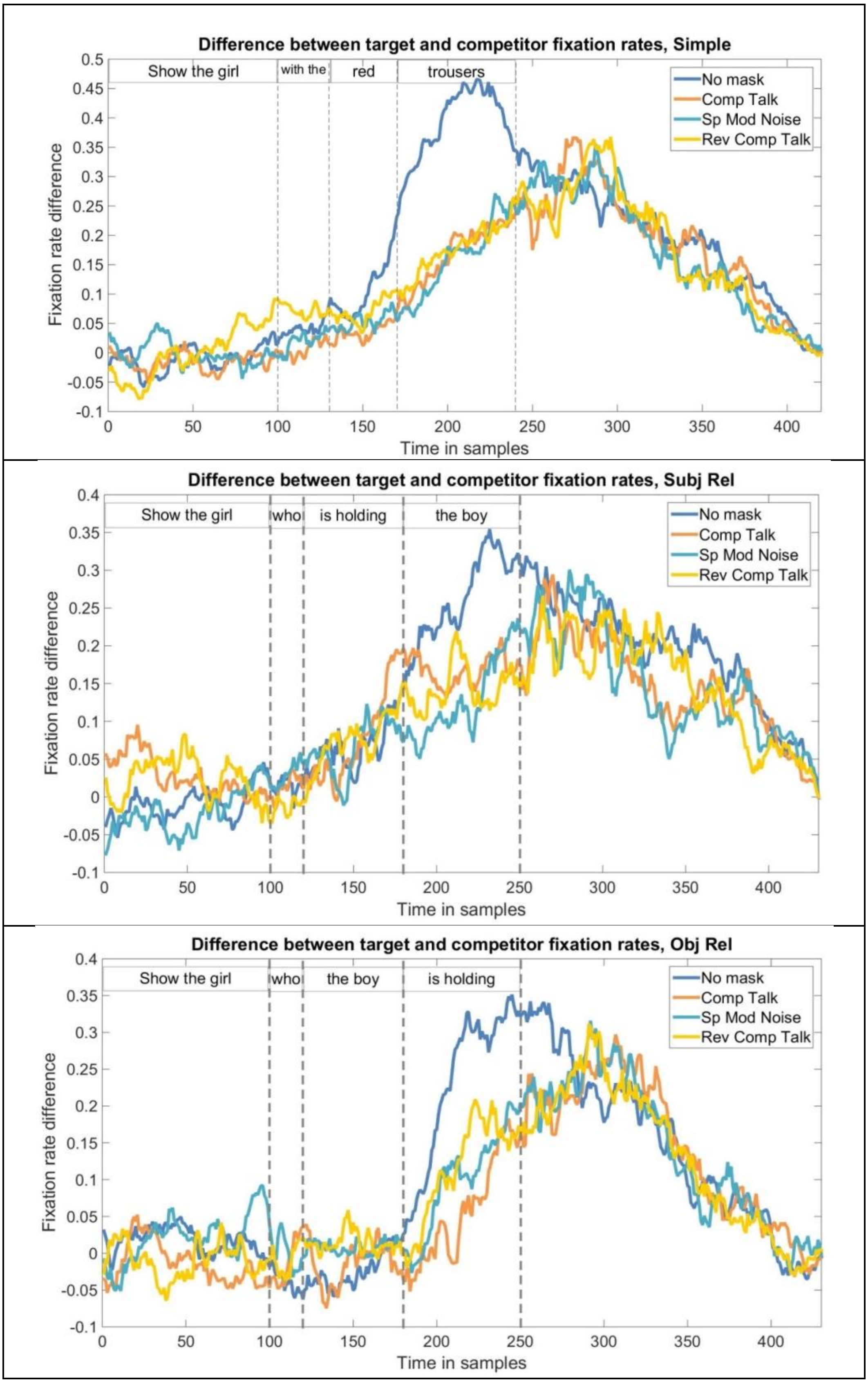


Figure 4.8. Experiment 5. Fixation rate differences between target and competitor characters for each mask condition (no mask, CT, RCT, SMN). The top panel shows the simple sentences, the middle panel the SR sentences, and the bottom panel the OR sentences.

Table 4.1 (decision points) and Figure 4.8 (fixation rate differences) provides complementary information about the time-course and degree of certainty of participants' sentence processing. The decision point can be interpreted as the average point in the sentence when participants had enough information to understand the sentence. The difference between target and competitor fixation rates may indicate the degree of uncertainty in sentence processing: presumably, the lower the fixation rate difference, the higher the uncertainty, since participants' eye gaze may be alternating between potential candidates. Although in the subject relative sentences the decision point was reached earlier for the competing talker condition than for the no mask condition, this does not tell the whole story. Indeed, the fixation rate data indicate that the competing talker condition led to greater uncertainty, which is reflected in the lower fixation rate difference between target and competitor overall compared to the no mask condition.

The main finding relating to the pattern of eye fixations as the sentence unfolds is that the no mask condition led to a significantly different pattern of fixations in the last segment of the sentence, compared to each of the masked conditions. This was confirmed by investigating the difference of the target vs competitor fixation rate differences with confidence intervals, for each mask pair. Figure 4.9 shows an example of the difference between the no mask condition and the CT condition for the simple sentences. In this example, the difference is negative over the course of the last sentence segment, indicating that the no mask condition led to earlier fixations to the target character, as well as higher certainty.

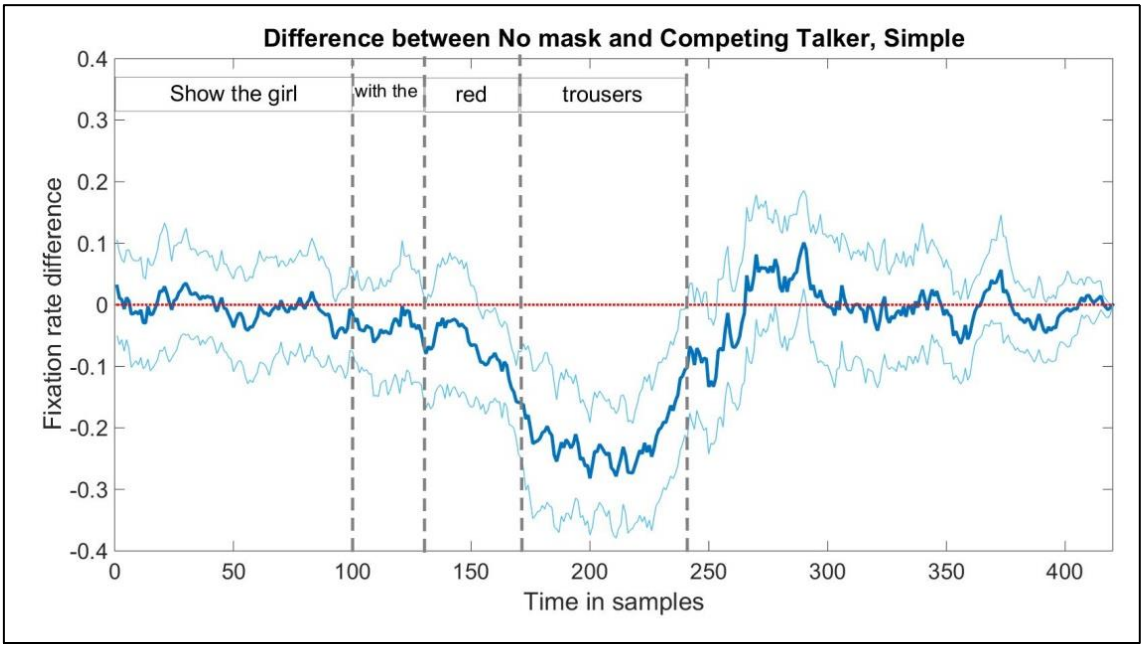


Figure 4.9. Experiment 5. Example of a significant fixation rate difference found during the last segment of the sentence, between the no mask condition and the CT condition for simple sentences.

All pairwise comparisons with 99.17% confidence intervals are reported in Appendix I section I.2 (Figure I.13 to Figure I.15). The difference between the no mask condition and the masked conditions was significantly different to 0 for at least 20 samples in the last segment of the sentence. This reflects the fact that fixation rate to the target character was higher during this segment in the no mask condition compared to the masked conditions, indicating a higher degree of certainty. None of the other pairwise comparisons were significantly different to 0 for at least 20 samples, with 99.17% confidence intervals correcting for the six pairwise comparisons.

4.3.2.5 Cognitive measures

The main goal of including these cognitive tests was to investigate individual differences and their relationship with the sentence comprehension task. Before reporting the correlations between cognitive test results and sentence comprehension performance, the following paragraph describes participants' results in the cognitive tests.

For the visual flanker task, the average difference between the inconsistent and the consistent condition was 52.04 ms (*SD* = 23.79), ranging from 5 ms to 104 ms. For both measures of memory, the mean standard score was around 100, which is expected for standardised scores from the same population. Descriptive statistics for participants' standardised scores in the phonological short-term memory test (non-word repetition subtest of the AWMA) and the verbal working memory test (listening recall subtest of the AWMA) are shown in Table 4.2.

	Range	Mean	Standard deviation
Non-word repetition	68 - 127	93.31	13.08
Listening recall	77 – 129	99.14	15.96
Listening recall processing	77 - 130	100.25	16.53

Table 4.2. Experiment 5. Range, mean and standard deviation for standardised scores in the non-word repetition task and the listening recall task including processing scores.

As in Experiment 2 (Chapter 2) and Experiment 3 (Chapter 3), we investigated individual differences in the detrimental effect of a mask (average of CT, SMN and RCT minus no mask condition), as well as the possible detrimental effect of the competing talker compared to the energetic mask controls (CT minus the average of RCT and SMN), and finally the detrimental effect of processing OR sentences compared to SR sentences (OR minus SR). Table 4.3 summarises Pearson's correlations between each of the cognitive measures and differences in accuracy and reaction times in the sentence comprehension task.

	Non-word repetition	Listening recall	Flanker task difference
Accuracy difference masked – unmasked	.009	-.340	-.051
Accuracy difference CT – (RCT + SMN)	.387	-.115	.012
Accuracy difference OR-SR	-.106	-.006	-.058
Reaction time difference masked – unmasked	-.009	-.072	-.030
Reaction time difference CT – (RCT + SMN)	-.060	.331	.076
Reaction time difference OR – SR	.064	-.168	.259

Table 4.3. Experiment 5. Bivariate correlations between each of the 3 cognitive measures and the difference in response accuracy and reaction times for the sentence comprehension task between masked and unmasked conditions, between the CT condition and energetic mask controls (SMN and RCT), and between OR and SR. No correlations were significant at $\alpha = .0056$ (Bonferroni-corrected).

Contrary to our hypotheses, none of the correlations were significant at $\alpha = .0056$ (Bonferroni-corrected). No conclusions can be drawn from these data with regard to the relationship between susceptibility to informational interference (or indeed masking) and memory and attention capacity.

4.4 Discussion

This chapter aimed to investigate the effect of reducing intelligibility of the target sentences on comprehension in the presence of a competing talker. Before the main sentence comprehension experiment (Experiment 5), a series of SNRs was tested in an intelligibility task to select the SNR at which each mask led to around 80% transcription accuracy (Experiment 4). With those less favourable SNRs, Experiment 4 revealed a main effect of sentence type where the simple sentences were significantly more accurate than each of the relative clause sentences. We did not expect a difference between sentence conditions in the intelligibility task, since participants did not necessarily have to process the sentences to correctly transcribe them. However, the low SNRs may have forced them to conduct some 'guesswork', thereby relying more on the relationship between the picture and the sentence to determine what had been said. This relationship was somewhat different for the simple sentences than for the relative clause sentences. Indeed, to correctly transcribe the simple sentences, hearing even a fragment of the word and visually analysing the picture sufficed to match the words to the characters and their accessories. Furthermore, both the nouns and adjectives used in these sentences were relatively unambiguous once mapped onto the picture. However, in the relative clause sentences, the actions performed by the characters were more ambiguous. For example, hitting, beating, hurting, or harming could all have been used to describe the same action, so if participants only heard a fragment of the verb, there was a higher probability of this being incorrectly transcribed. It is interesting to note however that even if the verb had been misheard, the comprehension task (such as in Experiment 5) would still have been correctly answered, since the agent and the patient of the action could be correctly identified irrespective of the exact verb.

In Experiment 4, the different energetic masking properties of speech-modulated noise compared to time-reversed speech and forward speech led to lower transcription accuracy for SMN compared to the two other masks, at a given SNR. This confirmed the need to include the RCT condition as an additional mask to control for energetic masking of the competing talker. By choosing the SNRs at which transcription accuracy was equal across masks, we ensured that any differences between masks in Experiment 5 would not be due to differences in intelligibility.

In Experiment 5, the main effect of sentence type followed the direction of our hypothesis, with lower accuracy and slower reaction times as syntactic complexity increased. This result was in line with the pattern found in Experiments 2 and 3. The general pattern of button presses for each sentence segment indicated that responses for Experiment 5 were situated after those in Experiment 2 (native listeners at -5 dB SNR) and before those in Experiment 3 (non-native listeners at -5 dB SNR). Participants' responses were spread across the last segment of the sentence and after the end of the sentence in Experiment 5, whereas they were given earlier for Experiment 2 and later for Experiment 3. Although it is not possible to directly compare these experiments due to the different methodologies, this pattern of responses follows the hypothesis that decreasing intelligibility increases overall task difficulty.

The main finding of Experiment 5 was that the no mask condition was the least demanding, as evidenced by higher accuracy, lower reaction times, earlier sentence resolution (average decision point based on eye fixations), and higher certainty (greater difference between target and competitor eye fixations). Studies that have investigated the effect of a competing talker on sentence identification rarely included a no mask condition (e.g. none of those reported in Appendix A). Furthermore, the most common measure in speech in noise experiments is the speech reception threshold, which relies on reporting a signal-to-noise ratio, which would of course be impossible to do for an unmasked target. In addition to this, no similar studies have investigated sentence comprehension using reaction times. It is therefore difficult to compare the effect size found in Experiment 5 of the current thesis to effect sizes of previous research.

Although a main effect of mask vs no mask was found,even at these low SNRs no effect of informational interference from a competing talker was found. In fact, when considering accuracy only, the competing talker condition was more accurate than the energetic mask controls. The fact that this (small) effect disappeared in the reaction time and eye-tracking data could be due to these measures reflecting different processes. Indeed, at such low SNRs, accuracy reflects both processing load and intelligibility, to a greater extent than with higher SNRs. If all three keywords essential to sentence comprehension are unintelligible, then the sentence simply can't be interpreted correctly. In this task, unintelligible keywords would lead to an inaccurate response, which would not then translate to slower reaction times given that these only take into account correct responses. Therefore, if the accuracy difference between masks was driven mainly by differences in intelligibility between the masks, then one could expect this difference to be unique to the accuracy measure. If the accuracy difference between masks was mainly driven by a difference in processing load, then this difference should also be apparent in the reaction time and eye-tracking measures. However, it is surprising that the competing talker was more intelligible than the reversed competing talker and the speech-modulated noise, given that we chose the SNRs to reach similar levels of intelligibility. It is possible to surmise that the energetic masking properties of the three masks played a role in this difference, despite them having been matched in energetic masking and intelligibility. Indeed, the intensity envelope of the RCT mask was reversed, which may have led to greater energetic masking of different segments of the target sentence, and as already mentioned, there is a difference in the frequency spectrum of the speech-modulated noise compared to the competing talker at a given point in time.

There is a caveat to the explanation that accuracy reflects intelligibility more than it does processing load, and that reaction times and eye fixations reflect processing load more than intelligibility. The difference between the no mask and the masked conditions can clearly be considered as driven by the difference in intelligibility, or perceptual load. Yet, this effect was evidenced in all measures, not only in the accuracy measure. Given that masking in general has been linked to individual differences in cognitive factors (Akeroyd, 2008), there might have been a general processing load induced by masking, reflected in the slower reaction times, later decision points and greater uncertainty in eye fixations. However, contrary to our hypothesis the competing talker did not lead to higher processing load than the other masks.

The final hypothesis for this chapter was that measures of memory and attention would be related to informational interference. Although there was no evidence of informational interference in the average accuracy, reaction time and eye fixation data, there were nonetheless individual differences with just under half of all participants showing slower reaction times for the CT condition than for the energetic mask conditions. However, none of the correlations allowed any conclusions to be reached regarding the relationship between susceptibility to informational interference and memory or attention.

In conclusion, this chapter confirmed that the lack of informational interference from a competing talker observed in the previous experiments was not due to the SNR being too high to reveal mask differences. Participants in Experiment 5 were affected by the low SNRs, as evidenced by the main effect of mask vs no mask, however there was still no effect of informational interference from the competing talker. There are several possible explanations to this. The first is that a competing talker does not actually lead to any measurable informational interference. Following this explanation, the studies that have shown a detrimental effect of a competing talker compared to energetic masks (e.g. Brungart, 2001; Trammell & Speaks, 1970) may actually have been capturing a difference at a lower level, in energetic masking properties or segregation difficulties, rather than at a higher cognitive level.

Another explanation is that a competing talker leads to informational interference under certain circumstances only. For example, the relevance of the competing talker utterances to the task or the listener may determine whether it will interfere with the task or not. If listeners know that the competing talker utterances are irrelevant to the task, they may completely block it out, effectively treating it the same way as noise. Indeed, in many of the studies that have found a detrimental effect of a competing talker compared to EM (e.g. Brungart, 2001; Trammell & Speaks, 1970), the competing talker maskers were always viable responses to the target utterances given that CRM sentences were used for both target and competitor utterances. Furthermore, in the classic "identification paradox" (Moray, 1959), listeners seem sensitive to the content of a competing signal. Around 33% of the listeners in Cherry (1953) and Moray (1959) were unable to report the content of a competing message, except when they heard their own name. Perhaps the participants in Experiments 2, 3 and 5 responded to all masks equally simply because the competing talker was systematically irrelevant to the main task. Intuitively, this idea echoes anecdotal accounts of listeners in multi-speaker environments being distracted by a conversation simply because it is relevant or interesting to them at that time. The next chapter will further explore the idea that the relevance of the semantic content of the competing talker determines whether it leads to informational interference.

Chapter 5: Effect of semantic content of the competing talker on informational interference

This chapter will explore the effect of the semantic content of the competing talker on informational interference and sentence comprehension. Experiments 2, 3, and 5 provided evidence that a competing talker does not always give rise to informational interference, at least in the sentence comprehension task developed for this thesis. This corroborates other studies that found no difference between a competing talker masker and matched energetic maskers, such as Bernstein and Grant (2009), Qin and Oxenham (2003), or Dirks and Bower (1969). The latter concluded that “the masking efficiency of speech by competing speech is due to the masking spectrum rather than the semantic properties of the competition”. However, other studies have shown that a competing talker can be more detrimental than equivalent energetic masking in an intelligibility task (Brungart, 2001; Trammell & Speaks, 1970). Several factors could contribute to explaining these apparent contradictions. One of these was mentioned by Trammell and Speaks (1970), who suggested that the meaning of the competing message and the interest that the listener has in that meaning might play a part.

In the final experiment of this thesis, I turned to the meaning of the competing talker sentences to determine the factors giving rise to informational interference. Indeed, I hypothesised that informational interference does not include low-level informational masking (e.g. segregation of two acoustically similar voices), and is not exacerbated by lowering intelligibility. When intelligibility is high (low EM) and when target and competitor are easy to segregate (low low-level IM), a competing talker may affect sentence comprehension due to its ‘attention-grabbing’ properties. For example, the ‘identification paradox’ refers to experiments where about one third of listeners are unaware of the content of competing utterances, except when they hear their own name (Cherry, 1953; Moray, 1959). Presumably, one’s own name is attention-grabbing enough to change the focus of attention to the competing utterance. The idea that the content of the competing utterances has an effect on its distracting properties is quite intuitive. Anecdotally, many people report being able to follow a conversation in the presence of other conversations, and may not be aware of the content of the competing speech, that is until they become relevant to them. Relevance can of course change depending on individuals’ preferences and personal characteristics, as was shown in the own name experiments. The hypothesis that drove Experiment 6 was that a competing talker is only attended to when the content is relevant to the task. In the previous experiments of this thesis, the content of the competing talker was chosen to be unrelated to the target speech, both in its syntactic structure and semantic content. If listeners allocate their attention to the competing stream based on its relevance, then a completely irrelevant competing talker would be unlikely to be attended to, and would not affect target sentence comprehension. By making the content relevant to the target speech, informational interference might arise, thereby affecting target sentence comprehension.

A series of experiments that varied the content of the competing message was described by Iyer, Brungart, and Simpson (2010). In one experiment of particular relevance to this chapter, the authors explored the relative contributions of energetic (specifically spectral overlap) and informational masking (in particular the confusability of the target and the masker) on the multimasker penalty[18]. Participants were asked to report a target sentence from the CRM corpus (Bolia et al., 2000), masked by one or two contextually relevant or irrelevant speech or noise maskers, at SNRs of -8, -4, 0, 4, or 8 dB. The target and masker were always voices of the same gender. The contextually relevant masker consisted of competing CRM sentences. The contextually irrelevant maskers were either speech or noise maskers. The speech maskers were competing sentences that were dissimilar to the CRM sentences both in their semantic content and their syntactic structure (randomly selected segments from 'The Rainbow Passage' (Fairbanks, 1960)), and in some cases a different language (Dutch and Spanish were used in addition to English). The noise maskers were reversed speech (English CRM, English Rainbow Passage, Dutch and Spanish), as well as speech-shaped noise and speech-modulated noise created from the contextually relevant speech maskers. Note that unlike in Experiments 1-5 of this thesis, the noise maskers in Iyer et al. (2010) were not matched to the corresponding speech masker, since they were randomly chosen at each trial. The authors hypothesised that the irrelevant competing speech maskers would increase performance, due to the decreased similarity between the target and the masker. Although the authors do not report statistics comparing each of the mask conditions, the average performance per condition was provided.

Performance was higher for the single-masker conditions compared to the two-masker conditions, and decreased rapidly with the SNR whereas the single-masker condition stayed relatively flat across SNRs. Of greater relevance was the comparison between the different single-masker conditions. Indeed, some of the conditions and the results found by Iyer et al.

[18] The multimasker penalty (e.g. Durlach, 2006) refers to the observation that the detrimental effect of competing speech does not follow a monotonous pattern, but rather increases dramatically between one and two competing talkers, before reaching a plateau at around three competing talkers.

(2010) are similar to those found in Experiments 1, 2, and 3 of the present thesis, in particular the irrelevant competing speech (Rainbow Passage), the speech-modulated noise and the reversed irrelevant competing speech conditions. I will report the results found by Iyer et al. (2010) at -4 dB SNR only, as this is the closest SNR to -5dB SNR in Experiments 2 and 3, but the patterns are comparable across the other SNRs. The irrelevant competing speech (English Rainbow Passage) led to similarly high performance (*M* = 95%, *SD* = 1.12) as the reversed English Rainbow Passage competing speech (*M* = 97%, *SD* = .98) and the speech-modulated noise (*M* = 94%, *SD* = .87). Iyer et al. (2010) conclude that the acoustic characteristics of the irrelevant maskers do not influence performance. However, this conclusion is nuanced by the fact that performance was at ceiling for these conditions. These results echo the null results found in Experiments 2 and 3 of the present thesis, which also compared irrelevant competing speech, reversed competing speech and speech-modulated noise at similar SNRs. The relevant competing speech in Iyer et al. (2010) led to lower performance (*M* = 79%, *SD* = 1.59) than the average of the irrelevant maskers (*M* = 98%, *SD* = .69). Surprisingly, the relevant competing speech led to lower performance than the speech-shaped noise (*M* = 90%, *SD* = 1.16), even though speech-shaped noise is a steady-state masker and as such creates more energetic masking (fewer spectro-temporal gaps) than a competing talker. This study points to the possibility that competing speech may lead to interference due to the relevance of its content to the target speech.

There are several differences between the experiments in this thesis and those reported in Iyer et al. (2010). The first of these is the choice of target sentences. The CRM corpus consists of very structured sentences, with the same word order throughout. To perform well in this task, it is sufficient to identify only the two colour and number keywords, without having to carry out any complex syntactic processing of the sentence. In contrast, the target sentences developed for this thesis, in particular the relative clause sentences, were designed to require syntactic processing of the sentence and could not be solved if participants simply put the keywords together in a random order. The second difference was that the target and masker voices were of the same gender in Iyer et al. (2010), whereas the target was male and the competitor was female in the experiments in this thesis. When target and masker are the same gender, this increases the spectral similarity between the two, thus leading to a greater difficulty in segregating masker and target. Informational interference should arise even when low-level IM is low, i.e. when the target and competitor are easy to segregate.

Experiment 6 of the current thesis aimed to determine the conditions in which competing speech leads to informational interference, and to shed some light on the issue of semantic and syntactic similarity of the masker and target speech vs. acoustic segregation. In addition to the unrelated competing talker HINT sentences presented in Experiments 1 to 5 (re-named 'neutral 1' in this chapter), three new categories of competing talker sentences were created, varying in relevance and similarity to the target sentence. Unlike the HINT sentences, all of these new sentences followed the same syntactic structure as the corresponding target sentence they were paired with. The new competing talker sentences were either unrelated to the target sentences ('neutral 2'), related but providing conflicting information (incongruent), or related and providing information consistent with the target sentence (congruent). If listeners are unaffected by the masker regardless of its content, there should be no difference between conditions. However, if the influence of the masker depends on its relevance to the target, the hypothesis was that the incongruent condition would lead to the slowest reaction times and lowest accuracy. This hypothesis is in line with the results found by Iyer et al. (2010). The hypothesis for the congruent condition was more complex, since it could lead to several outcomes depending on the mechanisms at play. The first possibility is that the similarity of the content could be interfering regardless of the congruence of the message, leading to slower reaction times and lower accuracy in both the congruent and incongruent compared to the neutral conditions. Indeed, the mere presence of semantically related words in the competing talker could attract listeners' attention, regardless of the meaning of the sentence. On the other hand, if participants monitor both competing streams more or less continuously, the similarity of the content and the congruence of the message could lead to a facilitation effect, where the information given in the competing message speeds up the processing of the target message. Finally, there were also several options for the neutral conditions. If syntactic similarity leads to heightened interference, then the neutral 2 condition (same syntax as target sentence) should be more distracting than the neutral 1 condition (different syntax). However, if only the semantic similarity of the competing speech determines interference, then the two neutral conditions should be similar to each other, but faster and more accurate than the incongruent condition.

Measures of selective attention, verbal short-term and working memory were once again administered to assess the relationship between individual differences in these cognitive measures and individual differences in the sentence comprehension task. Although previous research has investigated the relationship between cognition and masking, none of these have compared masks with varying semantic content in relation to cognition. There were several alternatives for this relationship. Participants with higher short-term memory capacity, working memory capacity and/or selective attention might be less prone to interference from the competing talker in any condition where the information is relevant, i.e. congruent and incongruent conditions. Participants with higher memory and attention capacity might be less prone to interference from the incongruent condition specifically. In particular, the inhibitory component of working memory may allow participants with higher WM to inhibit irrelevant information, whereas the short-term memory component of WM may allow participants to hold short segments of the competing speech in mind while they evaluate its relevance.

5.1 Method

5.1.1 Participants

Thirty-six University of York students took part in Experiment 6 for payment or course-credit. There were 10 male and 26 female participants, with an average age of 20;3 years (*SD* = 1;1). All were monolingual speakers of British English, and reported never having experienced hearing difficulties (including tinnitus) or speech-language impairments (including dyslexia). Participants had normal or corrected vision (contact lenses or glasses).

5.1.2 Materials

5.1.2.1 Sentence comprehension task

The target sentence recordings were the same as those used in Experiments 1 to 5, with 165 simple sentences (including 45 fillers), 63 subject relative sentences (including 3 practice sentences), and 63 object relative sentences (with 3 practice sentences).

The competing talker sentences fell into four conditions: neutral 1, neutral 2, incongruent, and congruent. The neutral 1 condition was exactly the same as the competing talker condition in Experiments 1 to 5. This consisted of the same 291 concatenated pairs of Hearing in Noise Test (HINT) sentences (Nilsson et al., 1994) used in Experiments 1 to 5. All of the individual HINT sentences were shorter than the target sentences, which is why they were concatenated in pairs to create longer competing talker utterances. These utterances were unrelated to the corresponding target sentences, both in their syntactic structure and their semantic content. For example, if the target sentence referred to a boy and a man, that sentence was not paired with HINT sentences referring to either a boy or a man.

The three other competing talker conditions were specific to Experiment 6, and were created for this thesis. Each of these conditions was comprised of 165 simple sentences, 63 SR and 63 OR sentences, allowing each competing talker sentence to be matched with one target sentence with the corresponding syntactic structure. The competing talker sentences were designed to be longer than the target sentences, to allow for a lag time of 100 ms (and a variable lead time). To achieve longer sentences while keeping the same syntactic structure, each of these new conditions included the auxiliary "is", followed by an adverb and an adjective at the end of the sentence, e.g. "The fox that the bear is biting is fairly strange" (with the underlined portion indicating the added words).

In the 'neutral 2' condition, the sentence content was unrelated to the associated target sentence. To ensure that the keywords of the target sentences were unrelated to the competing talker sentences, they were classified into categories based on whether the main characters were human or not. Whenever a target sentence was in the 'human' category (e.g. a man and a boy), the associated competing talker sentence contained non-human protagonists (e.g. a fish and a bubble), and vice-versa[19]. None of the nouns, verbs, and adjectives from the target sentences were included in any of the neutral 2 competing talker sentences. Just as for the target sentences, each SR sentence had a corresponding OR structure. Each neutral 2 competing talker SR sentence followed the structure below, where the square brackets denote the subject relative clause:

The *noun1* [that $verb_{aux+gerund}$ the *noun2*] is *adverb adjective*.
e.g. The *koala* that is *chewing* the *leaf* is *unbearably smelly*.

Each neutral 2 competing talker OR sentence followed the structure below, where the square brackets denote the object relative clause:

The *noun2* [that the *noun1* $verb_{aux+gerund}$] is *adverb adjective*.
e.g. The *leaf* that the *koala* is *chewing* is *unbearably smelly*.

Across all neutral 2 relative clause sentences, there were 30 different verbs (occurring one to three times across all sentences), 37 nouns as subjects (one or two occurrences), 43 nouns as objects (one or two occurrences), 38 adverbs (one to three occurrences), and 45 adjectives (one or two occurrences).

[19] The witch was classified as human, and the ghost as non-human, based on their physical appearance.

Each neutral 2 simple sentence followed the structure below, where noun1 was always an animate noun, and noun2 an inanimate noun:

The *noun1* with the *adjective1 noun2* is *adverb adjective2.*
e.g., The *dragon* with the *unique eyes* is *extraordinarily fierce*.

Across all neutral 2 simple sentences, there were 43 animate nouns (noun1) occurring one to five times, 18 adjectives in first position (1 to 11 occurrences), 14 inanimate nouns (5 to 11 occurrences), 57 adverbs (one to three occurrences), and 54 adjectives in second position (one to three occurrences).

In the ‘incongruent’ condition, the sentence content was chosen to contradict the target sentence, excluding the introductory word “Show”. In the subject and object relative conditions, this simply meant that the agent and patient of the sentence were inversed. For example, when the target SR sentence was “Show the dog that is biting the horse”, the incongruent sentence was “The horse that is biting the dog is sometimes violent”. Crucially, what was being described in the incongruent sentence was actually happening in the accompanying picture. In the previous example, the picture depicted a dog that was biting a horse biting another dog. Both target and competing talker sentences could have corresponded to the picture, creating a potential interference.

The simple sentences in the incongruent condition simply used the other character present in the picture, but with the same accessory and adjective. For example, for the target sentence “Show the witch with the small bag”, the accompanying picture showed a witch with a small bag, a man with a small bag, and a witch with a big bag. The corresponding neutral 2 sentence was “The man with the small bag is finally popular”.

In the ‘congruent’ condition, the sentence content was chosen to convey the same information as the target sentence. The only difference between the congruent condition sentences and the target sentences was the addition of “is adverb adjective” at the end, and the exclusion of “Show” at the beginning. For example, for the SR target sentence “Show the cat that is nudging the dog”, the corresponding congruent sentence was “The cat that is nudging the dog is unbearably funny”.

Table 5.1 shows examples of each of the competing talker conditions in Experiment 6 for each sentence type, including descriptive statistics for the length in syllables. For a full list of the stimuli, please refer to Appendix B.

Sentence type	Example target sentence with corresponding picture	Mask type	Example competing talker sentences	Mean syllable length (SD)	Range of syllable length
Simple	Show the witch with the small bag 	**Neutral 1**	The fruit is on the ground. They like orange marmalade.	13 (0.76)	11 -14
		Neutral 2	The dolphin with the original doll is actually exotic.	17 (2.24)	12 - 24
		Incongruent	The man with the small bag is finally popular.	13 (1.75)	10 -19
		Congruent	The witch with the small bag is finally popular.	13 (1.74)	10 - 19
SR	Show the boy who is kicking the girl 	**Neutral 1**	Big dogs can be dangerous. The towel fell on the floor.	13 (0.77)	12 - 14
		Neutral 2	The owl that is hunting the mouse is momentarily distracted.	17 (1.90)	14 - 22
		Incongruent	The girl who is kicking the boy is particularly rude.	15 (2.08)	12 - 20
		Congruent	The boy who is kicking the girl is particularly rude.	15 (2.08)	12 - 20
OR	Show the boy who the girl is kicking 	**Neutral 1**	He is washing his face with soap. The cleaner swept the floor.	13 (0.77)	11 -14
		Neutral 2	The mouse that the owl is hunting is momentarily distracted.	17 (1.89)	14 -22
		Incongruent	The girl who the boy is kicking is probably young.	15 (1.82)	12 - 20
		Congruent	The boy who the girl is kicking is probably young.	15(1.80)	12 – 20

Table 5.1. Experiment 6. Examples of target and competing talker sentence pairings by mask condition (neutral 1, neutral 2, incongruent, congruent) and sentence type (simple, SR, OR), with mean length, standard deviations and ranges of lengths in syllables across all sentences.

The sentences for the three new competing talker conditions were recorded by the same female monolingual native speaker of Standard Southern British English as for Experiments 1 to 5. The procedure for recording these stimuli was the same as for the neutral 1 condition (HINT sentences), except that recording was carried out over three sessions instead of one, using a Sennheiser battery-operated microphone, and the sentences were not concatenated since there was only one competing talker sentence per target sentence. The competing talker sentences were extracted from the stream using Cool Edit Pro (Version 2.0, 2002). One hundred ms of silence was manually inserted at the onset and offset of each sentence, and the root mean square level was normalised across sound files to an intensity of 68 dB, using Praat. As before, the average intensity of the target sentences was normalised to 63 dB, yielding a signal-to-noise ratio of -5 dB. The target sentences were then combined with the corresponding competing talker sentence in each of the conditions. Once again the alignment of the target sentences with the corresponding competing talker sentences was carried out from the end of the sound files, with the competing talker sentences ending 100 ms after the offset of the target sentence sound file. Table 5.2 shows the mean lengths, standard deviations and ranges in ms for each competing talker condition and sentence condition.

	Simple		**SR**		**OR**		**Total**	
	Mean (SD)	**Range**	**Mean (SD)**	**Range**	**Mean (SD)**	**Range**	**Mean (SD)**	**Range**
Neutral 1	3231 (260)	2730 – 3716	3397 (184)	3114 – 3781	3402 (181)	3126 – 3805	**3315 (239)**	**2730 – 3805**
Neutral 2	3461 (279)	2936 – 4539	3494 (263)	2838 – 4316	3511 (258)	2963 – 4151	**3482 (270)**	**2838 – 4539**
Incongruent	3076 (222)	2661 – 3821	3282 (243)	2862 – 3947	3272 (225)	2861 – 3777	**3177 (248)**	**2661 – 3947**
Congruent	3069 (221)	2676 – 3712	3289 (255)	2750 – 3988	3238 (226)	2824 – 3694	**3166 (251)**	**2676 – 3988**
Total	**3209 (253)**	**2661 – 4539**	**3365 (236)**	**2750 – 4316**	**3356 (222)**	**2824 – 4151**		

Table 5.2. Experiment 6. Mean lengths in ms (with standard deviations) and ranges of each of the sentence conditions, by mask type (neutral 1, neutral 2, incongruent, congruent) and sentence type (simple, SR, OR).

The lead times varied across sentences, so that participants would not be able to rely on the lead time as a cue. Average lead times with standard deviations and ranges are shown in Table 5.3. Given that the neutral 2 condition contained longer sentences on average, this condition was also the one with the longest lead times.

		Sentence							
		Simple		SR		OR		Total	
		Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (*SD*)	Range	Mean (SD)	Range
Mask	Neutral 1	553 (276)	7 – 1233	613 (229)	119 - 1313	529 (171)	261 - 1117	562 (243)	7 – 1313
	Neutral 2	783 (322)	66 – 1936	714 (282)	82 – 1532	638 (247)	251 – 1307	730 (300)	66 – 1936
	Incongruent	398 (232)	3 – 1086	506 (273)	77 – 1537	399 (211)	52 – 945	425 (241)	3 – 1537
	Congruent	391 (222)	38 – 965	513 (284)	10 – 1578	365 (200)	15 – 818	415 (239)	10 - 1578

Table 5.3. Experiment 6. Mean (with standard deviation) and range of lead times in ms between target sentence and corresponding competing talker sentence, by mask type (neutral 1, neutral 2, incongruent, congruent) and sentence type (simple, SR, OR).

5.1.2.2 Visual flanker task

The same flanker task was used as for Experiments 2, 3, and 5. Each stimulus was composed of a row of five arrows or diamonds. Participants' task was to indicate the direction of a central arrow flanked by distracters. In the consistent condition, the central arrow was flanked by arrows pointing in the same direction. In the inconsistent condition, the central arrow was flanked by arrows pointing in the opposite direction. In the neutral condition, the central arrow was flanked by diamonds. The reaction time difference between the inconsistent and the consistent conditions was used as a measure of susceptibility to visual interference (or visual selective attention).

5.1.2.3 Short-term and working memory tasks

A non-word repetition task and a forward digit span task were administered as measures of verbal short-term memory, and a backward digit span was administered as a measure of verbal working memory. The non-word repetition task used the same stimuli as in Experiments 2 and 5, but it was presented using DMDX. The stimuli consisted of one-syllable non-words in six blocks containing six strings of non-words each, increasing in length.

The forward and backward digit spans consisted of recordings of digits one to nine, spoken by a female speaker of Standard Southern British English (different to the female competing talker in the sentence comprehension task). Participants heard a string of digits and were asked to repeat the digits in the same order (forward digit span) or in reverse order (backward digit span). The string length increased from three to ten digits, with blocks of three trials for each length.

5.2 Design and Procedure

Participants were tested individually in a quiet room with blackout curtains to ensure that the lighting conditions were the same across participants. Each session lasted approximately one and a half hours. The sentence comprehension task was administered first (approximately one hour), followed by the visual flanker task, the non-word repetition task, the forward digit span and finally the backward digit span.

5.2.1 Sentence comprehension task with eye-tracking

Participants were seated approximately 60 cm from the screen, and their eye-movements were monitored using a SR Research Eyelink 1000 Plus desk-mounted camera at a sampling rate of 1000 Hz, with a chin-rest to minimise head movements. Viewing was binocular, but only the dominant eye was tracked[20]. The chin-rest and desk were adjusted to a comfortable height, and participants were instructed to keep their head still and their hand on the keyboard. After adjusting the focus of the camera, and ensuring the tracker parameters were acceptable (pupil and corneal reflection thresholds), a 9-point calibration was performed by the experimenter, and participants carried out seven practice trials including all mask conditions, to familiarise themselves with the task. Following the practice trials, participants' queries were answered and their position readjusted if necessary. Another 9-point calibration was carried out at the beginning of the main task, and after any head movements or breaks. Participants were encouraged to take breaks whenever they needed, to avoid lapses in attention. As in Experiments 2, 3, and 5, the picture was first presented for 1 second before the auditory stimulus was played. All auditory stimuli were presented via Sony MDR V700 headphones, on a Dell PC with a Samsung monitor with 1680 x 1050 resolution. Although the monitor was different, the size of the pictures displayed on the screen was the same as for Experiments 2, 3, and 5.

5.2.2 Visual flanker task

The visual flanker task was the same as in Experiments 2, 3, and 5. It was administered via DMDX (Forster & Forster, 2003) on a Lenovo L540 laptop, with a screen resolution of 1920 x 1080 pixels. Participants were shown a print-out of all the stimuli, and were verbally instructed to press on either the left or the right shift key of the keyboard depending on the direction of the middle arrow, as quickly and accurately as they could. After a series of 13

[20] For one participant with glasses, the non-dominant eye was tracked due to excessive glare.

practice items containing all conditions (consistent, inconsistent and neutral), participants could ask for clarification, and then moved on to the 120 experimental items. Expected response side (left or right) and condition (consistent, inconsistent and neutral) were randomised to avoid the use of strategies.

5.2.3 Short-term and working memory tasks

All three memory tasks were administered using DMDX and manually scored during test administration. Testing stopped when two strings in a block were incorrectly recalled. For the non-word repetition task, the final score was the non-word repetition span, corresponding to the length of non-words in each string of a correctly recalled block. To count as a correctly recalled block, at least two of the strings in that block had to be repeated correctly. The resulting non-word span was more intuitively comparable to the forward and backward digit spans than a standardised score (which was used in the previous experiments).

In the forward and backward digit span tasks, a string of digits was scored as correct if all of the digits in the string were repeated in the expected order. Two out of three strings of the same length had to be correct for the block to be counted as correct. The digit span was defined as the number of digits in the highest correct block.

5.3 Results

5.3.1 Sentence comprehension and speeded picture-selection task

5.3.1.1 Accuracy

The proportion of correct button presses was calculated for each condition. Response accuracy was high across all conditions, ranging from 91% to 97%. Figure 5.1 summarises the percent of accurate responses by sentence type (simple, SR, OR) and mask type (neutral 1, neutral 2, incongruent, congruent) for the sentence comprehension task.

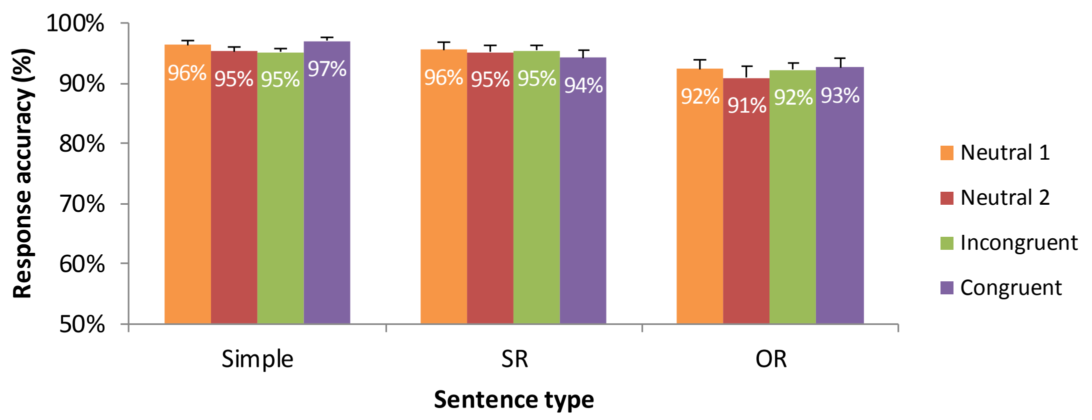


Figure 5.1. Experiment 6. Percent accurate button presses by sentence type (simple, SR, OR) and mask type (neutral 1, neutral 2, incongruent, congruent) in the sentence comprehension task. Error bars represent one standard error (by participants).

A two-way repeated-measures ANOVA by participants and items was conducted with percent correct responses as a dependent variable, and mask type (neutral 1, neutral 2, incongruent, congruent) and sentence type (simple, subject relative, object relative) as independent variables[21].

There was a main effect of sentence, $F_1(2, 70) = 20.30$, $p < .001$, $\eta_p^2 = .37$, $F_2(2, 236) = 12.65$, $p < .001$, $\eta_p^2 = .10$. Pairwise comparisons with Bonferroni corrections for multiple comparisons showed that the OR condition was significantly different from the simple condition ($p < .001$ by participants and by items) and from the SR condition ($p < .001$ by participants, $p = .006$ by items). The OR condition was the least accurate (M = 92%, SD = 9%), followed by the SR condition (M = 95%, SD = 7%) and the simple condition (M = 96%, SD = 4%).

There was no main effect of mask, $F_1(2.43, 85.15) = .39$, $p = .72$, $\eta_p^2 = .01$, $F_2(3, 708) = .75$, $p = .52$, $\eta_p^2 = .00$, and no mask by sentence interaction, $F_1(4.19, 146.68) = .65$, $p = .64$, $\eta_p^2 = .02$, $F_2(6, 708) = .80$, $p = .57$, $\eta_p^2 = .01$.

These accuracy data revealed once again the main effect of sentence type in the direction of our hypothesis, but revealed no differences between mask types. However, because accuracy was nonetheless very high, this may have led to a near-ceiling affect, thus not reflecting the actual processing cost of the task. Reaction time data on the other hand are

[21] Mauchly's test indicated that the assumption of sphericity had been violated in the by-participants analysis for the mask by sentence interaction, $X^2(20) = 47.11$, $p = .001$ and the main effect of mask, $X^2(5) = 18.25$, $p = .003$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

more sensitive to processing demands, therefore if a difference in processing cost between masks exists, it should be revealed in the reaction time data.

5.3.1.2 Reaction times

The same criteria were used for the reaction times in this experiment as in Experiments 2, 3, and 4. Reaction times included only correct responses and excluded outliers. Outliers were defined as reaction times greater than two standard deviations above the mean across all four masks and all sentence types (excluding familiarisation and filler items) on a subject by subject basis. Figure 5.2 shows the mean reaction times by sentence type and mask type.

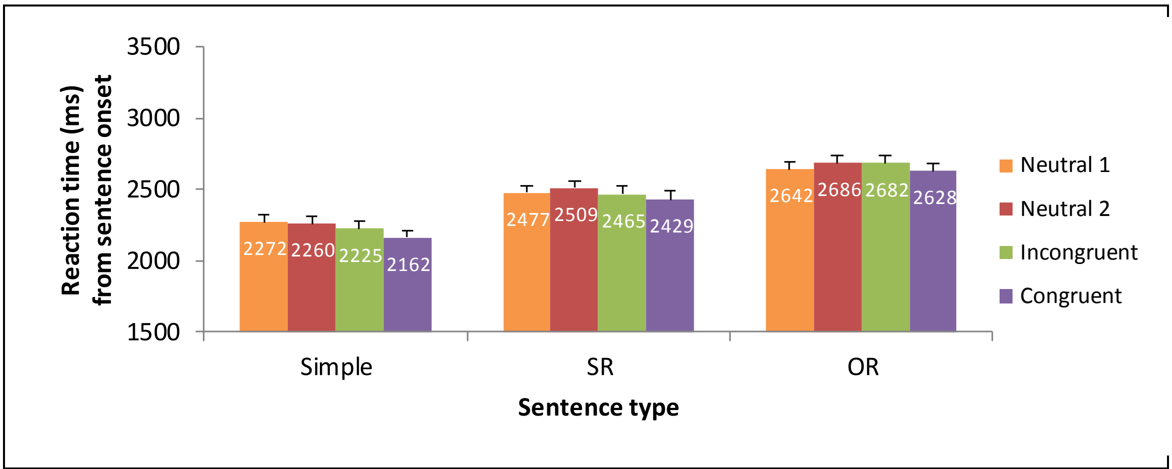


Figure 5.2. Experiment 6. Reaction time (ms) from sentence onset by sentence type (Simple, SR, OR) and mask type (neutral 1, neutral 2, incongruent, congruent). Error bars indicate one standard error, by participants.

Two-way repeated-measures ANOVAs by participants and items were conducted with reaction time as the dependent variable and mask type (neutral 1, neutral 2, incongruent, congruent) and sentence type (simple, subject relative, object relative) as independent variables[22].

There was a main effect of mask, $F_1(3, 105) = 11.30$, $p < .001$, $\eta_p^2 = .24$, $F_2(2.63, 621 .09) = 8.23$, $p < .001$, $\eta_p^2 = .03$. Pairwise comparisons with Bonferroni corrections for multiple comparisons showed that the congruent condition was significantly different to the neutral 1 condition ($p = .001$ by participants, $p = .004$ by items), to the neutral 2 condition ($p < .001$ by

[22] Mauchly's test indicated that the assumption of sphericity had been violated in the by-participants analysis for the mask by sentence interaction, $X^2(20) = 34.00$, $p = .027$ and the main effect of sentence, $X^2(2) = 8.98$, $p = .011$. In the by-items analysis the assumption of sphericity was violated for the main effect of mask, $X^2(5) = 66.55$, $p < .001$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

participants and by items) and to the incongruent condition (*p* = .007 by participants, *p* = .002 by items). The congruent condition was the fastest (*M* = 2406, *SD* = 325), followed by the incongruent condition (*M* = 2457, *SD* = 337), the neutral 1 condition (*M* = 2464, *SD* = 297), and the neutral 2 condition (*M* = 2485, *SD* = 301).

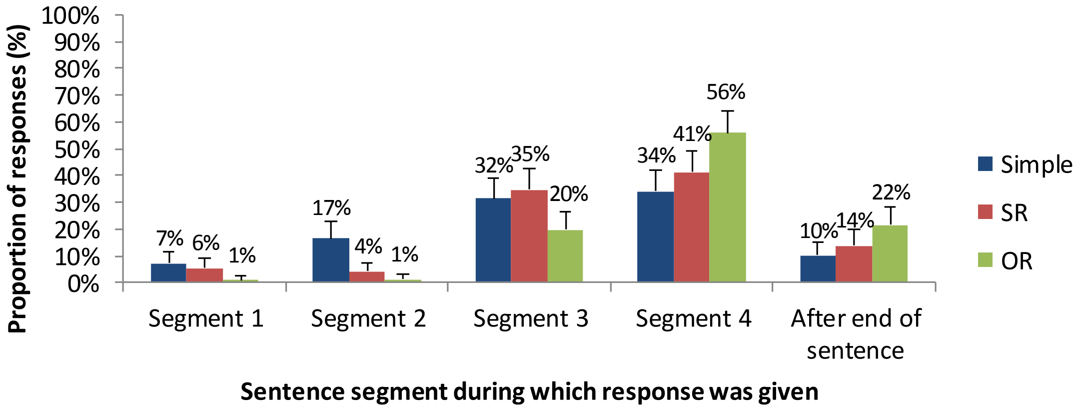
There was a main effect of sentence, $F_1(1.62, 56.81) = 235.95$, $p < .001$, $\eta_p^2 = .87$, $F_2(2, 236) = 218.94$, $p < .001$, $\eta_p^2 = .65$. All pairwise comparisons with Bonferroni corrections were significant (p < .001 by participants and items). The simple condition was the fastest (*M* = 2230, *SD* = 295), followed by the SR condition (*M* = 2470, *SD* = 329) and finally the OR condition (*M* = 2659, *SD* = 320).

There was no mask by sentence interaction, $F_1(4.59, 160.61) = 1.89$, $p = .11$, $\eta_p^2 = .05$, $F_2(5.26, 621.09) = 1.27$, $p = .27$, $\eta_p^2 = .01$.

To summarise the reaction time data, we found an effect of sentence type in the direction of the hypothesis that the simple sentences would be the fastest, followed by the SR sentences, and finally the OR sentences. The main effect of mask indicates that participants are sensitive to the semantic content of the competing talker utterances. However, the results did not follow the hypothesis that the incongruent condition would be the most challenging. Rather, the congruent condition was the fastest condition, and reaction times did not differ significantly between the other three masks.

5.3.1.2.1 Button presses for each sentence segment

In addition to the eye-tracking analysis, which allows a more fine-grained analysis of the time-course of sentence processing through eye fixations, I analysed the time course of the button presses. Figure 5.3 shows the breakdown of the proportion of responses per sentence segment concatenated across masks, as in Experiment 2, 3 and 4. Only correct responses were included in this analysis.



Examples	**Segment 1**	**Segment 2**	**Segment 3**	**Segment 4**
Simple	Show the girl	with the	red	trousers
SR	Show the girl	who	is holding	the boy
OR	Show the girl	who	the boy	is holding

Figure 5.3. Experiment 6. Proportion of responses (%) per segment by sentence type (simple, SR, OR), averaged across all mask conditions (neutral 1, neutral 2, incongruent, congruent). Error bars indicate one standard error (by participants). The bottom part of the figure shows examples of segm ents for each sentence type.

Responses were mostly produced during segments 3 and 4. However, there was a surprising proportion of responses given during segments 1 and 2, especially for the simple sentences (7% and 17% respectively). This was not expected, since there is not enough information during these segments to correctly identify the target character. The most likely explanation is that these answers reflect random responding, since they had one out of three chances of responding correctly. Another explanation could be that the congruent condition provided information that may have allowed participants to anticipate the correct answer. However, the breakdown by mask type showed that the congruent condition followed the same overall pattern: in the simple condition 3% and 14% of the answers were given during the first and second segments, respectively. Therefore this early button press pattern was not unique to the congruent condition.

5.3.1.2.2 Reaction times across the experiment: did participants improve over time?

To determine whether the mask effect found differed over the course of the experiment, items were divided into four equal categories depending on when they had been presented to the participant. Figure 5.4 shows reaction times broken down by presentation blocks (1^{st} quarter, 2^{nd} quarter, 3^{rd} quarter, final quarter) and averaged across all mask types and sentence types, Figure 5.5 shows reaction times broken down by presentation blocks (1^{st} quarter, 2^{nd} quarter, 3^{rd} quarter, final quarter) and mask type, and Figure 5.6 shows reaction times broken down by presentation blocks (1st quarter, 2nd quarter, 3rd quarter, final quarter) and sentence type.

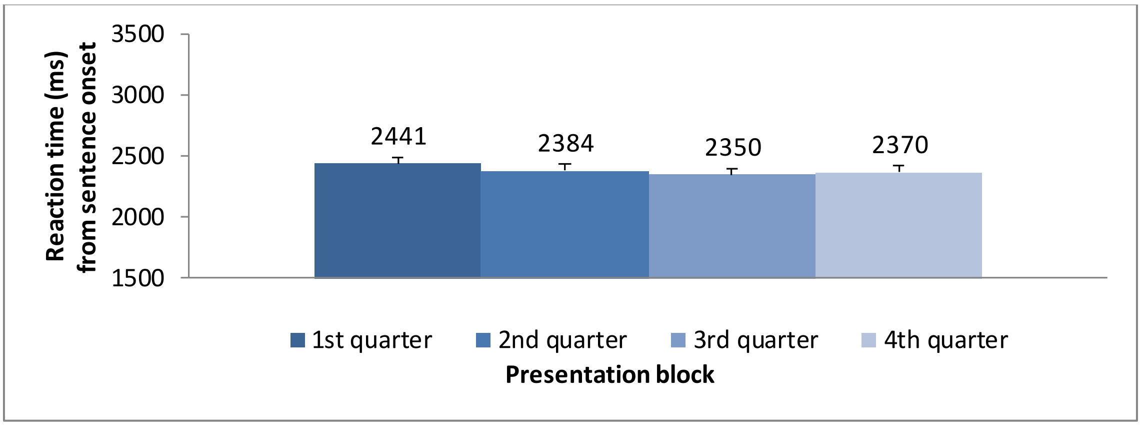


Figure 5.4. Experiment 6. Reaction times separated by block of presentation, averaged across masks and sentences. Error bars indicate one standard error above the mean, by participants.

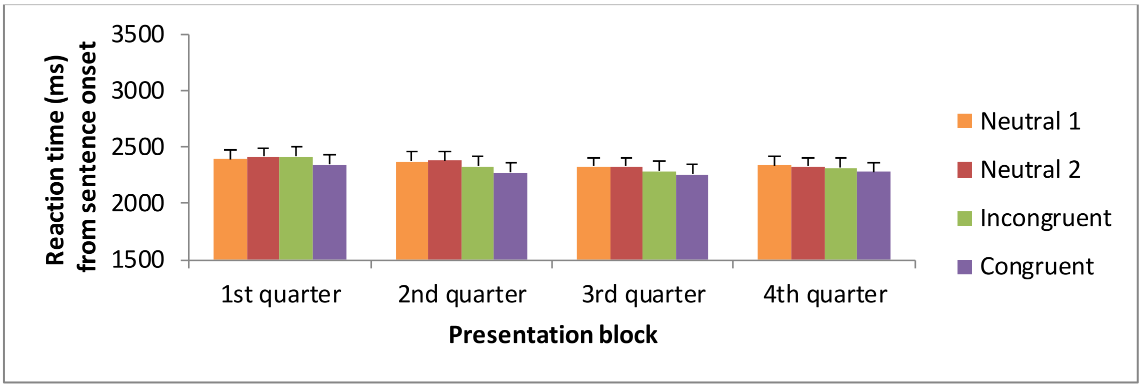


Figure 5.5. Experiment 6. Reaction times separated by block of presentation for each mask type (neutral 1, neutral 2, incongruent, congruent), averaged across all sentence types. Error bars indicate one standard error above the mean, by participants.

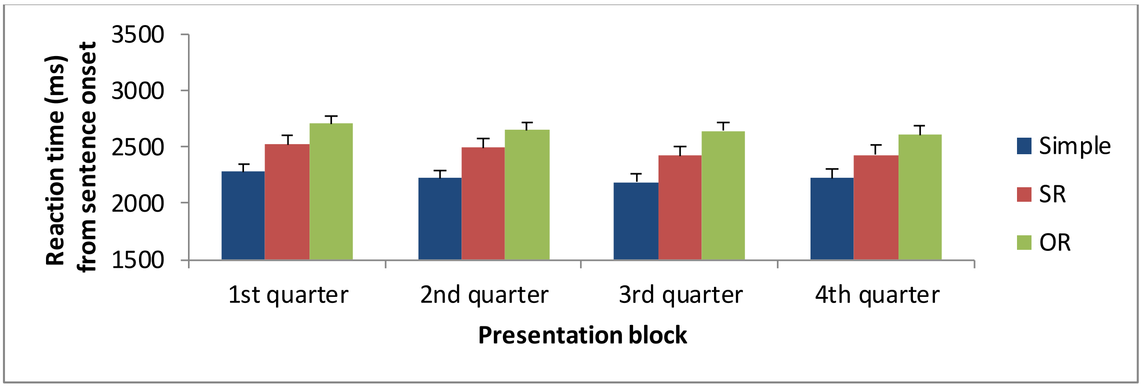


Figure 5.6. Experiment 6. Reaction times separated by block of presentation for each sentence type (simple, SR, OR), averaged across all mask types. Error bars indicate one standard error above the mean, by participants.

A two-way repeated measures ANOVA by participants was conducted, with reaction times as an independent variable, and presentation block (1st quarter, 2nd quarter, 3rd quarter and final quarter), and mask type (neutral 1, neutral 2, incongruent, congruent) as dependent variables[23].

There was a main effect of presentation block, $F(2.17, 75.95) = 5.84$, $p = .003$, $\eta_p^2 = .14$. Pairwise comparisons with Bonferroni corrections showed that the first quarter was significantly different to the second quarter ($p = .046$), and to the third quarter ($p = .012$), with the first quarter ($M = 2441$, $SD = 296$) being slower than the second ($M = 2384$, $SD = 327$) and the third quarter ($M = 2350$, $SD = 324$).

Unsurprisingly, the main effect of mask was once again apparent, $F(3, 105) = 19.59$, $p < .001$, $\eta_p^2 = .36$. Pairwise comparisons with Bonferroni corrections confirmed the previous findings, with the congruent condition significantly faster than each of the other masks ($p < .001$).

Crucially, there was no block by mask interaction, $F(9, 315) = 1.36$, $p = .21$, $\eta_p^2 = .04$. This confirmed that, although participants did get faster over blocks 2 and 3, the improvement was independent of mask type.

To determine whether participants' responses differed over time depending on the sentence type, a second two-way repeated measures ANOVA by participants[24] was conducted with reaction times as a dependent variable, and presentation block and mask type as independent variables.[25]

There was once again a main effect of block, $F(1.87, 105) = 6.17$, $p = .004$, $\eta_p^2 = .15$. This time the pairwise comparisons with Bonferroni corrections only revealed a significant difference between the first quarter and the third quarter, $p = .011$. There was also a main

---

[23] Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of block, $X^2(5) = 19.11$, $p = .002$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

[24] It was not possible to conduct a three-way repeated measures ANOVA, nor to conduct by-items analyses due to the randomisation of sentence and mask conditions. Indeed, each block contains different conditions for each participant, which would lead to many empty cells in a three-way ANOVA.

[25] Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of block, $X^2(5) = 29.59$, $p < .001$, the main effect of sentence, $X^2(2) = 7.50$, $p = .023$, and the block by sentence interaction, $X^2(20) = 33.06$, $p = .034$. Degrees of freedom were corrected with Greenhouse-Geisser estimates.

effect of sentence, $F(1.67, 58.43) = 247.48$, $p < .001$, $\eta_p^2 = .87$, with all pairwise comparisons significant at $p < .001$ with Bonferroni corrections for multiple comparisons.

There was no block by sentence interaction, $F(4.44, 155.31) = 1.45$, $p = .22$, $\eta_p^2 = .04$, indicating that the main effect of block type did not depend on the type of sentence.

Together, these analyses show that participants were slightly faster in the second and third quarters of the experiment, indicating that they had become accustomed to the task. However, this was not modulated either by the type of mask or the type of sentence.

5.3.1.3 Eye-tracking

The same procedure and analysis as in Experiments 3 and 5 (Chapters 3 and 4) was carried out for the eye-tracking data in Experiment 6. The proportion of eye fixations per sample falling within the regions of interest with the target and competitor characters (ROI1 and ROI3) was calculated. For the simple sentences, the items where the target character was in ROI2 were excluded. Only the sentences with correct button presses were included in the analysis.

Table 5.4 shows the decision points for each sentence type and mask type. The decision point corresponds to the point in the sentence when the target character was fixated significantly more than the competitor for at least 20 samples. These decision points are indicated with black crosses in Appendix L, section L.1, Figure L.1 to Figure L.12.

	Neutral 1	**Neutral 2**	**Incongruent**	**Congruent**	**AVERAGE**
Simple	146	141	112	85[26]	**121**
Subj Rel	155	169	166	151	**160**
Obj Rel	186	191	199	190	**192**
AVERAGE	**162**	**167**	**159**	**142**	

Table 5.4. Experiment 6. Point in time (expressed in samples) at which the target character was fixated significantly more than the competitor for at least 20 samples (corresponding to 200 ms).

The decision points were based on the results of the permutation tests comparing fixations to the target and competitor characters. It is not statistically possible to compare these decision points, since they were not obtained for each participant (standard deviations are therefore not possible either). Thus, although useful from a descriptive point of view, they cannot be

[26] For the congruent simple sentence condition, $p < .05$ from the 85th sample (end of the first segment of the sentence), but at the 101st sample (very beginning of the second segment), $p = .051$ and $t = 3.666$ (critical $t = 3.675$) for just one sample, before going back to $p < .05$ from the 102nd sample.

used to reach firm conclusions from an inferential point of view. The congruent condition led to the earliest average decision point, which is consistent with the reaction time data reported in section 5.3.1.2. However, this difference was only apparent in the simple sentence condition. The decision point for the simple condition with the congruent mask took place at the end of the first segment, which is surprisingly early. However, by this point, participants had enough information to determine whether the competing talker was providing the same information as the target talker, and thus rely on the information from the competing talker to make their decision. The decision point for the simple condition with the incongruent mask took place during the second segment, and during the third segment for the two neutral mask conditions. However, for the SR sentence condition, all decision moments took place during the third segment, and for the OR sentence condition all decision moments took place during the fourth (final) sentence segment.

Figure 5.7 shows the average fixation rate difference between the target and competitor characters for each mask type (separate lines) and sentence type (separate graphs).

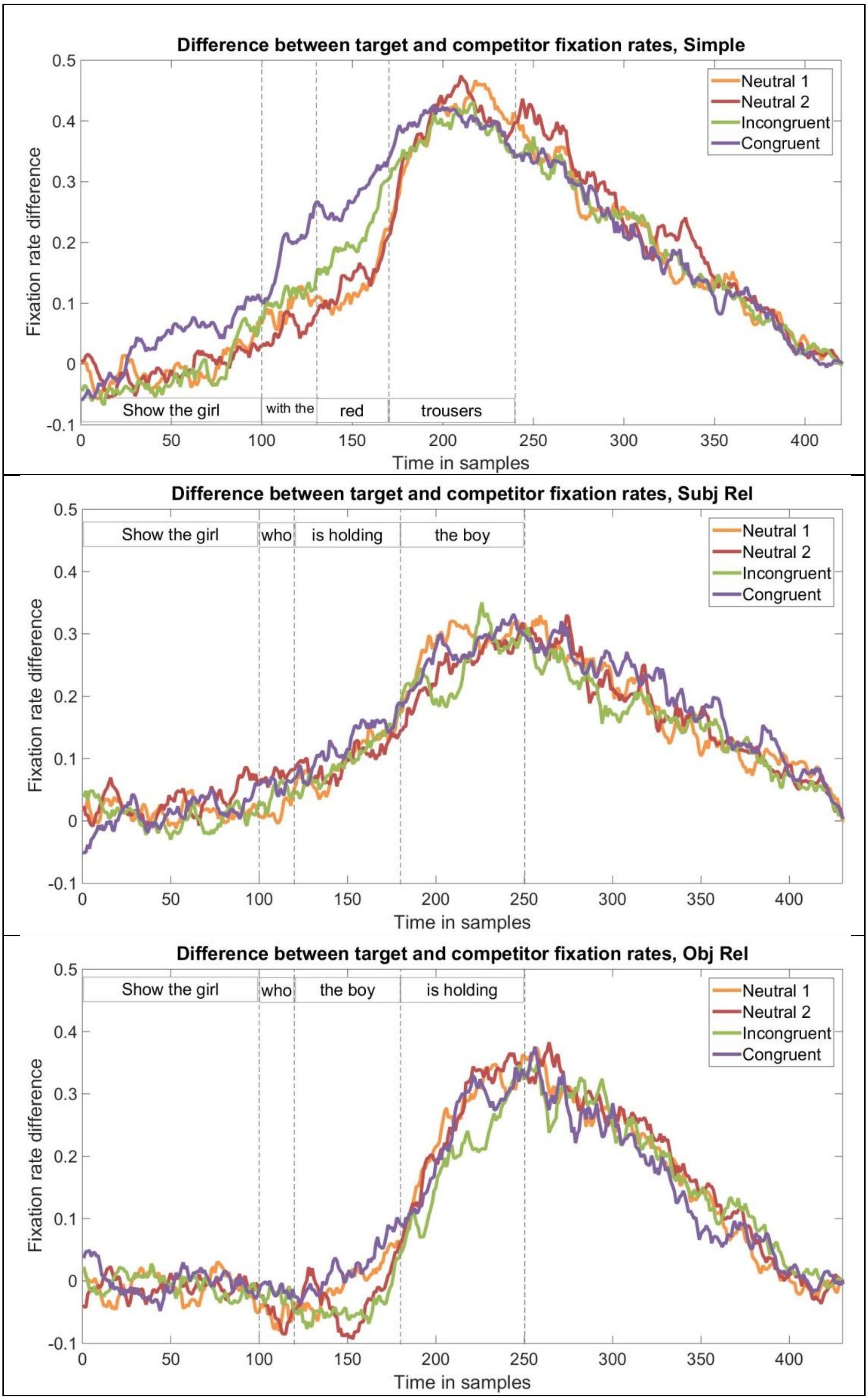


Figure 5.7. Experiment 6. Fixation rate differences between target and competitor characters for each mask (neutral 1, neutral 2, incongruent, congruent). The top panel shows the simple sentences, the middle panel the subject relative sentences, and the bottom panel the object relative sentences.

From these graphs, it seems once again that the only sentence type revealing a difference between mask conditions was the simple sentence type. Indeed, when the differences between fixation rate differences for each mask pair was calculated with 99.17% confidence intervals, the only pairwise comparisons where the confidence intervals did not overlap with 0 for at least 20 samples were the congruent condition compared to each of the other masks, for simple sentences only (Figure L.13 in Appendix L). The significant difference between the congruent condition and each of the other masks spanned segments 2 and 3, where the congruent condition led to a higher proportion of fixations to the target character than to the competitor. The fixation rate difference for the congruent condition then joined the other masks in segment 4, and there was no longer any difference between the masks at that point. This pattern indicates that participants were faster at identifying the target character in the simple sentences with the congruent mask, but that their overall certainty was not greater, given that the peaks in segment 4 were not different. There was no difference between the masks in the SR or in the OR conditions (Appendix L, Figure L.14 and Figure L.15), contrary to what was found in the reaction time data.

5.3.1.4 Cognitive measures

For the visual flanker task, the average difference between the inconsistent and the consistent conditions was 47 ms (*SD* = 26), ranging from -9 ms to 96 ms.

For technical reasons, only 28 participants (out of 36) completed the non-word repetition test, and the standard scores were not available for this task, unlike in Chapters 2 and 3. All participants completed the forward and backward digit span tests, which were unstandardized measures. The span for all these tests corresponds to the length of the longest correctly recalled block. Descriptive statistics for participants' spans are shown in Table 5.5.

	Range	Mean	Standard deviation
Non-word span	1 - 4	2.86	0.71
Forward digit span	3 - 8	6.33	1.29
Backward digit span	2 - 7	4.58	1.30

Table 5.5. Experiment 6. Range, mean and standard deviation for spans in the non-word repetition task, the forward digit recall task and the backward digit recall task.

The forward and backward digit spans were significantly positively correlated, $r(34) = .411$ ($p = .013$). Despite also being a measure of short-term memory, the non-word repetition span did not correlate as strongly with the forward digit span, $r(26) = .324$ ($p = .093$). The non-word span and the backward digit span were not significantly correlated either, $r(26) = .197$ (p

= .315). Based on these correlations, the forward and backward digit spans were combined as one composite digit span score (based on individual Z scores), and the non-word span was analysed separately in the correlational analyses reported in the next section[27].

5.3.1.4.1 Relationship between cognitive measures and sentence comprehension task

Table 5.6 summarises Pearson's correlations between the cognitive measures and differences in accuracy in the sentence comprehension task. The reaction time and accuracy differences were calculated between the congruent condition and each of the other mask conditions, and between the OR and the SR sentences. The correlational analysis focused on these differences in light of the findings reported in section 5.3.1.2. Indeed, given that there was a reaction time difference between the congruent condition and each of the other mask conditions, this difference might be explained, at least in part, by individual differences in memory and attention. Participants with higher memory spans and/or selective attention capacity might have been better able to take advantage of the information in the congruent competing talker utterances to respond more quickly.

[27] Some authors argue that digit span measures lead to better performance than non-word repetition tasks due to the higher frequency of random digits occurring in everyday conversation. The influence of long-term memory on performance in digit span measures (forward and backward) could be an explanation for the correlation observed between the two digit span measures but not with the non-word repetition task.

	Non-word repetition	Digit span composite	Flanker task difference
Accuracy difference Neutral 1 - Congruent	.376	.295	.159
Accuracy difference Neutral 2 - Congruent	-.030	.096	-.139
Accuracy difference Incongruent - Congruent	-.088	-.204	.008
Accuracy difference OR-SR	.310	.094	-.073
Reaction time difference Neutral 1 - Congruent	-.159	- .108	-.068
Reaction time difference Neutral 2 - Congruent	.435	.034	-.077
Reaction time difference Incongruent - Congruent	.193	- .123	.201
Reaction time difference OR - SR	.136	.022	-.192

Table 5.6. Experiment 6. Bivariate correlations between each of the cognitive tests (non-word repetition, the two digit spans together, and the flanker task difference), the response accuracy and reaction time differences between the congruent condition and each of the other masks, and the response accuracy and reaction time differences between the OR and the SR sentences. No correlations were significant at $\alpha = .0042$ (Bonferroni-corrected).

None of the correlations were significant at $\alpha = .0042$ (Bonferroni-corrected). Based on these results, the relationship between memory, attention and susceptibility to informational interference (or indeed facilitation) is still unclear.

5.4 Discussion

Similarly to the previous experiments in this thesis, Experiment 6 revealed an effect of syntactic complexity across all measures. However, the main finding of this experiment concerned the hypothesis that the semantic relevance of the competing talker utterances would affect sentence comprehension, which was indeed the case as there was an effect of mask. The effect of mask was present for all sentences in the reaction time data and for the simple sentences only in the eye-tracking data. This effect reflected facilitation in the congruent condition: faster reaction times (for all sentences) and earlier sentence resolution as measured by eye fixations (for simple sentences only), compared to each of the other masks. The congruent condition may thus have led to semantic priming which in turn allowed participants to understand the sentence earlier, indicating that even though sentence comprehension may not be adversely affected by a competing talker, the mask is not necessarily ignored. Given that the masks were randomly presented, participants did not know whether they could rely on the competing talker to answer until they had heard at least segment 1 of the target sentences. Thus, they must have been attending to both streams for at least the beginning of the target sentence.

Contrary to Iyer et al. (2010), the incongruent condition did not affect participants' performance more or less than the other masks. Participants in Experiment 6 were able to successfully ignore the contradicting information at no apparent additional cost compared to the irrelevant information given in the neutral conditions. Furthermore, the differing syntax of the neutral 1 and neutral 2 conditions did not lead to different performance, indicating that when the content of the competing talker is irrelevant, so is the syntax. However, confirming this possibility would require manipulating syntactic structure and semantic content orthogonally. A number of consequences of these results with relation to models of attention will be explored in Chapter 6.

An intriguing finding was that although the reaction time data showed a main effect of mask type with no mask by sentence interaction, the fixation rate data indicated that the facilitation effect of the congruent mask was only present for the simple sentences. Furthermore, 17% of the button presses for the simple sentences were made during segment 2 of the sentence, which is surprisingly high given that there shouldn't be enough information by that point to determine which character is correct (although these could have reflected random response). Given these findings, there may be something unique about the simple sentence condition which led participants to develop a strategy over the course of the experiment, despite the use of filler items. Indeed, the filler items were designed to decrease the possibility of participants relying on a strategy consisting of excluding the only character which was represented once in the picture. For example, if a picture depicted two men and a boy, the filler items ensured that the boy would sometimes be the target. However, in the simple sentence conditions and in the filler items, the target character was always the one with the same accessory as the single character. In the example in Table 5.1, where there were two witches and a man, the target witch was bound to be the one with the same size bag as the man, since there were no filler pictures where the target character was the other witch. It is nonetheless surprising that this strategy would have emerged only in Experiment 6, and not in the previous experiments.

Despite the use of filler sentences, participants may have used different strategies for the simple sentences compared to the SR and OR sentences simply because many of the accompanying pictures for the simple sentences included colour, whereas all the SR/OR pictures were black and white only. Furthermore, all pictures for the simple sentences depicted accessories, whereas none of the SR/OR pictures did. Therefore, participants may have used visual characteristics of the pictures to determine whether they would hear a simple sentence or a relative clause before they heard the beginning of the sentence. This in turn could lead to facilitation in the simple condition, which may have exacerbated the facilitation effect of the congruent condition. This strategy could only be applied to differentiate the simple sentences from the relative clauses, so there should have been no compounded facilitation effect in the SR and OR sentences, since participants did not know in advance whether a SR or a OR would follow. Thus, it is possible that the facilitation effect of the congruent condition was only visible in the simple condition for eye-fixations due to the additional pictorial cue. It is important to note that despite this possible bias in favour of the simple sentences, participants were nonetheless unable to determine which of the three characters the target was before hearing the sentence. Furthermore, the SR/OR contrast was not affected by this additional cue, since there was no difference between these conditions other than the actual target sentence. In future research utilising these stimuli, it would be interesting to add accessories to all the SR/OR pictures, and to use these same pictures for simple sentences rather than create new pictures.

One way of determining whether participants were indeed applying a strategy for the simple sentences is to look at their responses over the course of the experiment. If participants improve significantly over time for the simple sentences only, then they may indeed be using a strategy. However, as reported in section 5.3.1.2.2, there was no interaction between the block of presentation and the type of sentence, indicating that the sentence type did not affect participants' reaction time pattern over the course of the experiment. This suggests that participants did not use a strategy unique to the simple sentences.

Finally, the relationship between cognition and the ability to monitor a competing talker stream and benefit from its content is not clear from these data, as none of the correlations were particularly strong or significant.

Chapter 6: General discussion

6.1 Research aims

The central aim of this thesis was to investigate the conditions under which informational interference arises and the cognitive factors which underpin it. Informational interference was defined as the higher-order cognitive and linguistic aspects of IM, induced by competing speech. The main hypothesis was that informational interference depletes central processing resources that could otherwise be allocated to recognising and understanding target speech. By increasing processing demands imposed by either task (syntactic complexity, intelligibility), or listener characteristics (proficiency), informational interference should therefore be magnified.

To explore this hypothesis, the effect of a competing talker was compared to that of energetic mask controls across six experiments. Each experiment aimed to investigate specific aspects of informational interference, thus allowing energetic masking (EM) to be isolated from informational masking (IM).

Experiments 2, 3, 5 and 6 all shared the same method, with varying SNRs and mask types. Participants carried out a picture-selection task aiming to assess their online comprehension of target sentences under various masking conditions. All experiments used target sentences which varied in syntactic complexity, and thus in the degree of processing resources required: from simple syntactic structures to more complex subject relatives (SR) and even more complex object relatives (OR). Experiments 1 and 4 were designed to select the SNRs for the other experiments.

Regarding the cognitive factors involved in informational interference, I hypothesised that in addition to selective attention, short-term and working memory would be particularly recruited in dealing with the challenge posed by a competing talker. All sentence comprehension experiments (Experiments 2, 3, 5 and 6) therefore also included measures of visual selective attention, verbal short-term memory (STM) and verbal working memory (WM), to determine whether individual differences in these cognitive tests were correlated with individual differences in susceptibility to informational interference from a competing talker. The hypothesis was that participants with better scores in these tests would show less informational interference.

The following paragraphs summarise the specific objectives for each of the empirical chapters of this thesis (Chapters 2 -5).

6.1.1 Chapter 2: Is informational interference influenced by the syntactic complexity of the target utterance?

In addition to the main hypotheses outlined above, the hypothesis in Chapter 2 (Experiments 1 and 2) was that syntactic complexity of the target utterance would magnify the effect of informational interference. If processing resources are limited, and if informational interference increases reliance on processing resources, then target sentences that increase demands on processing resources should also increase informational interference. The prediction was that the more complex the target syntactic structure, the greater the detrimental effect of a competing talker, compared to an EM control (speech-modulated noise). Masked sentences were presented at a -5dB SNR, based on the results of a transcription task (Experiment 1).

6.1.2 Chapter 3: Is informational interference influenced by the language proficiency of the listeners?

The hypothesis specific to Chapter 3 (Experiment 3) was that informational interference would increase for non-native listeners. Indeed, non-native listeners expend more processing resources to deal with speech in adverse conditions and to process complex syntax. This greater reliance on limited processing resources should lead to lesser availability of these resources to deal with informational interference from a competing talker. This in turn should result in lower performance in target sentence comprehension with a competing talker than with EM controls (speech-modulated noise and reversed competing talker). As in the previous chapter, an interaction between syntax and informational interference was also expected: the more complex the target syntax, the greater the detrimental effect of the competing talker, compared to EM controls. Masked sentences were once again presented at -5 dB SNR.

6.1.3 Chapter 4: Is informational interference influenced by the intelligibility of the target utterance?

The hypothesis specific to Chapter 4 (Experiments 4 and 5) was that informational interference would increase as SNR decreases. Indeed, a decrease in SNR should increase the overall difficulty of the task by decreasing intelligibility of target sentences, which should in turn increase reliance on processing resources, thus leading to greater informational interference. Once again, we expected an interaction between syntactic complexity and the detrimental effect of the competing talker compared to the two EM controls (SMN and RCT). Masked sentences were played at -22 dB SNR for the SMN condition, and -25 dB SNR for the CT and RCT conditions, based on a transcription task (Experiment 4).

6.1.4 Chapter 5: Is informational interference influenced by the semantic content of the competing talker utterance?

The main hypothesis in Chapter 5 (Experiment 6) was that informational interference would increase when the content of the competing speech contained attention-grabbing information. The content of the competing talker utterances was manipulated to be either semantically congruent, incongruent or unrelated (neutral) to the target utterance. The prediction was that semantically related (congruent or incongruent) competing speech would be more difficult to inhibit than unrelated utterances, leading to greater informational interference. Furthermore, the incongruent competing talker utterances were expected to lead to lower sentence comprehension performance, due to simultaneously presented conflicting information. These detrimental effects should also be more visible as syntactic complexity increased.

The following section summarises the results of each experiment.

6.2 Summary of findings

The main findings of each of the four sentence comprehension experiments can be found in Table 6.1 below, and detailed in the next paragraphs.

		Effect of syntax?	Effect of mask type?	Evidence of informational interference?	Cognitive measures related to susceptibility to masking and informational interference?
Experiment 2: **effect of syntax and a competing talker** • -5 dB SNR • No mask, CT, SMN (blocked) • Native listeners	**Accuracy**	No	No	No	No significant correlations
	Reaction times	Yes • OR > SR • SR > Simple	No	No	
Experiment 3: effect of language proficiency • -5 dB SNR • No mask, CT, SMN, RCT (randomised) • Non-native listeners	**Accuracy**	Yes • Simple > OR • SR > OR	No	No	No significant correlations
	Reaction times	Yes • Simple < SR < OR	Main effect but no significant pairwise comparisons	No	
	Eye-tracking	Yes*	No	No	
Experiment 5: effect of intelligibility • -22 dB (SMN) and -25 dB (CT & RCT) • No mask, CT, SMN, RCT (randomised) • Native listeners	**Accuracy**	Yes • Simple > OR • SR > OR	Yes • Unmasked > masked • CT > RCT • CT > SMN (by-participants analysis only)	No	No significant correlations
	Reaction times	Yes • Simple < SR < OR	Yes • Unmasked < masked	No	
	Eye-tracking	Yes*	Yes • Unmasked vs masked: earlier and higher certainty (peak)	No	

Experiment 6: effect of semantic content • -5 dB SNR • CT only: neutral 1, neutral 2, incongruent, congruent (randomised) • Native listeners	**Accuracy**	Yes • Simple > OR • SR > OR	No	No, but evidence of facilitation from congruent condition	No significant correlations
	Reaction times	Yes • Simple < SR < OR	Yes • Congruent < neutral 1 • Congruent < neutral 2 • Congruent < incongruent		
	Eye-tracking	Yes*	Yes • Simple only: congruent earlier than each of the other masks		

Table 6.1. Summary of the main findings in Experiments 2, 3, 5 and 6, for each of the sentence comprehension measures (accuracy, reaction times, and eye-tracking where applicable) and for the relationship between the cognitive measures and the sentence comprehension task. *Note that, for the eye-tracking measures, the effect of syntax was not statistically verified and is based only on descriptive values.

6.2.1 Chapter 2 (Experiments 1 & 2)

Experiment 1 was conducted to select the signal-to-noise ratio (SNR) for Experiment 2. In Experiment 1, participants were asked to transcribe the masked sentences without the accompanying pictures. The aim was to choose a SNR at which transcription accuracy was equally high across mask conditions (competing talker and speech-modulated noise), while still presenting a challenge for listeners. Of the three SNRs tested (0 dB, -5 dB, -10 dB), -5 dB was chosen for Experiment 2 because performance was comparable across mask conditions without reaching ceiling.

Experiment 2 investigated sentence comprehension in the presence of no mask, a competing talker (CT) or speech-modulated noise (SMN) at -5 dB SNR. Participants' performance was measured with accuracy and reaction times to assess the cost of processing sentences with a CT compared to an EM control (SMN). Across both accuracy and reaction times there was no difference between masks, and indeed no difference was found between the no mask condition and either of the masked conditions. An effect of syntactic complexity was found for reaction times, in the direction of our hypothesis: the most complex syntactic structures (OR) were the slowest, followed by the SR sentences and finally the least complex syntactic structures (simple).

In Experiment 2 the relationship between individual differences in reaction times to the sentence comprehension task and individual differences in STM, WM and selective attention were investigated. No significant correlations were found between reaction time differences to the sentence comprehension task (masked – unmasked; CT – SMN) and scores in the cognitive tests (non-word repetition, listening recall, visual flanker task).

Thus, although reaction times were sensitive enough to evidence a difference between syntactic structures, they did not reveal an effect of informational interference, let alone an effect of mask vs no mask. The lack of main effect of mask makes it difficult to conclude that there was no added detriment of a competing talker compared to EM, since this lack of difference may simply have been caused by a ceiling effect. Indeed, the masked conditions may not have been challenging enough compared to the unmasked condition for these native participants. The following experiment addressed this possibility.

6.2.2 Chapter 3 (Experiment 3)

In Experiment 3, a series of methodological changes was made to enable the emergence of an effect of informational interference, if indeed such an effect exists. Given that the conditions in Experiment 2 may not have been challenging enough for native listeners, a group of non-native listeners (Danish L1, English L2) was tested, based on the assumption that they would expend more processing resources than native listeners. By increasing reliance on shared cognitive resources due to their L2 status, informational interference should be more likely to emerge in this group than in the native listener group.

In addition to changing the population, I carried out a series of modifications to the task design and the measures. Time-reversed speech was added as a second energetic mask control based on the competing talker. As mentioned in Chapter 1, time-reversed speech and speech-modulated noise have different acoustic properties, which taken together provide a better control for the EM generated by the competing talker.

Furthermore, whereas Experiment 2 used a blocked design for mask conditions (one mask type per block), in Experiment 3 the masks were randomised on a trial-by-trial basis. This reduced habituation effects and increased uncertainty, thus adding an extra demand on listeners. Finally, Experiment 3 introduced the use of eye-tracking as an online measure of sentence processing, complementing the information from accuracy and reaction times.

None of the three measures revealed an effect of mask type on sentence comprehension. All measures did however reveal an effect of syntactic complexity, and the fact that this effect was already apparent in the accuracy data indicated that the non-native listeners in this experiment were more sensitive to syntactic complexity than the native listeners in Experiment 2. Furthermore, the non-native participants in Experiment 3 were slower across all conditions than participants in Experiment 2 (although it is not possible to make a direct comparison given the methodological differences between the experiments).

A series of correlations addressed the hypothesis that individual differences in susceptibility to informational interference from a competing talker are related to individual differences in language proficiency, short-term and working memory, and selective attention. Once again, none of the cognitive tests or proficiency measures was related to the difference between masks (masked – no mask; CT – EM controls).

In conclusion, Experiment 3 revealed that although the measures used were sensitive enough to highlight syntactic complexity differences, no effect of informational interference from a competing talker was evidenced. Indeed no effect of mask compared to no mask was found. This was surprising given that the listeners were non-native, with varying levels of proficiency, and as such they were expected to perform less efficiently with a mask.

Before concluding that there was no informational interference at all in this experiment, it is important to ascertain that the lack of difference was once again not due to a ceiling effect. Indeed, the lack of effect of mask vs no mask could be due to a relatively unchallenging SNR. The next chapter aimed to investigate this possibility by reducing the SNR.

6.2.3 Chapter 4 (Experiments 4 & 5)

Experiment 4 consisted of a transcription task that was very similar to Experiment 1, but this time the masked sentences were played with the accompanying pictures. This ensured that performance included the benefit of seeing the pictures, which restrict the number of lexical candidates. The target level of performance was 80% correctly transcribed keywords. Six SNRs were tested: -13 dB, -16 dB, -19 dB, -22 dB, -25 dB, and -28 dB SNR. For a given SNR, performance was lower in the SMN condition compared to the CT and RCT conditions. This was most likely due to the different EM properties of the SMN mask compared to the CT and RCT masks. To counteract this variation, different SNRs were chosen for the SMN mask (-22 dB SNR) and for the CT and RCT masks (-25 dB SNR), based on the value that led to 82% average transcription accuracy.

Experiment 5 consisted of a sentence comprehension task identical to Experiment 3 except for the SNRs, and participants were native listeners. This time, there was an effect of mask across all measures (accuracy, reaction times and eye-tracking), reflecting the greater challenge imposed by the SNR in the masked conditions (CT, SMN, RCT) compared to the no mask condition. However, there was no evidence of informational interference from the competing talker. Indeed, participants were more accurate in the CT condition than in the EM conditions, although this difference was not apparent in the reaction times or eye-fixation data. The effect of syntactic complexity in the direction of our hypothesis was once again found across all measures.

Correlations between the cognitive tests and performance in the sentence comprehension task were non-significant, which does not allow any conclusions to be made with regard to the relationship between cognition and susceptibility to informational interference or masking.

The findings of Experiment 5 indicate that the lack of informational interference observed in Experiments 2 and 3 cannot simply be explained by an unchallenging SNR. Indeed, the accuracy data in this experiment even point to a possible release from masking in the CT condition. However, the fact that the difference between the CT and the EM masks was only found in the accuracy data suggests that the difficulty posed by the RCT and SMN conditions may have been due to slightly reduced intelligibility because of higher EM from the RCT and SMN conditions. Indeed, if the difficulty were due to an increased cognitive processing load (not due to EM) then this should be reflected in reaction times and eye-fixations. A complementary explanation for the detrimental effect of the RCT and SMN conditions compared to the CT condition is that although the RCT condition may not have created additional EM compared to the CT condition, it was more attention-grabbing than the CT condition due to its unusual acoustic characteristics. Perhaps the CT condition did not lead to informational interference in this experiment because the content of the utterances was not attention-grabbing enough. Indeed, the competing sentences were chosen to be as unrelated as possible to the target sentences in terms of their semantic content and structure. The next chapter investigated the influence of semantic content on the emergence of informational interference.

6.2.4 Chapter 5 (Experiment 6)

In Experiment 6, participants carried out the same sentence comprehension task with the same target sentences as in Experiments 2, 3, and 5, but this time the sentences were only masked by competing speech (at -5 dB SNR). The competing talker conditions varied in the similarity/relevance and congruence of their content in relation to the target sentences. In the neutral 1 and neutral 2 conditions, the sentences were unrelated in their semantic content. The neutral 1 condition was identical to the CT condition in the previous experiments, and consisted of pairs of HINT sentences with simple syntactic structures. The neutral 2 condition consisted of sentences that followed the same syntactic structure as the target sentence they were paired with. The congruent and incongruent conditions also followed the syntactic structure of the corresponding target sentence, and both were relevant to the target as they contained the same words but in different orders. The congruent condition followed the same message as the target sentence, whereas the incongruent condition consisted of a contradicting message. If the attention-grabbing nature of a competing talker is due to the relevance of its semantic content, then the neutral conditions should be less attention-grabbing than the congruent and incongruent conditions. Furthermore, if participants monitor the semantic content of the competing talker, then the contradicting information in the incongruent condition should lead to lower performance than the congruent condition in the sentence comprehension task, if indeed it is attention-grabbing at all.

Contrary to any of these hypotheses, the only difference between masks was the facilitation effect of the congruent condition compared to each of the other conditions. This difference was observed across all sentences in the reaction time data, and for the simple sentences in the eye-fixation data. This suggests that although participants were not completely blocking out the mask indiscriminately, the content does not seem to have had a detrimental effect on target sentence comprehension. It is interesting to note that the incongruent condition did not affect participants' ability to understand the target sentence, unlike Iyer et al. (2010). Participants were able to successfully ignore the contradicting information at no apparent additional cost compared to the irrelevant information in the neutral conditions. As in the previous sentence comprehension experiments, an effect of syntactic complexity in the direction of our hypothesis was found, across all measures.

Finally, correlations between the cognitive test scores and sentence comprehension performance were once again inconclusive.

6.3 General discussion

The main hypothesis was that a competing talker leads to informational interference by depleting central processing resources that could otherwise be allocated to processing the target sentence. None of the results supported this hypothesis. Indeed, there was no detrimental effect of the competing talker in any of the conditions designed to increase reliance on processing resources (syntactic complexity, proficiency, SNR). Each of these conditions did however give rise to main effects: syntactic complexity influenced general performance across experiments, the non-native listeners were slower overall than the native listeners, and the low SNRs affected performance in the masked conditions compared to the unmasked condition. Furthermore, although there were individual differences in sentence comprehension performance and cognitive test results, none of the cognitive tests were conclusive in showing a link between STM, WM or selective attention and the ability to deal with informational interference from a competing talker.

How do these findings relate to previous studies that have investigated the effect of a competing talker compared to energetic mask controls? Table A.1 (Appendix A) summarised some of the main characteristics of various studies investigating the effect of a competing talker on sentence intelligibility with normal-hearing young native listeners. The main differences between the experiments in this thesis and the ones reported in Table A.1 are the type of task and the measures. Only one of the studies reported in Table A.1 used a measure of listening effort or processing load (Koelewijn et al., 2012), and only one required listeners to process the syntax of the sentences presented, assessed by a sentence comprehension task (Sörqvist & Rönnberg, 2012). The authors in this latter study found a detrimental effect of the competing talker compared to spectrally rotated speech, however this effect may have been due to low-level IM. Indeed, both the target and the competitor voices were male, which increases segregation difficulties.

Three other studies assessed sentence comprehension in the presence of a competing talker (Brungart, 2001; Brungart et al., 2013; Iyer et al., 2010), but the stimuli were from the CRM corpus, for which it is sufficient to identify the keywords without establishing syntactic dependencies between words (unlike the relative clause sentences in my experiment). Of these three studies, one did not find a detrimental effect of the competing talker (Iyer et al., 2010) but the other two did. It is not possible to disentangle the lower-level components of IM from the higher-level components in Brungart et al. (2013) and Iyer et al. (2010), because voices of the same gender were once again used. However, Brungart (2001) investigated all combinations of male and female target and masker voices and found that when voices of different genders were used, the competing talker was more detrimental than modulated noise at SNRs between +15 dB and -6 dB. In the context of these results, it is surprising that the experiments in this thesis did not reveal a detrimental effect of the competing talker on sentence comprehension. A few differences between the experiments in this thesis and Brungart (2001) may partly explain the discrepancy in results. In Brungart (2001), the target and masker both consisted of CRM sentences. Given the structure of these sentences ("Ready <call sign> go to <colour> <number> now"), each of the words in the competitor sentence overlaps with the corresponding word of the same category in the target sentence. This probably leads to fewer opportunities of dip-listening and less time to build up the separate auditory streams than in my experiments. Indeed, I introduced a lag between the start of the competing talker sentence and the start of the target talker sentence, which enabled the auditory stream to build up for the competing talker before the target started. Furthermore, although the content of the competing talker was manipulated in the final experiment to be similar or dissimilar to the target, there was always a delay between hearing the related word in the competing talker sentence and the target talker, contrary to Brungart (2001) where the target and competing words of the same category were presented at the same time. Another possible explanation for the difference between my results and Brungart (2001) is that although the CRM is also a closed-set task (only eight possible colours and eight possible numbers), the number of possibilities is greater for a given item in the CRM than for a given item in my experiments, where participants only have a choice of three characters.

A number of experiments requiring participants to repeat the target sentences (but not necessarily process the meaning) have also shown a detrimental effect of a competing talker compared to energetic mask controls (Francart et al., 2011; Helfer & Freyman, 2014; Kidd et al., 2014; Koelewijn et al., 2012; Rhebergen et al., 2005; Trammell & Speaks, 1970). Of these experiments, only three used voices of different genders (Francart et al., 2011; Koelewijn et al., 2012; Rhebergen et al., 2005), thus reducing the effect of low-level IM. Koelewijn et al. (2012) found a detrimental effect of a competing talker compared to speech-modulated noise when analysing the pupil dilation, but not with SRTs, indicating that listening effort was not captured by SRTs. It is not clear why this study and the findings in my experiments do not show similar patterns. The answer may lie in the use of a competing talker with the same long-term average frequency as the target, which may have increased low-level IM in Koelewijn et al., (2012). Francart et al. (2011) compared a competing talker in a native language to a competing talker in a non-native language and found that the non-native competing talker led to lower (better) SRTs than the native competing talker. However, it is possible that these results were due to different EM properties of the two conditions, especially as there were no matched EM controls for each of these conditions. Finally, Rhebergen et al. (2005) found a detrimental effect of a competing talker compared to a reversed competing talker. Once again the reason for the discrepancy between these results and the lack of a detrimental effect of a competing talker in my experiments is unclear. The difference may lie in the type of task (sentence comprehension vs. intelligibility), however a sentence comprehension task should require more processing resources than a repetition task. The contextual information given to participants in my experiments was greater than that typically given in repetition tasks, thus reducing reliance on acoustic input. Further research is needed to determine the source of the differences (developed in section 6.4).

The results reported in this thesis give rise to a number of additional questions, which are the focus of the following sections:

1. Does competing speech ever lead to informational interference?
2. Is competing speech actually less demanding than EM controls?
3. Do listeners selectively attend to the target voice and inhibit the mask at early listening stages?
4. What role does cognition play in sentence comprehension with a competing talker?

6.3.1 Does competing speech ever lead to informational interference?

The main finding reported in this thesis is that there was no detrimental effect of a competing talker on sentence comprehension, beyond its EM and low-level IM. One possible explanation for this finding is that a competing talker does not lead to informational interference in young normal-hearing listeners when low-level IM and EM are controlled for. To conclude this however, the following alternative explanations (addressed in the next paragraphs) must first be discounted:

1. The energetic mask controls created additional EM that counteracted the effect of informational interference, despite the use of two types of energetic mask controls that have been widely used in intelligibility studies.
2. The task was not resource-demanding enough for an effect to show, despite the use of online measures such as reaction times and eye-tracking, and the various manipulations to increase processing load.

6.3.1.1 Were the energetic mask controls optimal?

The first possible explanation is that the energetic mask controls (SMN and RCT) generated more EM than the competing talker. If this is true, and if the CT has a small detrimental effect beyond its EM, then this effect would be counteracted by the added EM in the SMN and RCT conditions. I have already mentioned that although SMN has the same temporal amplitude modulations as the CT from which it was created, it does not have the same spectro-temporal structure as speech. In particular, SMN does not vary in its spectral characteristics like speech does, which can lead to more EM at a given point in time compared to speech. Furthermore, SMN does not have the same periodicity profile as speech, however the ability to exploit periodicity cues that are present in speech but not in SMN has been shown to determine performance in speech-in-noise tasks (Steinmetzger & Rosen, 2015). Despite these differences, SMN has been effectively used as an EM control in many experiments, and some of these studies have shown a detrimental effect of a competing talker compared to SMN (Brungart, 2001; Brungart et al., 2013; Koelewijn, Zekveld, Festen, & Kramer, 2014).

As a complementary approach to controlling for EM of the competing talker, time-reversed speech was introduced from Experiment 3 onwards. As mentioned in previous chapters, time-reversed speech preserves aspects of speech that SMN does not, such as formants and harmonic structure, while removing the semantic content of speech. However, time-reversed speech also creates more forward masking, and the amplitude contour does not follow the original speech from which it was created. Thus, it is possible that although the average opportunities for glimpsing are equivalent, different portions of the target sentence are masked to different degrees with competing speech compared to reversed competing speech. In a study comparing the intelligibility of target sentences presented with native and non-native reversed speech and native and non-native competing speech, it was estimated that the decrease in intelligibility due to forward masking corresponds to an increase in SRTs of around 2.3 dB, whereas the cost of interference from the native competing speech may correspond to an increase in SRTs of around 6.6 dB (Rhebergen et al., 2005). Although these authors used a sentence repetition task and I used a sentence comprehension task, it is reasonable to assume that the cost of forward masking should not be greater in my experiments than in Rhebergen et al. (2005). Therefore, although forward masking may have contributed to attenuating the effect of the competing talker in my experiments, this effect would have had to be just as small as the effect of forward masking from the reversed competing talker.

Thus, although EM may have been greater in the SMN and RCT conditions than in the CT condition, previous research suggests that these masks can be used as effective controls for the EM of competing speech (Brungart, 2001; Brungart et al., 2013; Koelewijn et al., 2012; Rhebergen et al., 2005; Trammell & Speaks, 1970). There is still a possibility that the effect of the competing talker in my experiments was too small to counteract the additional masking induced by the SMN and RCT. I will address ways to explore this possibility in section 6.4, where a number of future directions for this research are envisaged.

6.3.1.2 Was the task resource-intensive enough?

The second alternative explanation to the lack of the existence of informational interference is that the task was not resource-demanding enough for informational interference to arise. Indeed, it can be argued that showing participants the pictures before and during target sentence presentation greatly reduced the number of lexical candidates, thus facilitating the task (compared to a task where participants do not know what the content will be about before hearing the sentence). Thus, participants could rely less on the finer acoustic details of the target speech and more on word order, especially in the relative clause sentences. This characteristic of the task resembles everyday conversations, where the context is often given, and conversational partners may see the objects they are referring to. Although it is possible that the reduction in lexical candidates may have facilitated the task, it was nonetheless surprising that no effect of mask vs no mask was found at -5 dB SNR. Indeed, previous research using eye-tracking (Wendt et al., 2015) used very similar pictures with sentences varying in syntactic complexity, and found an effect of mask (speech-shaped noise or speech-modulated noise vs no mask). The 80% SRTs in Wendt et al. (2015) ranged from -9.8 dB to -3.6 dB for the normal-hearing listeners. Thus, in my experiments at -5 dB SNR using similar stimuli it was expected that participants would show an effect of mask vs no mask. One major difference that could partly explain the discrepancy in results between the experiments is the language of presentation. Indeed, Wendt et al. used German stimuli from the OLACS corpus (Uslar et al., 2013), whereas the stimuli in this thesis were in English. In German, the contrast between the different sentence structures in the OLACS corpus (e.g. SR vs OR or SVO vs OVS) relies on subtle morphological differences in the case marking of the article ("der" for nominative, "den" for accusative) and the case marking of the adjective ("kleine" for nominative, "kleinen" for accusative). The word order is the same for two contrasting structures (e.g. SR vs OR) in German, whereas the word order is different for the same two contrasting structures in English, which does not have case marking. Thus, listeners rely on more subtle acoustic differences in German compared to English, leading to a possible increased detrimental effect of EM in German compared to English.

I addressed the possible issue of the task not being cognitively demanding enough in Experiment 3, by testing non-native participants who should expend more processing resources than native listeners by virtue of their reduced proficiency. Although participants all had relatively good command of the English language, the range of proficiencies was such that we expected to see a possible effect of proficiency on performance. Indeed, participants who showed the smallest difference between the OR and the SR sentences were those with the highest proficiency, indicating that proficiency did affect complex syntax comprehension. However, proficiency did not correlate with the difference in performance between the masked and unmasked conditions. This was surprising given that non-native listeners have been shown to expend more processing resources in speech in noise tasks (Lecumberri et al., 2010). The fact that non-native listeners were not more sensitive to EM than native listeners, and that neither of the two groups showed an effect of mask vs no mask at a SNR that has previously shown effects points to the possibility that the sentence comprehension task used throughout the experiments is particularly robust to EM. This was further demonstrated in Experiment 5, where the SNRs had to be reduced to very low levels that are rarely used in intelligibility experiments, even at SRTs of 50% (one exception is Lew & Jerger, 1991, who used SNRs as low as -30 dB). Although robustness to EM may seem like a shortcoming, it could actually be beneficial when studying the unique effects of competing speech beyond EM. Indeed, if performance in the task is relatively unaffected by EM, but it is affected by a manipulation in higher-level processing resource demands such as syntactic complexity, then one can expect to observe the effect of informational interference if it imposes additional demands on higher-level processing resources.

Finally, performance was not at ceiling for all conditions, since there was a main effect of sentence complexity across all sentence comprehension experiments. The fact that a difference between sentence types was found indicates that the task was demanding enough, at least for the effect of syntax to emerge.

Although one cannot discard the possibilities that the EM controls were not optimal and that the task was not resource-demanding enough, there still is a possibility that a competing talker does not lead to informational interference in a sentence comprehension task (with normal-hearing typically-developing young adults), at least under certain conditions: when the target and competitor voices are acoustically dissimilar enough to lead to successful streaming and object formation, and when the context provides disambiguating information allowing to reduce the number of lexical candidates.

6.3.2 Is competing speech less demanding than energetic mask controls?

The second question arising from the results of this thesis was whether competing speech was in fact less demanding than EM controls. Indeed, in Experiment 5 (low SNR), participants' accuracy in the sentence comprehension task was significantly higher for the CT condition than for the EM control conditions. Does this mean that competing speech is in fact less demanding than EM alone? The answer may lie in the different demands imposed on EM and IM by low SNRs. Results of an experiment conducted by Brungart (2001) comparing accuracy in the CRM task with competing speech, speech-shaped noise and modulated noise at different SNRs (+15 dB to -21 dB SNR decreasing in 3 dB steps) provide additional insight into my results. In particular, Brungart showed that as SNR decreased, performance with the noise maskers decreased monotonically between -3 dB and -21 dB SNR, whereas performance with the speech maskers reached a plateau between 0 dB down to -12 dB SNR (the lowest SNRs were not tested with speech maskers). Accuracy for the noise maskers was higher than for the speech maskers from around +6 dB SNR until around -6 dB or -9 dB, but fell below the speech maskers in the lower SNRs. In Experiment 5 of this thesis investigating very low SNRs of -22 dB and -25 dB, accuracy with the competing talker was higher than accuracy with the EM controls, whereas there had been no difference at -5 dB SNR. These results can be understood in the context of Brungart (2001) who suggested that performance is predominantly influenced by EM at the lowest SNRs, whereas it is predominantly influenced by IM in the higher SNRs. The fact that there was no difference between masks using reaction times or eye-tracking measures indicates that the difference in accuracy is predominantly driven by EM. Indeed, reaction times and eye-tracking are sensitive measures of the processing resources involved, whereas accuracy is not as sensitive to assess processing resources. Thus accuracy may be a better reflection of perceptual degradation (EM) instead. To conclude this section, the difference between CT and EM controls revealed in Experiment 5 can be explained by the dominance of EM with low SNRs.

6.3.3 Do listeners selectively attend to the target voice and inhibit the mask at early listening stages?

The third question arising from the results reported in this thesis concerned the time-course of the focus of attention. Chapter 1 introduced several models of attention that provide a framework for understanding the level at which informational interference may operate. Broadly speaking, auditory attention may be allocated at an early stage or a late stage depending on the theoretical viewpoint. From the results of Experiments 2, 3, and 5, it is tempting to conclude that participants took advantage of the low-level acoustic cues differentiating the target voice from the masker at a very early stage in listening, thus completely blocking out the masker. This could explain why there was no difference between the types of mask. Under this view, the focus of attention operates at a very early stage in cases where stream formation and object selection are easily accomplished based on distinct acoustic cues in each stream (as in the case of the target male and competing talker female in my experiments). This would be in accordance with early selection theories such as Broadbent's (1958) early filter model or Treisman's (1969) attenuation model. It is also in line with neuroimaging studies suggesting that competing speech is actively suppressed at early stages of auditory processing (Evans, McGettigan, Agnew, Rosen, & Scott, 2016; Zion Golumbic et al., 2013). The results of these imaging studies suggest that higher-order aspects of competing speech (such as syntax) may in fact not be processed.

However, Experiment 6 provided evidence that participants did access higher-order aspects of the competing speech. Indeed, they monitored both streams of speech at least until after the first segment of the target sentence, given that they were aided by the congruent information in the competing talker sentences. Similarly, previous research has found that the content of a competing utterance can influence processing of target utterances (Iyer et al., 2010).

Although models of attention provide a framework within which informational interference can be studied, given that no evidence for informational interference was found in this thesis, these results may be explained by both early and late models of attention, while providing support for neither early nor late models of attention. As mentioned above, the fact that participants accessed the information from the congruent condition indicates that selection did not occur at an early stage as Broadbent’s early filter model would suggest. These results may however be consistent with Treisman’s attenuation model, also an early selection model. It is possible that although the competing talker information was attenuated, the task was not resource-demanding enough for the attenuation to be complete. In other words, the activation threshold of the competing words was relatively low due to the additional available resources. Furthermore, the words in the congruent and incongruent conditions should have even lower thresholds due to their relevance to the target sentences. Participants could therefore access the semantic content of the competing talker to aid their responses when relevant. The fact that the incongruent condition did not lead to slower reaction times than the neutral conditions, despite the possibly lower thresholds might once again have been due to the relatively low demands of the task. Although participants had access to the information in the congruent and incongruent conditions, their attentional capacity was not depleted enough for either of these conditions to have a detrimental effect on performance.

The results of my experiments may also be explained by late-selection models such as Deutsch and Deutsch (1963). Selection may have occurred at a later stage, after both the target and the competitor had been fully perceived. The time-course of this mechanism cannot however be inferred from my results. It is possible that participants' attention was captured by both the competing talker and the target throughout the full sentence presentation, but it is also possible that their attention focused on the target sentence once they had determined whether the competitor was congruent with the target.

The middle-ground view posited by Lavie's load theory of attention would predict that the competing talker interferes with target sentence comprehension only under low perceptual load and/or high cognitive load. It is thus possible that the competing talker did not interfere with sentence comprehension because cognitive load was not high enough. It is unlikely that the lack of interference was due to high perceptual load, although it is unclear how perceptual load would be instantiated in the context of my experiments.

Finally, Shinn-Cunningham's model could pinpoint the level at which a competing talker may interfere with target processing (object formation or object selection). However, in the experiments at -5 dB (Experiments 2, 3, 6) there was no failure of either object formation or selection, since no detrimental effect of a mask was found. In the low SNR experiment (Experiment 5), the difference between the masked and unmasked condition was most likely due to an increase in EM, and not because of a failure of object formation or selection. The task was designed to facilitate object formation by using two voices of different genders, and by introducing a lag between the mask and the target speech.

Thus, from the results of this thesis it is not possible to give definitive answers about the time-course of attentional focus, except that listeners do not inhibit the mask indiscriminately at the very early stage of voice segregation.

6.3.4 What role does cognition play in sentence comprehension with a competing talker?

I found no evidence that STM, WM, or selective attention were significantly correlated with susceptibility to informational interference. Although there were individual differences in participants' susceptibility to informational interference, on average there was no difference between masks. Therefore it is possible that the lack of correlation was simply due to the lack of informational interference.

Regarding the lack of correlation for the visual flanker task, perhaps the answer lies in the different modalities of presentation. Although Rönnberg et al. (2013) argue that language understanding is multimodal and have used visual cognitive tasks (e.g. reading span) to explore the relationship between cognition and speech-in-noise tasks, the cognitive tasks usually involve language in some form. In contrast, the visual flanker task used in my experiments presented only non-linguistic stimuli (shapes), which could arguably be processed in a modality-specific way.

Although previous research has linked various cognitive factors, in particular WM, to susceptibility to masking (Akeroyd, 2008; Rönnberg et al., 2013), the results of this thesis do not provide additional evidence for the involvement of STM, WM and selective attention. Indeed, a recent review suggests that the link between WM as measured by the reading span task and speech-in-noise processing may not systematically hold for normal-hearing young listeners (Füllgrabe & Rosen, 2016). These authors argue that the predictive value of WM in speech-in-noise performance is mainly observed with hearing-impaired and older listeners. Given that the participants in all my experiments were young, normal-hearing listeners, the lack of association between cognition and performance in the sentence comprehension task is perhaps unsurprising in light of the aforementioned review. The following section will address possibilities for future research, including further exploring the link between informational interference and cognition.

6.4 Future directions

The previous sections presented a number of questions arising from the results of my experiments. Although there is some evidence contributing to possible answers, these questions are still mostly open. Future research could provide further answers and add to our understanding of the underlying mechanisms behind informational interference. In the following paragraphs I will briefly outline some of the possible directions this research may take.

6.4.1 Exploring the limits of informational interference

It is possible that informational interference from a competing talker only arises given specific circumstances. What are these circumstances, and how can one go about revealing the unique contribution of informational interference beyond EM and low-level IM? Although the experiments in this thesis explored some of the possible conditions under which informational interference can arise, a number of additional conditions can still be explored.

6.4.1.1 Controlling for energetic masking

As mentioned in section 6.3.1, one cannot discard the possibility that the SMN and RCT conditions created additional masking that may have counteracted a possible detrimental effect of the competing talker. One way of determining whether this was indeed the case would be to use spectrally-rotated speech (created from the competing talker) as an additional masker. Indeed, it has been argued that spectrally-rotated speech is a better control for EM of a competing talker since it preserves pitch variation, rhythm and differences in periodicity while being unintelligible to the untrained ear (Green et al., 2013). However, the spectral shape of the mask is by definition different, and as such is not a perfect energetic mask control either (as mentioned previously, only the original speech would be a perfect match).

To bypass the thorny issue of imperfect energetic mask controls, another option would be to introduce a mask that can be used both as a competing talker and as its own energetic mask control. For example, vocoded speech (or indeed spectrally-rotated speech) is unintelligible to the untrained ear, thus acting as an energetic mask only. After presenting the target sentences with vocoded speech in the first part of the experiment to two groups of participants, one group could then be trained to understand the vocoded speech, which would thus become the competing talker in the second part of the experiment. A recent unpublished study used a similar training paradigm (Dai, Kösem, McQueen, & Hagoort, 2016) with untrained and trained 4-band noise vocoded speech and untrained 2-band noise-vocoded speech, both in dichotic and diotic presentations. Results pointed to a possible interference effect of the trained 4-band noise vocoded speech compared to untrained 2-band noise vocoded speech presented dichotically, however the effect did not persist when comparing trained and untrained 4-band noise vocoded speech nor when target and competitor were presented diotically. One disadvantage of using this method is that mask types cannot be randomised within the experiment, thus increasing a possibility of habituation to the mask. Furthermore, there are large individual differences in the ability to understand vocoded speech, and for some individuals it can be quite difficult to show training effects at all. One way of bypassing this issue is to spectrally shift the vocoded speech, thus decreasing the intelligibility of the masks and increasing the potential for responsiveness to training (Rosen, Faulkner, & Wilkinson, 1999).

Another alternative to the issue of controlling for EM would be to present target and competitor sentences dichotically. Indeed, dichotic presentation eliminates acoustic overlap at the periphery, thus eliminating EM. In fact the original 'cocktail party' experiments used dichotic listening tasks (Cherry, 1953; Moray, 1959). Future experiments could compare diotic listening to dichotic listening, to factor out the contribution of EM.

6.4.1.2 Manipulating low-level informational masking

By using two voices of different genders, we aimed to reduce the difficulty due to segregation of the voices, i.e. low-level IM. However, informational interference may be exacerbated by low-level IM, simply due to the increase in general task demands. To further explore this possibility, the same sentence comprehension task could be administered but with different combinations of voices: same talker, different talkers but same gender, different talkers and different genders. Furthermore, the similarity of the spectral characteristics of the voices could be manipulated, thus allowing low-level IM to be teased apart from high-level IM. If participants are only affected by the competing talker when the voices are acoustically similar, then the definition of informational interference must be revised to include low-level IM. However, if this were the case it might also be because the demands of the task were increased by increasing low-level IM. The next section explores this possibility.

6.4.1.3 Manipulating task demands and measures of processing load

One of the major outstanding questions of this thesis is the issue of task complexity and the processing load required of participants. Indeed, the lack of mask effect may simply be due to the fact that the pictures reduce the number of lexical candidates, thus reducing task difficulty. As previously mentioned, the main effect of syntax found across all experiments indicates that the task did not lead to ceiling effects across all conditions. However, it would be interesting to increase the difficulty of the task in the following ways. The pictures could be presented after the sentence was heard, to ensure that participants did not already form expectations about the sentences. A disadvantage of using this method would be the loss of information about online sentence processing (reaction times and eye-tracking). However, pupil dilation could instead be measured as an indicator of processing load (Koelewijn et al., 2012; Wendt, Dau, & Hjortkjær, 2016; Zekveld & Kramer, 2014).

Another issue related to measures of processing load is the choice of accompanying cognitive tests. Indeed, the underlying cognitive factors of informational interference are still unclear. Administering a broader range of cognitive tests thought to involve different cognitive functions would be useful. In particular, a test of inhibition (auditory Stroop) would be ideal, as I have hypothesised that interference from a competing talker may be related to WMC through its toll on inhibitory processes. A test of auditory selective attention would also be a suitable addition, since it may be tapping into modality-specific mechanisms not captured by the visual flanker task.

6.4.2 Applications to other populations and clinical relevance

Several aspects of the task used in this thesis could be of interest when studying informational interference in different populations, in particular for children. The picture-selection task was based on a similar task that has been used to study subject and object relative acquisition in children (Adani, 2011; Arosio, Adani, & Guasti, 2009). The stimuli were designed to appeal to younger audiences through the use of cartoon-like pictures, familiar animals and characters. The vocabulary of the target sentences was chosen to be as high-frequency as possible, so the task can be used with younger children and populations with low vocabulary. Furthermore, no verbal response is required, which makes this task accessible to groups of listeners for whom spoken language production is atypical or impaired (e.g. dysarthria, dyspraxia, fluency disorders, aphasia).

Informational interference is particularly important to study in children, since this group is often exposed to classroom environments with competing speech, and academic performance depends on children’s ability to focus their attention in adverse listening conditions. Children are more affected than adults by noisy environments. However, it is unclear whether they are more or less affected by IM than adults. It would seem that the developmental trajectory for dealing with stationary maskers is different to that of modulated maskers, since adult-like performance in speech intelligibility with stationary maskers is around the age of 6 years (Schneider, Trehub, Morrongiello, & Thorpe, 1989), whereas adult-like performance with competing speech is only reached by age 10 (Wightman & Kistler, 2005). It is possible that the added difficulty of a competing talker would be particularly resource-demanding for younger children, who already struggle with modulated masks. Furthermore, as with adult studies, fewer studies have focused on children’s speech comprehension in adverse conditions (e.g. Klatte, Lachmann, & Meis, 2010; Lewis, Manninen, Valente, & Smith, 2014; Sullivan, Osman, & Schafer, 2015) than speech intelligibility in adverse conditions. Thus, studying children’s performance in the sentence comprehension task presented in this thesis could contribute to our understanding of informational interference in children, as well as the underlying cognitive factors at play.

The mechanisms behind informational interference can be further explored and pinpointed by assessing listeners who find speech-in-speech tasks particularly demanding. For example, individuals with autism spectrum disorder (ASD) may be particularly affected by competing speech. Indeed, auditory processing is reported to be impaired and/or qualitatively different in ASD (O'Connor, 2012). Language delay and language impairment can also be central for many individuals. However the nature of the interaction between language and auditory processing in ASD is still under scrutiny. In particular, the tasks that have been most widely used to study auditory processing assess speech *intelligibility* in background noise, and do not assess comprehension specifically. Furthermore, the use of background noise, whether fluctuating or stationary, does not take into account the type of masking that children are most often exposed to, i.e. speech in background speech. Individuals with ASD show particularly impaired performance with modulated maskers as compared to steady-state maskers (e.g. Alcántara, Weisblatt, Moore, & Bolton (2004), Mair (2013)). This profile is different to specific language impairment (Ziegler, Pech-Georgel, George, & Lorenzi, 2011) and dyslexia (Ziegler, Pech-Georgel, George, & Lorenzi, 2009), where individuals perform poorly regardless of the mask type. It is conceivable that the difference between these groups lies in higher-order cognitive factors, such as executive functions and WM. Most studies investigating speech perception in adverse conditions in ASD have focused on 'high-functioning' adults. However, as children are often required to learn in environments with background speech, and as the social demands of a classroom are particularly difficult to deal with for ASD children, it is conceivable that those with better performance for speech-in-speech tasks have more resources available for learning and socializing with their peers. In addition to informing our understanding of the mechanisms of informational interference, identifying the cognitive factors involved in dealing with informational interference could allow educators and therapists to tailor interventions based on specific cognitive hearing profiles.

6.5 General conclusions

The results of the first five experiments indicated that under certain circumstances listeners are remarkably robust at understanding sentences in the presence of a competing talker, compared to energetic mask controls. Despite a number of manipulations designed to increase cognitive processing load (syntactic complexity, proficiency, SNR), participants' sentence comprehension was just as effective with a competing talker as with speech-modulated noise or time-reversed speech. Although from these results it seemed that participants indiscriminately blocked out the mask, Experiment 6 suggested that listeners monitored both the target and competing streams, thus leading to faster responses when the content of the competing talker aided their response.

Considering that informational interference did not occur in any of the experiments in this thesis, the conditions under which informational interference can be observed are still not clear. Informational interference did not arise for normal-hearing typically developing young adults when the target and competitor were acoustically distinct enough to lead to successful streaming and object formation and when visual information reduced the number of lexical candidates. The latter condition resembles many everyday conversations, where contextual information is often given.

Future research could explore the conditions under which informational interference might appear, by using different energetic mask controls, manipulating low-level informational masking and/or increasing task demands. Additionally, the task developed for this thesis is ideal to focus on exploring informational interference in populations with impaired or atypical language and/or auditory processing, who may experience difficulty in following conversations in the presence of competing speech. The cognitive processes involved in informational interference could be pinpointed by identifying the underlying cognitive factors that explain individual differences.

Appendix A Summary of studies comparing a competing talker to EM controls

Studies that have found a detrimental effect of a competing talker compared to energetic mask controls						
Reference	**Materials**	**Task**	**Gender of target and competing talker**	**Other masks**	**SNRs /SRTs**	**Additional comments**
Brungart (2001)	Coordinate Response Measure (CRM) (Moore, 1981)	Following instructions	Male and female, with all combinations.	Gaussian noise, Speech-modulated noise based on the competing speech	-12 to +15 dB in 3 dB steps for CT ; -21 to 0 dB in 3 dB steps for SMN	Depended on SNR: lowest SNRs flipped difference (CT more accurate than noise)
Brungart et al. (2013) Experiment 1	CRM	Detection (least complex), discrimination or identification (most complex).	Target male Competing talker male	Stationary speech-shaped noise	-56 dB to 8 dB in 4 dB steps. SRT_{75}: -8 dB SSN, -18 dB CT	Effect found only in the identification task (most complex)
Brungart et al. (2013) Experiment 2	CRM Competing talker was irrelevant continuous speech.	Following instructions with either monaural presentation, binaural with target in known ear, binaural with target in unknown ear, or binaural with response to both ears.	Target male Competing talker male (different talker)	Stationary speech-shaped noise Reversed competing talker Speech-modulated noise based on the competing speech Four-talker babble (2 male, 2 female)	19 SNRs ranging from -27 dB to 21 dB. SRT_{80} (nearest integer): -16dB for CT and RCT; -12 dB for SMN, -5 dB SSN, -3 dB for babble.	Effect found in the relative decrease in performance between least complex and most complex tasks for CT vs SMN; Effect not found in overall SRTs; No difference between CT and RCT
Brungart et al. (2013) Experiment 3	Highly probable (HP) or anomalous (AP) sentences created from R-SPIN sentences (Bilger et al., 1984)	Sentence repetition and 1-back repetition	Target female 2-talker competing speech female (same talker), fairy tale passages.	Stationary speech-shaped noise based on the competing speech	SRT_{80} (nearest integer): -1 dB (HP) and +2 dB (AP) SNR for CT; -3 dB (HP) and 0 dB (AP) SNR for SSN.	Effect found in the 1-back task (most complex). Not found in simple repetition

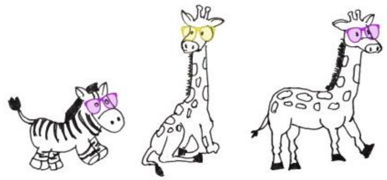
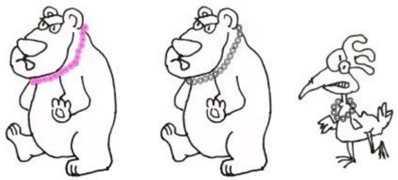
Francart, van Wieringen, and Wouters (2011)	Everyday Dutch sentences (Versfeld, Daalder, Festen, & Houtgast, 2000) and Leuven intelligibility sentence test (LIST) sentences in Dutch.	Sentence repetition	Target male and female Competing talker male	Stationary speech-shaped noise, 20-talker English babble, Fluctuating noise (ICRA-250 and kICRA), Unintelligible speech (ISTS), Swedish competing talker	SRT_{50} -15.2 dB SNR (LIST target, Swedish CT) to -0.1 dB SNR (male target, 20-talker babble).	Native CT worse than non-native CT
Helfer and Freyman (2014)	Revised version of the Theo-Victor-Michael (TVM) sentences (Helfer & Freyman, 2009) Competing talker: TVM sentences or Rainbow passage	Sentence repetition	Target female Competing talker female (different voices)	Stationary speech-shaped noise, 2 competing talkers (TVM and Rainbow passage)	-3, 0, +3 dB SNR	Effect found for TVM vs SSN; Not found for Rainbow Passage vs SSN
Kidd et al., (2014)	BU corpus (Kidd et al., 2008): five-word sentences, syntactically correct or random order	Selecting words from a list of eight alternatives for each word in the sentence.	Target female Competing talkers female	Speech-shaped noise bursts		Greater difference between syntactic and random sentences for CT than for noise.
Koelewijn, Zekveld, Festen and Kramer (2012)	Everyday Dutch sentences (Versfeld et al., 2000)	Sentence repetition	Target female Competing talker male with same long-term average frequency spectrum as target	Stationary speech-shaped noise, Speech-modulated noise based on the competing talker	SRT_{50} : -12.2 dB SNR for SMN, -13.2 dB SNR for CT. SRT_{84} : -6.4 dB SNR for SMN and CT.	Effect found for pupil dilation; Not found for SRTs
Rhebergen, Versfeld, and Dreschler (2005)	Everyday Dutch sentences (Versfeld et al., 2000)	Sentence repetition	Target male Competing talker female	Reversed competing talker (RCT) in native language (Dutch), CT in unknown language (Swedish), RCT in unknown language (Swedish)	SRT_{50} : -15.2 dB SNR for RCT in native language (Dutch), -13.9 SNR for unknown (Swedish) CT, -11.6 dB SNR for unknown (Swedish) RCT, -10.9 dB SNR for native (Dutch) CT	Effect found for Dutch CT vs Dutch RCT; Swedish CT same as Swedish RCT

Sörqvist and Rönnberg (2012)	Stories about fictitious cultures (2 stories, 10 short paragraphs each)	Listen to a sentence, choose which sentence was presented (4 AFC recognition); Listen to stories and answer comprehension questions.	Target male Competing talker male (different voice)	Spectrally-rotated speech	+5 dB SNR	Effect found for both recognition and comprehension.
Trammell and Speaks (1970)	Not specified	Sentence repetition	No information	Reversed competing talker	SRT_{50}: -10.9 dB for CT, -16.1 dB for RCT	
Studies that have not found a detrimental effect of a competing talker compared to energetic mask controls						
Reference	**Materials**	**Task**	**Gender of target and competing talker**	**Other masks**	**SNRs / SRTs**	**Additional comments**
Boebinger et al. (2015)	Bamford-Kowal-Bench (BKB) sentences (Bench, Kowal, & Bamford, 1979)	Sentence repetition	Target female Competing talker male	Stationary speech-shaped noise, Speech-modulated noise, Spectrally rotated speech based on the CT	SRT_{50}: -11.9 dB SNR for CT, -10.3 dB SNR for spectrally-rotated speech, -7.8 dB SNR for SMN.	
Dirks and Bower (1969)	Synthetic sentences by Speaks and Jerger (1965)	Sentence repetition	Target male Competing talker male	Reversed competing talker, Competing talker in Latin (forward and reversed)	-12, -18, -24 and -30 dB; 20% accuracy at -30 dB SNR and 45% at -24 dB SNR.	
Duquesnoy (1983)	Sentences by Plomp and Mimpen (1979)	Sentence repetition (binaural/ monaural/ different spatial configurations)	Target female Competing talker male	Speech-shaped noise, Reversed competing talker	For binaural with both sources presented at the front, SRT_{50}: -10.7 dB (CT), -17.6 dB (RCT and SSN)	CT better than RCT and SSN
Festen and Plomp (1990)	Short everyday sentences	Sentence repetition	Male and female, with all combinations	Stationary speech-shaped noise Single-band modulated noise, Two-band	SRT_{50}: -11.4 dB SNR for CT, -8.4 and -9.5 dB SNR for SMN, and -2.3 dB for the RCT when target and	CT better than other masks

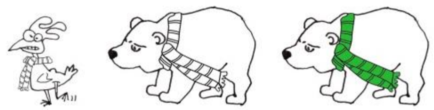
				modulated noise, RCT	masker were both female.	
Hygge et al. (1992)	Target and competitor material: continuous fiction story, with and without visual information (lip-reading)	Adjust the sound level of the target to "the subjective level at which it was just possible to understand"	Target female Competing talker male	Stationary speech-spectrum random noise, Reversed competing talker	Mean level in audio-visual: -2.3 dB (SSN), -12.1 dB (CT), -12.5 dB SNR (RCT), Mean level in visual: -0.6 dB (SSN), -9.2 dB (CT), -10.1 dB (RCT)	
Iyer, Brungart, and Simpson (2010)	Coordinate Response Measure (Bolia et al., 2000)	Following instructions	Male and female, but target and competing talker were always the same gender	Stationary speech-shaped noise, Speech-modulated noise, Reversed competing talker	-8, -4, 0, +4, +8 dB SNR	
Scott, Rosen, Wickham, and Wise (2004)	Bamford-Kowal-Bench (BKB) sentences (Bench et al., 1979)	Sentence repetition (behavioural measures); listening for meaning (PET)	Target male Competing talker male	Speech-shaped noise based on the competing talker	-6, -3, 0, +3 dB SNR for CT, -3, 0, +3, +6 dB SNR for SSN	PET data showed a difference between CT and SSN brain responses
Qin and Oxenham (2003)	H.I.N.T. sentences (Nilsson et al., 1994)	Sentence repetition	Target male Competing talker male or female	Stationary speech-shaped noise, Speech-modulated noise	-20 to +5 dB in 5 dB steps. SRT_{50}: -10.3 dB SNR for female CT, -11.3 dB SNR for male competing talker, -9.1 dB SNR for speech-modulated noise	

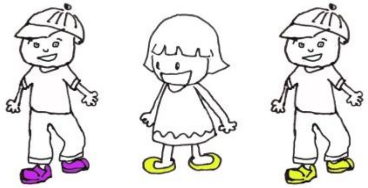
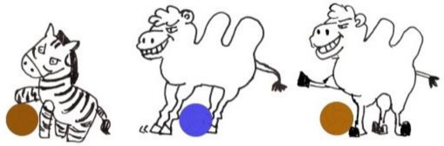
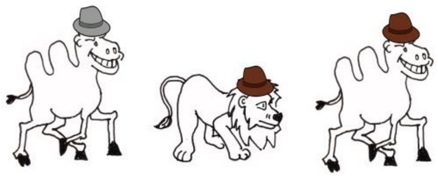
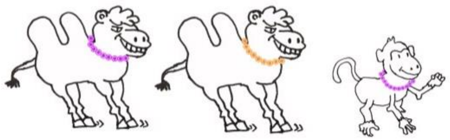
Table A.1. Summary of main characteristics (task, voice, other masks in addition to competing talker, SNRs, findings) for various studies investigating the effect of a competing talker on sentence intelligibility, with normal-hearing young native listeners. SNRs are only reported for modulated masks. Studies using stationary masks are included in this table for completeness, but the relevant comparisons between these studies and the current thesis are for fluctuating makers (SMN, spectrally-rotated speech, reversed or forward speech). Studies that have found an effect of a competing talker compared to energetic mask controls are reported in the top part of the table.

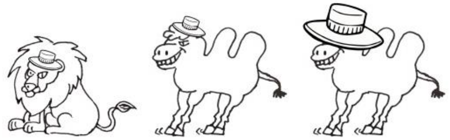
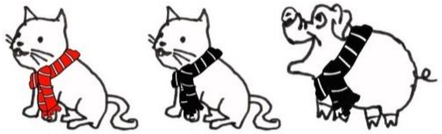
Appendix B Target sentences with corresponding pictures[28] and competing talker sentences

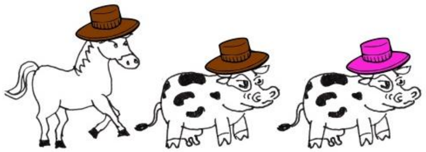
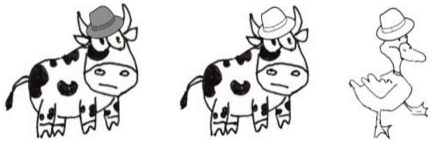
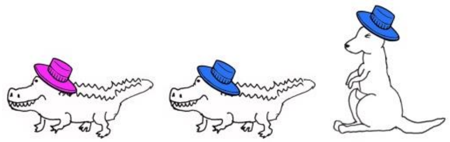
Target type	Target sentence and corresponding picture	Competing talker condition[29]	Competing talker sentence
Simple	Show the zebra with the purple glasses 	Neutral 1[30]	The old woman is at home. The front garden is pretty
		Neutral 2	The manager with the unusual paper is sadly wrong
		Incongruent	The giraffe with the yellow glasses is unhealthily emotional
		Congruent	The zebra with the purple glasses is unhealthily emotional
Simple	Show the bear with the grey ball 	Neutral 1	They're playing in the park. She paid for the bread
		Neutral 2	The manager with the colourful doll is unbearably messy
		Incongruent	The fox with the grey ball is ordinarily strong
		Congruent	The bear with the grey ball is ordinarily strong
Simple	Show the bear with the white necklace 	Neutral 1	They're going out tonight. They're watching the train go by
		Neutral 2	The lawyer with the beautiful painting is terribly loud
		Incongruent	The hen with the white necklace is perfectly frozen
		Congruent	The bear with the white necklace is perfectly frozen

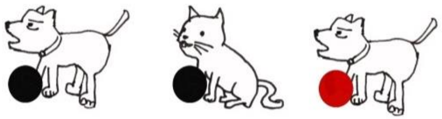
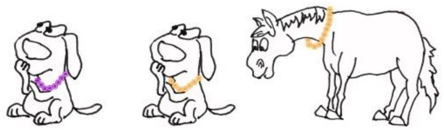
[28] Pictures were presented in Experiments 2 - 6.
[29] For Experiments 1-5 (Chapters 2 - 4), only the “Neutral 1” condition was presented. For Experiment 6 (Chapter 5), all conditions were presented.
[30] All Neutral 1 sentences were HINT sentences taken from Nilsson et al. (1994), with modifications reported in Appendix C

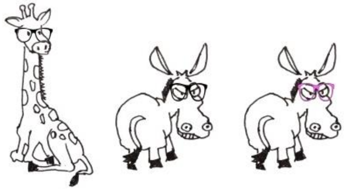
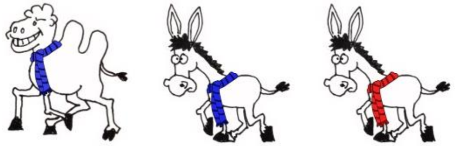
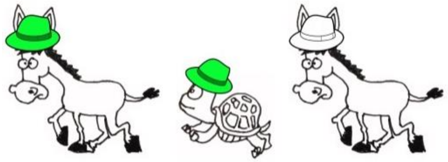
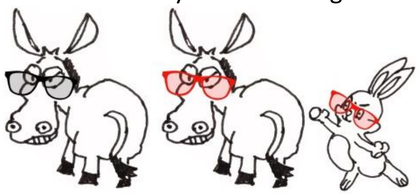
Simple	Show the bear with the white ribbon 	Neutral 1	They knocked on the window. He cut his index finger
		Neutral 2	The athlete with the crispy apple is especially greedy
		Incongruent	The ghost with the white ribbon is ominously dangerous
		Congruent	The bear with the white ribbon is ominously dangerous
Simple	Show the bear with the white scarf 	Neutral 1	They are coming for dinner. The bakery is open
		Neutral 2	The reporter with the colourful apple is uncharacteristically emotional
		Incongruent	The hen with the white scarf is never emotional
		Congruent	The bear with the white scarf is never emotional
Simple	Show the boy with the green gloves 	Neutral 1	They hear a funny noise. The engine is running
		Neutral 2	The hedgehog with the rotten character is fairly smelly
		Incongruent	The lady with the green gloves is fairly gentle
		Congruent	The boy with the green gloves is fairly gentle
Simple	Show the boy with the pink ball 	Neutral 1	The team is playing well. The driver waited for me
		Neutral 2	The hamster with the unusual apple is already old
		Incongruent	The cat with the pink ball is probably heavy
		Congruent	The boy with the pink ball is probably heavy
Simple	Show the boy with the red trousers 	Neutral 1	They laughed at his story. They are crossing the street
		Neutral 2	The owl with the piercing eyes is completely round
		Incongruent	The girl with the red trousers is naturally sweet
		Congruent	The boy with the red trousers is naturally sweet

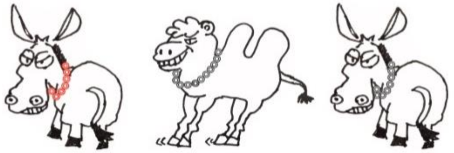
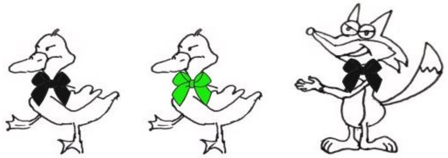
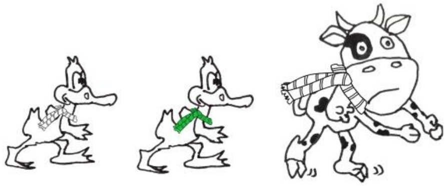
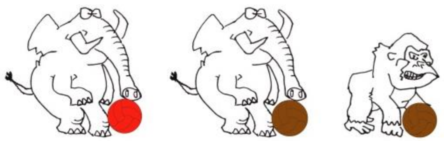
Simple	Show the boy with the yellow shoes 	Neutral 1	The buckets fill up quickly. Flowers grow in the garden
		Neutral 2	The octopus with the unusual flower is thoroughly sticky
		Incongruent	The girl with the yellow shoes is exceedingly lively
		Congruent	The boy with the yellow shoes is exceedingly lively
Simple	Show the camel with the brown ball 	Neutral 1	They called an ambulance. The sun melted the snow
		Neutral 2	The teacher with the astonishing character is decidedly strange
		Incongruent	The zebra with the brown ball is extremely expensive
		Congruent	The camel with the brown ball is extremely expensive
Simple	Show the camel with the brown hat 	Neutral 1	The family likes fish. The sweet shop is empty
		Neutral 2	The professor with the delicious cereal is probably old
		Incongruent	The lion with the brown hat is very rude
		Congruent	The camel with the brown hat is very rude
Simple	Show the camel with the pink necklace 	Neutral 1	The dinner plate is hot. She made her bed and left
		Neutral 2	The photographer with the abundant money is sadly violent
		Incongruent	The donkey with the pink necklace is sadly boring
		Congruent	The camel with the pink necklace is sadly boring
Simple	Show the camel with the purple necklace 	Neutral 1	The girl is washing her hair. She took off her fur coat
		Neutral 2	The painter with the difficult character is offensively corrupt
		Incongruent	The monkey with the purple necklace is reasonably fast
		Congruent	The camel with the purple necklace is reasonably fast

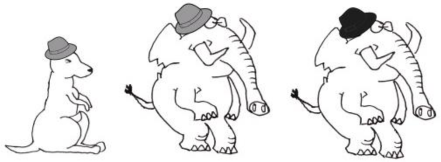
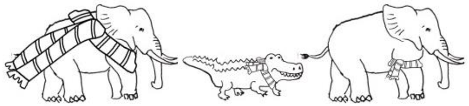
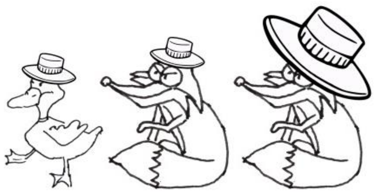
Simple	Show the camel with the small hat 	Neutral 1	The player lost a shoe. Someone is crossing the road
		Neutral 2	The therapist with the crispy apple is uncharacteristically positive
		Incongruent	The lion with the small hat is interestingly grateful
		Congruent	The camel with the small hat is interestingly grateful
Simple	Show the cat with the black scarf 	Neutral 1	Somebody stole the money. It's getting cold in here
		Neutral 2	The waiter with the bitter character is slightly scary
		Incongruent	The pig with the black scarf is quite different
		Congruent	The cat with the black scarf is quite different
Simple	Show the cat with the red ribbon 	Neutral 1	The towel fell on the floor. Mother cut the birthday cake
		Neutral 2	The architect with the wrong personality is very expensive
		Incongruent	The pig with the red ribbon is unfortunately young
		Congruent	The cat with the red ribbon is unfortunately young
Simple	Show the cat with the white scar  f	Neutral 1	The mother heard the baby. They are coming for dinner
		Neutral 2	The artist with the rotten money is ominously scary
		Incongruent	The dog with the white scarf is thoroughly funny
		Congruent	The cat with the white scarf is thoroughly funny
Simple	Show the cat with the yellow scarf 	Neutral 1	Mother got a sauce pan. A sharp knife is dangerous
		Neutral 2	The doctor with the astonishing doll is unusually excellent
		Incongruent	The boy with the yellow scarf is horrifyingly noisy
		Congruent	The cat with the yellow scarf is horrifyingly noisy

Simple	Show the cow with the blue necklace 	Neutral 1	The bus stopped suddenly. They finished dinner on time
		Neutral 2	The officer with the unlimited money is sadly dirty
		Incongruent	The dog with the blue necklace is really bright
		Congruent	The cow with the blue necklace is really bright
Simple	Show the cow with the brown hat 	Neutral 1	The jam jar is full. The baby is pretty
		Neutral 2	The pilot with the difficult personality is brutally fierce
		Incongruent	The horse with the brown hat is frankly rare
		Congruent	The cow with the brown hat is frankly rare
Simple	Show the cow with the pink scarf 	Neutral 1	The boy is running away. Mother shut the window
		Neutral 2	The artist with the unusual canvas is extraordinarily ancient
		Incongruent	The pig with the pink scarf is very worried
		Congruent	The cow with the pink scarf is very worried
Simple	Show the cow with the white hat 	Neutral 1	There was a bad train wreck. The milk is in the pitcher
		Neutral 2	The professor with the difficult puzzle is possibly soft
		Incongruent	The duck with the white hat is unhealthily lazy
		Congruent	The cow with the white hat is unhealthily lazy
Simple	Show the crocodile with the blue hat 	Neutral 1	Her sister stayed for lunch. The cat drank from the saucer
		Neutral 2	The photographer with the tiny apple is luckily precise
		Incongruent	The kangaroo with the blue hat is possibly hot
		Congruent	The crocodile with the blue hat is possibly hot

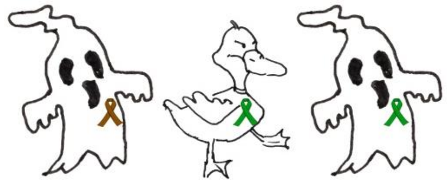
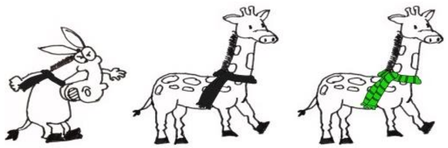
Simple	Show the crocodile with the blue necklace 	Neutral 1	School got out early today. They took some food outside
		Neutral 2	The magician with the mysterious canvas is somewhat interesting
		Incongruent	The elephant with the blue necklace is unusually fierce
		Congruent	The crocodile with the blue necklace is unusually fierce
Simple	Show the crocodile with the green ball 	Neutral 1	She looked in her mirror. The kitchen window was clean
		Neutral 2	The singer with the unusual money is wonderfully emotional
		Incongruent	The gorilla with the green ball is really soft
		Congruent	The crocodile with the green ball is really soft
Simple	Show the dog with the black ball 	Neutral 1	The road goes up a hill. They went on holidays
		Neutral 2	The baby with the tiny doll is especially soft
		Incongruent	The cat with the black ball is momentarily violent
		Congruent	The dog with the black ball is momentarily violent
Simple	Show the dog with the green ball 	Neutral 1	The book tells a story. They're going out tonight
		Neutral 2	The baby with the beautiful flower is truly warm
		Incongruent	The cow with the green ball is systematically polite
		Congruent	The dog with the green ball is systematically polite
Simple	Show the dog with the orange necklace 	Neutral 1	The woman cleaned her house. They had two empty bottles
		Neutral 2	The architect with the beautiful paper is unusually ancient
		Incongruent	The horse with the orange necklace is intensely sour
		Congruent	The dog with the orange necklace is intensely sour

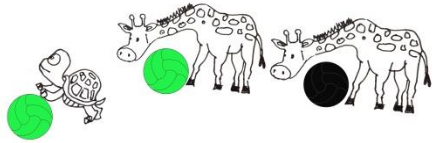
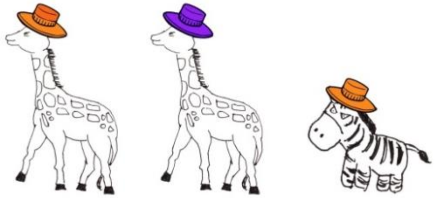
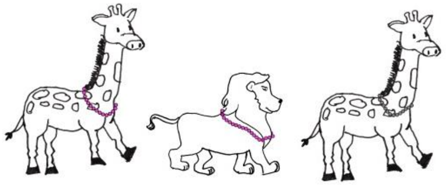
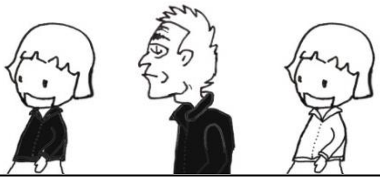
Simple	Show the dog with the white necklace 	Neutral 1	The big boy kicked the ball. The milk is by the front door
		Neutral 2	The architect with the piercing eyes is interestingly emotional
		Incongruent	The cat with the white necklace is irrefutably great
		Congruent	The dog with the white necklace is irrefutably great
Simple	Show the donkey with the black glasses 	Neutral 1	The broom is in the corner. The house had a nice garden
		Neutral 2	The artist with the rotten cereal is annoyingly greedy
		Incongruent	The giraffe with the black glasses is somewhat greedy
		Congruent	The donkey with the black glasses is somewhat greedy
Simple	Show the donkey with the blue scarf 	Neutral 1	The policeman knows the way. The bananas were too ripe
		Neutral 2	The artist with the original attitude is extremely creative
		Incongruent	The camel with the blue scarf is luckily positive
		Congruent	The donkey with the blue scarf is luckily positive
Simple	Show the donkey with the green hat 	Neutral 1	They rode their bicycles. A boy ran down the path
		Neutral 2	The athlete with the unusual character is utterly creative
		Incongruent	The turtle with the green hat is finally dry
		Congruent	The donkey with the green hat is finally dry
Simple	Show the donkey with the red glasses 	Neutral 1	The bag fell off the shelf. The shop closes for lunch
		Neutral 2	The athlete with the wrong paper is excessively great
		Incongruent	The rabbit with the red glasses is really old
		Congruent	The donkey with the red glasses is really old

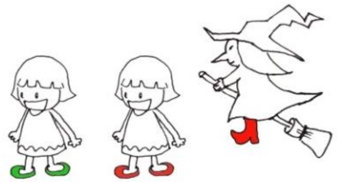
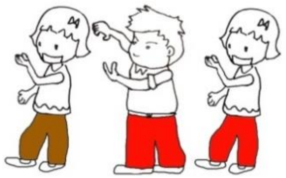
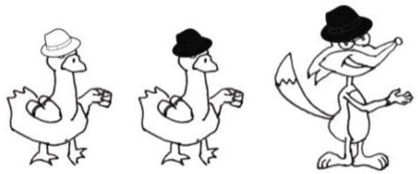
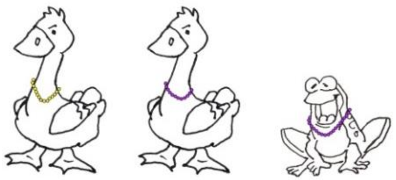
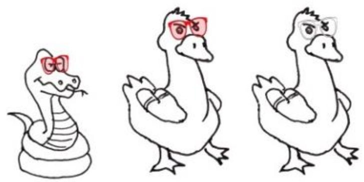
Simple	Show the donkey with the white necklace 	Neutral 1	The scissors are very sharp. The man is painting the sign
		Neutral 2	The baby with the difficult puzzle is unbearably delicate
		Incongruent	The camel with the white necklace is unbearably wet
		Congruent	The donkey with the white necklace is unbearably wet
Simple	Show the duck with the black ribbon 	Neutral 1	They are drinking coffee. The apple pie is baking
		Neutral 2	The baby with the astonishing eyes is somewhat worried
		Incongruent	The fox with the black ribbon is nearly hot
		Congruent	The duck with the black ribbon is nearly hot
Simple	Show the duck with the pink ribbon 	Neutral 1	The lady washed the shirt. He found his brother hiding
		Neutral 2	The builder with the abundant money is ominously corrupt
		Incongruent	The ghost with the pink ribbon is particularly warm
		Congruent	The duck with the pink ribbon is particularly warm
Simple	Show the duck with the white scarf 	Neutral 1	The little girl is happy. He is sucking his thumb
		Neutral 2	The builder with the wrong painting is exceedingly sour
		Incongruent	The cow with the white scarf is utterly cunning
		Congruent	The duck with the white scarf is utterly cunning
Simple	Show the elephant with the brown ball 	Neutral 1	Flowers grow in the garden. The dishcloth is soaking wet
		Neutral 2	The cook with the rotten apple is momentarily fierce
		Incongruent	The gorilla with the brown ball is terribly colourful
		Congruent	The elephant with the brown ball is terribly colourful

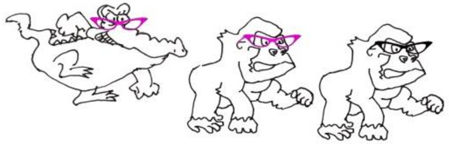
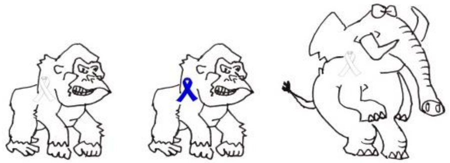
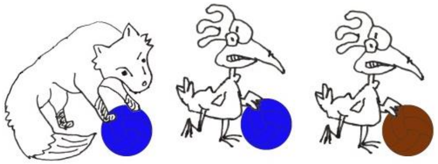
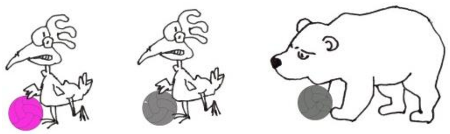
Simple	Show the elephant with the grey hat 	Neutral 1	They waited for an hour. The two farmers were talking
		Neutral 2	The cook with the fresh cereal is intensely terrified
		Incongruent	The kangaroo with the grey hat is extraordinarily complex
		Congruent	The elephant with the grey hat is extraordinarily complex
Simple	Show the elephant with the small scarf 	Neutral 1	The match boxes are empty. The teapot is very hot
		Neutral 2	The doctor with the unlimited money is momentarily bad
		Incongruent	The crocodile with the small scarf is exceedingly corrupt
		Congruent	The elephant with the small scarf is exceedingly corrupt
Simple	Show the fox with the green scarf 	Neutral 1	The boy did a handstand. The baby slept all night
		Neutral 2	The doctor with the difficult attitude is sometimes smelly
		Incongruent	The goose with the green scarf is completely ill
		Congruent	The fox with the green scarf is completely ill
Simple	Show the fox with the orange glasses 	Neutral 1	The silly boy was hiding. The football hit the goalpost
		Neutral 2	The driver with the unique personality is unhealthily smelly
		Incongruent	The bear with the orange glasses is disgustingly dirty
		Congruent	The fox with the orange glasses is disgustingly dirty
Simple	Show the fox with the small hat 	Neutral 1	He got mud on his shoes. The rice pudding is ready
		Neutral 2	The driver with the original character is interestingly popular
		Incongruent	The duck with the small hat is excessively scary
		Congruent	The fox with the small hat is excessively scary

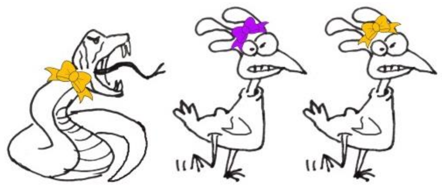
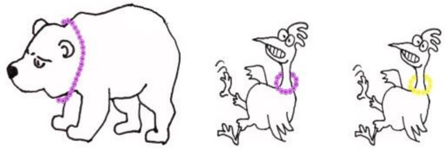
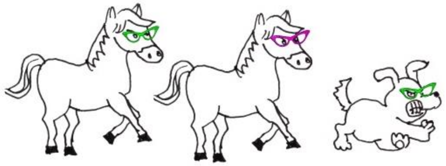
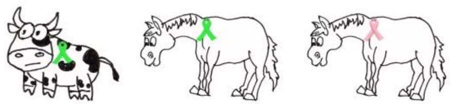
Simple	Show the fox with the white glasses 	Neutral 1	The little boy left home. The towel is near the sink
		Neutral 2	The farmer with the unique painting is precisely good
		Incongruent	The hen with white glasses is dismally bad
		Congruent	The fox with the white glasses is dismally bad
Simple	Show the frog with the brown ball 	Neutral 1	They ate the lemon pie. The dinner plate is hot
		Neutral 2	The farmer with the unusual personality is naturally cunning
		Incongruent	The duck with the brown ball is positively round
		Congruent	The frog with the brown ball is positively round
Simple	Show the frog with the pink glasses 	Neutral 1	They're clearing the table. Her shoes were very dirty
		Neutral 2	The fireman with the delicious cereal is thoroughly confused
		Incongruent	The snake with the pink glasses is especially interesting
		Congruent	The frog with the pink glasses is especially interesting
Simple	Show the frog with the pink necklace 	Neutral 1	They had a wonderful day. The boy slipped on the stairs
		Neutral 2	The fireman with the unique puzzle is nearly warm
		Incongruent	The goose with the pink necklace is already distracted
		Congruent	The frog with the pink necklace is already distracted
Simple	Show the frog with the white necklace 	Neutral 1	The baby wants his bottle. The puppy played with the ball
		Neutral 2	The judge with the unlimited money is offensively loud
		Incongruent	The hen with the white necklace is beautifully polite
		Congruent	The frog with the white necklace is beautifully polite

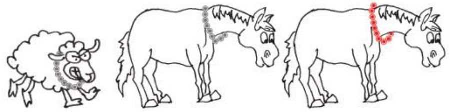
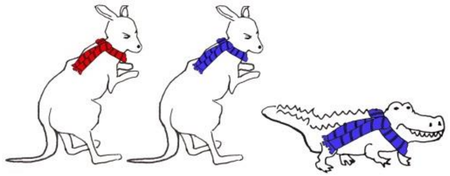
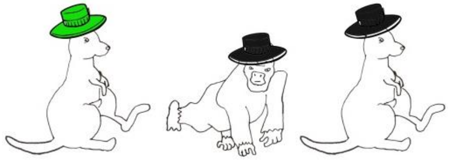
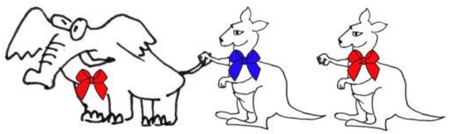
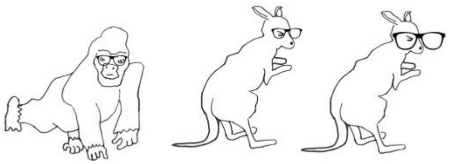
Simple	Show the ghost with the big bag 	Neutral 1	The boy went to bed early. The two children are laughing
		Neutral 2	The judge with the piercing attitude is finally done
		Incongruent	The bear with the big bag is finally cold
		Congruent	The ghost with the big bag is finally cold
Simple	Show the ghost with the black ball 	Neutral 1	A grocer sells butter. The three girls are listening
		Neutral 2	The lawyer with the difficult personality is decidedly interesting
		Incongruent	The sheep with the pink ball is disappointingly important
		Congruent	The ghost with the black ball is disappointingly important
Simple	Show the ghost with the brown hat 	Neutral 1	The cups are on the table. They knocked on the window
		Neutral 2	The lawyer with the unlimited cereal is horrifyingly wrong
		Incongruent	The witch with the brown hat is annoyingly loud
		Congruent	The ghost with the brown hat is annoyingly loud
Simple	Show the ghost with the green ribbon 	Neutral 1	The dog jumped on the chair. The shoes are very dirty
		Neutral 2	The magician with the beautiful imagination is horrifyingly messy
		Incongruent	The duck with the green ribbon is always wrong
		Congruent	The ghost with the green ribbon is always wrong
Simple	Show the giraffe with the black scarf 	Neutral 1	My mother stirred her tea. A mouse ran into the hole
		Neutral 2	The magician with the colourful cereal is systematically sticky
		Incongruent	The donkey with the black scarf is positively delicate
		Congruent	The giraffe with the black scarf is positively delicate

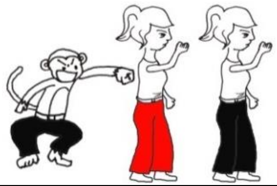
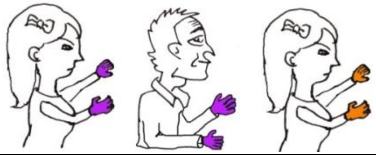
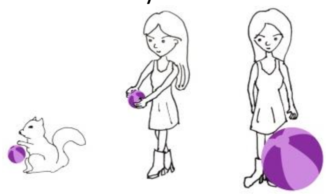
Simple	Show the giraffe with the green ball 	Neutral 1	The food is expensive. He really scared his sister
		Neutral 2	The manager with the unique flower is systematically funny
		Incongruent	The turtle with the green ball is actually rich
		Congruent	The giraffe with the green ball is actually rich
Simple	Show the giraffe with the orange hat 	Neutral 1	The sweet shop is empty. The children washed the plates
		Neutral 2	The manager with the wrong money is exceedingly gentle
		Incongruent	The zebra with the orange hat is luckily done
		Congruent	The giraffe with the orange hat is luckily done
Simple	Show the giraffe with the pink necklace 	Neutral 1	The towel is near the sink. The children helped their teacher
		Neutral 2	The musician with the astonishing painting is obviously cheerful
		Incongruent	The lion with the pink necklace is slightly sticky
		Congruent	The giraffe with the pink necklace is slightly sticky
Simple	Show the girl with the black shirt 	Neutral 1	The oven is too hot. The rancher has a bull
		Neutral 2	The dolphin with the mysterious imagination is depressingly heavy
		Incongruent	The man with the black shirt is systematically cheerful
		Congruent	The girl with the black shirt is systematically cheerful
Simple	Show the girl with the brown bag 	Neutral 1	They met some friends at dinner. The old woman is at home
		Neutral 2	The penguin with the original flower is ordinarily dirty
		Incongruent	The boy with the brown bag is disgustingly messy
		Congruent	The girl with the brown bag is disgustingly messy

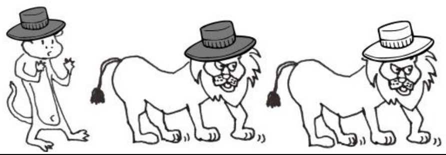
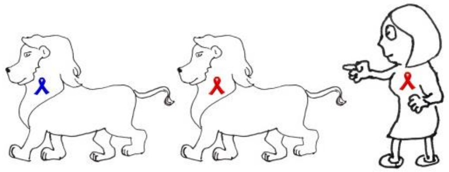
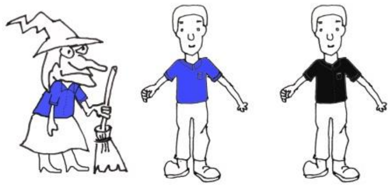
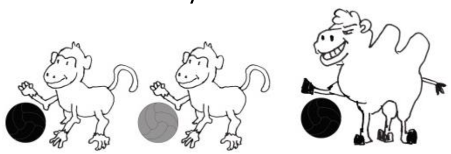
Simple	Show the girl with the red shoes 	Neutral 1	The cat lay on the bed. Mother read the instructions
		Neutral 2	The jellyfish with the beautiful doll is obviously exotic
		Incongruent	The witch with the red shoes is naturally sticky
		Congruent	The girl with the red shoes is naturally sticky
Simple	Show the girl with the red trousers 	Neutral 1	The dog played with a stick. The broom is in the corner
		Neutral 2	The bee with the colourful flower is precisely round
		Incongruent	The boy with the red trousers is ominously glum
		Congruent	The girl with the red trousers is ominously glum
Simple	Show the goose with the black hat 	Neutral 1	The football game is over. The police helped the driver
		Neutral 2	The musician with the tiny eyes is particularly excellent
		Incongruent	The fox with the black hat is disappointingly smelly
		Congruent	The goose with the black hat is disappointingly smelly
Simple	Show the goose with the purple necklace 	Neutral 1	Mother cut the birthday cake. The dog is eating some meat
		Neutral 2	The nurse with the abundant cereal is offensively sour
		Incongruent	The frog with the purple necklace is terribly itchy
		Congruent	The goose with the purple necklace is terribly itchy
Simple	Show the goose with the red glasses 	Neutral 1	The exit was well lit. The ball bounced very high
		Neutral 2	The nurse with the mysterious attitude is nevertheless great
		Incongruent	The snake with the red glasses is really slimy
		Congruent	The goose with the red glasses is really slimy

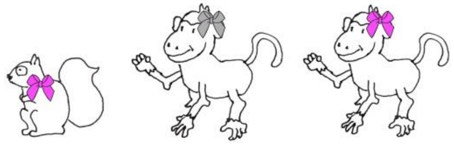
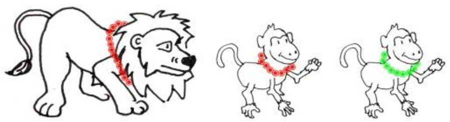
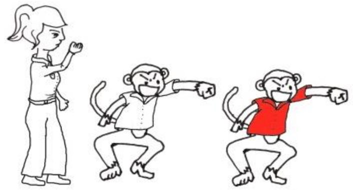
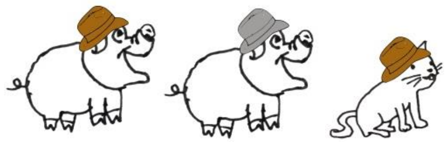
Simple	Show the gorilla with the pink glasses 	Neutral 1	The train stops at the station. They carried some shopping bags
		Neutral 2	The officer with the tiny paper is depressingly corrupt
		Incongruent	The crocodile with the pink glasses is frankly strong
		Congruent	The gorilla with the pink glasses is frankly strong
Simple	Show the gorilla with the purple ribbon 	Neutral 1	Potatoes grow in the ground. They followed the garden path
		Neutral 2	The officer with the bitter apple is extremely heavy
		Incongruent	The kangaroo with the purple ribbon is horrifyingly frozen
		Congruent	The gorilla with the purple ribbon is horrifyingly frozen
Simple	Show the gorilla with the white ribbon 	Neutral 1	The cook is baking a cake. Swimmers can hold their breath
		Neutral 2	The painter with the colourful canvas is beautifully precise
		Incongruent	The elephant with the white ribbon is ordinarily creative
		Congruent	The gorilla with the white ribbon is ordinarily creative
Simple	Show the hen with the blue ball 	Neutral 1	She's calling her daughter. Mother got a sauce pan
		Neutral 2	The painter with the unique attitude is excessively complex
		Incongruent	The fox with the blue ball is frankly popular
		Congruent	The hen with the blue ball is frankly popular
Simple	Show the hen with the grey ball 	Neutral 1	The cows are in the pasture. His father will come home soon
		Neutral 2	The photographer with the mysterious paper is disappointingly loud
		Incongruent	The bear with the grey ball is sadly dangerous
		Congruent	The hen with the grey ball is sadly dangerous

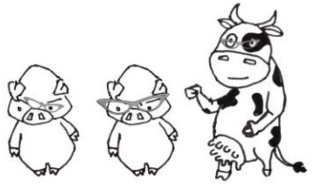
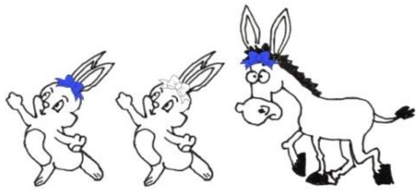
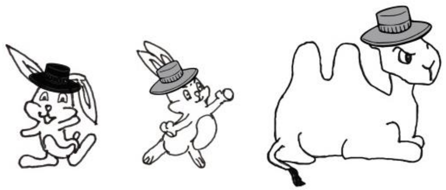
Simple	Show the hen with the orange ribbon 	Neutral 1	They are crossing the street. He played with his toy train
		Neutral 2	The photographer with the original painting is nearly old
		Incongruent	The snake with the orange ribbon is beautifully gentle
		Congruent	The hen with the orange ribbon is beautifully gentle
Simple	Show the hen with the purple necklace 	Neutral 1	The driver waited for me. The train is moving fast
		Neutral 2	The pilot with the delicious character is depressingly strong
		Incongruent	The bear with the purple necklace is unfortunately heavy
		Congruent	The hen with the purple necklace is unfortunately heavy
Simple	Show the horse with the brown scarf 	Neutral 1	The oven door was open. She looked in her mirror
		Neutral 2	The pilot with the unique canvas is brutally burnt
		Incongruent	The dog with the brown scarf is exceedingly fast
		Congruent	The horse with the brown scarf is exceedingly fast
Simple	Show the horse with the green glasses 	Neutral 1	They lost all their money. They wanted some potatoes
		Neutral 2	The professor with the mysterious doll is perfectly tender
		Incongruent	The dog with the green glasses is fairly sweet
		Congruent	The horse with the green glasses is fairly sweet
Simple	Show the horse with the green ribbon 	Neutral 1	He hung up his raincoat. The silly boy was hiding
		Neutral 2	The professor with the rotten flower is never distracted
		Incongruent	The cow with the green ribbon is nevertheless rude
		Congruent	The horse with the green ribbon is nevertheless rude

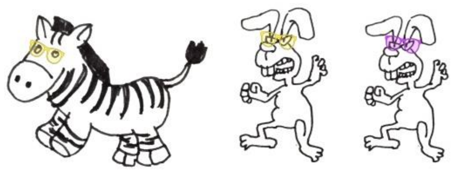
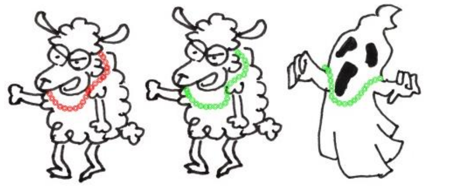
Simple	Show the horse with the grey necklace 	Neutral 1	The picture came from a book. The children are walking home
		Neutral 2	The reporter with the mysterious paper is terribly strong
		Incongruent	The sheep with the grey necklace is probably expensive
		Congruent	The horse with the grey necklace is probably expensive
Simple	Show the kangaroo with the blue scarf 	Neutral 1	A boy ran down the path. The train stops at the station
		Neutral 2	The reporter with the wrong painting is dismally boring
		Incongruent	The crocodile with the blue scarf is completely wet
		Congruent	The kangaroo with the blue scarf is completely wet
Simple	Show the kangaroo with the green hat 	Neutral 1	She argues with her sister. The tall man tied his shoes
		Neutral 2	The salesperson with the colourful attitude is sometimes violent
		Incongruent	The gorilla with the green hat is decidedly noisy
		Congruent	The kangaroo with the green hat is decidedly noisy
Simple	Show the kangaroo with the red ribbon 	Neutral 1	The clown has a funny face. They're shopping for school clothes
		Neutral 2	The salesperson with the unlimited imagination is uncharacteristically sweet
		Incongruent	The elephant with the red ribbon is irrefutably different
		Congruent	The kangaroo with the red ribbon is irrefutably different
Simple	Show the kangaroo with the small glasses 	Neutral 1	The table has three legs. Potatoes grow in the ground
		Neutral 2	The singer with the unique paper is unhealthily dirty
		Incongruent	The gorilla with the small glasses is unusually young
		Congruent	The kangaroo with the small glasses is unusually young

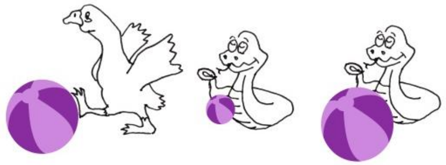
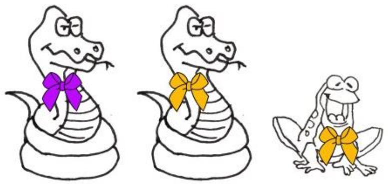
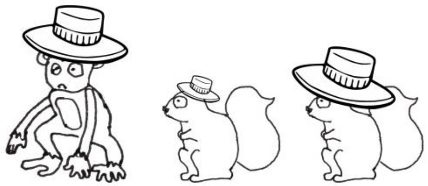
Simple	Show the lady with the black trousers 	Neutral 1	It's getting cold in here. He grew lots of vegetables
		Neutral 2	The eagle with the unusual eyes is really delicate
		Incongruent	The monkey with the black trousers is unbearably precise
		Congruent	The lady with the black trousers is unbearably precise
Simple	Show the lady with the purple glasses 	Neutral 1	They carried some shopping bags. She uses her spoon to eat
		Neutral 2	The koala with the tiny flower is irrefutably tender
		Incongruent	The lion with the purple glasses is reasonably bright
		Congruent	The lady with the purple glasses is reasonably bright
Simple	Show the lady with the purple gloves 	Neutral 1	The dog's chasing the cat. The ground was very hard
		Neutral 2	The tiger with the unusual eyes is especially confused
		Incongruent	The man with the purple gloves is momentarily smelly
		Congruent	The lady with the purple gloves is momentarily smelly
Simple	Show the lady with the small ball 	Neutral 1	She stands near the window. She wore her yellow shirt
		Neutral 2	The panda with the unique doll is naturally heavy
		Incongruent	The squirrel with the small ball is annoyingly rare
		Congruent	The lady with the small ball is annoyingly rare
Simple	Show the lion with the big ribbon 	Neutral 1	The milk is in the pitcher. The girl played with the baby
		Neutral 2	The singer with the unusual painting is perfectly intriguing
		Incongruent	The camel with the big ribbon is officially violent
		Congruent	The lion with the big ribbon is officially violent

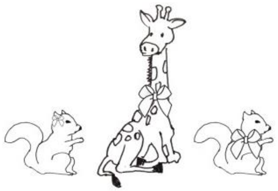
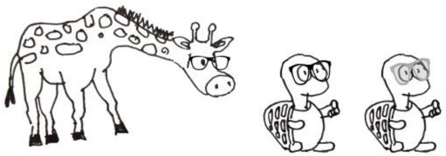
Simple	Show the lion with the grey hat	Neutral 1	They're pushing an old car. She argues with her sister
		Neutral 2	The teacher with the bitter attitude is disgustingly expensive
		Incongruent	The monkey with the grey hat is momentarily terrified
		Congruent	The lion with the grey hat is momentarily terrified
Simple	Show the lion with the red ribbon	Neutral 1	He found his brother hiding. The table has three legs
		Neutral 2	The teacher with the rotten imagination is quite intriguing
		Incongruent	The lady with the red ribbon is perfectly beautiful
		Congruent	The lion with the red ribbon is perfectly beautiful
Simple	Show the man with the blue shirt	Neutral 1	She's paying for her bread. They are drinking coffee
		Neutral 2	The rhinoceros with the bitter character is always hot
		Incongruent	The witch with the blue shirt is ominously great
		Congruent	The man with the blue shirt is ominously great
Simple	Show the man with the red shoes	Neutral 1	He climbed up the ladder. The dog growled at the neighbours
		Neutral 2	The lizard with the unusual painting is unfortunately dangerous
		Incongruent	The witch with the red shoes is especially sour
		Congruent	The man with the red shoes is especially sour
Simple	Show the monkey with the black ball	Neutral 1	She paid for the bread. A boy fell from the window
		Neutral 2	The technician with the wrong doll is very positive
		Incongruent	The camel with the black ball is utterly dry
		Congruent	The monkey with the black ball is utterly dry

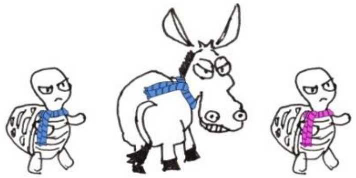
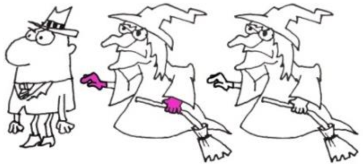
Simple	Show the monkey with the pink ribbon 	Neutral 1	He is sucking his thumb. Yesterday he lost his hat
		Neutral 2	The technician with the rotten flower is unusually precise
		Incongruent	The squirrel with the pink ribbon is very light
		Congruent	The monkey with the pink ribbon is very light
Simple	Show the monkey with the red necklace 	Neutral 1	Mother shut the window. Strawberry jam is sweet
		Neutral 2	The therapist with the tiny painting is momentarily scary
		Incongruent	The lion with the red necklace is never positive
		Congruent	The monkey with the red necklace is never positive
Simple	Show the monkey with the white shirt 	Neutral 1	The ball broke the window. The driver started the car
		Neutral 2	The therapist with the wrong personality is particularly ancient
		Incongruent	The lady with the white shirt is disappointingly complex
		Congruent	The monkey with the white shirt is disappointingly complex
Simple	Show the pig with the big ribbon 	Neutral 1	They had some chocolate pudding. They sat on a wooden bench
		Neutral 2	The waiter with the astonishing attitude is always soft
		Incongruent	The cat with the big ribbon is actually cunning
		Congruent	The pig with the big ribbon is actually cunning
Simple	Show the pig with the brown hat 	Neutral 1	The lady packed her bag. The painter uses a brush
		Neutral 2	The waiter with the colourful imagination is thoroughly greedy
		Incongruent	The cat with the brown hat is excessively colourful
		Congruent	The pig with the brown hat is excessively colourful

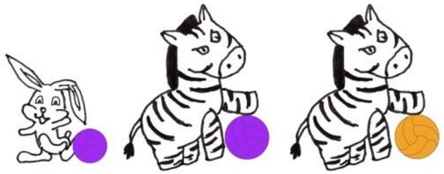
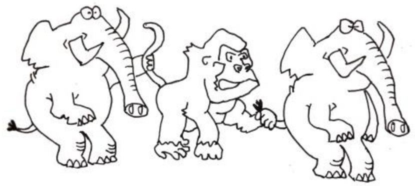
Simple	Show the pig with the small glasses 	Neutral 1	The engine is running. He paid his bill in full
		Neutral 2	The architect with the colourful painting is irrefutably clueless
		Incongruent	The cow with the small glasses is extraordinarily strange
		Congruent	The pig with the small glasses is extraordinarily strange
Simple	Show the rabbit with the black scarf 	Neutral 1	The waiter brought the cream. She bumped her head on the door
		Neutral 2	The artist with the unique imagination is dismally dark
		Incongruent	The monkey with the black scarf is possibly done
		Congruent	The rabbit with the black scarf is possibly done
Simple	Show the rabbit with the blue ball 	Neutral 1	The fruit came in a box. The orange is very sweet
		Neutral 2	The athlete with the colourful eyes is terribly dark
		Incongruent	The lion with the blue ball is irrefutably dirty
		Congruent	The rabbit with the blue ball is irrefutably dirty
Simple	Show the rabbit with the blue ribbon 	Neutral 1	The puppy played with the ball. The clown has a funny face
		Neutral 2	The baby with the crispy cereal is disgustingly slimy
		Incongruent	The donkey with the blue ribbon is particularly early
		Congruent	The rabbit with the blue ribbon is particularly early
Simple	Show the rabbit with the grey hat 	Neutral 1	The rancher has a bull. The little girl is happy
		Neutral 2	The builder with the original canvas is really worried
		Incongruent	The camel with the grey hat is unbearably violent
		Congruent	The rabbit with the grey hat is unbearably violent

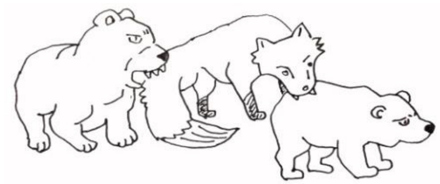
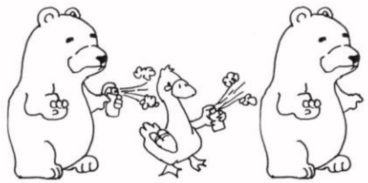
Simple	Show the rabbit with the yellow glasses 	Neutral 1	Flowers can grow in a pot. She's helping her friend move
		Neutral 2	The waiter with the tiny cereal is luckily excellent
		Incongruent	The zebra with the yellow glasses is brutally fierce
		Congruent	The rabbit with the yellow glasses is brutally fierce
Simple	Show the sheep with the green necklace 	Neutral 1	They went on holidays. They washed in cold water
		Neutral 2	The doctor with the mysterious painting is ominously strange
		Incongruent	The ghost with the green necklace is wonderfully light
		Congruent	The sheep with the green necklace is wonderfully light
Simple	Show the sheep with the grey ball 	Neutral 1	There were branches everywhere. The lady went to the store
		Neutral 2	The driver with the unique paper is exceedingly complex
		Incongruent	The dog with the grey ball is disgustingly slimy
		Congruent	The sheep with the grey ball is disgustingly slimy
Simple	Show the sheep with the pink glasses 	Neutral 1	The old gloves are dirty. Mother picked some flowers
		Neutral 2	The farmer with the bitter attitude is probably terrified
		Incongruent	The horse with the pink glasses is uncharacteristically ill
		Congruent	The sheep with the pink glasses is uncharacteristically ill
Simple	Show the sheep with the red scarf 	Neutral 1	Milk comes in a carton. They're clearing the table
		Neutral 2	The fireman with the abundant money is systematically bad
		Incongruent	The horse with the red scarf is nearly burnt
		Congruent	The sheep with the red scarf is nearly burnt

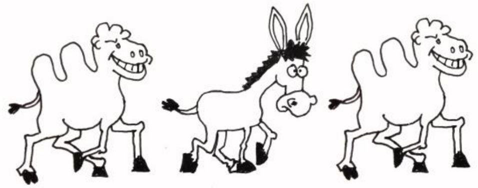
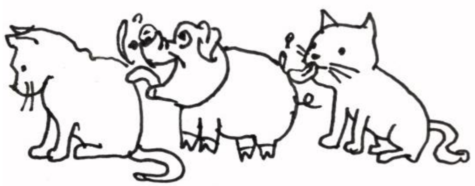
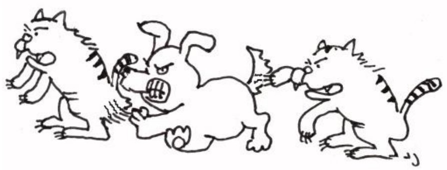
Simple	Show the snake with the big ball 	Neutral 1	The angry man shouted. They had a wonderful day
		Neutral 2	The judge with the beautiful canvas is quite delicate
		Incongruent	The goose with the big ball is interestingly wrong
		Congruent	The snake with the big ball is interestingly wrong
Simple	Show the snake with the grey glasses 	Neutral 1	The tree fell on a house. The dog came home at last
		Neutral 2	The lawyer with the wrong personality is disappointingly distracted
		Incongruent	The hen with the grey glasses is thoroughly grateful
		Congruent	The snake with the grey glasses is thoroughly grateful
Simple	Show the snake with the orange ribbon 	Neutral 1	The shoes are very dirty. Flowers can grow in a pot
		Neutral 2	The magician with the beautiful puzzle is actually popular
		Incongruent	The frog with the orange ribbon is dismally dark
		Congruent	The snake with the orange ribbon is dismally dark
Simple	Show the snake with the red scarf 	Neutral 1	She lost her credit card. They waited for an hour
		Neutral 2	The manager with the beautiful apple is finally good
		Incongruent	The goose with the red scarf is sadly clueless
		Congruent	The snake with the red scarf is sadly clueless
Simple	Show the squirrel with the big hat 	Neutral 1	The truck made it up the hill. Her sister stayed for lunch
		Neutral 2	The musician with the wrong canvas is intensely cunning
		Incongruent	The monkey with the big hat is offensively noisy
		Congruent	The squirrel with the big hat is offensively noisy

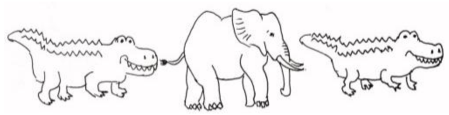
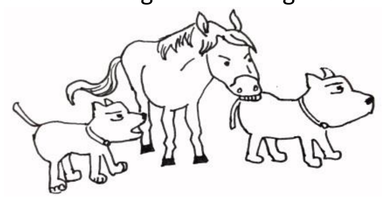
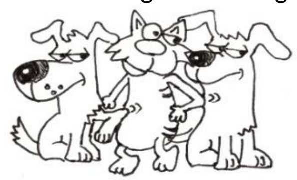
Simple	Show the squirrel with the big ribbon 	Neutral 1	The dog came home at last. The nervous driver got lost
		Neutral 2	The nurse with the astonishing doll is reasonably strong
		Incongruent	The giraffe with the big ribbon is depressingly fragile
		Congruent	The squirrel with the big ribbon is depressingly fragile
Simple	Show the squirrel with the green glasses 	Neutral 1	They sat on a wooden bench. School got out early today
		Neutral 2	The officer with the original personality is ordinarily warm
		Incongruent	The turtle with the green glasses is beautifully round
		Congruent	The squirrel with the green glasses is beautifully round
Simple	Show the squirrel with the grey hat 	Neutral 1	The lady sat in her chair. They're buying some fresh bread
		Neutral 2	The painter with the tiny imagination is reasonably boring
		Incongruent	The monkey with the grey hat is intensely funny
		Congruent	The squirrel with the grey hat is intensely funny
Simple	Show the turtle with the black glasses 	Neutral 1	The rice pudding is ready. Big dogs can be dangerous
		Neutral 2	The photographer with the piercing eyes is nearly done
		Incongruent	The giraffe with the black glasses is terribly loud
		Congruent	The turtle with the black glasses is terribly loud
Simple	Show the turtle with the blue necklace 	Neutral 1	He took the dogs for a walk. The girl is washing her hair
		Neutral 2	The pilot with the original imagination is really interesting
		Incongruent	The squirrel with the blue necklace is reasonably good
		Congruent	The turtle with the blue necklace is reasonably good

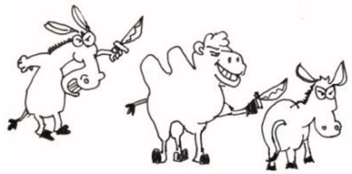
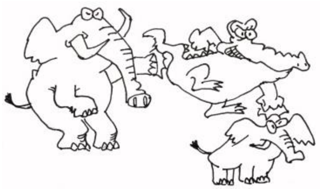
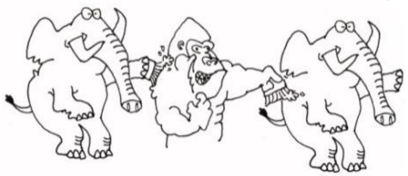
Simple	Show the turtle with the blue scarf 	Neutral 1	He needs his holiday. He took the dogs for a walk
		Neutral 2	The professor with the unique money is disappointingly wrong
		Incongruent	The donkey with the blue scarf is nearly icy
		Congruent	The turtle with the blue scarf is nearly icy
Simple	Show the witch with the pink shoes 	Neutral 1	The baby is on the rug. The bag fell off the shelf
		Neutral 2	The wolf with the rotten apple is somewhat dark
		Incongruent	The monkey with the pink shoes is positively polite
		Congruent	The witch with the pink shoes is positively polite
Simple	Show the witch with the small bag 	Neutral 1	The fruit is on the ground. They like orange marmalade
		Neutral 2	The dolphin with the original doll is actually exotic
		Incongruent	The man with the small bag is finally popular
		Congruent	The witch with the small bag is finally popular
Simple	Show the witch with the white gloves 	Neutral 1	They stared at the picture. They lost all their money
		Neutral 2	The hamster with the abundant imagination is dismally rare
		Incongruent	The man with the white gloves is perfectly sour
		Congruent	The witch with the white gloves is perfectly sour
Simple	Show the zebra with the pink necklace 	Neutral 1	She bumped her head on the door. The cook is baking a cake
		Neutral 2	The reporter with the difficult personality is ordinarily messy
		Incongruent	The giraffe with the pink necklace is offensively challenging
		Congruent	The zebra with the pink necklace is offensively challenging

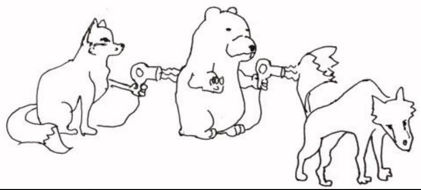
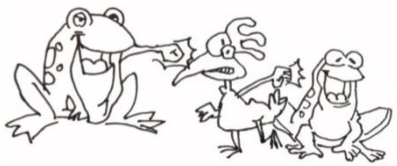
Simple	Show the zebra with the pink scarf 	Neutral 1	The lady wore a coat. The woman cleaned her house
		Neutral 2	The salesperson with the tiny puzzle is unfortunately hot
		Incongruent	The camel with the pink scarf is sadly rare
		Congruent	The zebra with the pink scarf is sadly rare
Simple	Show the zebra with the purple ball 	Neutral 1	She stood near the window. He's skating with his friend
		Neutral 2	The singer with the original personality is fairly funny
		Incongruent	The rabbit with the purple ball is uncharacteristically beautiful
		Congruent	The zebra with the purple ball is uncharacteristically beautiful
Simple	Show the zebra with the small hat 	Neutral 1	The baby is pretty. The paint dripped on the ground
		Neutral 2	The technician with the colourful money is possibly gentle
		Incongruent	The turtle with the small hat is nevertheless colourful
		Congruent	The zebra with the small hat is nevertheless colourful
Familiarisation SR	Show the elephant that is holding the gorilla 	Neutral 1	He broke his leg again. The ice cream was melting
		Neutral 2	The nurse who is performing the surgery is particularly delicate
		Incongruent	The gorilla that is holding the elephant is brutally challenging
		Congruent	The elephant that is holding the gorilla is brutally challenging
Familiarisation SR	Show the man who is combing the boy 	Neutral 1	The kitchen clock was wrong. The pond water is dirty
		Neutral 2	The fish that is checking the bubble is frankly strange
		Incongruent	The boy who is combing the man is quite grateful
		Congruent	The man who is combing the boy is quite grateful

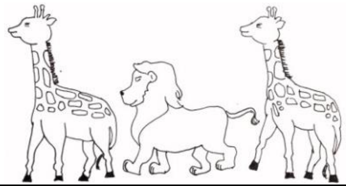
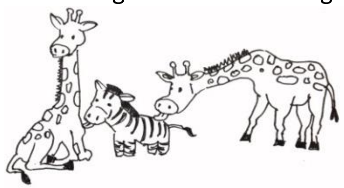
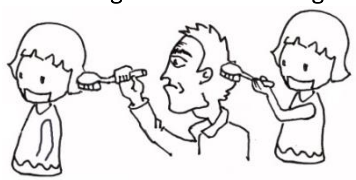
Familiarisation SR	Show the snake that is spraying the frog 	Neutral 1	Mother opened the drawer. The postman brought a letter
		Neutral 2	The cook who is tasting the honey is quite bad
		Incongruent	The frog who is spraying the snake is already dry
		Congruent	The snake that is spraying the frog is already dry
SR	Show the bear that is biting the fox 	Neutral 1	She wore her yellow shirt. It's time to go to bed
		Neutral 2	The waiter who is cancelling the steak is thoroughly burnt
		Incongruent	The fox that is biting the bear is annoyingly greedy
		Congruent	The bear that is biting the fox is annoyingly greedy
SR	Show the bear that is spraying the goose 	Neutral 1	The kitchen window was clean. The big boy kicked the ball
		Neutral 2	The salesperson who is hating the teddy is probably soft
		Incongruent	The goose that is spraying the bear is dismally strange
		Congruent	The bear that is spraying the goose is dismally strange
SR	Show the boy who is combing the girl 	Neutral 1	The rain came pouring down. The cups are on the table
		Neutral 2	The wolf that is tasting the meat is wonderfully tender
		Incongruent	The girl who is combing the boy is never great
		Congruent	The boy who is combing the girl is never great
SR	Show the boy who is kicking the girl 	Neutral 1	Big dogs can be dangerous. The towel fell on the floor
		Neutral 2	The owl that is hunting the mouse is momentarily distracted
		Incongruent	The girl who is kicking the boy is particularly rude
		Congruent	The boy who is kicking the girl is particularly rude

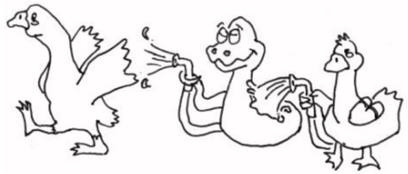
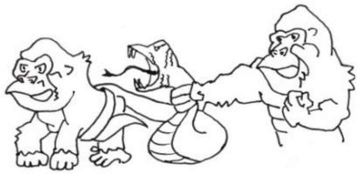
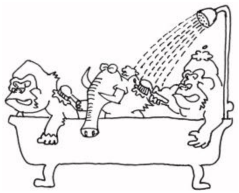
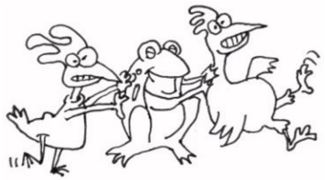
SR	Show the camel that is dirtying the lion 	Neutral 1	The child drank some fresh milk. They're running past the house
		Neutral 2	The singer who is performing the piece is excessively emotional
		Incongruent	The lion that is dirtying the camel is terribly worried
		Congruent	The camel that is dirtying the lion is terribly worried
SR	Show the camel that is pursuing the donkey 	Neutral 1	The floor looks clean and shiny. A girl came into the room
		Neutral 2	The waiter who is advertising the tea is extraordinarily sour
		Incongruent	The donkey that is pursuing the camel is wonderfully exotic
		Congruent	The camel that is pursuing the donkey is wonderfully exotic
SR	Show the cat that is licking the pig 	Neutral 1	Mother read the the instructions. The oven is too hot
		Neutral 2	The teacher who is admiring the answer is completely wrong
		Incongruent	The pig that is licking the cat is unhealthily messy
		Congruent	The cat that is licking the pig is unhealthily messy
SR	Show the cat that is scratching the dog 	Neutral 1	They're buying some fresh bread. He got mud on his shoes
		Neutral 2	The architect who is checking the cave is unbearably dark
		Incongruent	The dog that is scratching the cat is actually different
		Congruent	The cat that is scratching the dog is actually different
SR	Show the cow that is watching the sheep 	Neutral 1	The girl played with the baby. The exit was well lit
		Neutral 2	The farmer who is spoiling the case is exceedingly heavy
		Incongruent	The sheep that is watching the cow is sometimes hot
		Congruent	The cow that is watching the sheep is sometimes hot

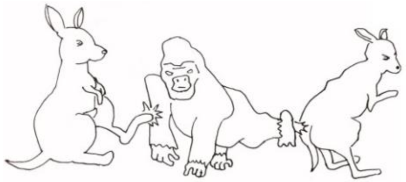
SR	Show the cow that is wetting the pig 	Neutral 1	The nervous driver got lost. They're pushing an old car
		Neutral 2	The builder who is throwing the rake is disgustingly dirty
		Incongruent	The pig that is wetting the cow is quite exotic
		Congruent	The cow that is wetting the pig is quite exotic
SR	Show the crocodile that is following the elephant 	Neutral 1	Men normally wear long trousers. The boy forgot his book
		Neutral 2	The magician who is repairing the furniture is disappointingly expensive
		Incongruent	The elephant that is following the crocodile is systematically wrong
		Congruent	The crocodile that is following the elephant is systematically wrong
SR	Show the dog that is biting the horse 	Neutral 1	They finished dinner on time. The machine is noisy
		Neutral 2	The cook who is cancelling the order is finally hot
		Incongruent	The horse that is biting the dog is sometimes violent
		Congruent	The dog that is biting the horse is sometimes violent
SR	Show the dog that is nudging the cat 	Neutral 1	The salt shaker was empty. The angry man shouted
		Neutral 2	The lawyer who is advertising the book is slightly boring
		Incongruent	The cat that is nudging the dog is unbearably funny
		Congruent	The dog that is nudging the cat is unbearably funny
SR	Show the donkey that is attacking the turtle 	Neutral 1	The boy broke the wooden fence. The postman shut the gate
		Neutral 2	The officer who is imagining the crime is extremely violent
		Incongruent	The turtle that is attacking the donkey is unusually terrified
		Congruent	The donkey that is attacking the turtle is unusually terrified

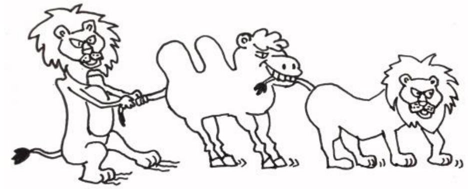
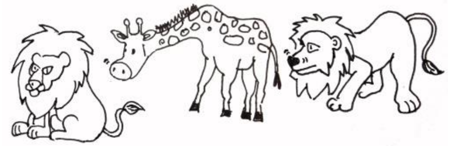
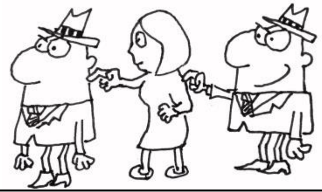
SR	Show the donkey that is harming the camel 	Neutral 1	She washed her new silk dress. The tub tap is leaking
		Neutral 2	The judge who is imitating the answer is naturally precise
		Incongruent	The camel that is harming the donkey is excessively loud
		Congruent	The donkey that is harming the camel is excessively loud
SR	Show the duck that is chasing the ghost 	Neutral 1	The house had a nice garden. He is washing his car
		Neutral 2	The manager who is chewing the chocolate is sadly bad
		Incongruent	The ghost that is chasing the duck is excessively positive
		Congruent	The duck that is chasing the ghost is excessively positive
SR	Show the duck that is washing the frog 	Neutral 1	He played with his toy train. My mother stirred her tea
		Neutral 2	The fireman who is moving the furniture is fairly burnt
		Incongruent	The frog that is washing the duck is possibly emotional
		Congruent	The duck that is washing the frog is possibly emotional
SR	Show the elephant that is hurting the crocodile 	Neutral 1	The cat caught a little mouse. Father paid at the gate
		Neutral 2	The doctor who is planning the surgery is extraordinarily precise
		Incongruent	The crocodile that is hurting the elephant is obviously dangerous
		Congruent	The elephant that is hurting the crocodile is obviously dangerous
SR	Show the elephant that is scrubbing the gorilla 	Neutral 1	The cow was milked every day. The baby broke his cup
		Neutral 2	The musician who is imagining the piece is ordinarily popular
		Incongruent	The gorilla that is scrubbing the elephant is offensively smelly
		Congruent	The elephant that is scrubbing the gorilla is offensively smelly

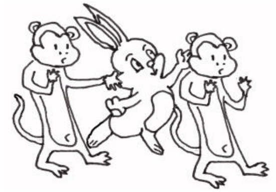
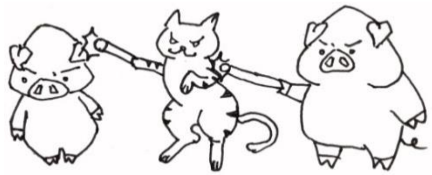
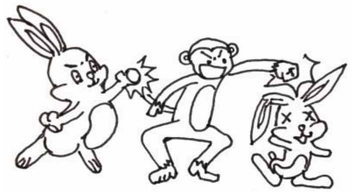
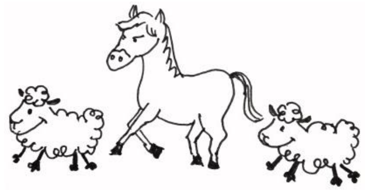
SR	Show the fox that is drying the bear 	Neutral 1	They watched a scary movie. Milk comes in a carton
		Neutral 2	The judge who is saving the case is interestingly complex
		Incongruent	The bear who is drying the fox is very wet
		Congruent	The fox that is drying the bear is very wet
SR	Show the fox that is touching the witch 	Neutral 1	The neighbour's boy has black hair. The policeman knows the way
		Neutral 2	The snail that is munching the salad is disgustingly slimy
		Incongruent	The witch that is touching the fox is unusually polite
		Congruent	The fox that is touching the witch is unusually polite
SR	Show the frog that is cleaning the duck 	Neutral 1	The girl caught a head cold. The book tells a story
		Neutral 2	The painter who is saving the ship is terribly slimy
		Incongruent	The duck that is cleaning the frog is unbearably slimy
		Congruent	The frog that is cleaning the duck is unbearably slimy
SR	Show the frog that is hurting the hen 	Neutral 1	The two children are laughing. The road goes up a hill
		Neutral 2	The driver who is solving the crime is decidedly strange
		Incongruent	The hen that is hurting the frog is somewhat fragile
		Congruent	The frog that is hurting the hen is somewhat fragile
SR	Show the ghost that is dirtying the duck 	Neutral 1	The football hit the goalpost. The lady wore a coat
		Neutral 2	The magician who is cancelling the trick is very interesting
		Incongruent	The duck that is dirtying the ghost is especially colourful
		Congruent	The ghost that is dirtying the duck is especially colourful

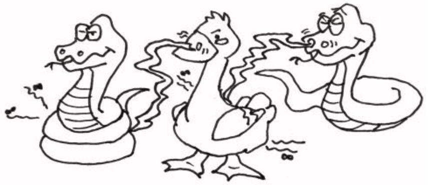
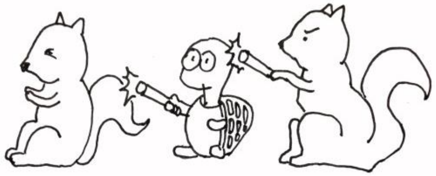
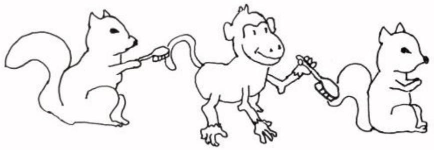
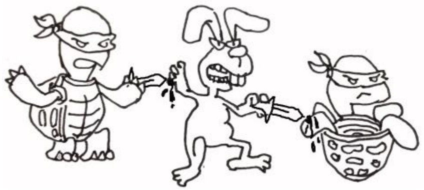
SR	Show the ghost that is poking the witch 	Neutral 1	They like orange marmalade. They laughed at his story
		Neutral 2	The tiger that is cuddling the cub is uncharacteristically gentle
		Incongruent	The witch that is poking the ghost is extraordinarily sour
		Congruent	The ghost that is poking the witch is extraordinarily sour
SR	Show the giraffe that is following the lion 	Neutral 1	Mother tied the string too tight. They're watching the cuckoo clock
		Neutral 2	The technician who is repairing the file is disappointingly corrupt
		Incongruent	The lion that is following the giraffe is depressingly icy
		Congruent	The giraffe that is following the lion is depressingly icy
SR	Show the giraffe that is licking the zebra 	Neutral 1	Swimmers can hold their breath. The bath water is warm
		Neutral 2	The artist who is creating the speech is unbearably violent
		Incongruent	The zebra that is licking the giraffe is somewhat fast
		Congruent	The giraffe that is licking the zebra is somewhat fast
SR	Show the girl who is brushing the man 	Neutral 1	A mouse ran into the hole. The match fell on the floor
		Neutral 2	The lizard that is hunting the fly is very worried
		Incongruent	The man who is brushing the girl is always early
		Congruent	The girl who is brushing the man is always early
SR	Show the girl who is holding the boy 	Neutral 1	The black dog was hungry. The yellow pears taste good
		Neutral 2	The koala that is chewing the leaf is unbearably smelly
		Incongruent	The boy who is holding the girl is unfortunately poor
		Congruent	The girl who is holding the boy is unfortunately poor

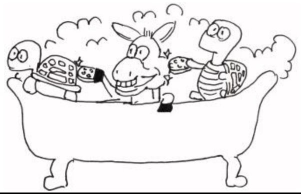
SR	Show the goose that is splashing the bear 	Neutral 1	The small tomatoes are green. They called an ambulance
		Neutral 2	The professor who is flying the helicopter is somewhat old
		Incongruent	The bear that is splashing the goose is horrifyingly boring
		Congruent	The goose that is splashing the bear is horrifyingly boring
SR	Show the goose that is wetting the snake 	Neutral 1	The teapot is very hot. The truck made it up the hill
		Neutral 2	The therapist who is admiring the sunshine is beautifully warm
		Incongruent	The snake that is wetting the goose is extremely cheerful
		Congruent	The goose that is wetting the snake is extremely cheerful
SR	Show the gorilla that is squeezing the snake 	Neutral 1	The bus leaves before the train. A fish swam in the pond
		Neutral 2	The salesperson who is repairing the book is exceedingly funny
		Incongruent	The snake that is squeezing the gorilla is already done
		Congruent	The gorilla that is squeezing the snake is already done
SR	Show the gorilla that is washing the elephant 	Neutral 1	They had some cold meat for lunch. New neighbours are moving in
		Neutral 2	The officer who is observing the village is unfortunately dangerous
		Incongruent	The elephant that is washing the gorilla is thoroughly itchy
		Congruent	The gorilla that is washing the elephant is thoroughly itchy
SR	Show the hen that is pushing the frog 	Neutral 1	Mother picked some flowers. The dog played with a stick
		Neutral 2	The builder who is answering the question is actually great
		Incongruent	The frog that is pushing the hen is luckily light
		Congruent	The hen that is pushing the frog is luckily light

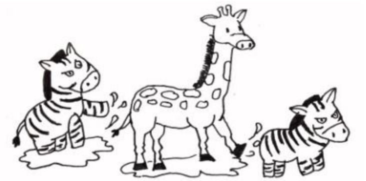
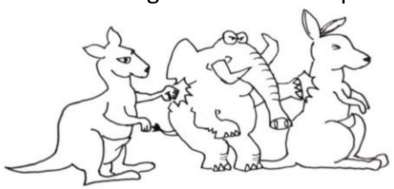
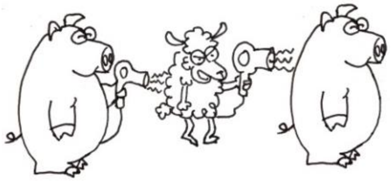
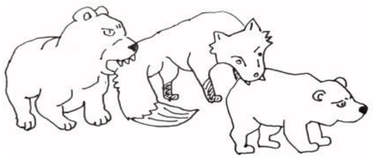
SR	Show the horse that is smelling the sheep 	Neutral 1	The boy slipped on the stairs. He climbed up the ladder
		Neutral 2	The lawyer who is creating the problem is quite expensive
		Incongruent	The sheep that is smelling the horse is slightly dirty
		Congruent	The horse that is smelling the sheep is slightly dirty
SR	Show the horse that is watching the dog 	Neutral 1	Yesterday he lost his hat. The ball broke the window
		Neutral 2	The painter who is spoiling the art is reasonably cheerful
		Incongruent	The dog that is watching the horse is ominously dark
		Congruent	The horse that is watching the dog is ominously dark
SR	Show the kangaroo that is kicking the gorilla 	Neutral 1	She writes to her friend daily. A cat jumped over the fence
		Neutral 2	The musician who is appreciating the piano is particularly good
		Incongruent	The gorilla that is kicking the kangaroo is uncharacteristically fierce
		Congruent	The kangaroo that is kicking the gorilla is uncharacteristically fierce
SR	Show the lady who is grabbing the man 	Neutral 1	The ground was very hard. The family likes fish
		Neutral 2	The shark that is tasting the steak is decidedly sweet
		Incongruent	The man who is grabbing the lady is reasonably sweet
		Congruent	The lady who is grabbing the man is reasonably sweet
SR	Show the lady who is painting the boy 	Neutral 1	The shop closes for lunch. The dog jumped on the chair
		Neutral 2	The bat that is spoiling the cave is horrifyingly scary
		Incongruent	The boy who is painting the lady is momentarily bad
		Congruent	The lady who is painting the boy is momentarily bad

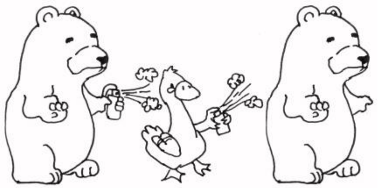
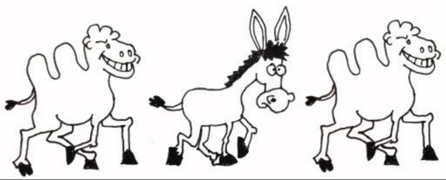
SR	Show the lion that is pulling the camel 	Neutral 1	She uses her spoon to eat. The sky was very blue
		Neutral 2	The teacher who is imagining the problem is somewhat interesting
		Incongruent	The camel that is pulling the lion is frankly heavy
		Congruent	The lion that is pulling the camel is frankly heavy
SR	Show the lion that is sniffing the giraffe 	Neutral 1	Her husband brought some flowers. The shirts are in the closet
		Neutral 2	The farmer who is moving the rake is extraordinarily ancient
		Incongruent	The giraffe that is sniffing the lion is beautifully tender
		Congruent	The lion that is sniffing the giraffe is beautifully tender
SR	Show the man who is painting the girl 	Neutral 1	The dog growled at the neighbours. The baby is on the rug
		Neutral 2	The penguin that is moving the fish is ordinarily greedy
		Incongruent	The girl who is painting the man is ordinarily gentle
		Congruent	The man who is painting the girl is ordinarily gentle
SR	Show the man who is poking the lady 	Neutral 1	The dishcloth is soaking wet. The baby wants his bottle
		Neutral 2	The bamboo that the panda is throwing is unhealthily dirty
		Incongruent	The lady who is poking the man is probably important
		Congruent	The man who is poking the lady is probably important
SR	Show the monkey that is pushing the squirrel 	Neutral 1	The boy ran away from school. The bottle is on the shelf
		Neutral 2	The therapist who is improvising the speech is thoroughly boring
		Incongruent	The squirrel that is pushing the monkey is brutally burnt
		Congruent	The monkey that is pushing the squirrel is brutally burnt

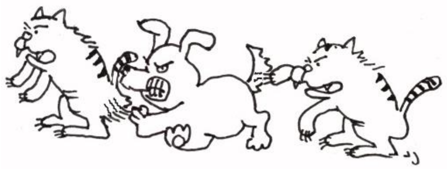
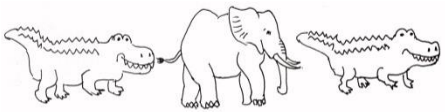
SR	Show the monkey that is touching the rabbit 	Neutral 1	The girl was fixing her dress. The old man is worried
		Neutral 2	The fireman who is saving the swimmer is completely terrified
		Incongruent	The rabbit that is touching the monkey is irrefutably noisy
		Congruent	The monkey that is touching the rabbit is irrefutably noisy
SR	Show the pig that is beating the cat 	Neutral 1	Someone is crossing the road. She stood near the window
		Neutral 2	The manager who is saving the file is sadly wrong
		Incongruent	The cat that is beating the pig is actually round
		Congruent	The pig that is beating the cat is actually round
SR	Show the pig that is sniffing the cow 	Neutral 1	A boy fell from the window. The fruit is on the ground
		Neutral 2	The nurse who is enjoying the moment is always positive
		Incongruent	The cow that is sniffing the pig is utterly ill
		Congruent	The pig that is sniffing the cow is utterly ill
SR	Show the rabbit that is hitting the monkey 	Neutral 1	He closed his eyes and jumped. They walked across the grass
		Neutral 2	The reporter who is observing the party is probably done
		Incongruent	The monkey that is hitting the rabbit is decidedly intriguing
		Congruent	The rabbit that is hitting the monkey is decidedly intriguing
SR	Show the sheep that is chasing the horse 	Neutral 1	The front garden is pretty. Father forgot the bread
		Neutral 2	The pilot who is investigating the car is unhealthily loud
		Incongruent	The horse that is chasing the sheep is fairly young
		Congruent	The sheep that is chasing the horse is fairly young

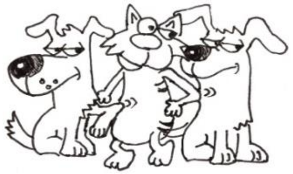
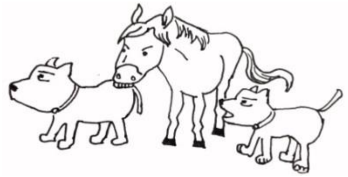
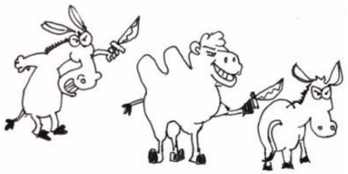
SR	Show the sheep that is pulling the cow 	Neutral 1	The painter uses a brush. Somebody stole the money
		Neutral 2	The driver who is advertising the car is naturally strong
		Incongruent	The cow that is pulling the sheep is completely frozen
		Congruent	The sheep that is pulling the cow is completely frozen
SR	Show the snake that is smelling the goose 	Neutral 1	Her shoes were very dirty. The cows are in the pasture
		Neutral 2	The singer who is checking the sound is wonderfully intriguing
		Incongruent	The goose that is smelling the snake is disappointingly expensive
		Congruent	The snake that is smelling the goose is disappointingly expensive
SR	Show the squirrel that is beating the turtle 	Neutral 1	They're shopping for school clothes. The waiter brought the cream
		Neutral 2	The technician who is creating the puzzle is fairly complex
		Incongruent	The turtle that is beating the squirrel is already old
		Congruent	The squirrel that is beating the turtle is already old
SR	Show the squirrel that is brushing the monkey 	Neutral 1	She's helping her friend move. The old gloves are dirty
		Neutral 2	The reporter who is examining the hut is sometimes smelly
		Incongruent	The monkey that is brushing the squirrel is extraordinarily rare
		Congruent	The squirrel that is brushing the monkey is extraordinarily rare
SR	Show the turtle that is attacking the rabbit 	Neutral 1	She's drinking from her own cup. The silly boy is hiding
		Neutral 2	The professor who is improvising the question is unusually good
		Incongruent	The rabbit that is attacking the turtle is always cold
		Congruent	The turtle that is attacking the rabbit is always cold

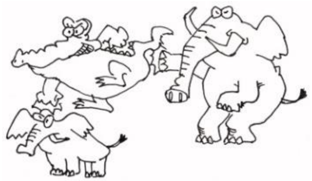
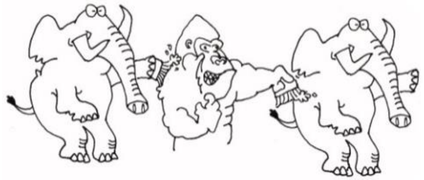
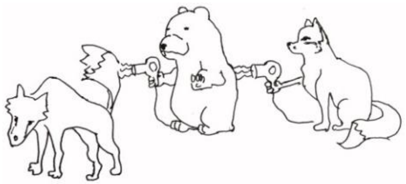
SR	Show the turtle that is cleaning the donkey 	Neutral 1	The house had nine bedrooms. Her coat is on the chair
		Neutral 2	The pilot who is flying the helicopter is luckily precise
		Incongruent	The donkey that is cleaning the turtle is terribly ancient
		Congruent	The turtle that is cleaning the donkey is terribly ancient
SR	Show the turtle that is pursuing the zebra 	Neutral 1	She injured four of her fingers. Snow falls in the winter
		Neutral 2	The artist who is improvising the music is unusually creative
		Incongruent	The zebra that is pursuing the turtle is never scary
		Congruent	The turtle that is pursuing the zebra is never scary
SR	Show the witch that is scratching the fox 	Neutral 1	They're watching the train go by. A grocer sells butter
		Neutral 2	The prey that the eagle is examining is sometimes cunning
		Incongruent	The fox that is scratching the witch is intensely sticky
		Congruent	The witch that is scratching the fox is intensely sticky
SR	Show the witch that is scrubbing the ghost 	Neutral 1	He cut his index finger. They met some friends at dinner
		Neutral 2	The hedgehog that is solving the puzzle is irrefutably intriguing
		Incongruent	The ghost that is scrubbing the witch is naturally beautiful
		Congruent	The witch that is scrubbing the ghost is naturally beautiful
SR	Show the zebra that is harming the turtle 	Neutral 1	The car is going too fast. The fire is very hot
		Neutral 2	The doctor who is investigating the illness is unfortunately fierce
		Incongruent	The turtle that is harming the zebra is perfectly creative
		Congruent	The zebra that is harming the turtle is perfectly creative

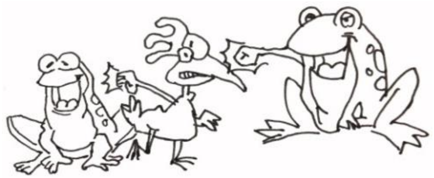
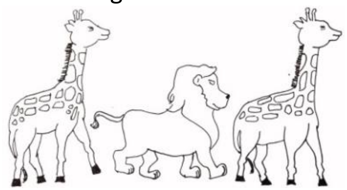
SR	Show the zebra that is splashing the giraffe 	Neutral 1	The dog sleeps in a basket. The man called the police
		Neutral 2	The architect who is planning the village is already popular
		Incongruent	The giraffe that is splashing the zebra is disgustingly complex
		Congruent	The zebra that is splashing the giraffe is disgustingly complex
Familiarisation OR	Show the kangaroo that the elephant is hitting 	Neutral 1	The match fell on the floor. They watched a scary movie
		Neutral 2	The honey that the cook is tasting is quite bad
		Incongruent	The elephant that the kangaroo is hitting is systematically distracted
		Congruent	The kangaroo that the elephant is hitting is systematically distracted
Familiarisation OR	Show the pig that the sheep is drying 	Neutral 1	The machine is noisy. The girl caught a head cold
		Neutral 2	The surgery that the nurse is performing is particularly delicate
		Incongruent	The sheep that the pig is drying is nearly cold
		Congruent	The pig that the sheep is drying is nearly cold
Familiarisation OR	Show the witch who the girl is grabbing 	Neutral 1	It's time to go to bed. The rain came pouring down
		Neutral 2	The whale that is swimming the oceans is always warm
		Incongruent	The girl who the witch is grabbing is unhealthily itchy
		Congruent	The witch who the girl is grabbing is unhealthily itchy
OR	Show the bear that the fox is biting 	Neutral 1	The lady went to the store. They rode their bicycles
		Neutral 2	The steak that the waiter is cancelling is thoroughly burnt
		Incongruent	The fox that the bear is biting is fairly strange
		Congruent	The bear that the fox is biting is fairly strange

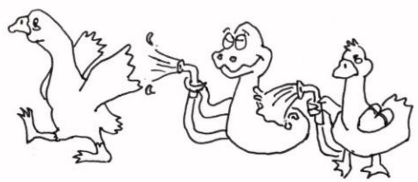
OR	Show the bear that the goose is spraying 	Neutral 1	The two farmers were talking. There was a bad train wreck
		Neutral 2	The teddy that the salesperson is hating is probably soft
		Incongruent	The goose that the bear is spraying is thoroughly clueless
		Congruent	The bear that the goose is spraying is thoroughly clueless
OR	Show the boy who the girl is combing 	Neutral 1	The ball bounced very high. The match boxes are empty
		Neutral 2	The meat that the wolf is tasting is wonderfully tender
		Incongruent	The girl that the boy is combing is disappointingly rich
		Congruent	The boy who the girl is combing is disappointingly rich
OR	Show the boy who the girl is kicking 	Neutral 1	He is washing his face with soap. The cleaner swept the floor
		Neutral 2	The mouse that the owl is hunting is momentarily distracted
		Incongruent	The girl that the boy is kicking is probably young
		Congruent	The boy who the girl is kicking is probably young
OR	Show the camel that the donkey is pursuing 	Neutral 1	The truck carries fresh fruit. He wore his yellow shirt
		Neutral 2	The piece that the singer is performing is excessively emotional
		Incongruent	The donkey that the camel is pursuing is depressingly complex
		Congruent	The camel that the donkey is pursuing is depressingly complex
OR	Show the camel that the lion is dirtying 	Neutral 1	She's washing her new silk dress. The fire was very hot
		Neutral 2	The tea that the waiter is advertising is extraordinarily sour
		Incongruent	The lion that the camel is dirtying is always precise
		Congruent	The camel that the lion is dirtying is always precise

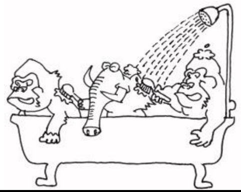
OR	Show the cat that the dog is scratching 	Neutral 1	A sharp knife is dangerous. She's paying for her bread
		Neutral 2	The answer that the teacher is admiring is completely wrong
		Incongruent	The dog that the cat is scratching is obviously dirty
		Congruent	The cat that the dog is scratching is obviously dirty
OR	Show the cat that the pig is licking 	Neutral 1	The three girls are listening. He broke his leg again
		Neutral 2	The cave that the architect is checking is unbearably dark
		Incongruent	The pig that the cat is licking is perfectly quiet
		Congruent	The cat that the pig is licking is perfectly quiet
OR	Show the cow that the pig is wetting 	Neutral 1	The man is painting the sign. The football game is over
		Neutral 2	The case that the farmer is spoiling is exceedingly heavy
		Incongruent	The pig that the cow is wetting is especially delicate
		Congruent	The cow that the pig is wetting is especially delicate
OR	Show the cow that the sheep is watching 	Neutral 1	The pond water is dirty. The lady packed her bag
		Neutral 2	The rake that the builder is throwing is disgustingly dirty
		Incongruent	The sheep that the cow is watching is particularly interesting
		Congruent	The cow that the sheep is watching is particularly interesting
OR	Show the crocodile that the elephant is following 	Neutral 1	The man cleaned his suede shoes. The baby has blue eyes
		Neutral 2	The furniture that the magician is repairing is disappointingly expensive
		Incongruent	The elephant that the crocodile is following is sadly tender
		Congruent	The crocodile that the elephant is following is sadly tender

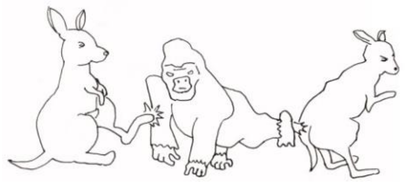
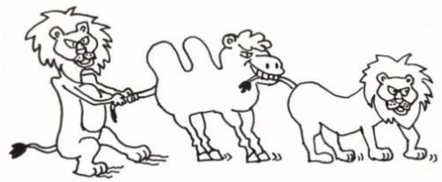
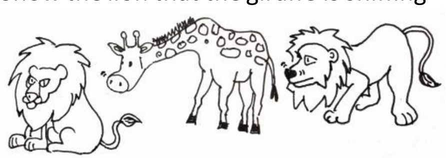
OR	Show the dog that the cat is nudging 	Neutral 1	The postman brought a letter. Sugar is very sweet
		Neutral 2	The order that the cook is cancelling is finally hot
		Incongruent	The cat that the dog is nudging is decidedly heavy
		Congruent	The dog that the cat is nudging is decidedly heavy
OR	Show the dog that the horse is biting 	Neutral 1	The police helped the driver. The buckets fill up quickly
		Neutral 2	The book that the lawyer is advertising is slightly boring
		Incongruent	The horse that the dog is biting is extremely cunning
		Congruent	The dog that the horse is biting is extremely cunning
OR	Show the donkey that the camel is harming 	Neutral 1	The girl ran along the fence. The tree fell on a house
		Neutral 2	The crime that the officer is imagining is extremely violent
		Incongruent	The camel that the donkey is harming is somewhat lively
		Congruent	The donkey that the camel is harming is somewhat lively
OR	Show the donkey that the turtle is attacking 	Neutral 1	A field mouse found the cheese. A fire engine is coming
		Neutral 2	The answer that the judge is imitating is naturally precise
		Incongruent	The turtle that the donkey is attacking is systematically fast
		Congruent	The donkey that the turtle is attacking is systematically fast
OR	Show the duck that the frog is washing 	Neutral 1	He grew lots of vegetables. The boy did a handstand
		Neutral 2	The chocolate that the manager is chewing is sadly bad
		Incongruent	The frog that the duck is washing is already warm
		Congruent	The duck that the frog is washing is already warm

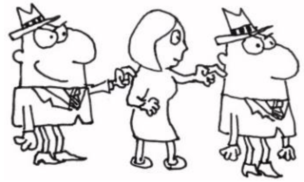
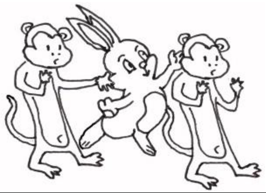
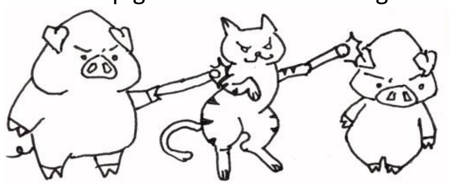
OR	Show the duck that the ghost is chasing	Neutral 1	The milk is by the front door. They had some chocolate pudding
		Neutral 2	The furniture that the fireman is moving is fairly burnt
		Incongruent	The ghost that the duck is chasing is especially warm
		Congruent	The duck that the ghost is chasing is especially warm
OR	Show the elephant that the crocodile is hurting	Neutral 1	Dad stopped to pick some pears. The police cleared the road
		Neutral 2	The surgery that the doctor is planning is extraordinarily precise
		Incongruent	The crocodile that the elephant is hurting is brutally scary
		Congruent	The elephant that the crocodile is hurting is brutally scary
OR	Show the elephant that the gorilla is scrubbing	Neutral 1	The milkman drives a small truck. The little girl is shouting
		Neutral 2	The piece that the musician is imagining is ordinarily popular
		Incongruent	The gorilla that the elephant is scrubbing is luckily bright
		Congruent	The elephant that the gorilla is scrubbing is luckily bright
OR	Show the fox that the bear is drying	Neutral 1	The children are walking home. The boy got into trouble
		Neutral 2	The case that the judge is saving is interestingly complex
		Incongruent	The bear that the fox is drying is slightly intriguing
		Congruent	The fox that the bear is drying is slightly intriguing
OR	Show the fox that the witch is touching	Neutral 1	The train is moving fast. The team is playing well
		Neutral 2	The salad that the snail is munching is disgustingly slimy
		Incongruent	The witch that the fox is touching is interestingly corrupt
		Congruent	The fox that the witch is touching is interestingly corrupt

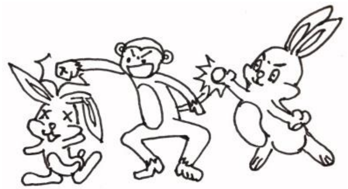
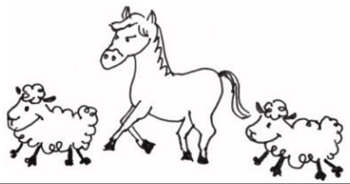
OR	Show the frog that the duck is cleaning 	Neutral 1	He really scared his sister. The scissors are very sharp
		Neutral 2	The ship that the painter is saving is terribly slimy
		Incongruent	The duck that the frog is cleaning is officially important
		Congruent	The frog that the duck is cleaning is officially important
OR	Show the frog that the hen is hurting 	Neutral 1	The ice cream was melting. Mother opened the drawer
		Neutral 2	The crime that the driver is solving is decidedly strange
		Incongruent	The hen that the frog is hurting is utterly challenging
		Congruent	The frog that the hen is hurting is utterly challenging
OR	Show the ghost that the duck is dirtying 	Neutral 1	The tall man tied his shoes. There were branches everywhere
		Neutral 2	The trick that the magician is cancelling is very interesting
		Incongruent	The duck that the ghost is dirtying is possibly dangerous
		Congruent	The ghost that the duck is dirtying is possibly dangerous
OR	Show the ghost that the witch is poking 	Neutral 1	The chicken laid some eggs. The mother heard the baby
		Neutral 2	The cub that the tiger is cuddling is uncharacteristically gentle
		Incongruent	The witch that the ghost is poking is reasonably popular
		Congruent	The ghost that the witch is poking is reasonably popular
OR	Show the giraffe that the lion is following 	Neutral 1	The child ripped open the bag. The apple pie was good
		Neutral 2	The file that the technician is repairing is disappointingly corrupt
		Incongruent	The lion that the giraffe is following is nevertheless fierce
		Congruent	The giraffe that the lion is following is nevertheless fierce

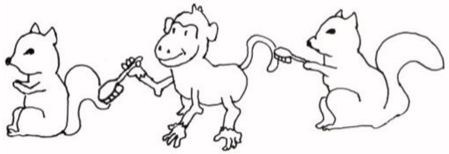
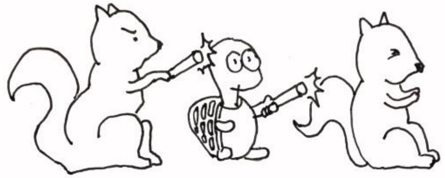
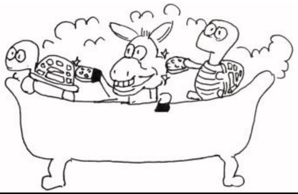
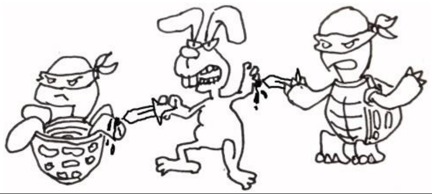
OR	Show the giraffe that the zebra is licking 	Neutral 1	The yellow pears taste good. Children like strawberries
		Neutral 2	The speech that the artist is creating is unbearably violent
		Incongruent	The zebra that the giraffe is licking is frankly smelly
		Congruent	The giraffe that the zebra is licking is frankly smelly
OR	Show the girl who the boy is holding 	Neutral 1	Strawberry jam is sweet. The kitchen clock was wrong
		Neutral 2	The leaf that the koala is chewing is unbearably smelly
		Incongruent	The boy that the girl is holding is interestingly great
		Congruent	The girl who the boy is holding is interestingly great
OR	Show the girl who the man is brushing 	Neutral 1	The sun melted the snow. The fruit came in a box
		Neutral 2	The fly that the lizard is hunting is very worried
		Incongruent	The man who the girl is brushing is positively creative
		Congruent	The girl who the man is brushing is positively creative
OR	Show the goose that the bear is splashing 	Neutral 1	She made her bed and left. The boy is running away
		Neutral 2	The helicopter that the professor is flying is somewhat old
		Incongruent	The bear that the goose is splashing is momentarily distracted
		Congruent	The goose that the bear is splashing is momentarily distracted
OR	Show the goose that the snake is wetting 	Neutral 1	The children helped their teacher. The boy went to bed early
		Neutral 2	The sunshine that the therapist is admiring is beautifully warm
		Incongruent	The snake that the goose is wetting is extremely good
		Congruent	The goose that the snake is wetting is extremely good

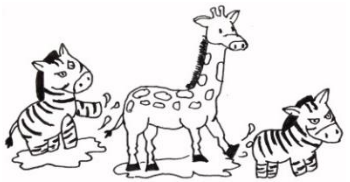
OR	Show the gorilla that the elephant is washing 	Neutral 1	She spoke to her eldest son. The black dog was hungry
		Neutral 2	The book that the salesperson is repairing is exceedingly funny
		Incongruent	The elephant that the gorilla is washing is utterly lazy
		Congruent	The gorilla that the elephant is washing is utterly lazy
OR	Show the gorilla that the snake is squeezing 	Neutral 1	They painted the wall white. The truck drove up the road
		Neutral 2	The village that the officer is observing is unfortunately dangerous
		Incongruent	The snake that the gorilla is squeezing is irrefutably strong
		Congruent	The gorilla that the snake is squeezing is irrefutably strong
OR	Show the hen that the frog is pushing 	Neutral 1	The orange is very sweet. She's calling her daughter
		Neutral 2	The question that the builder is answering is actually great
		Incongruent	The frog that the hen is pushing is naturally wet
		Congruent	The hen that the frog is pushing is naturally wet
OR	Show the horse that the dog is watching 	Neutral 1	He's skating with his friend. The little boy left home
		Neutral 2	The problem that the lawyer is creating is quite expensive
		Incongruent	The dog that the horse is watching is dismally glum
		Congruent	The horse that the dog is watching is dismally glum
OR	Show the horse that the sheep is smelling 	Neutral 1	He paid his bill in full. They ate the lemon pie
		Neutral 2	The art that the painter is spoiling is reasonably cheerful
		Incongruent	The sheep that the horse is smelling is slightly delicate
		Congruent	The horse that the sheep is smelling is slightly delicate

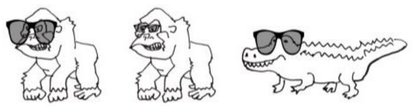
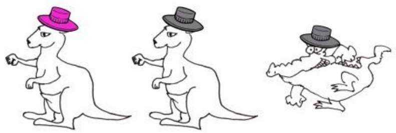
OR	Show the kangaroo that the gorilla is kicking 	Neutral 1	He's washing his face with soap. The cleaner swept the floor
		Neutral 2	The piano that the musician is appreciating is particularly good
		Incongruent	The gorilla that the kangaroo is kicking is officially expensive
		Congruent	The kangaroo that the gorilla is kicking is officially expensive
OR	Show the lady who the boy is painting 	Neutral 1	The children washed the plates. The cat lay on the bed
		Neutral 2	The cave that the bat is spoiling is horrifyingly scary
		Incongruent	The boy that the lady is painting is quite sweet
		Congruent	The lady who the boy is painting is quite sweet
OR	Show the lady who the man is grabbing 	Neutral 1	The cat drank from the saucer. The player lost a shoe
		Neutral 2	The steak that the shark is tasting is decidedly sweet
		Incongruent	The man who the lady is grabbing is horrifyingly corrupt
		Congruent	The lady who the man is grabbing is horrifyingly corrupt
OR	Show the lion that the camel is pulling 	Neutral 1	The wife helped her husband. Rain is good for the trees
		Neutral 2	The problem that the teacher is imagining is somewhat interesting
		Incongruent	The camel that the lion is pulling is very boring
		Congruent	The lion that the camel is pulling is very boring
OR	Show the lion that the giraffe is sniffing 	Neutral 1	They're climbing the old oak tree. The big fish got away
		Neutral 2	The rake that the farmer is moving is extraordinarily ancient
		Incongruent	The giraffe that the lion is sniffing is ominously quiet
		Congruent	The lion that the giraffe is sniffing is ominously quiet

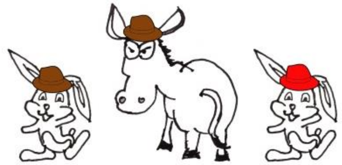
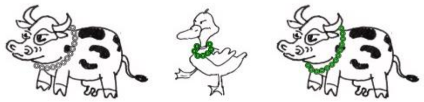
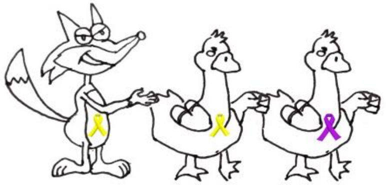
OR	Show the man who the girl is painting 	Neutral 1	The baby has blue eyes. The bakery is open
		Neutral 2	The fish that the penguin is moving is ordinarily greedy
		Incongruent	The girl who the man is painting is never cheerful
		Congruent	The man who the girl is painting is never cheerful
OR	Show the man who the lady is poking 	Neutral 1	The dog is eating some meat. The food is expensive
		Neutral 2	The panda that is throwing the bamboo is unhealthily dirty
		Incongruent	The lady who the man is poking is disgustingly rich
		Congruent	The man who the lady is poking is disgustingly rich
OR	Show the monkey that the rabbit is touching 	Neutral 1	She found her purse in the trash. A tree fell on the house
		Neutral 2	The speech that the therapist is improvising is thoroughly boring
		Incongruent	The rabbit that the monkey is touching is wonderfully soft
		Congruent	The monkey that the rabbit is touching is wonderfully soft
OR	Show the monkey that the squirrel is pushing 	Neutral 1	The raincoat was dripping wet. The young people are dancing
		Neutral 2	The swimmer who the fireman is saving is completely terrified
		Incongruent	The squirrel that the monkey is pushing is unfortunately glum
		Congruent	The monkey that the squirrel is pushing is unfortunately glum
OR	Show the pig that the cat is beating 	Neutral 1	His father will come home soon. The small tomatoes are green
		Neutral 2	The file that the manager is saving is sadly wrong
		Incongruent	The cat that the pig is beating is extraordinarily interesting
		Congruent	The pig that the cat is beating is extraordinarily interesting

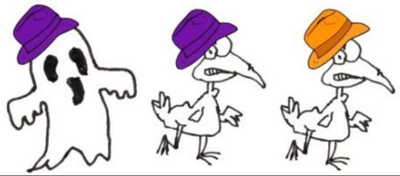
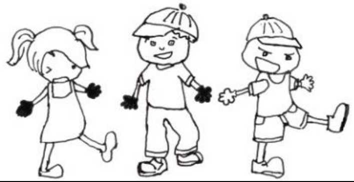
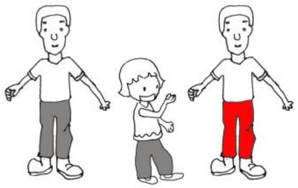
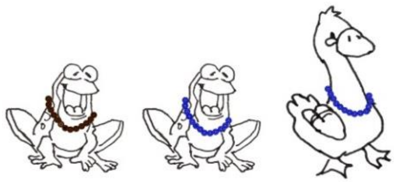
OR	Show the pig that the cow is sniffing 	Neutral 1	They washed in cold water. She lost her credit card
		Neutral 2	The moment that the nurse is enjoying is always positive
		Incongruent	The cow that the pig is sniffing is beautifully precise
		Congruent	The pig that the cow is sniffing is beautifully precise
OR	Show the rabbit that the monkey is hitting 	Neutral 1	Some animals sleep on straw. A letter fell on the floor
		Neutral 2	The party that the reporter is observing is probably done
		Incongruent	The monkey that the rabbit is hitting is exceedingly old
		Congruent	The rabbit that the monkey is hitting is exceedingly old
OR	Show the sheep that the cow is pulling 	Neutral 1	The driver started the car. The jam jar is full
		Neutral 2	The car that the pilot is investigating is unhealthily loud
		Incongruent	The cow that the sheep is pulling is always greedy
		Congruent	The sheep that the cow is pulling is always greedy
OR	Show the sheep that the horse is chasing 	Neutral 1	The bananas were too ripe. The lady washed the shirt
		Neutral 2	The car that the driver is advertising is naturally strong
		Incongruent	The horse that the sheep is chasing is obviously worried
		Congruent	The sheep that the horse is chasing is obviously worried
OR	Show the snake that the goose is smelling 	Neutral 1	They wanted some potatoes. The oven door was open
		Neutral 2	The sound that the singer is checking is wonderfully intriguing
		Incongruent	The goose that the snake is smelling is sometimes cunning
		Congruent	The snake that the goose is smelling is sometimes cunning

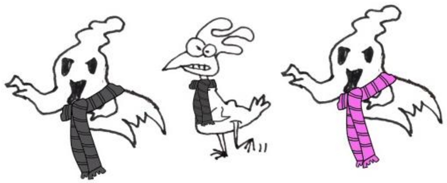
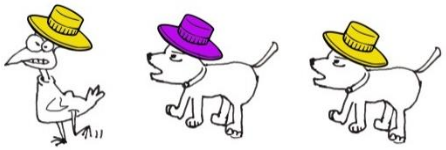
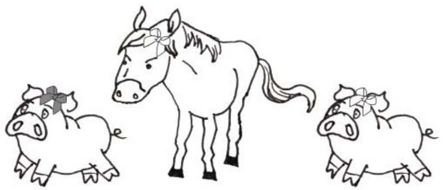
OR	Show the squirrel that the monkey is brushing 	Neutral 1	She took off her fur coat. He hung up his raincoat
		Neutral 2	The puzzle that the technician is creating is fairly complex
		Incongruent	The monkey that the squirrel is brushing is luckily soft
		Congruent	The squirrel that the monkey is brushing is luckily soft
OR	Show the squirrel that the turtle is beating 	Neutral 1	They took some food outside. The picture came from a book
		Neutral 2	The hut that the reporter is examining is sometimes smelly
		Incongruent	The turtle that the squirrel is beating is completely terrified
		Congruent	The squirrel that the turtle is beating is completely terrified
OR	Show the turtle that the donkey is cleaning 	Neutral 1	They followed the garden path. She stands near the window
		Neutral 2	The question that the professor is improvising is unusually good
		Incongruent	The donkey that the turtle is cleaning is completely scary
		Congruent	The turtle that the donkey is cleaning is completely scary
OR	Show the turtle that the rabbit is attacking 	Neutral 1	She cut the steak with her knife. The kitchen sink is empty
		Neutral 2	The helicopter that the pilot is flying is luckily precise
		Incongruent	The rabbit that the turtle is attacking is obviously different
		Congruent	The turtle that the rabbit is attacking is obviously different
OR	Show the turtle that the zebra is pursuing 	Neutral 1	They broke all the brown eggs. The matches are on the shelf
		Neutral 2	The music that the artist is improvising is unusually creative
		Incongruent	The zebra that the turtle is pursuing is fairly exotic
		Congruent	The turtle that the zebra is pursuing is fairly exotic

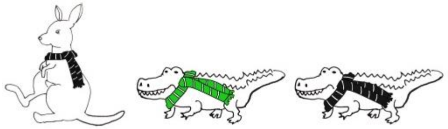
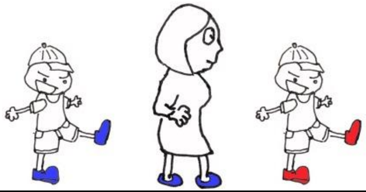
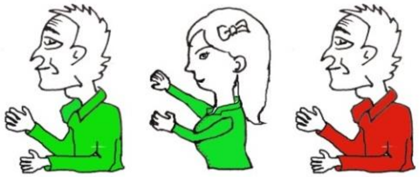
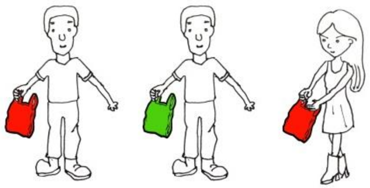
OR	Show the witch that the fox is scratching 	Neutral 1	The paint dripped on the ground. They stared at the picture
		Neutral 2	The eagle that is examining the prey is sometimes cunning
		Incongruent	The fox that the witch is scratching is intensely positive
		Congruent	The witch that the fox is scratching is intensely positive
OR	Show the witch that the ghost is scrubbing 	Neutral 1	The apple pie is baking. They're playing in the park
		Neutral 2	The puzzle that the hedgehog is solving is irrefutably intriguing
		Incongruent	The ghost that the witch is scrubbing is somewhat gentle
		Congruent	The witch that the ghost is scrubbing is somewhat gentle
OR	Show the zebra that the giraffe is splashing 	Neutral 1	The car was going too fast. The park is near the road
		Neutral 2	The illness that the doctor is investigating is unfortunately fierce
		Incongruent	The giraffe that the zebra is splashing is very lively
		Congruent	The zebra that the giraffe is splashing is very lively
OR	Show the zebra that the turtle is harming 	Neutral 1	The children waved at the train. The new road is on the map
		Neutral 2	The village that the architect is planning is already popular
		Incongruent	The turtle that the zebra is harming is ordinarily early
		Congruent	The zebra that the turtle is harming is ordinarily early
Filler	Show the bear with the pink ribbon 	Neutral 1	The apple pie was good. The truck carries fresh fruit
		Neutral 2	The builder with the astonishing imagination is utterly strange
		Incongruent	The cow with the pink ribbon is wonderfully popular
		Congruent	The bear with the pink ribbon is wonderfully popular

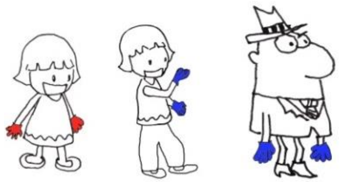
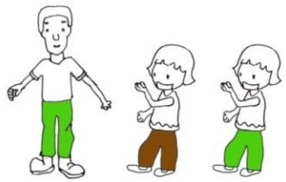
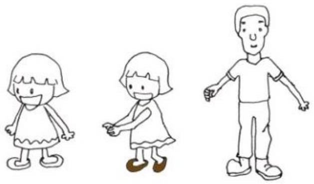
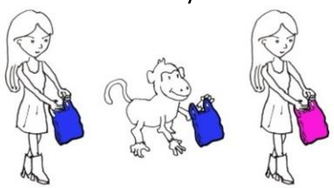
Filler	Show the boy with the yellow shirt 	Neutral 1	The bottle is on the shelf. They're climbing the old oak tree
		Neutral 2	The whale with the tiny doll is obviously rare
		Incongruent	The girl with the yellow shirt is slightly challenging
		Congruent	The boy with the yellow shirt is slightly challenging
Filler	Show the cat with the pink scarf 	Neutral 1	The boy got into trouble. Her husband brought some flowers
		Neutral 2	The architect with the beautiful character is intensely positive
		Incongruent	The sheep with the pink scarf is really quiet
		Congruent	The cat with the pink scarf is really quiet
Filler	Show the cow with the black ribbon 	Neutral 1	A girl came into the room. He's washing his face with soap
		Neutral 2	The nurse with the abundant imagination is beautifully gentle
		Incongruent	The bear with the black ribbon is quite beautiful
		Congruent	The cow with the black ribbon is quite beautiful
Filler	Show the crocodile with the big glasses 	Neutral 1	The family bought a house. The neighbour's boy has black hair
		Neutral 2	The therapist with the difficult paper is never intriguing
		Incongruent	The gorilla with the big glasses is sometimes precise
		Congruent	The crocodile with the big glasses is sometimes precise
Filler	Show the crocodile with the grey hat 	Neutral 1	The little girl is shouting. She spoke to her eldest son
		Neutral 2	The teacher with the beautiful canvas is very sweet
		Incongruent	The kangaroo with the grey hat is obviously terrified
		Congruent	The crocodile with the grey hat is obviously terrified

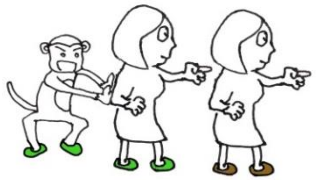
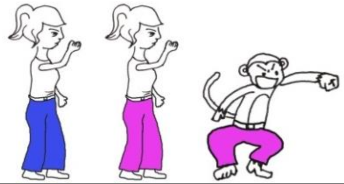
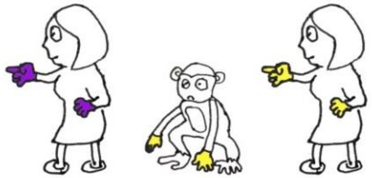
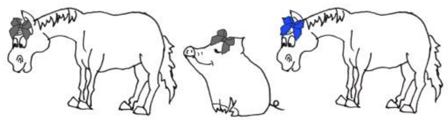
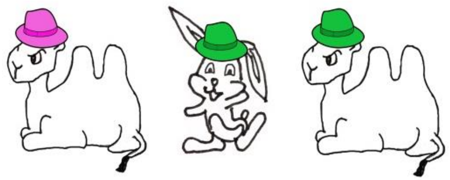
Filler	Show the donkey with the brown hat 	Neutral 1	The man called the police. She's drinking from her own cup
		Neutral 2	The fireman with the unlimited imagination is wonderfully popular
		Incongruent	The rabbit with the brown hat is nevertheless intriguing
		Congruent	The donkey with the brown hat is nevertheless intriguing
Filler	Show the duck with the green necklace 	Neutral 1	A tree fell on the house. A field mouse found the cheese
		Neutral 2	The technician with the bitter cereal is sometimes expensive
		Incongruent	The cow with the green necklace is officially light
		Congruent	The duck with the green necklace is officially light
Filler	Show the elephant with the purple bag 	Neutral 1	The boy forgot his book. The salt shaker was empty
		Neutral 2	The driver with the tiny canvas is unbearably terrified
		Incongruent	The gorilla with the purple bag is offensively strange
		Congruent	The elephant with the purple bag is offensively strange
Filler	Show the fox with the yellow ribbon 	Neutral 1	Father paid at the gate. She writes to her friend daily
		Neutral 2	The magician with the wrong attitude is already done
		Incongruent	The goose with the yellow ribbon is depressingly dark
		Congruent	The fox with the yellow ribbon is depressingly dark
Filler	Show the ghost with the green scarf 	Neutral 1	Her coat is on the chair. The girl ran along the fence
		Neutral 2	The builder with the difficult personality is nevertheless confused
		Incongruent	The bear with the green scarf is intensely good
		Congruent	The ghost with the green scarf is intensely good

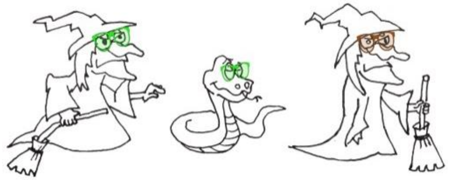
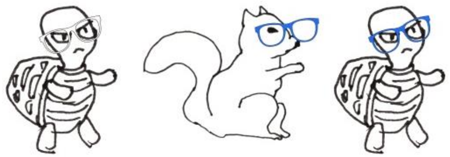
Filler	Show the ghost with the purple hat 	Neutral 1	The shirts are in the closet. The house had nine bedrooms
		Neutral 2	The salesperson with the wrong canvas is probably distracted
		Incongruent	The hen with the purple hat is decidedly poor
		Congruent	The ghost with the purple hat is decidedly poor
Filler	Show the giraffe with the big glasses 	Neutral 1	The truck drove up the road. They broke all the brown eggs
		Neutral 2	The farmer with the colourful character is utterly hot
		Incongruent	The zebra with the big glasses is interestingly ancient
		Congruent	The giraffe with the big glasses is interestingly ancient
Filler	Show the girl with the black gloves 	Neutral 1	The tub tap is leaking. Some animals sleep on straw
		Neutral 2	The parrot with the original puzzle is actually sweet
		Incongruent	The boy with the black gloves is excessively clueless
		Congruent	The girl with the black gloves is excessively clueless
Filler	Show the girl with the grey trousers 	Neutral 1	The postman shut the gate. She cut the steak with her knife
		Neutral 2	The butterfly with the astonishing flower is particularly round
		Incongruent	The man with the grey trousers is brutally burnt
		Congruent	The girl with the grey trousers is brutally burnt
Filler	Show the goose with the blue necklace 	Neutral 1	Snow falls in the winter. The bus leaves before the train
		Neutral 2	The singer with the delicious apple is annoyingly boring
		Incongruent	The frog with the blue necklace is probably icy
		Congruent	The goose with the blue necklace is probably icy

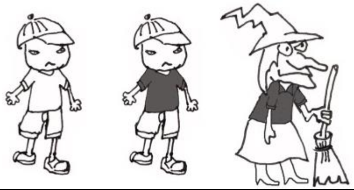
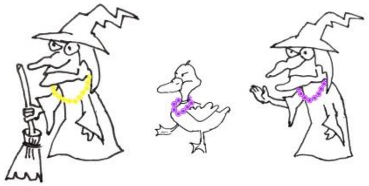
Filler	Show the goose with the green ribbon 	Neutral 1	The young people are dancing. She washed her new silk dress
		Neutral 2	The farmer with the mysterious canvas is annoyingly slimy
		Incongruent	The frog with the green ribbon is exceedingly fragile
		Congruent	The goose with the green ribbon is exceedingly fragile
Filler	Show the hen with the black scarf 	Neutral 1	The new road is on the map. He closed his eyes and jumped
		Neutral 2	The doctor with the difficult paper is excessively sour
		Incongruent	The fox with the black scarf is nevertheless lively
		Congruent	The hen with the black scarf is nevertheless lively
Filler	Show the hen with the grey scarf 	Neutral 1	Sugar is very sweet. The children waved at the train
		Neutral 2	The waiter with the beautiful canvas is completely clueless
		Incongruent	The ghost with the grey scarf is uncharacteristically ancient
		Congruent	The hen with the grey scarf is uncharacteristically ancient
Filler	Show the hen with the yellow hat 	Neutral 1	The big fish got away. The wife helped her husband
		Neutral 2	The officer with the tiny painting is horrifyingly dangerous
		Incongruent	The dog with the yellow hat is annoyingly emotional
		Congruent	The hen with the yellow hat is annoyingly emotional
Filler	Show the horse with the white ribbon 	Neutral 1	The silly boy is hiding. The child drank some fresh milk
		Neutral 2	The athlete with the fresh flower is frankly clueless
		Incongruent	The pig with the white ribbon is ordinarily grateful
		Congruent	The horse with the white ribbon is ordinarily grateful

Filler	Show the kangaroo with the black scarf 	Neutral 1	The police cleared the road. The cow was milked every day
		Neutral 2	The fireman with the mysterious doll is reasonably good
		Incongruent	The crocodile with the black scarf is dismally quiet
		Congruent	The kangaroo with the black scarf is dismally quiet
Filler	Show the kangaroo with the orange ball 	Neutral 1	The baby slept all night. They hear a funny noise
		Neutral 2	The technician with the astonishing puzzle is completely frozen
		Incongruent	The crocodile with the orange ball is depressingly boring
		Congruent	The kangaroo with the orange ball is depressingly boring
Filler	Show the lady with the blue shoes 	Neutral 1	They walked across the grass. The car is going too fast
		Neutral 2	The dragon with the unique eyes is extraordinarily fierce
		Incongruent	The boy with the blue shoes is utterly soft
		Congruent	The lady with the blue shoes is utterly soft
Filler	Show the lady with the green shirt 	Neutral 1	The fire is very hot. The boy ran away from school
		Neutral 2	The shark with the wrong flower is frankly frozen
		Incongruent	The man with the green shirt is unfortunately worried
		Congruent	The lady with the green shirt is unfortunately worried
Filler	Show the lady with the red bag 	Neutral 1	A letter fell on the floor. The dog's chasing the cat
		Neutral 2	The butterfly with the delicious apple is irrefutably exotic
		Incongruent	The man with the red bag is annoyingly lazy
		Congruent	The lady with the red bag is annoyingly lazy

Filler	Show the lion with the brown ball 	Neutral 1	He wore his yellow shirt. The cat caught a little mouse
		Neutral 2	The cook with the difficult attitude is quite worried
		Incongruent	The donkey with the brown ball is possibly exotic
		Congruent	The lion with the brown ball is possibly exotic
Filler	Show the man with the blue gloves 	Neutral 1	The bath water is warm. The car was going too fast
		Neutral 2	The snail with the tiny eyes is frankly slimy
		Incongruent	The girl with the blue gloves is positively funny
		Congruent	The man with the blue gloves is positively funny
Filler	Show the man with the green trousers 	Neutral 1	The park is near the road. She found her purse in the trash
		Neutral 2	The seal with the colourful doll is perfectly frozen
		Incongruent	The girl with the green trousers is actually greedy
		Congruent	The man with the green trousers is actually greedy
Filler	Show the man with the white shoes 	Neutral 1	The sky was very blue. The dog sleeps in a basket
		Neutral 2	The ant with the fresh apple is luckily cheerful
		Incongruent	The girl with the white shoes is decidedly tender
		Congruent	The man with the white shoes is decidedly tender
Filler	Show the monkey with the blue bag 	Neutral 1	A fire engine is coming. Mother tied the string too tight
		Neutral 2	The judge with the astonishing character is decidedly cunning
		Incongruent	The lady with the blue bag is horrifyingly old
		Congruent	The monkey with the blue bag is horrifyingly old

Filler	Show the monkey with the green shoes 	Neutral 1	The old man is worried. The boy broke the wooden fence
		Neutral 2	The pilot with the rotten attitude is possibly complex
		Incongruent	The lady with the green shoes is nearly warm
		Congruent	The monkey with the green shoes is nearly warm
Filler	Show the monkey with the pink trousers 	Neutral 1	New neighbours are moving in. The man cleaned his suede shoes
		Neutral 2	The driver with the unusual puzzle is disgustingly bad
		Incongruent	The lady with the pink trousers is sometimes corrupt
		Congruent	The monkey with the pink trousers is sometimes corrupt
Filler	Show the monkey with the yellow gloves 	Neutral 1	The baby broke his cup. They had some cold meat for lunch
		Neutral 2	The judge with the mysterious paper is extremely sticky
		Incongruent	The lady with the yellow gloves is wonderfully intriguing
		Congruent	The monkey with the yellow gloves is wonderfully intriguing
Filler	Show the pig with the grey ribbon 	Neutral 1	The kitchen sink is empty. She's washing her new silk dress
		Neutral 2	The musician with the astonishing personality is naturally tender
		Incongruent	The horse with the grey ribbon is unhealthily hot
		Congruent	The pig with the grey ribbon is unhealthily hot
Filler	Show the rabbit with the green hat 	Neutral 1	They're watching the cuckoo clock. The floor looks clean and shiny
		Neutral 2	The musician with the mysterious character is slightly funny
		Incongruent	The camel with the pink hat is officially poor
		Congruent	The rabbit with the green hat is officially poor

Filler	Show the rabbit with the red ball 	Neutral 1	They're running past the house. The child ripped open the bag
		Neutral 2	The nurse with the colourful puzzle is fairly cheerful
		Incongruent	The lion with the red ball is particularly bad
		Congruent	The rabbit with the red ball is particularly bad
Filler	Show the sheep with the grey glasses 	Neutral 1	The matches are on the shelf. They painted the wall white
		Neutral 2	The reporter with the original eyes is slightly burnt
		Incongruent	The pig with the grey glasses is thoroughly dry
		Congruent	The sheep with the grey glasses is thoroughly dry
Filler	Show the sheep with the white scarf 	Neutral 1	Children like strawberries. The girl was fixing her dress
		Neutral 2	The salesperson with the fresh cereal is unfortunately violent
		Incongruent	The dog with the white scarf is unusually messy
		Congruent	The sheep with the white scarf is unusually messy
Filler	Show the snake with the green glasses 	Neutral 1	The fire was very hot. She injured four of her fingers
		Neutral 2	The lawyer with the tiny puzzle is always creative
		Incongruent	The witch with the green glasses is naturally strong
		Congruent	The snake with the green glasses is naturally strong
Filler	Show the squirrel with the blue glasses 	Neutral 1	The cleaner swept the floor. Men normally wear long trousers
		Neutral 2	The cook with the original puzzle is precisely burnt
		Incongruent	The turtle with the blue glasses is finally frozen
		Congruent	The squirrel with the blue glasses is finally frozen

Filler	Show the witch with the grey shirt 	Neutral 1	Rain is good for the trees. The raincoat was dripping wet
		Neutral 2	The fish with the rotten paper is extraordinarily rare
		Incongruent	The boy with the grey shirt is uncharacteristically rude
		Congruent	The witch with the grey shirt is uncharacteristically rude
Filler	Show the witch with the yellow necklace 	Neutral 1	A cat jumped over the fence. The milkman drives a small truck
		Neutral 2	The bat with the beautiful eyes is officially dangerous
		Incongruent	The duck with the purple necklace is offensively lazy
		Congruent	The witch with the yellow necklace is offensively lazy
Filler	Show the zebra with the blue hat 	Neutral 1	A fish swam in the pond. Dad stopped to pick some pears
		Neutral 2	The painter with the bitter cereal is officially great
		Incongruent	The squirrel with the blue hat is extremely sticky
		Congruent	The zebra with the blue hat is extremely sticky

Appendix C List of modified HINT sentences

Original sentences were published in Nilsson et al., (1994). The sentences listed below were modified to correspond to British English usage.

List	Item	Sentence presented	Original HINT sentence
2	10	The bath tap is leaking	The tub faucet is leaking
4	6	He needs his holiday	He needs his vacation
6	6	They went on holidays	They went on vacation
6	8	The postman shut the gate	The mailman shut the gate
12	5	They had some cold meat for lunch	They had some cold cuts for lunch
18	5	The cleaner swept the floor	The janitor swept the floor
23	9	The postman brought a letter	The mailman brought a letter
24	4	The jam jar is full	The jelly jar is full
25	9	The sweet shop is empty	The candy shop is empty
Practice 1	3	The front garden is pretty	The front yard is pretty
Practice 2	1	Men normally wear long trousers	Men normally wear long pants

In the original paper there were 25 lists with 10 sentences each and three lists with 12 sentences each. However, sentence number 3 from the third practice list was not reported in the original paper.

The following sentences were added to the list of HINT sentences, so that the overall number of HINT sentences would be at least equal to the number of target sentences. Each of these sentences was based on a pre-existing HINT sentence.

New sentence	Based on sentence from NEUTRAL 1
The shoes are very dirty	List 1 item 4: Her shoes were very dirty
She stands near the window	List 8 item 4: She stood near the window
The fire was very hot	List 1 item 7: The fire is very hot
The car was going too fast	List 1 item 10: The car is going too fast
She wore her yellow shirt	List 4 item 4: He wore his yellow shirt
The lady sat in her chair	List 4 item 5: The lady sits in her chair
She washed her new silk dress	List 4 item 7: She’s washing her new silk dress
The silly boy was hiding	List 6 item 1: The silly boy is hiding
The tree fell on a house	List 6 item 3: A tree fell on the house
She paid for the bread	List 7 item 6: She’s paying for her bread

Appendix D Experiment 2, individual results for cognitive measures

Participant	Non-word repetition	Listening recall	Flanker difference (ms)
1	95	108	34
2	98	113	31
3	91	70	52
4	95	122	14
5	92	91	60
6	91	80	37
7	104	101	24
8	100	80	37
9	113	119	66
10	104	98	45
11	109	108	73
12	104	101	36
13	91	73	29
14	77	98	31
15	91	77	23
16	100	105	34
17	95	115	63
18	95	80	27
19	118	115	32
20	100	105	47
21	82	70	48
22	86	91	78
23	100	105	44
24	104	119	74
25	77	91	61
26	104	80	36
27	113	87	56
28	100	80	34
29	98	91	42
30	104	101	3
31	95	101	81
32	104	101	53
33	100	115	43
34	73	80	34
35	109	84	54
36	118	101	10

Table D.1. Experiment 2. Individual results for cognitive measures (non-word standard score, listening recall standard score, flanker task difference between incongruent and congruent.

Appendix E Proficiency questionnaire[31] for Experiment 3

E.1 English version

To help us understand how to interpret our results, please fill in the information below.

1. Have you always lived in Denmark? YES / NO
   If you have lived elsewhere, where and for how long?
2. Do you speak any other languages in addition to Danish and English? YES / NO
   If you speak another language, which one(s) and for how long have you spoken them?
   On a scale of 1 (limited knowledge) to 7 (native or near-native), how would you rate your proficiency in this/these additional language(s)?
3. Have you ever needed speech and language therapy? YES / NO
4. Have you ever been diagnosed with dyslexia? YES / NO
5. For how many years did you study English? ____ *years*
6. Have you studied English outside of compulsory school classes? YES / NO
   If you answered yes to question 6, where else did you study English and for how long?
7. Have you ever taken a standardised English language test? (e.g. IELTS, TOEFL)
   If you have taken a standardised English language test, which one and what were your (approximate) results?
8. On a scale of 0 (not at all) to 10 (all the time), how often do you use English in the following situations:
   With friends
   1 2 3 4 5 6 7 8 9 10
   With family
   1 2 3 4 5 6 7 8 9 10
   At work
   - Reading/writing
   1 2 3 4 5 6 7 8 9 10
   - Speaking
   1 2 3 4 5 6 7 8 9 10
   For your studies
   - Reading/writing
   1 2 3 4 5 6 7 8 9 10
   - Speaking
   1 2 3 4 5 6 7 8 9 10
   During leisure activities (e.g. part of a club or a team)
   1 2 3 4 5 6 7 8 9 10
   Reading for fun (e.g. books, novels, magazines)

[31] Based on the questionnaire developed by MacIntyre, Noels, & Clément, 1997. Participants were given the Danish translation of this questionnaire.

1 2 3 4 5 6 7 8 9 10

Watching television or movies

1 2 3 4 5 6 7 8 9 10

Listening to the radio/podcasts

1 2 3 4 5 6 7 8 9 10

9. On a scale of 0 (impossible) to 10 (very comfortable), how comfortable would you feel in the following situations. Circle as appropriate.

On the telephone, understand a native English speaker who is speaking slowly and carefully (i.e., deliberately adapting his or her speech to suit you)

1 2 3 4 5 6 7 8 9 10

Understand two native English speakers when they are talking rapidly with one another.

1 2 3 4 5 6 7 8 9 10

In face-to-face conversation, understand a native English speaker who is speaking slowly and carefully (i.e., deliberately adapting his or her speech to suit you).

1 2 3 4 5 6 7 8 9 10

In face-to-face conversation, understand native English speakers who are talking to you as quickly and colloquially as they would to another English speaker.

1 2 3 4 5 6 7 8 9 10

Understand English movies without subtitles

1 2 3 4 5 6 7 8 9 10

Understand news broadcasts on the radio

1 2 3 4 5 6 7 8 9 10

Describe the Danish educational system in some detail.

1 2 3 4 5 6 7 8 9 10

Talk about your favourite hobby at some length, using appropriate vocabulary.

1 2 3 4 5 6 7 8 9 10

Give a brief description of a picture (e.g., photograph or picture in an art gallery) while looking at it.

1 2 3 4 5 6 7 8 9 10

Fill out a job application form requiring information about your interests and qualifications.

1 2 3 4 5 6 7 8 9 10

E.2 Proficiency questionnaire in Danish[32]

Udfyld venligst nedenstående for at hjælpe os med at tolke vores resultater.

1. Har du altid boet i Danmark? JA / NEJ
   Hvis du har boet andre steder, angiv her hvor og hvor længe.
2. Taler du andre sprog udover dansk og engelsk? JA / NEJ
   Hvis du taler andre sprog, angiv her hvilke og hvor længe du har talt dem.
   Angiv på en skala fra 1 (begrænset viden) til 7 (modersmål eller lignende) dine sprogkyndigheder i dette/disse ekstra sprog.
3. Har du nogensinde haft brug for en sprog- eller talepædagog? JA / NEJ
4. Er du nogensinde blevet diagnosticeret med ordblindhed? JA / NEJ
5. I hvor mange år har du modtaget engelsk-undervisning? ____ *år*
6. Har du studeret engelsk ud over den obligatoriske skole/gymnasieundervisning?
   JA / NEJ
   Hvis du har svaret "ja" til spørgsmål 6, angiv her hvor du har studeret engelsk og hvor længe.
7. Har du nogensinde taget en standardiseret engelsk-sprogprøve? (f.eks. IELTS, TOEFL)
   JA / NEJ
   Hvis du har taget en standardiseret engelsk-sprogprøve, angiv her hvilken og hvad dine (omtrentlige) resultater var.
8. Angiv på en skala fra 0 (aldrig) til 10 (altid) hvor ofte du burger engelsk i de følgende situationer:
   Med venner
   1 2 3 4 5 6 7 8 9 10
   Med familie
   1 2 3 4 5 6 7 8 9 10
   På arbejde
   - Læser/skriver
   1 2 3 4 5 6 7 8 9 10
   - Taler
   1 2 3 4 5 6 7 8 9 10
   På studiet
   - Læser/skriver
   1 2 3 4 5 6 7 8 9 10
   - Taler
   1 2 3 4 5 6 7 8 9 10
   I forbindelse med fritidsaktiviteter (f.eks. del af en klub eller et hold)
   1 2 3 4 5 6 7 8 9 10
   Læsning for underholdningens skyld (f.eks. bøger, noveller, magasiner/blade)
   1 2 3 4 5 6 7 8 9 10
   I forbindelse med TV / film
   1 2 3 4 5 6 7 8 9 10

---

[32] Translated into Danish by Simon Krogholt Christiansen, Technical University of Denmark.

I forbindelse med radio/podcasts

1 2 3 4 5 6 7 8 9 10

9. På en skala fra 0 (umuligt) til 10 (meget tilpas), angiv hvor tilpas du ville føle dig i de følgende situationer. Tegn cirkler som du finder passende.

   I en telefonsamtale hvor du skal forstå en person der taler engelsk som modersmål, og som taler langsomt og omhyggeligt (dvs. med vilje tilpasser hans/hendes tale til dig).

   1 2 3 4 5 6 7 8 9 10

   Forstå en samtale mellem to personer hvis modersmål er engelsk og som taler (hurtigt) sammen.

   1 2 3 4 5 6 7 8 9 10

   I en samtale, ansigt-til-ansigt, hvor du skal forstå en person der taler engelsk som modersmål, og som taler langsomt og omhyggeligt (dvs. med vilje tilpasser hans/hendes tale til dig).

   1 2 3 4 5 6 7 8 9 10

   I en samtale, ansigt-til-ansigt, hvor du skal forstå en person der taler engelsk som modersmål, og som taler til dig med samme hastighed og bruger udtryk som han/hun ville gøre til en anden person hvis modersmål var engelsk.

   1 2 3 4 5 6 7 8 9 10

   Forstå engelske film uden undertekster.

   1 2 3 4 5 6 7 8 9 10

   Forstå engelske nyhedsudsendelser i radioen.

   1 2 3 4 5 6 7 8 9 10

   Beskrive det danske uddannelsessystem i en vis grad.

   1 2 3 4 5 6 7 8 9 10

   Føre en længerevarende samtale om din yndlingshobby ved brug af et passende ordforråd.

   1 2 3 4 5 6 7 8 9 10

   Give en kort beskrivelse af et billede (f.eks. et fotografi eller et billede i et kunstgalleri) mens du kigger på det.

   1 2 3 4 5 6 7 8 9 10

   Udfyldning af en jobansøgningsformular der kræver information omkring dine interesser og kvalifikationer.

   1 2 3 4 5 6 7 8 9 10

Appendix F Individual results for additional measures in Experiment 3

F.1 English language proficiency: Lextale

Participant	Lextale score in %	Participant	Lextale score in %
1	86	11	88
2	74	12	54
3	92	13	96
4	75	14	83
5	76	15	68
6	63	16	85
7	47	17	87
8	34	18	84
9	52	19	53
10	73		

Table F.1. Experiment 3. Lextale scores by participant.

F.2 English language proficiency: self-report questionnaire (sections 8 and 9 only)

Participant	Friends	Family	Studies reading/ writing	Studies speaking	Leisure	Reading	TV/ movies	Radio/ podcasts
1	9	6	9	9	5	10	9	5
2	4	6	10	3	1	3	9	1
3	5	2	8	4	2	9	8	3
4	6	1	8	1	2	1	6	2
5	5	2	6	7	8	8	10	7
7	2	1	9	3	2	1	10	1
8	2	5	10	1	1	10	10	1
9	3	1	8	3	7	5	9	2
10	4	2	7	4	3	6	8	8
11	5	3	9	4	2	8	9	1
12	1	1	8	3	2	8	9	4
13	8.5	5	7	7	5	9	10	9
14	2	2	10	5	1	7	9	1
15	3	1	8	1	1	4	9	3
16	7	7	6	5	8	10	10	9
17	7	5	9	8	NA	8	8	6
18	5	1	9	7	3	10	10	10
19	3	3	1	1	1	1	8	1

Table F.2. Experiment 3. Individual ratings (1 to 10) for frequency of use of English in different contexts.

F.3 Short-term and working memory spans for Experiment 3

Participant	Forward span	Backward span
2	6	5
3	8	8
4	8	3
5	8	4
6	7	7
7	6	5
8	5	3
9	5	5
10	6	5
11	6	5
12	5	3
13	8	6
14	6	6
15	6	5
16	8	7
17	6	5
18	6	4

Table F.3. Experiment 3. Results for the forward and backward digit span tasks.

Participant	Rspan letter	Rspan meaning
2	3	6
4	2	7
5	5	9
6	4	6
7	5	8
8	4	7
9	5	8
10	5	9
11	6	5
12	6	5
13	5	10
14	3	7
15	4	7
16	5	8
17	6	10
18	5	7
19	5	5

Table F.4. Experiment 3. Reading span results for maximum number of letters recalled (Rspan letter), and accuracy in the truth-value judgments (Rspan meaning) of the sentences

F.4 Selective attention: flanker task results for Experiment 3

Participant	Consistent	Inconsistent	Neutral	Inconsistent-Consistent
1	357	442	358	85
2	369	447	390	78
3	395	446	405	52
4	454	511	451	57
5	475	501	489	27
6	519	574	486	55
7	391	472	424	81
8	503	530	533	27
11	382	425	368	42
12	395	430	425	35
13	336	403	374	67
15	487	485	470	-2
16	499	521	486	22
17	396	458	415	62
18	470	503	477	34
19	525	564	547	39

Table F.5. Experiment 3. Reaction times in ms per participant for each of the flanker task conditions. RTs include only accurate responses

Appendix G Pure-tone audiometry thresholds for Experiment 3

Right ear	**Frequency (Hz)**									
Participant	**125**	**250**	**500**	**1000**	**2000**	**3000**	**4000**	**6000**	**8000**	**Audiometer**
1	10	-5	0	-5	-10	5	5	5	0	AA222
2	5	15	15	10	0	0	-5	0	0	AA222
3	0	-5	-5	0	5	0	0	-5	-5	AA222
4	0	0	-5	0	-5	0	5	0	20	AA222
5	5	0	0	5	0	5	5	-5	10	AA222
6	30	30	30	20	10	25	15	15	20	AA222
7	10	-5	0	0	0	0	0	0	0	AA222
8	5	10	10	0	5	10	15	-5	0	AS216
9	5	-5	-5	0	0	-5	0	0	-10	AA222
10	0	-5	-5	-5	-5	0	-10	-5	0	AA222
11	0	0	-5	0	10	10	5	15	0	AS216
12	0	5	0	0	5	0	5	5	15	AS216
13	0	0	0	0	0	0	-5	10	10	AS216
15	-5	0	0	-5	0	0	5	5	5	AS216
16	5	5	0	10	5	5	5	20	10	AS216
17	-5	5	5	5	15	0	10	15	25	AS216
18	0	-5	0	-5	-5	-5	5	5	10	AS216
19	0	5	5	0	5	5	0	10	25	AS216

Left ear	**Frequency (Hz)**									
Participant	**125**	**250**	**500**	**1000**	**2000**	**3000**	**4000**	**6000**	**8000**	**Audiometer**
1	0	-5	-5	-5	0	0	5	0	0	AA222
2	5	10	-5	10	-10	-10	-5	0	0	AA222
3	0	-5	0	0	-5	5	5	0	5	AA222
4	5	0	5	0	-5	5	0	-5	15	AA222
5	0	-5	0	0	-5	0	0	-5	5	AA222
6	20	20	20	10	-5	20	5	5	10	AA222
7	0	0	-10	0	-10	-10	0	0	0	AA222
8	10	5	5	5	15	10	10	-5	5	AS216
9	10	0	5	-10	-10	5	0	-5	5	AA222
10	0	-5	0	0	-5	0	0	0	-5	AA222
11	5	5	10	10	15	10	0	5	25	AS216
12	0	0	0	0	10	0	10	15	-	AS216
13	0	0	-5	0	5	0	0	5	10	AS216
15	-5	0	0	0	0	0	0	5	10	AS216
16	0	-5	0	5	0	10	5	15	15	AS216
17	5	-5	5	5	30	20	20	35	40	AS216
18	-5	0	-5	-5	-5	-5	5	5	20	AS216
19	0	0	5	5	0	10	10	5	20	AS216

Appendix H Eye-fixation graphs for Experiment 3

H.1 Average fixation rates and *t*-statistics between target and competitor for Experiment 3

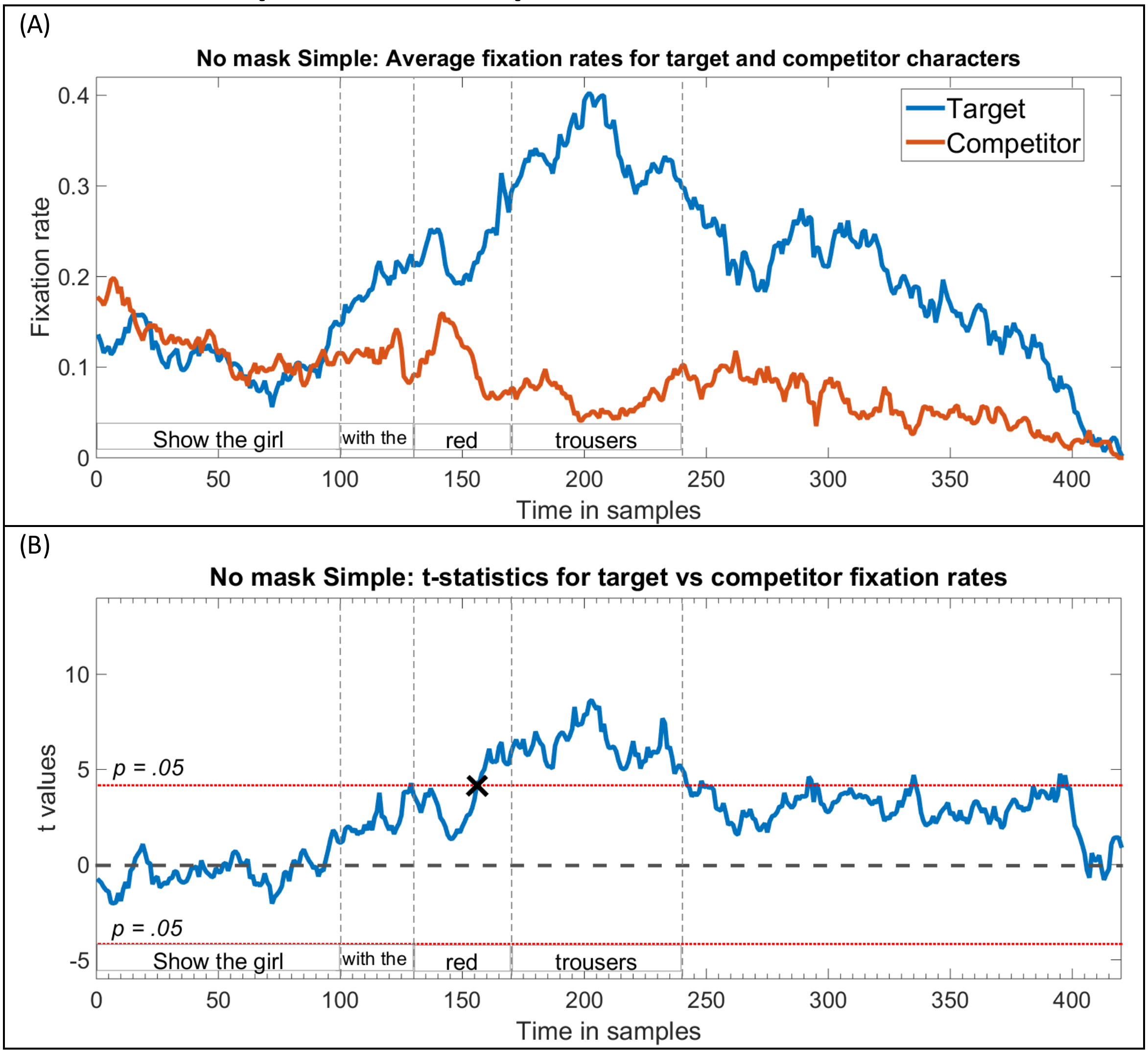


Figure H.1. Experiment 3. Average fixation rates to the target and the competitor (A), and values of *t*-statistics for the difference between target and competitor (B), for the no mask condition with simple sentences. In both panels, the horizontal dashed lines indicate the borders of the sentence segments, with an example sentence above the X-axis. In panel (B), the red dotted lines above and below 0 are plotted at the critical value of t where $p = .05$, and the black cross indicates the average point at which participants fixated the target more than the competitor for 20 samples or more.

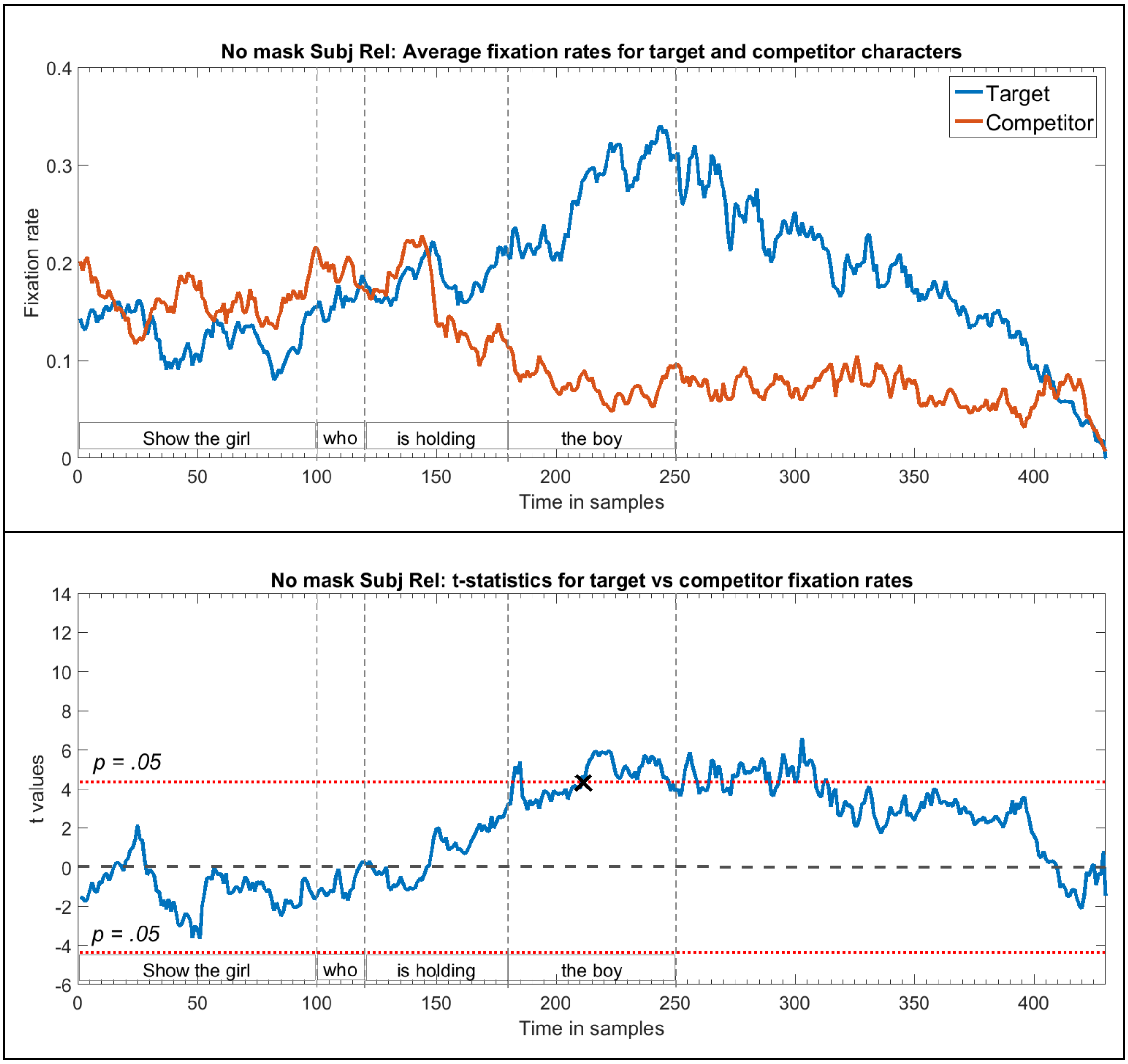


Figure H.2. Experiment 3. Average fixation rates and *t*-statistics for the no mask condition with subject relative sentences.

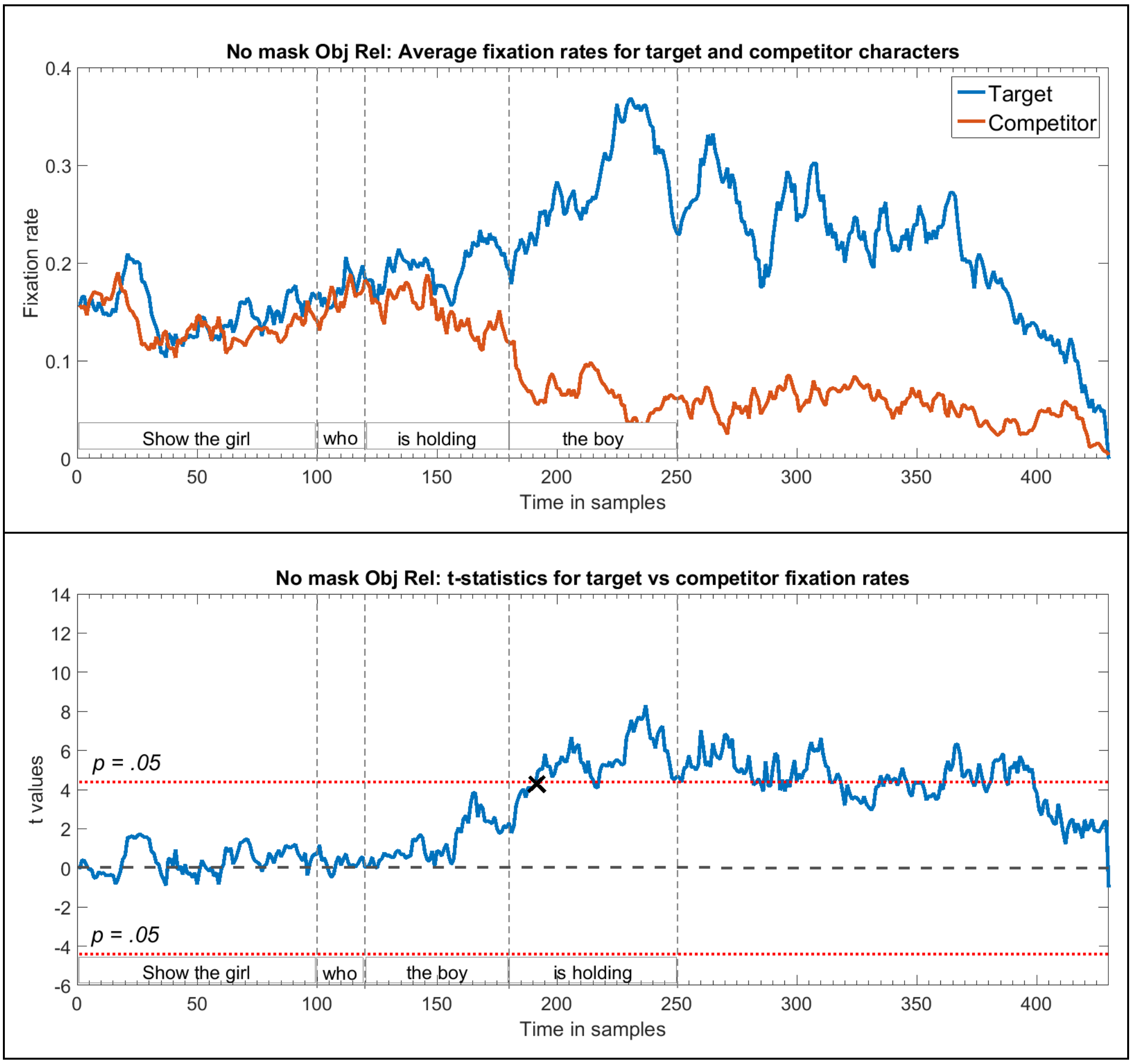


Figure H.3. Experiment 3. Average fixation rates and *t*-statistics for the no mask condition with object relative sentences.

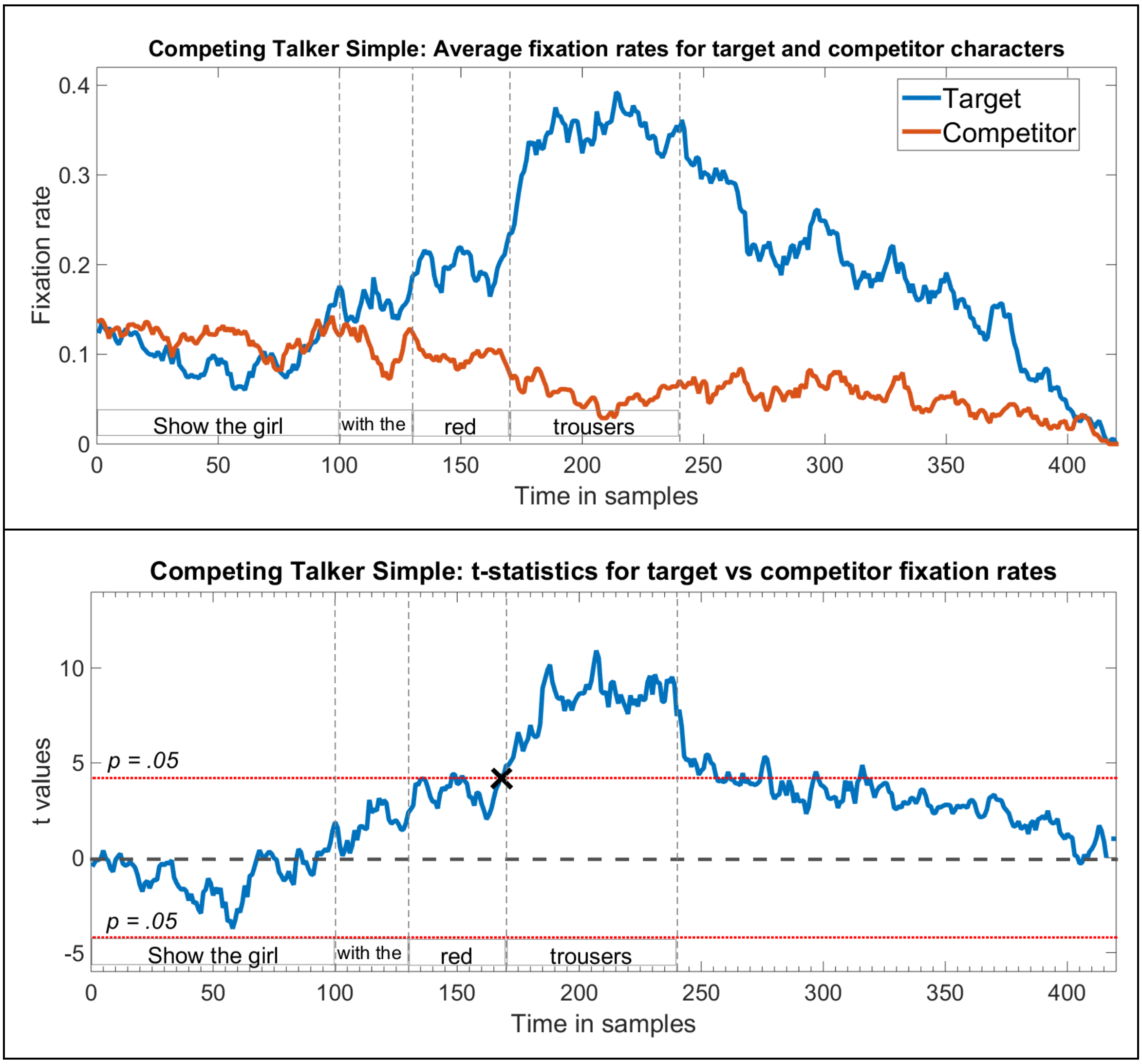


Figure H.4. Experiment 3. Average fixation rates and *t*-statistics for the competing talker condition with simple sentences.

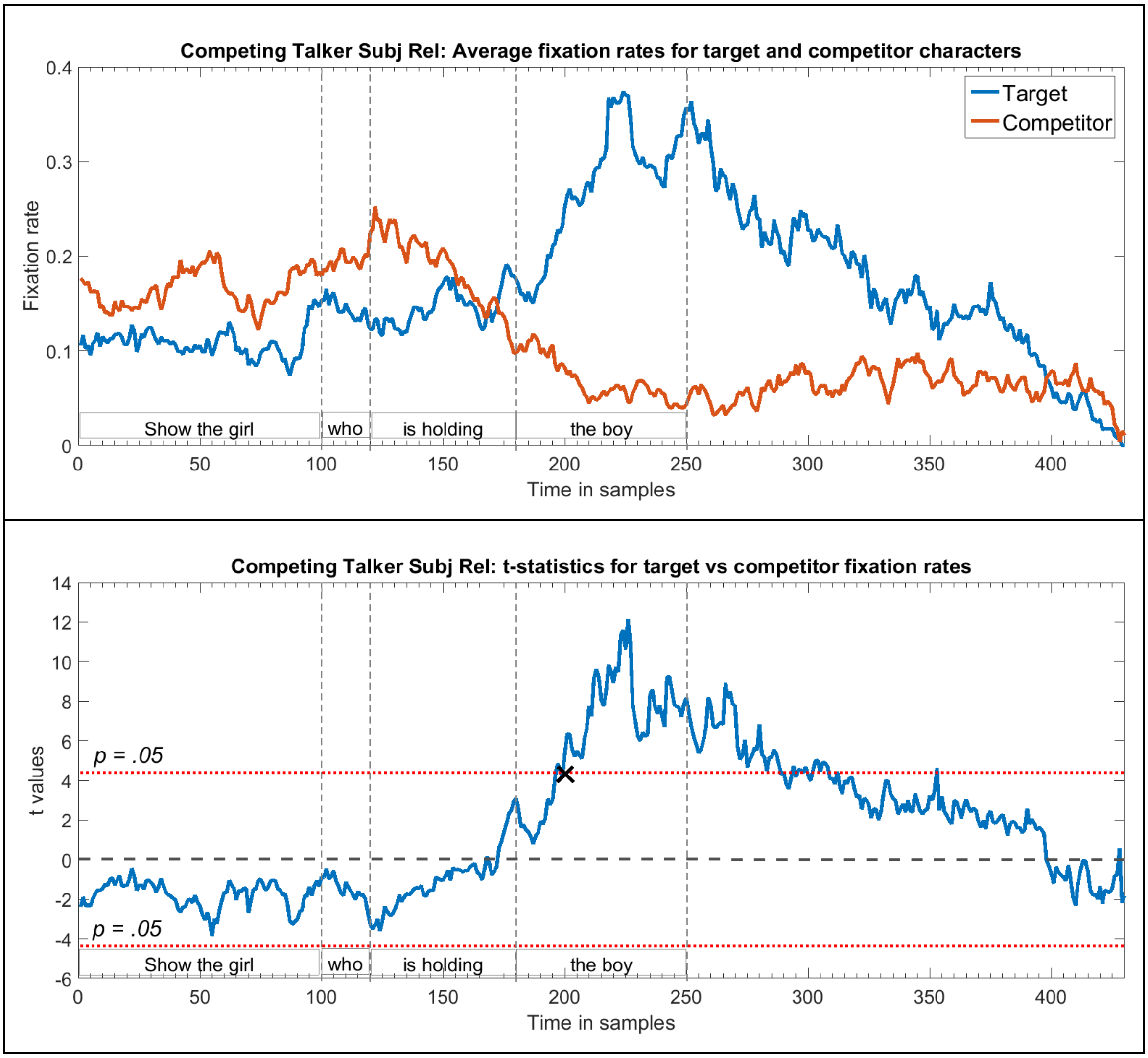


Figure H.5. Experiment 3. Average fixation rates and *t*-statistics for the competing talker condition with subject relative sentences.

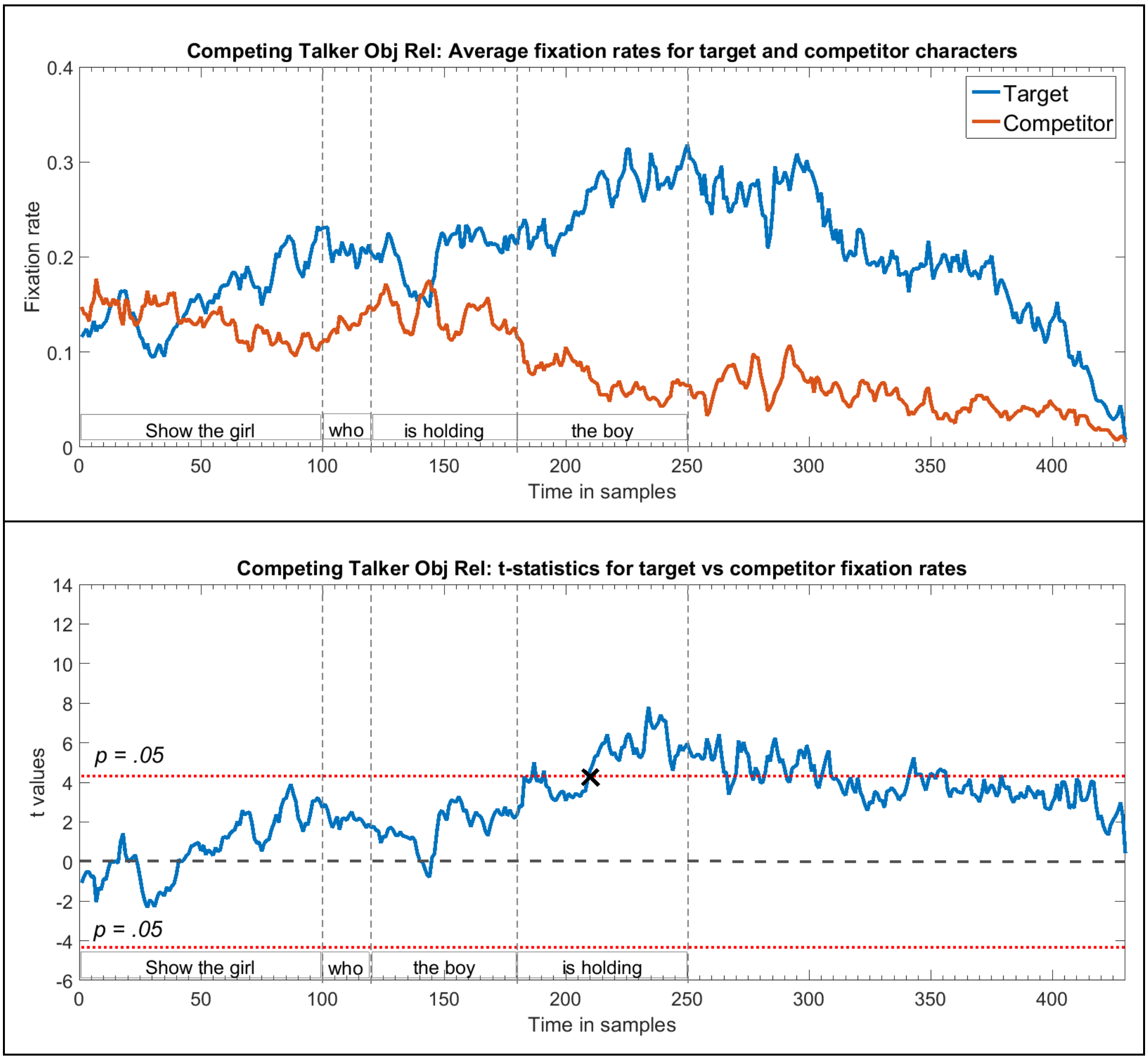


Figure H.6. Experiment 3. Average fixation rates and *t*-statistics for the competing talker condition with object relative sentences.

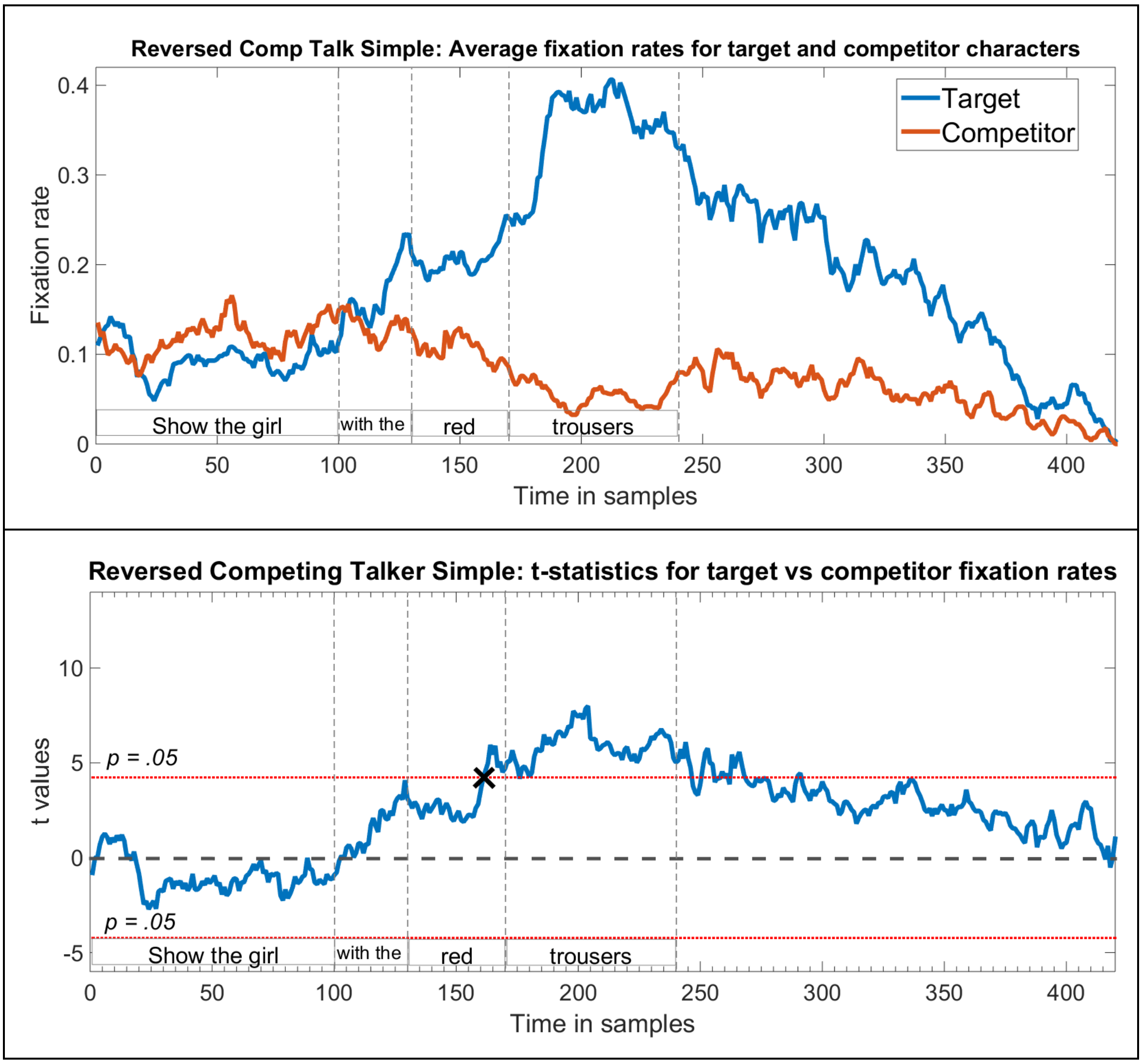


Figure H.7. Experiment 3. Average fixation rates and *t*-statistics for the reversed competing talker condition with simple sentences.

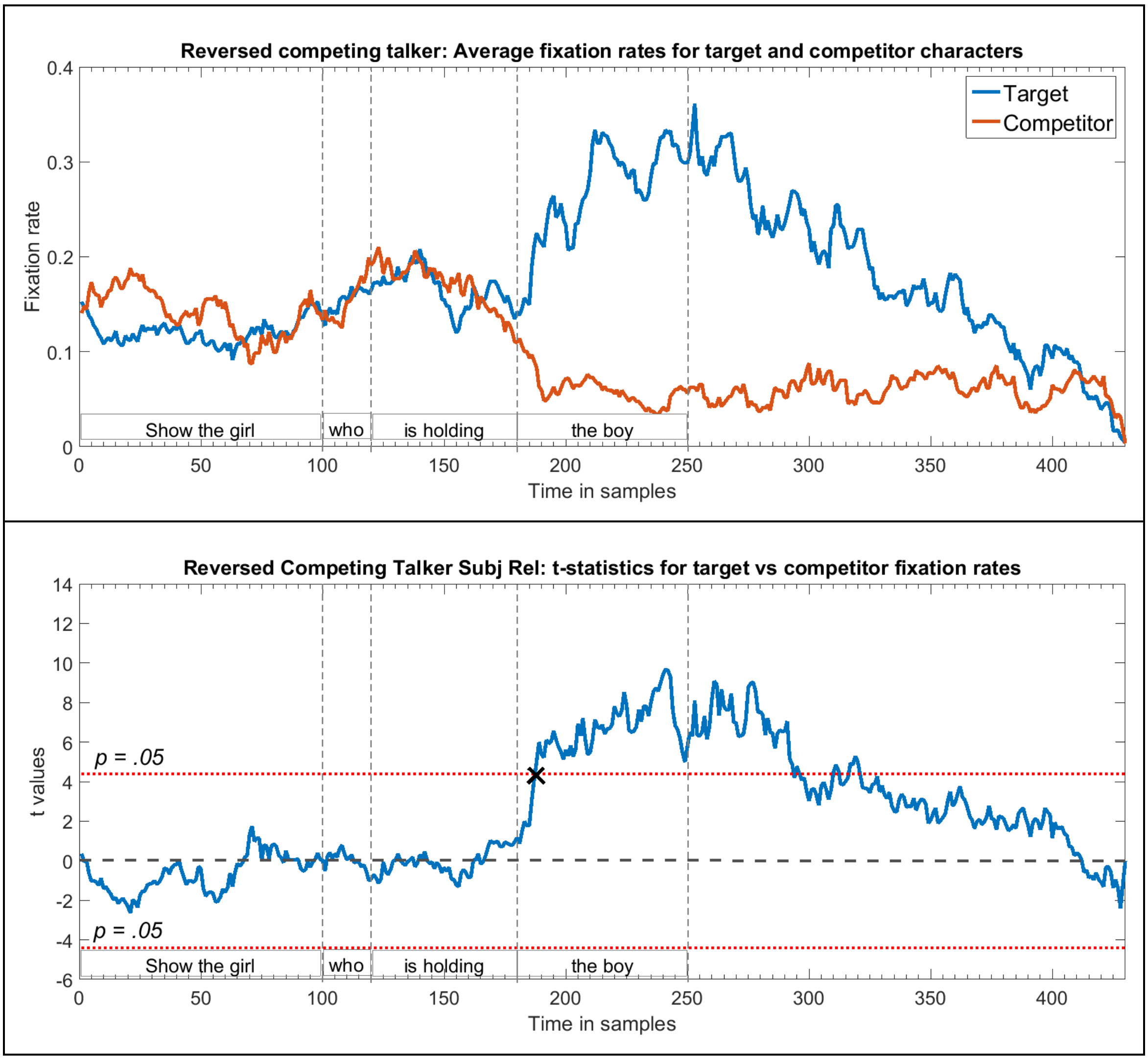


Figure H.8. Experiment 3. Average fixation rates and *t*-statistics for the reversed competing talker condition with subject relative sentences.

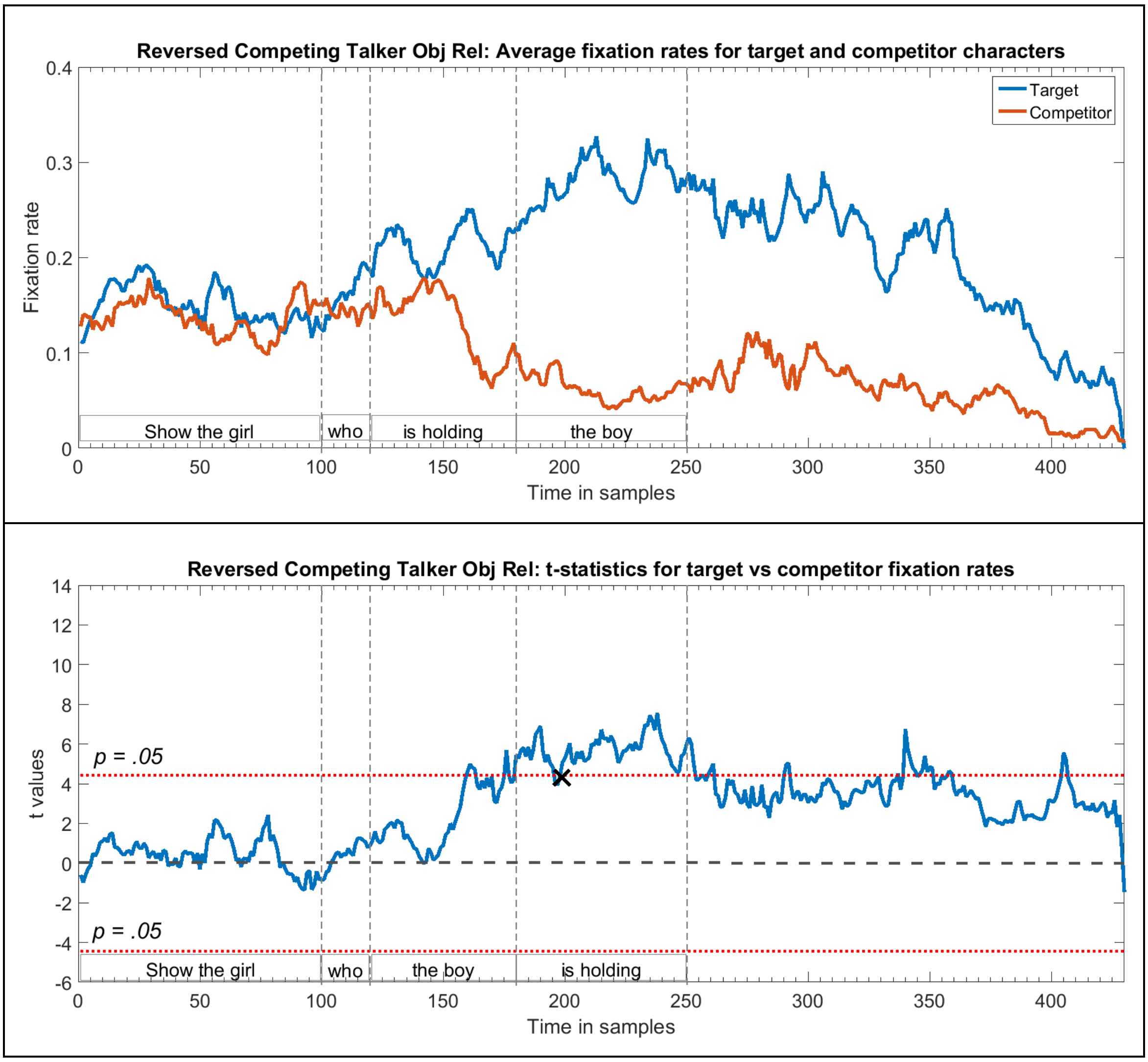


Figure H.9. Experiment 3. Average fixation rates and *t*-statistics for the reversed competing talker condition with object relative sentences.

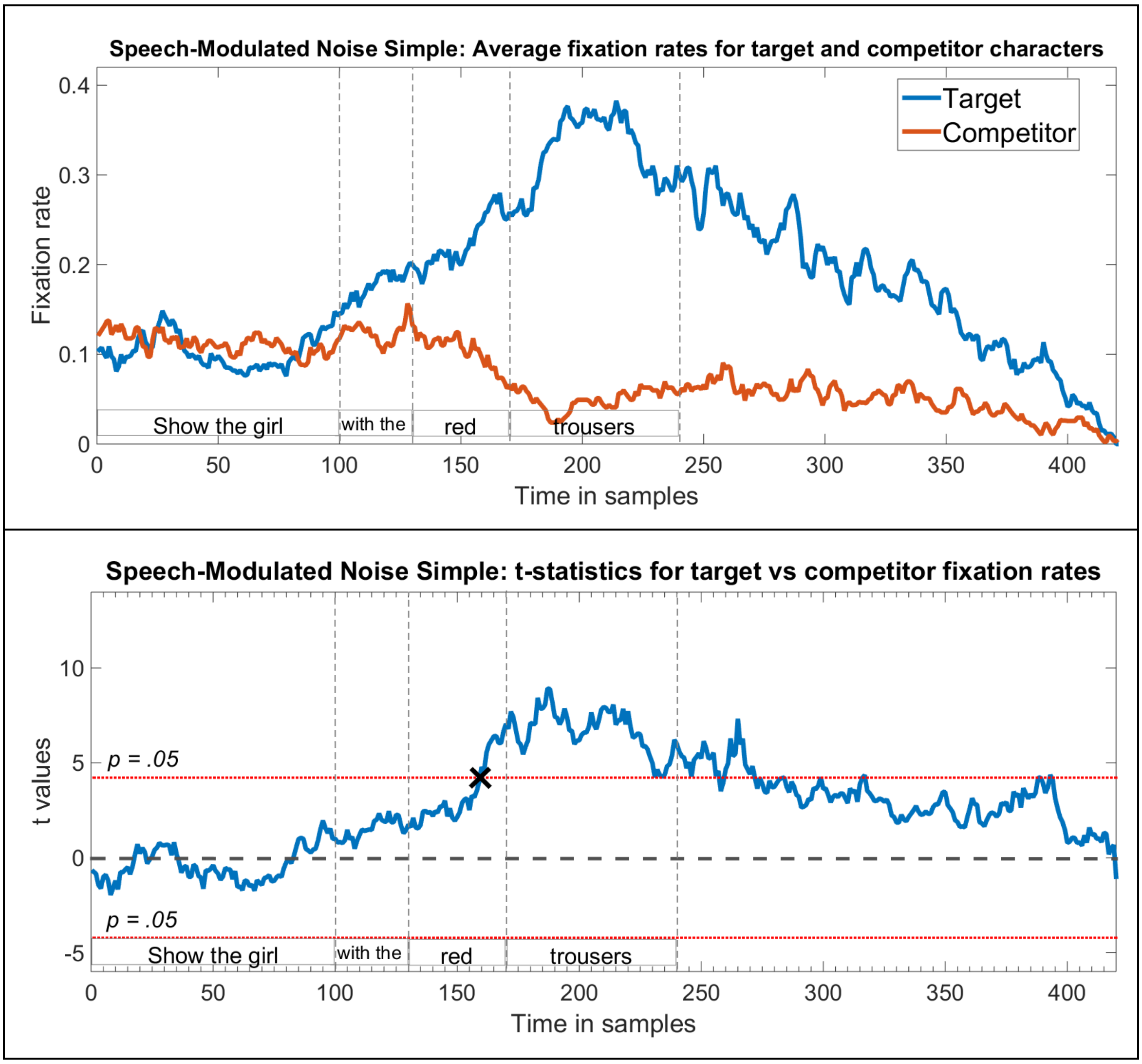


Figure H.10. Experiment 3. Average fixation rates and *t*-statistics for the speech-modulated noise condition with simple sentences.

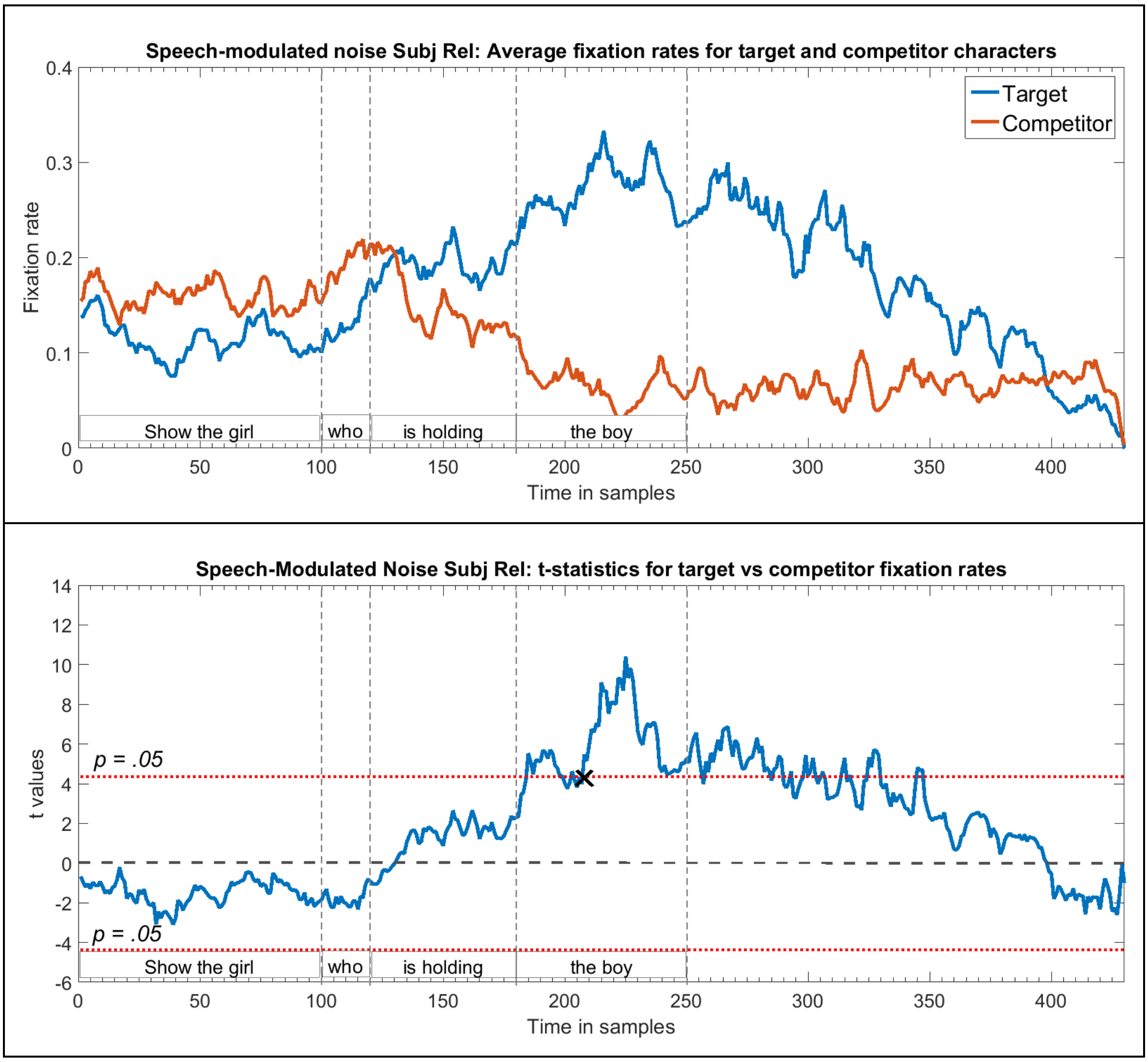


Figure H.11. Experiment 3. Average fixation rates and *t*-statistics for the speech-modulated noise condition with subject relative sentences.

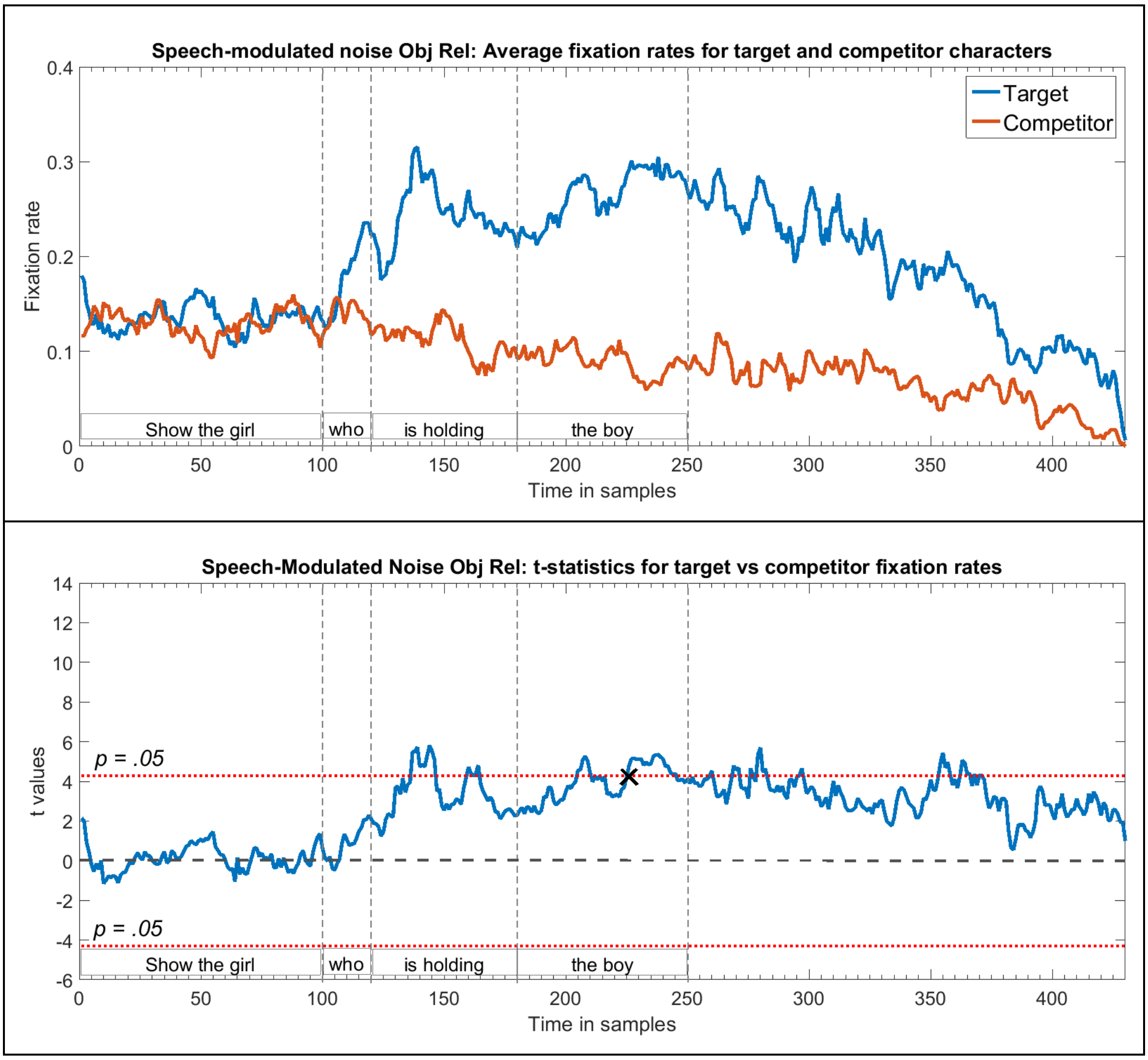
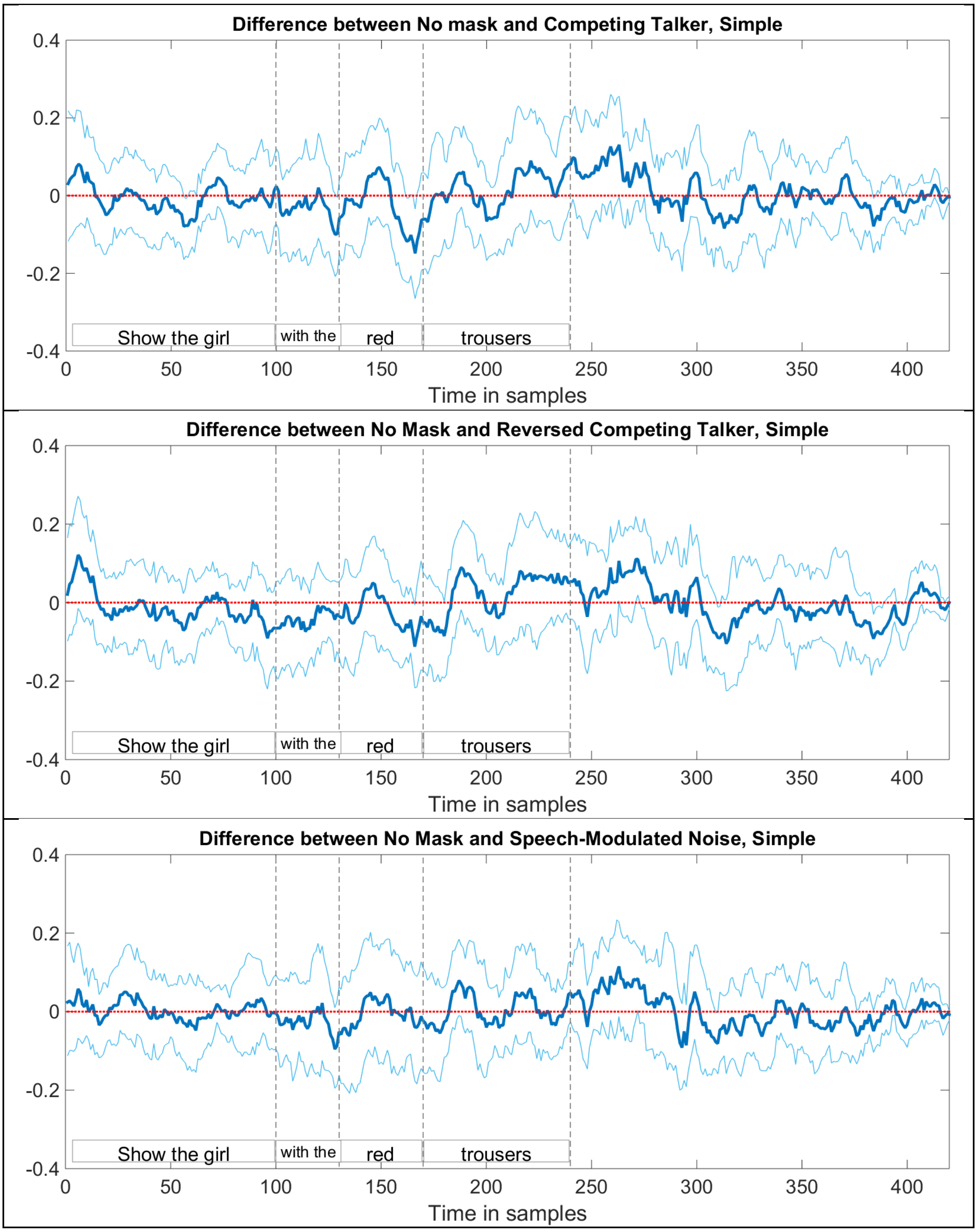


Figure H.12. Experiment 3. Average fixation rates and *t*-statistics for the speech-modulated noise condition with object relative sentences.

H.2 Pairwise comparisons between masks for Experiment 3



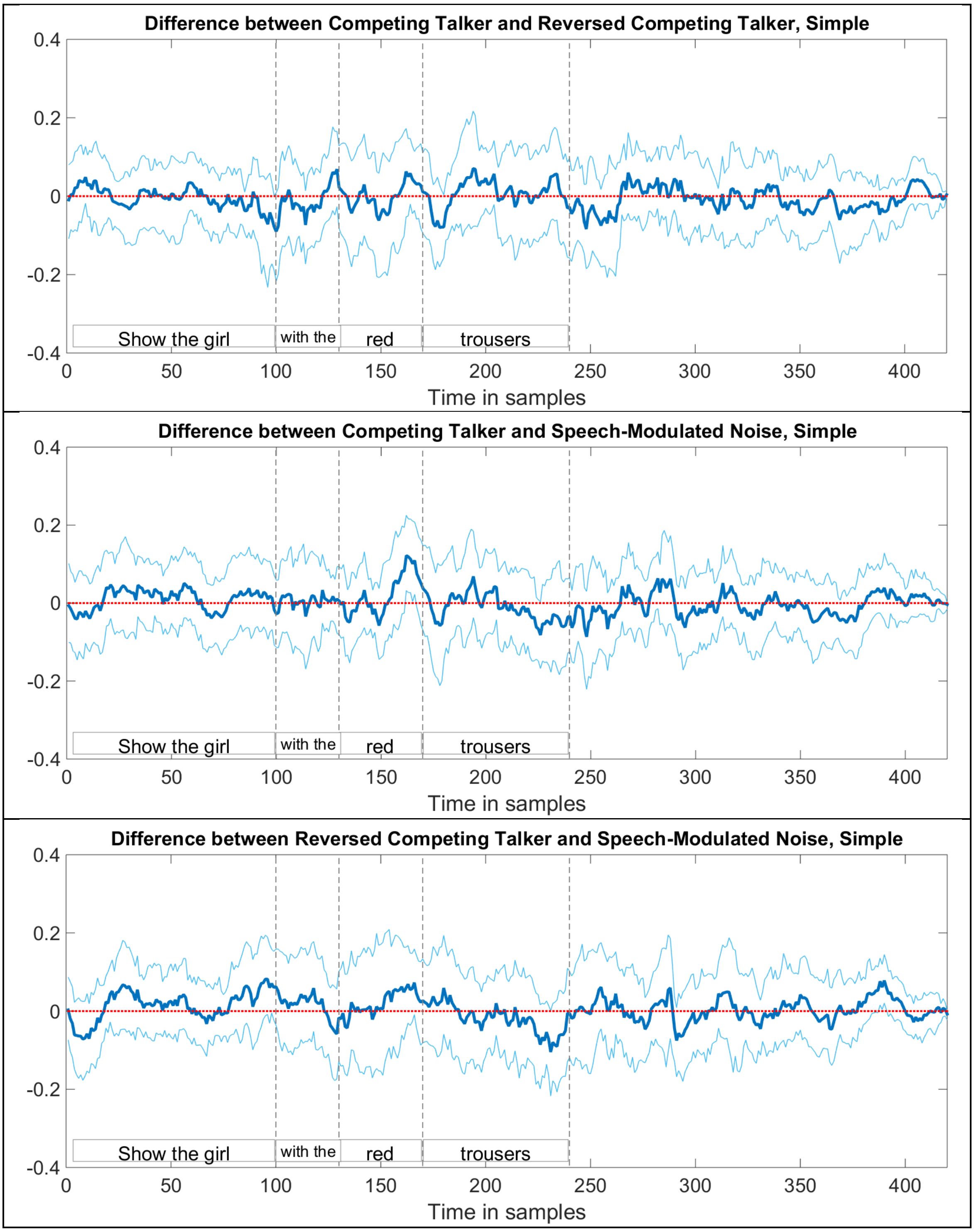
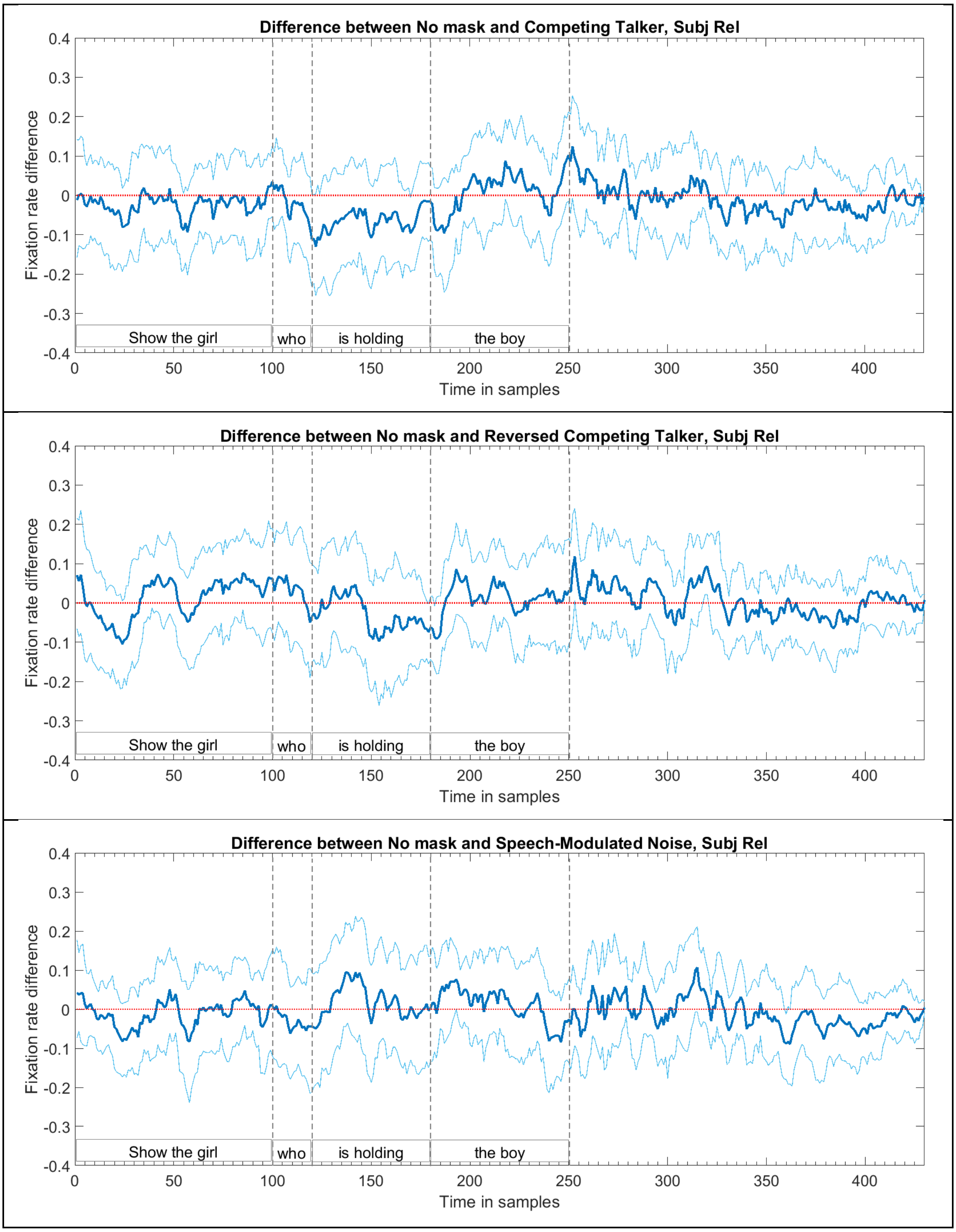


Figure H.13. Experiment 3. Differences between each of the mask conditions for simple sentences, with 99.17% confidence intervals to correct for multiple comparisons.


Difference between No mask and Competing Talker, Subj Rel
Fixation rate difference
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Show the girl
who
is holding
the boy
0
50
100
150
200
250
300
350
400
Time in samples
Difference between No mask and Reversed Competing Talker, Subj Rel
Fixation rate difference
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Show the girl
who
is holding
the boy
0
50
100
150
200
250
300
350
400
Time in samples
Difference between No mask and Speech-Modulated Noise, Subj Rel
Fixation rate difference
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Show the girl
who
is holding
the boy
0
50
100
150
200
250
300
350
400
Time in samples

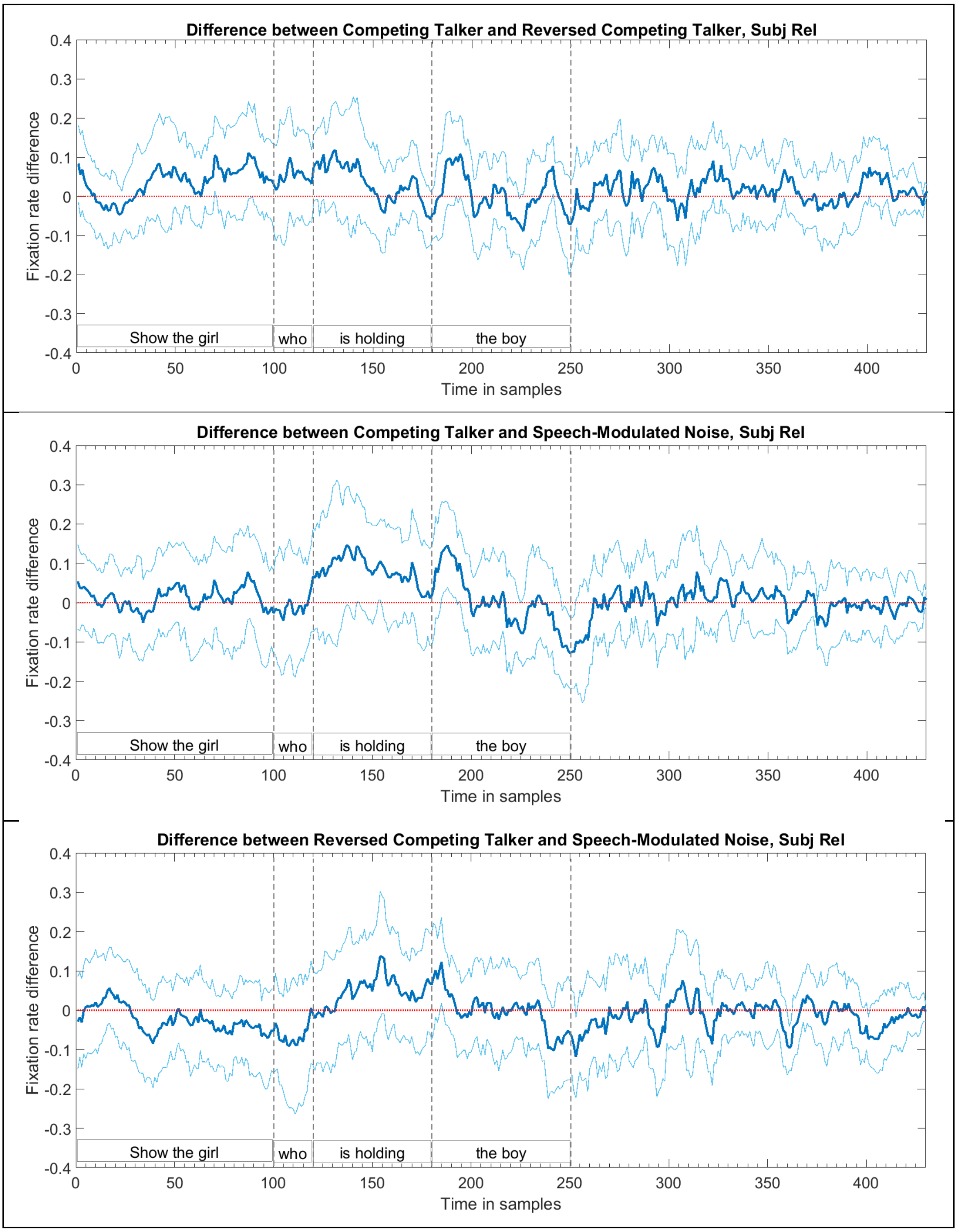
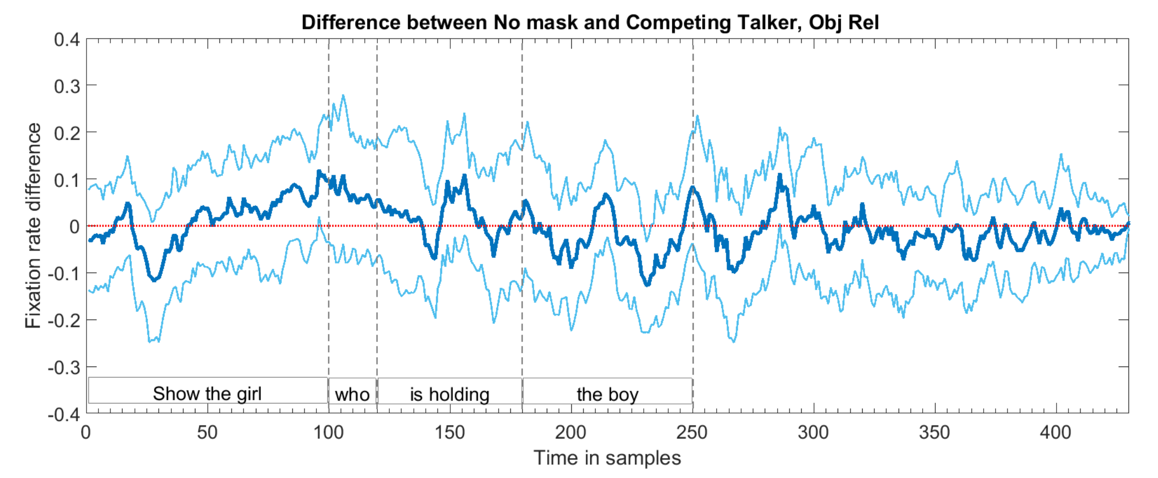
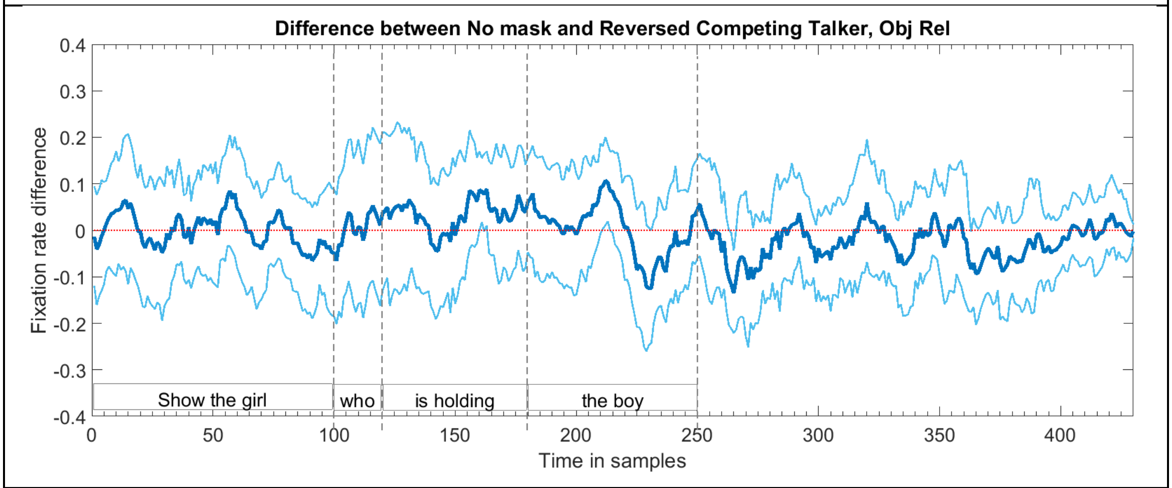
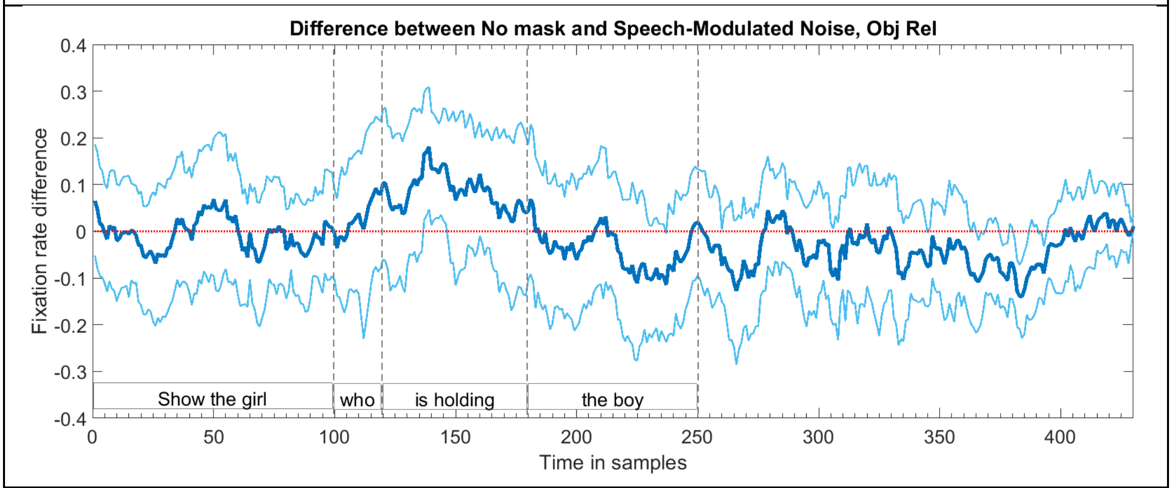


Figure H.14. Experiment 3. Differences between each of the mask conditions for subject relative sentences, with 99.17% confidence intervals to correct for multiple comparisons.


Difference between No mask and Competing Talker, Obj Rel
Fixation rate difference
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Show the girl
who
is holding
the boy
0
50
100
150
200
250
300
350
400
Time in samples


Difference between No mask and Reversed Competing Talker, Obj Rel
Fixation rate difference
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Show the girl
who
is holding
the boy
0
50
100
150
200
250
300
350
400
Time in samples


Difference between No mask and Speech-Modulated Noise, Obj Rel
Fixation rate difference
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Show the girl
who
is holding
the boy
0
50
100
150
200
250
300
350
400
Time in samples

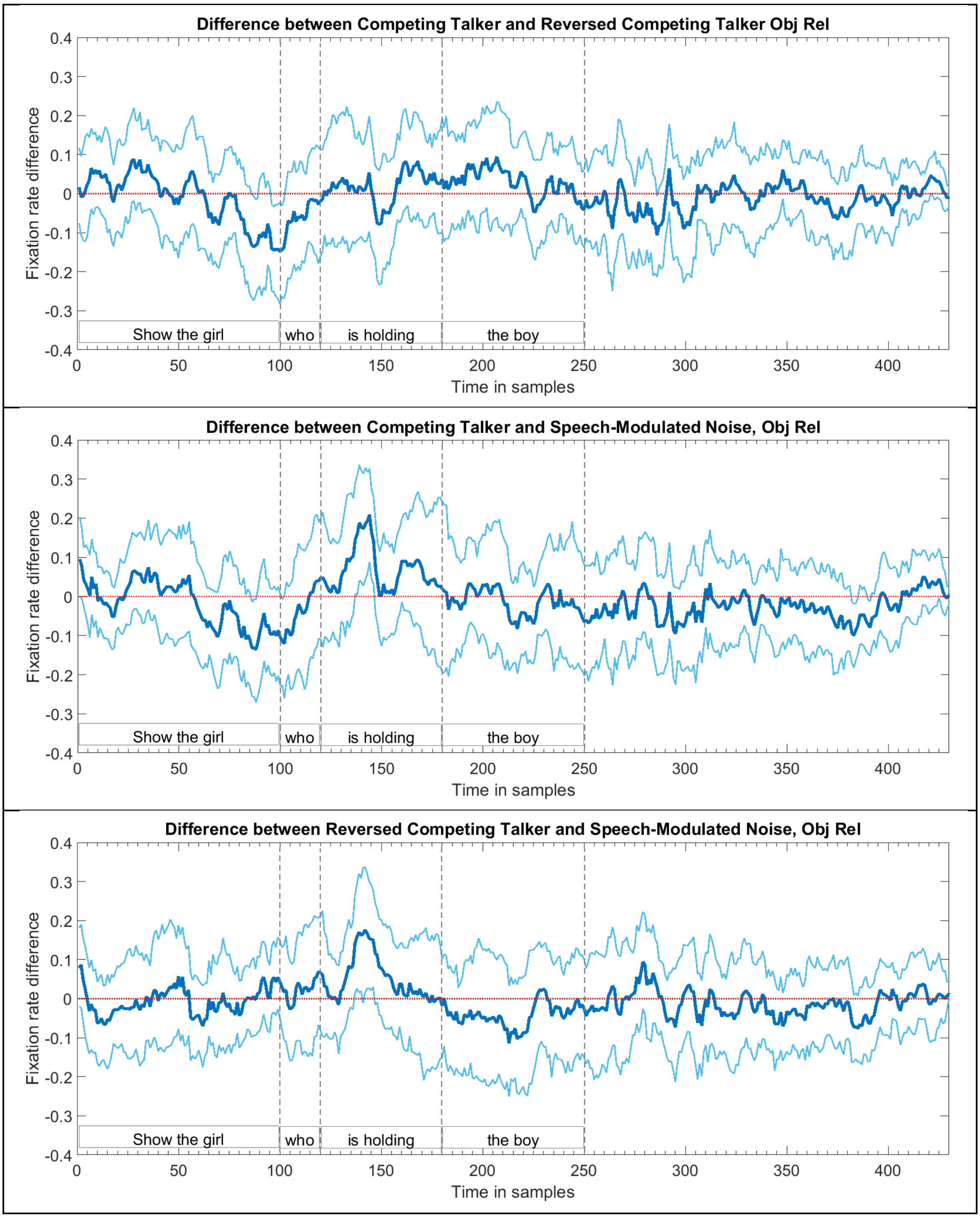


Figure H.15. Experiment 3. Differences between each of the mask conditions for object relative sentences, with 99.17% confidence intervals to correct for multiple comparisons.

Appendix I Eye-fixation graphs for Experiment 5

I.1 Average fixation rates and *t*-statistics between target and competitor for Experiment 5

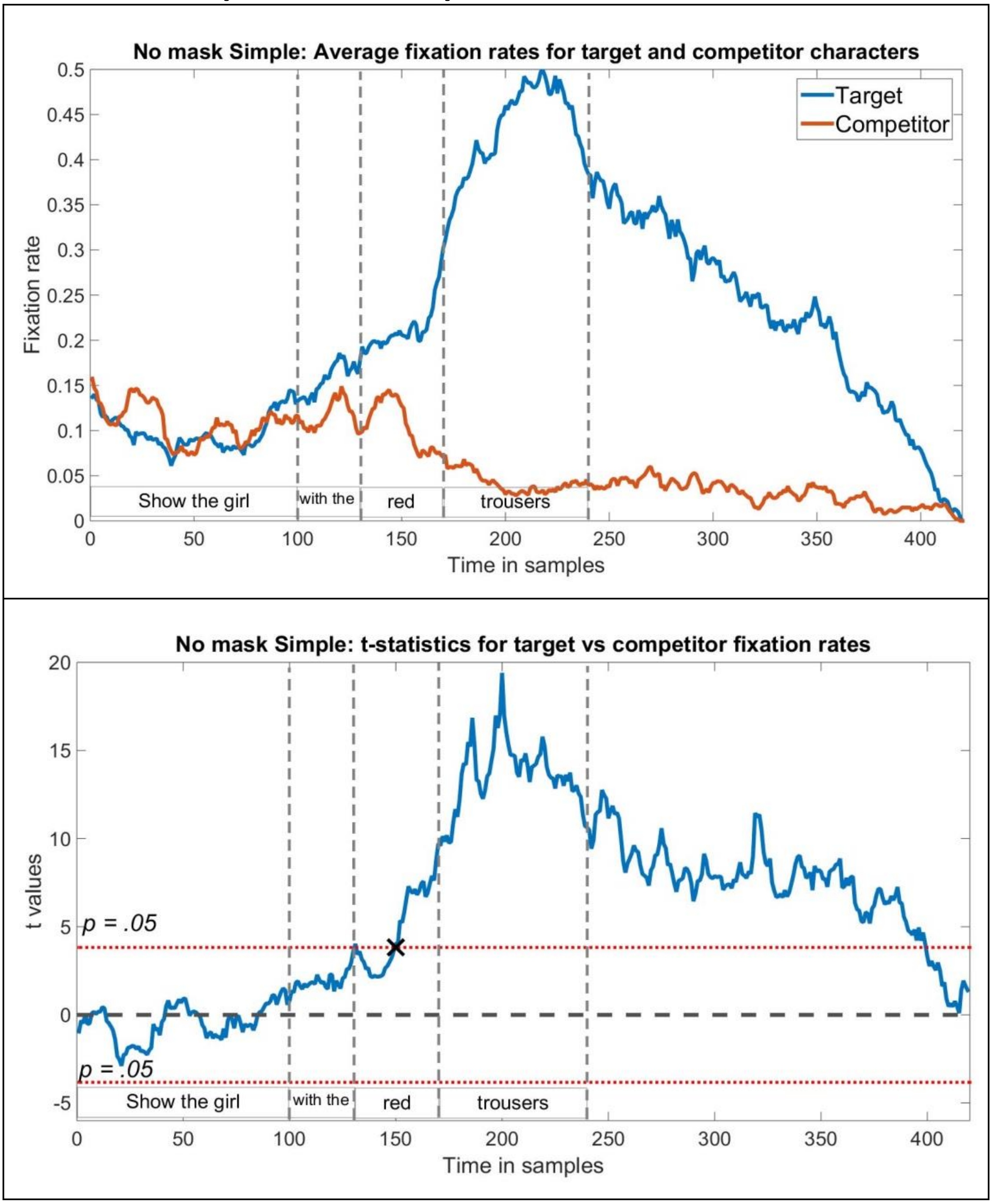


Figure I.1. Experiment 5. Average fixation rates and *t*-statistics for the no mask condition with simple sentences.

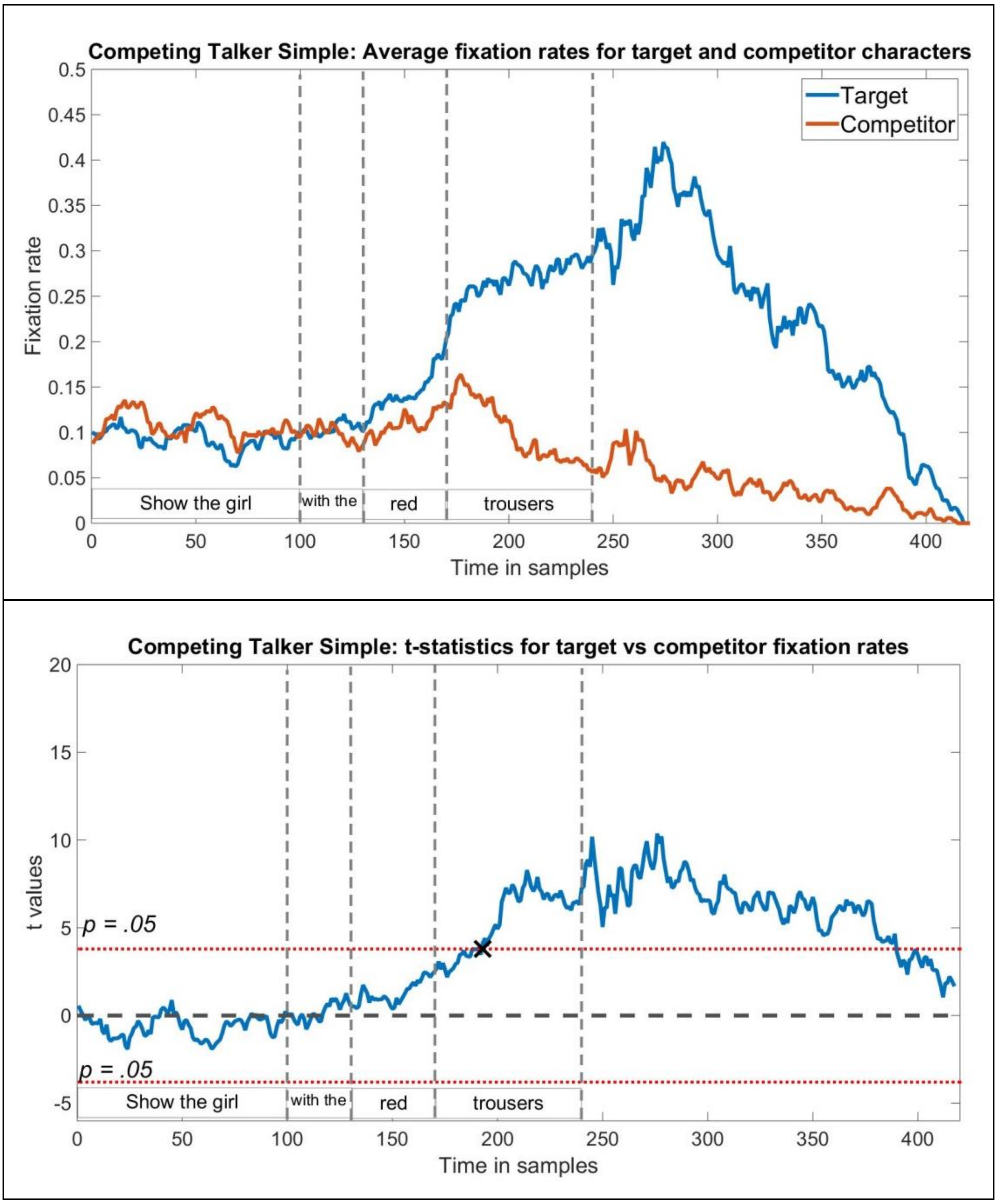


Figure I.2. Experiment 5. Average fixation rates and *t*-statistics for the competing talker condition with simple sentences.

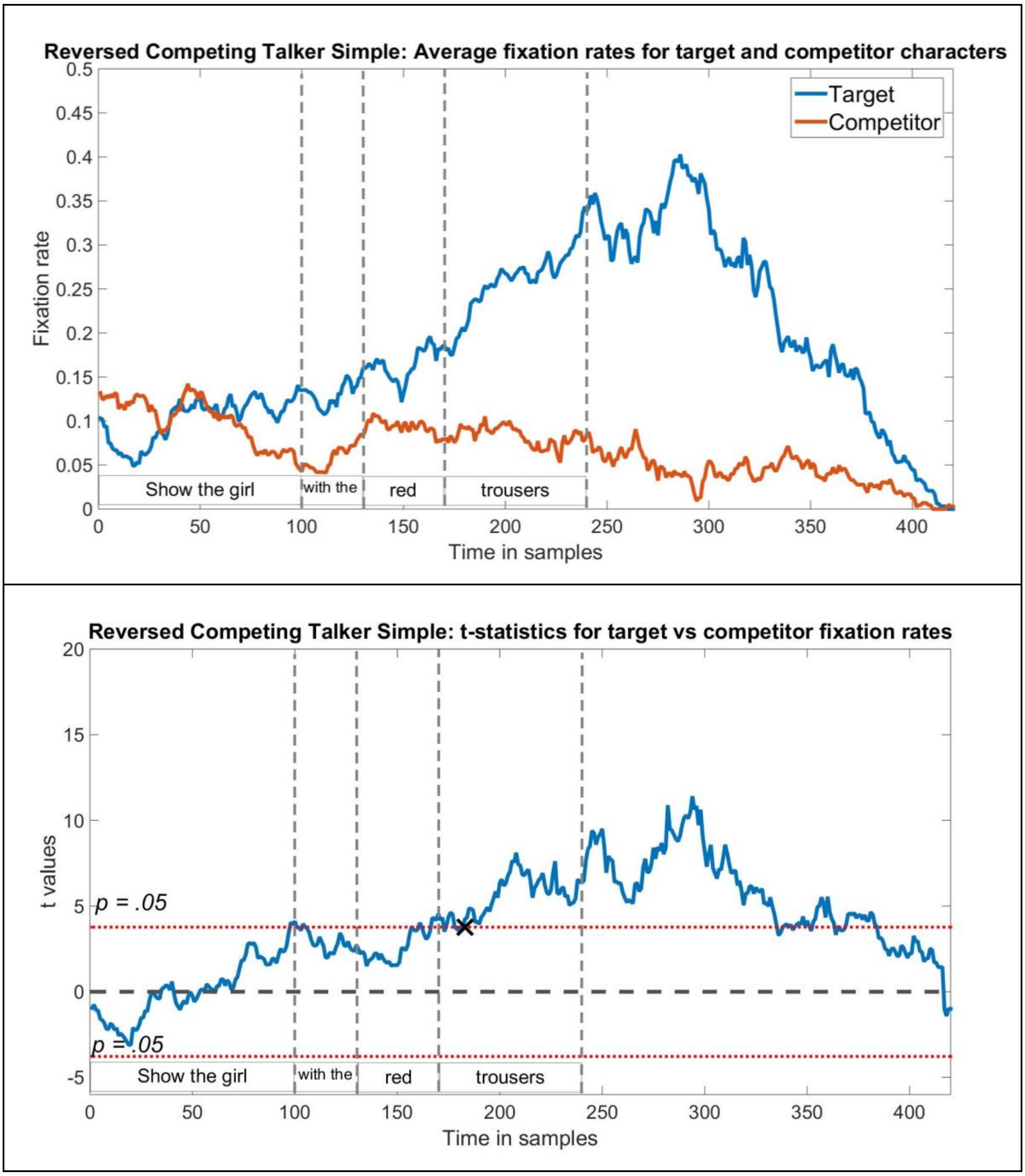


Figure I.3. Experiment 5. Average fixation rates and *t*-statistics for the reversed competing talker condition with simple sentences.

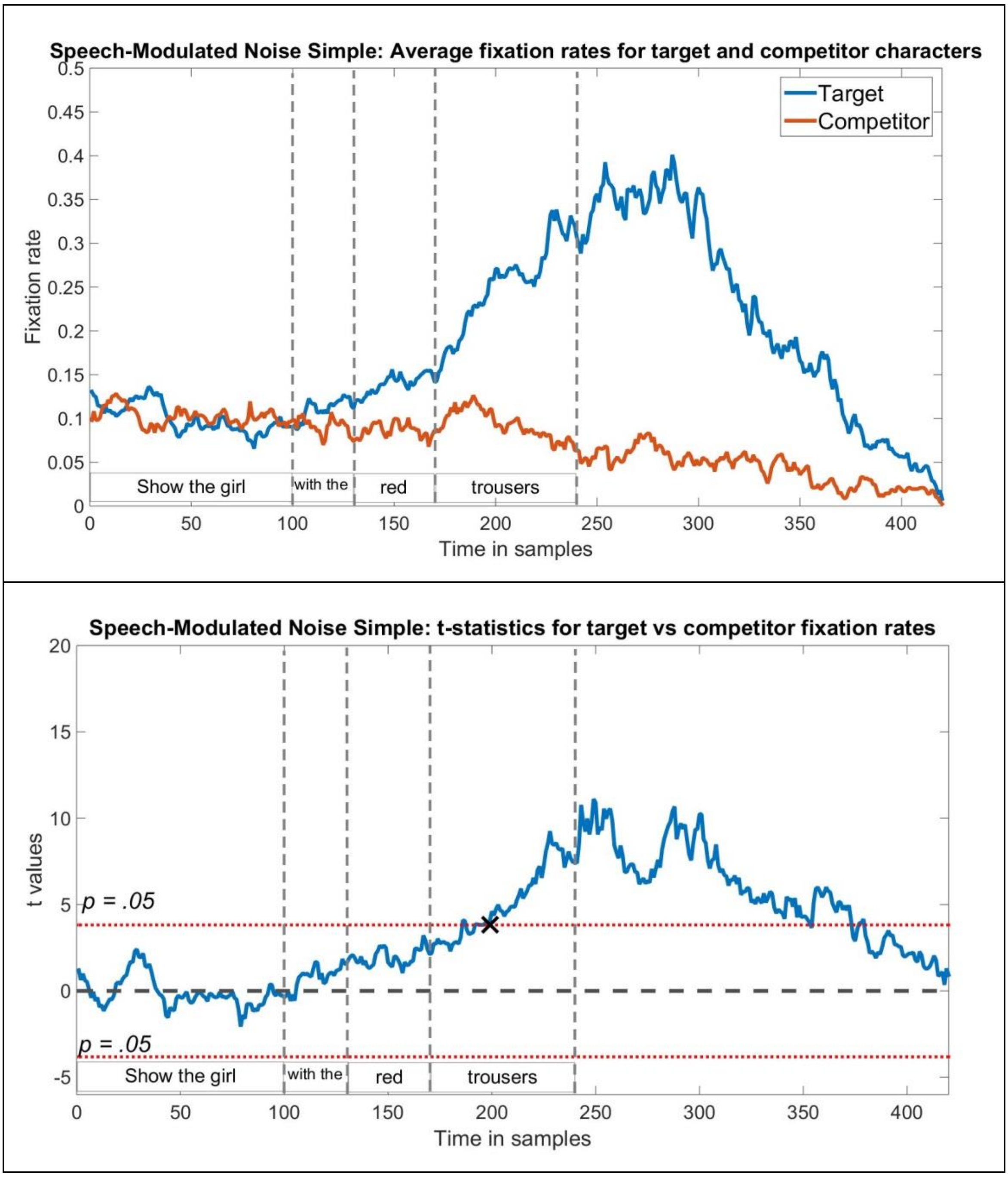


Figure I.4. Experiment 5. Average fixation rates and *t*-statistics for the speech-modulated noise condition with simple sentences.

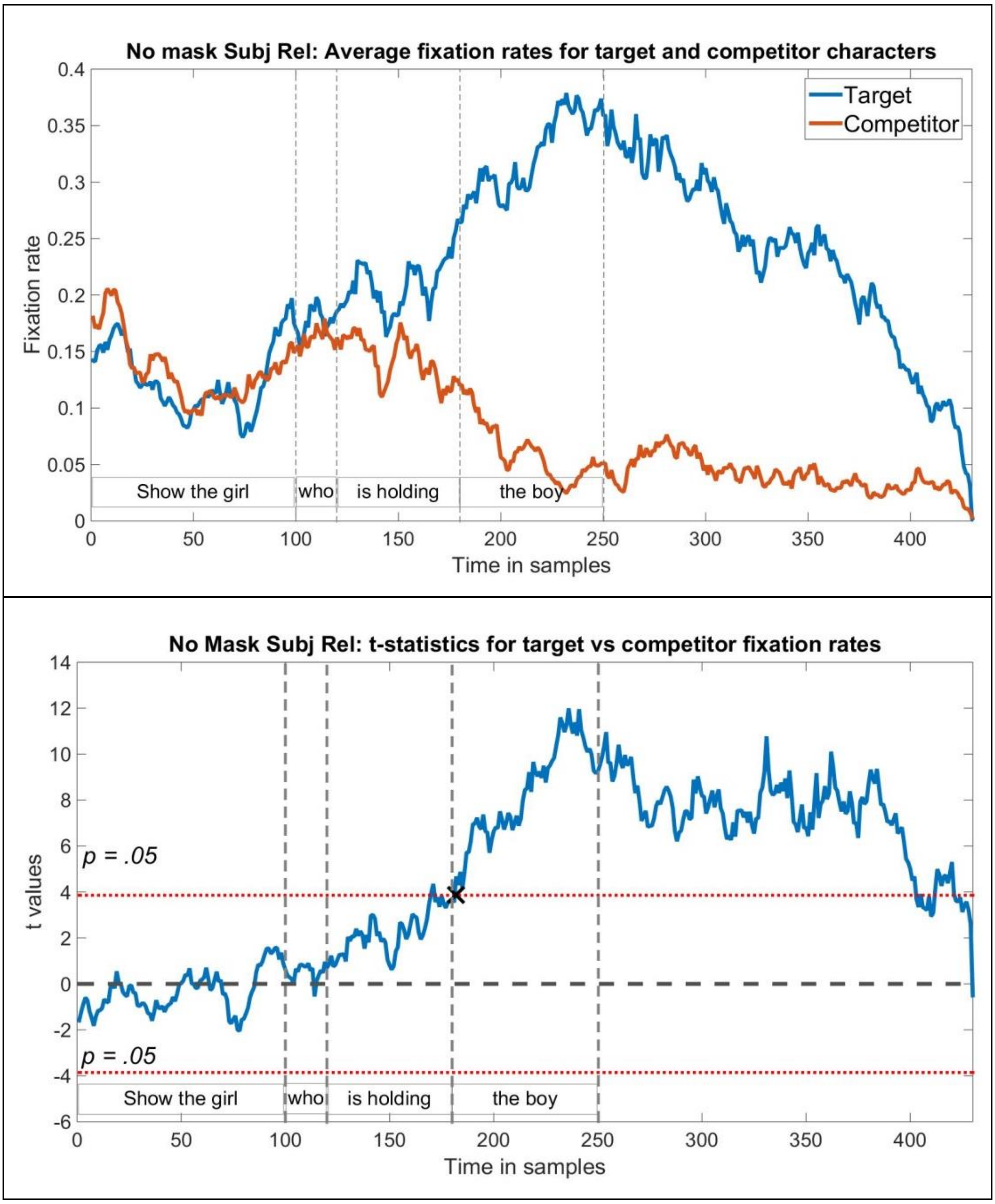


Figure I.5. Experiment 5. Average fixation rates and *t*-statistics for the no mask condition with subject relative sentences.

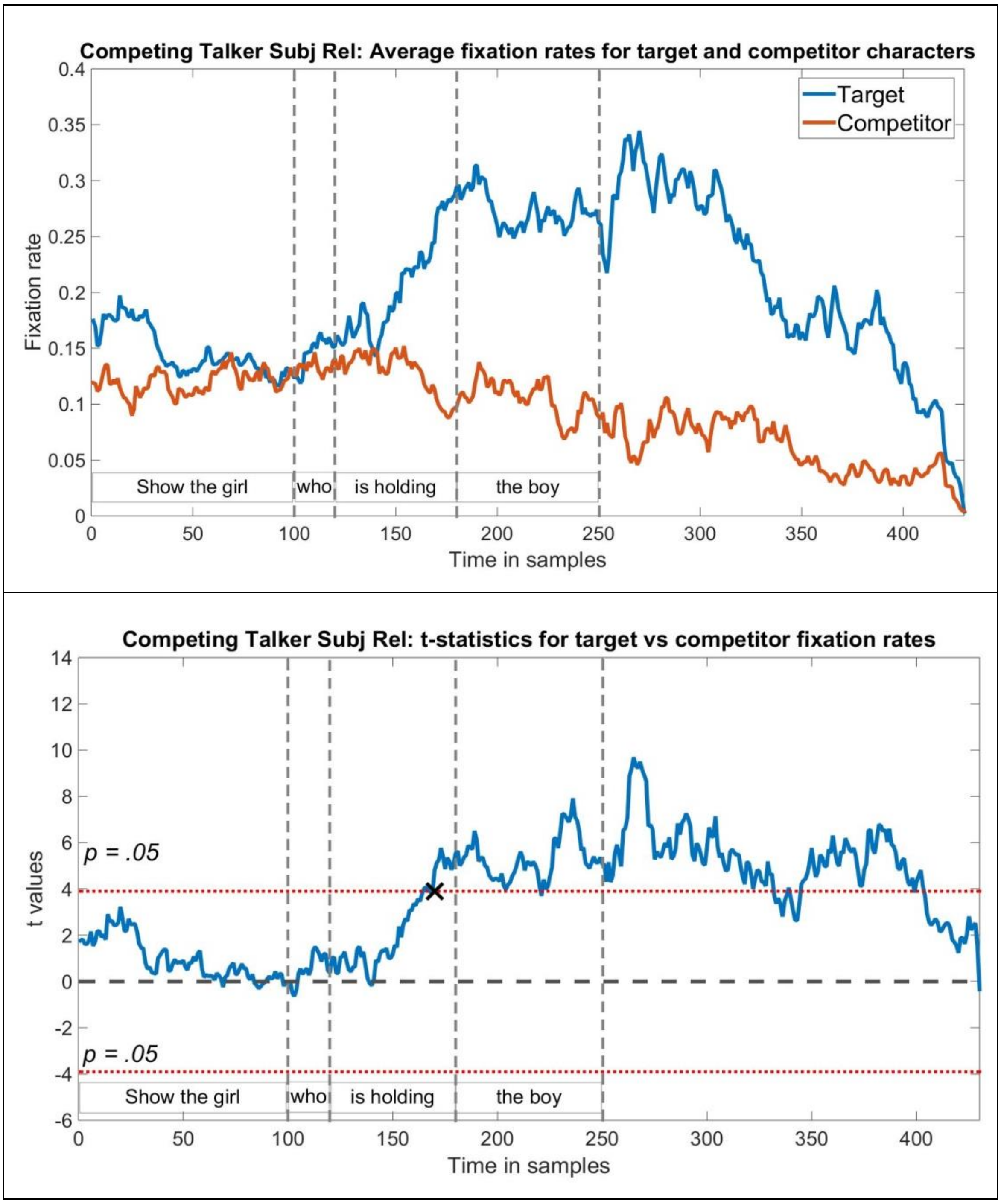


Figure I.6. Experiment 5. Average fixation rates and *t*-statistics for the competing talker condition with subject relative sentences.

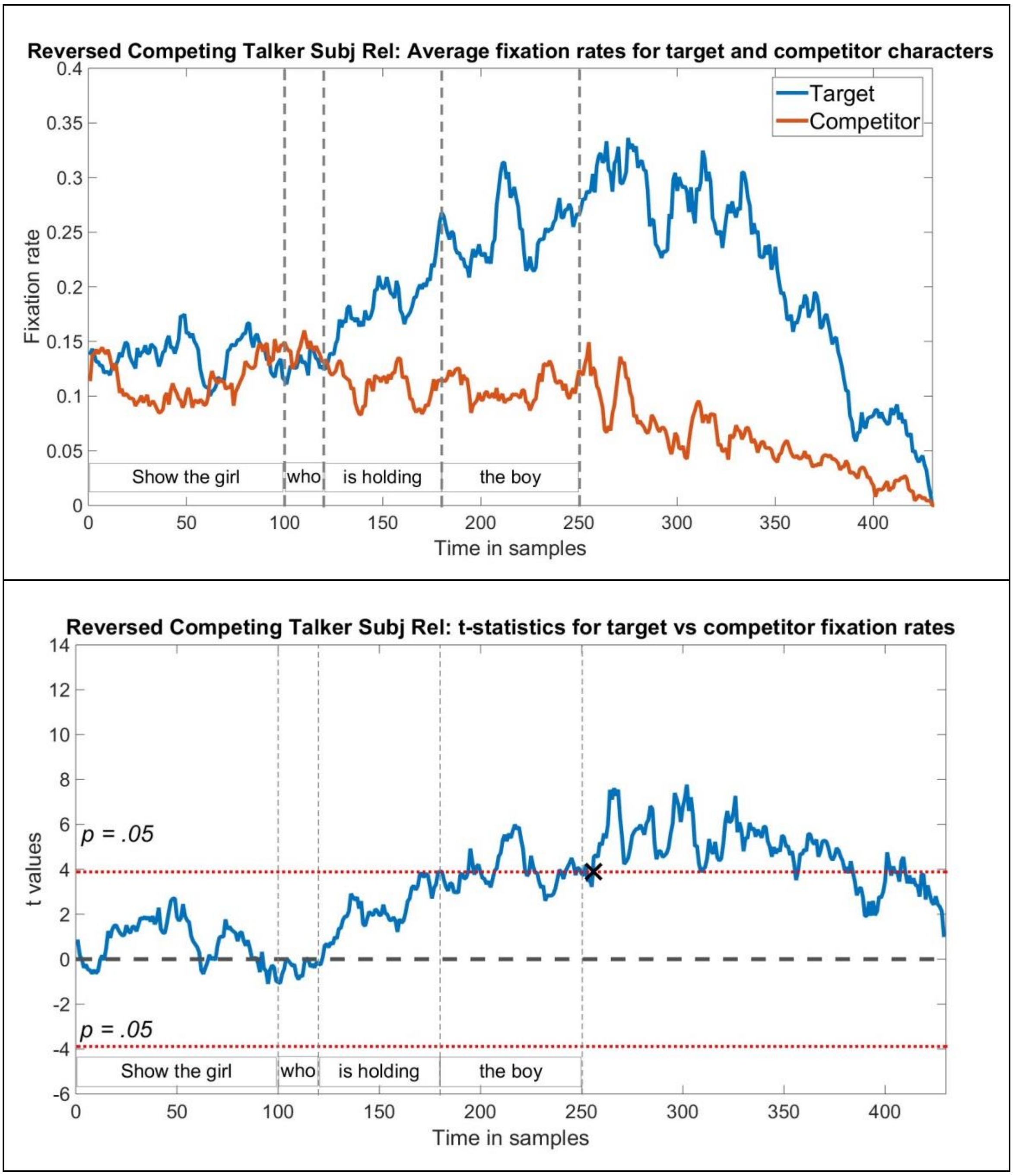


Figure I.7. Experiment 5. Average fixation rates and *t*-statistics for the reversed competing talker condition with subject relative sentences.

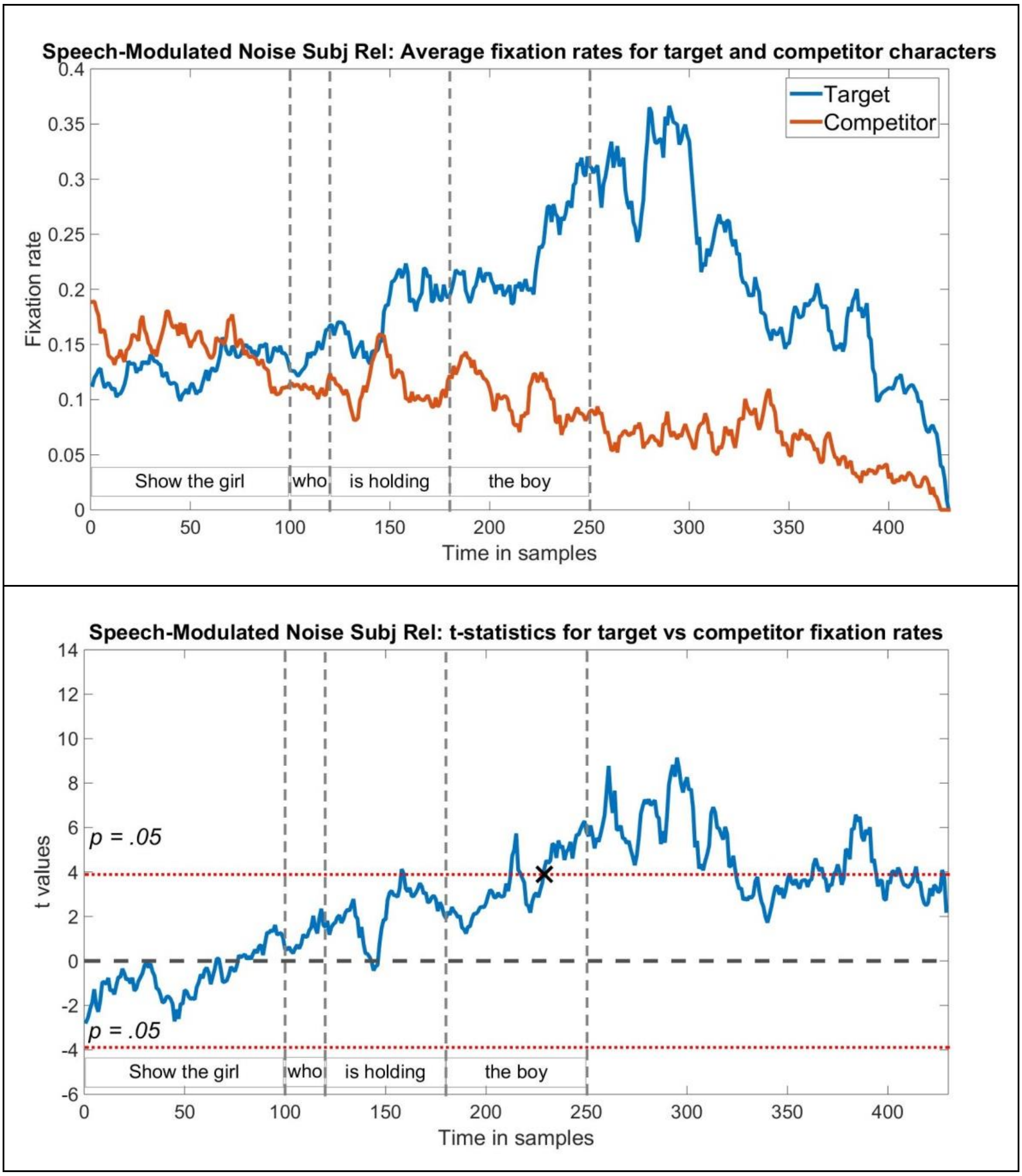


Figure I.8. Experiment 5. Average fixation rates and *t*-statistics for the speech-modulated noise condition with subject relative sentences.

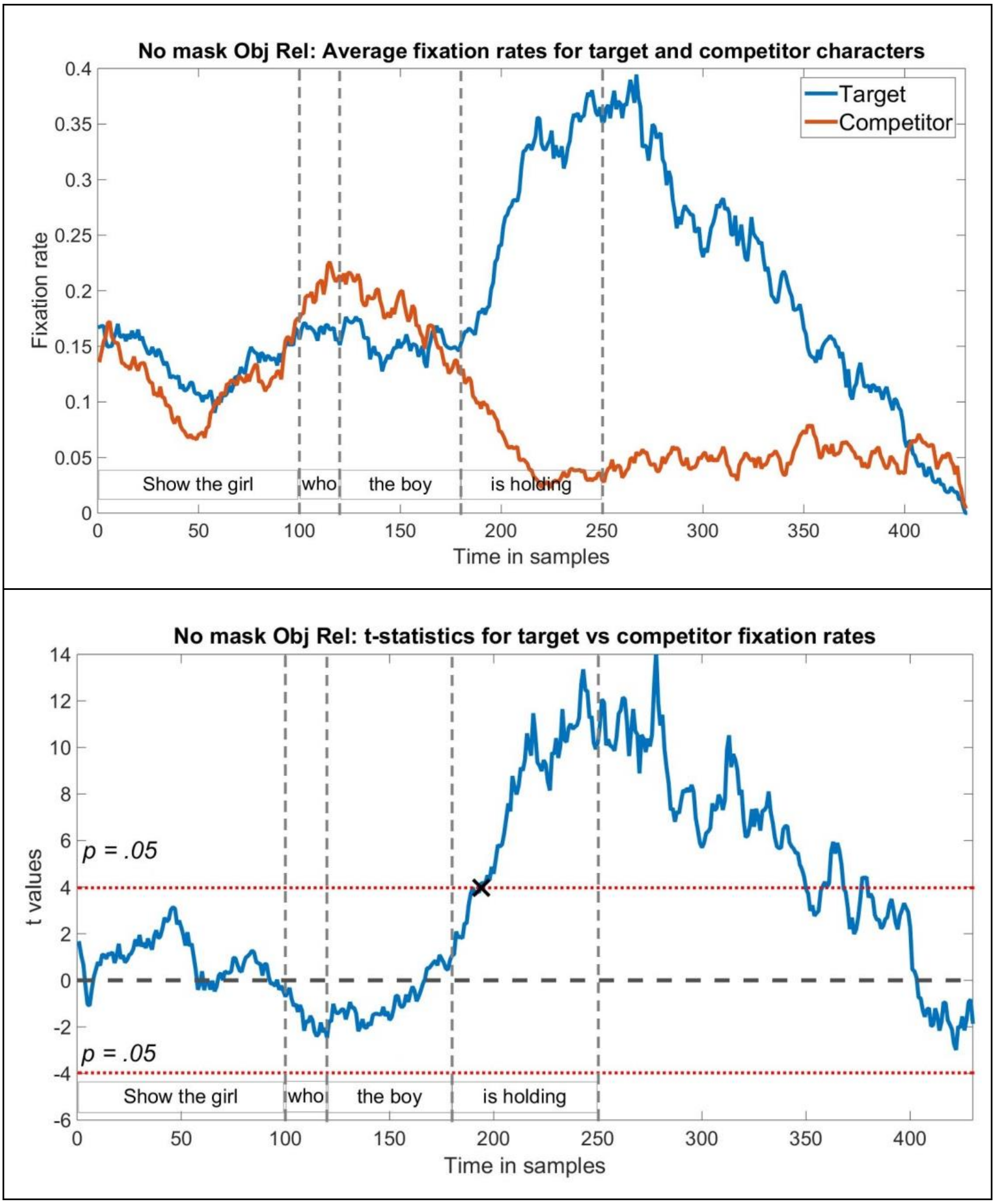


Figure I.9. Experiment 5. Average fixation rates and *t*-statistics for the no mask condition with object relative sentences.

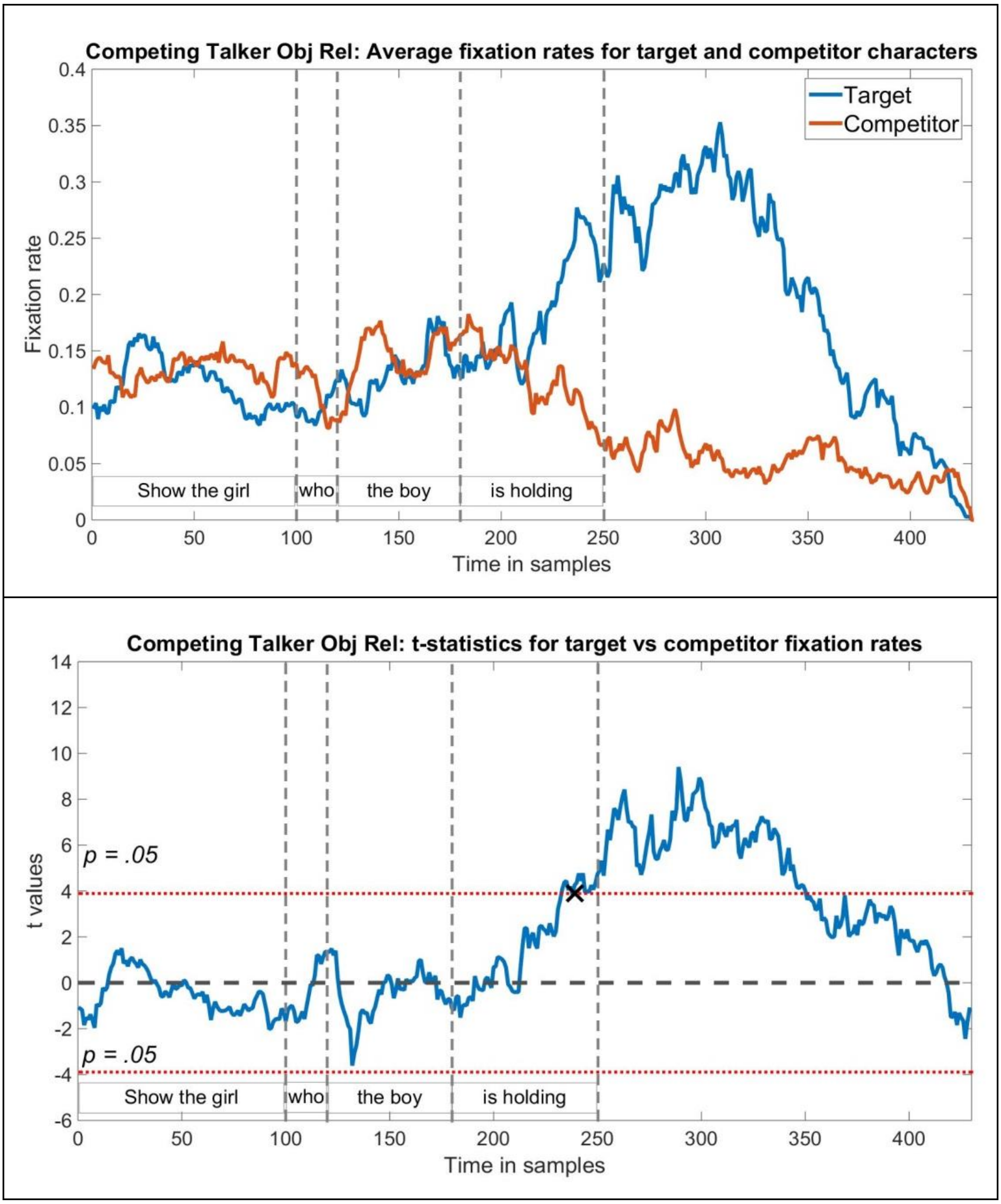


Figure I.10. Experiment 5. Average fixation rates and *t*-statistics for the competing talker condition with object relative sentences.

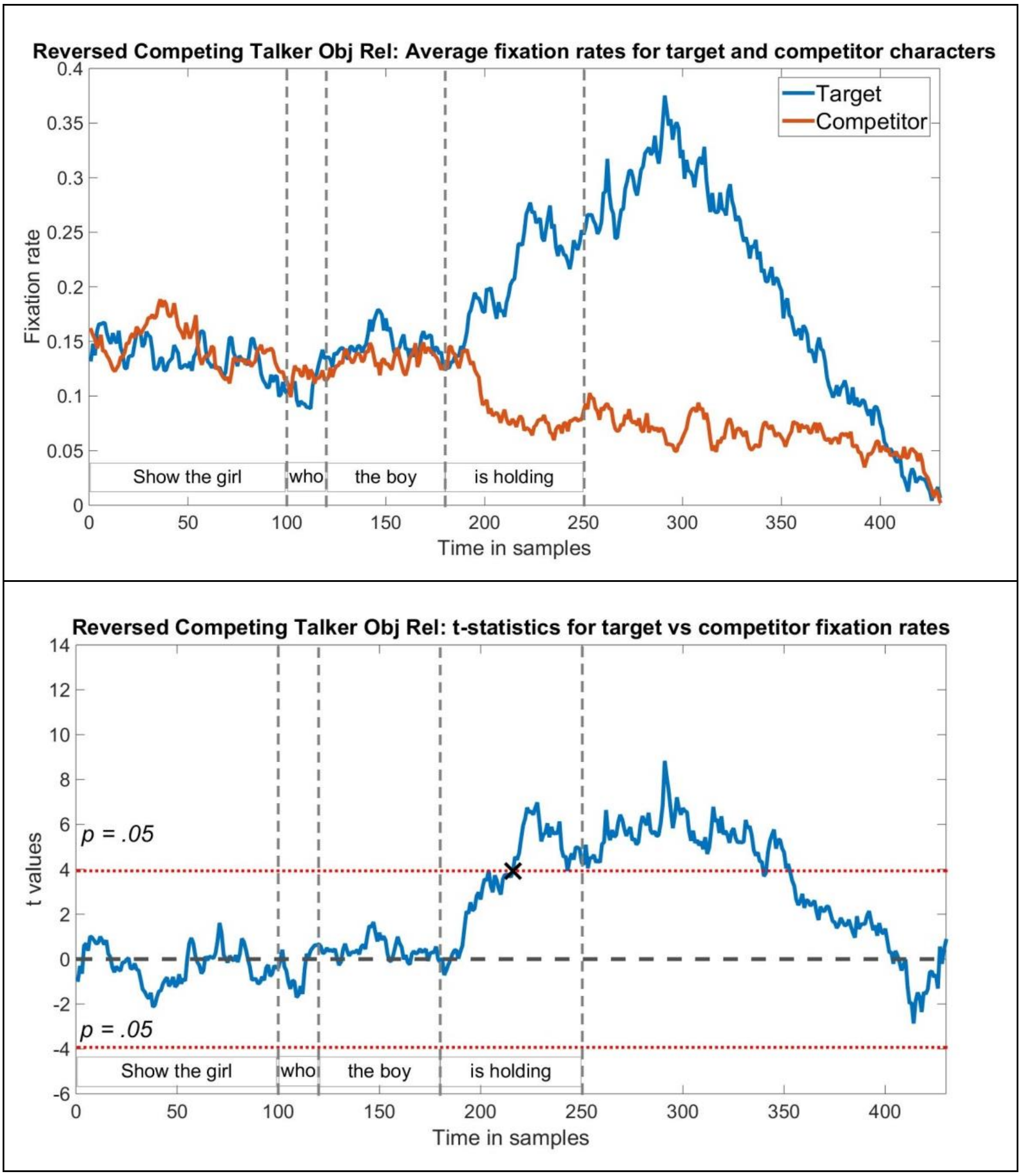


Figure I.11. Experiment 5. Average fixation rates and *t*-statistics for the reversed competing talker condition with object relative sentences.

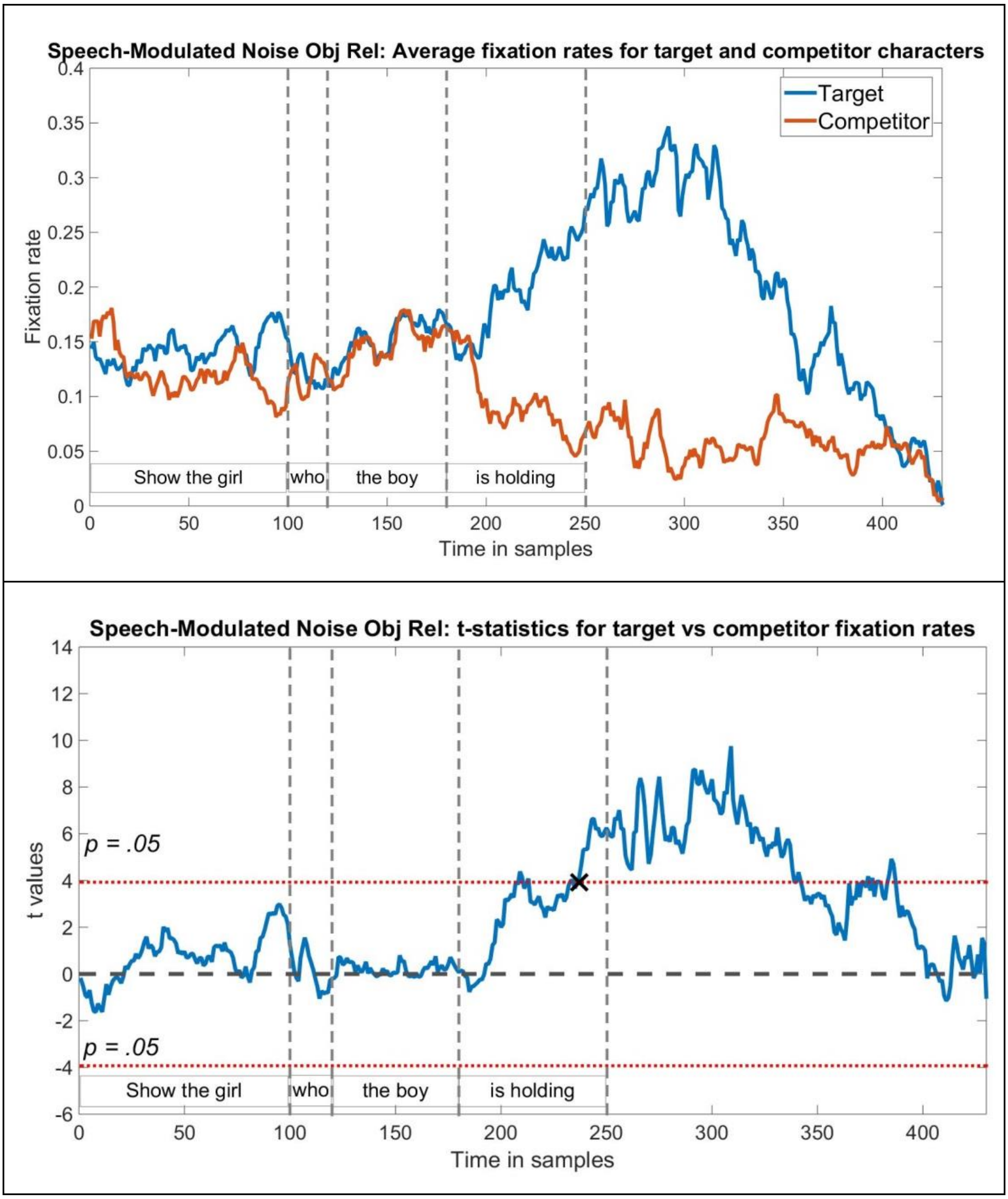
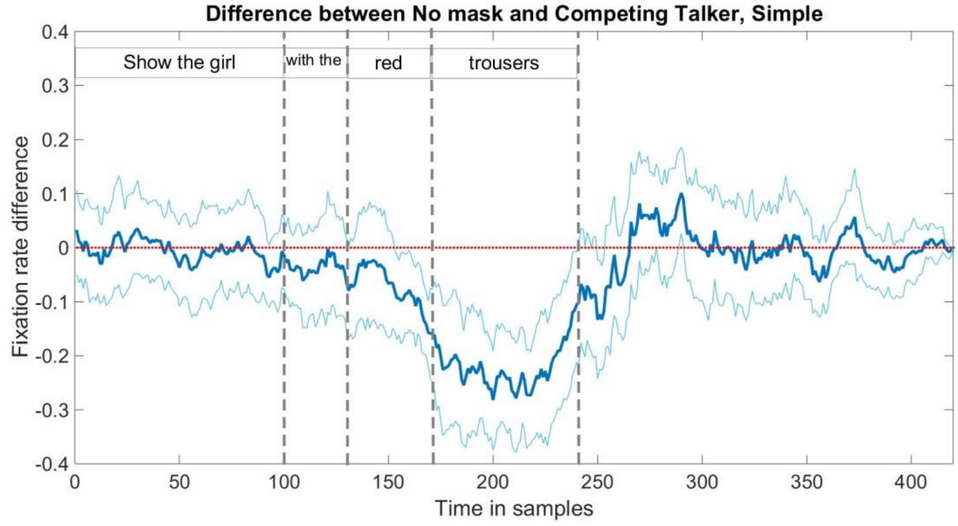
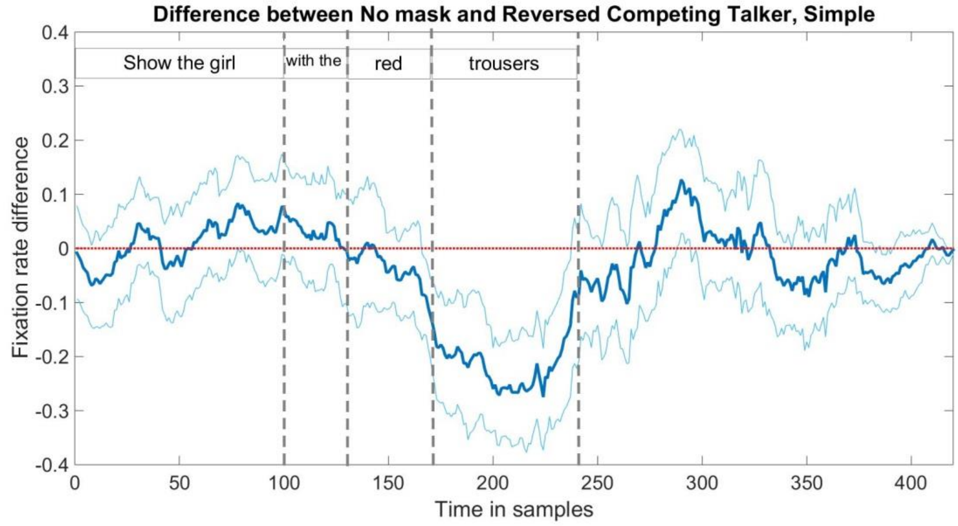
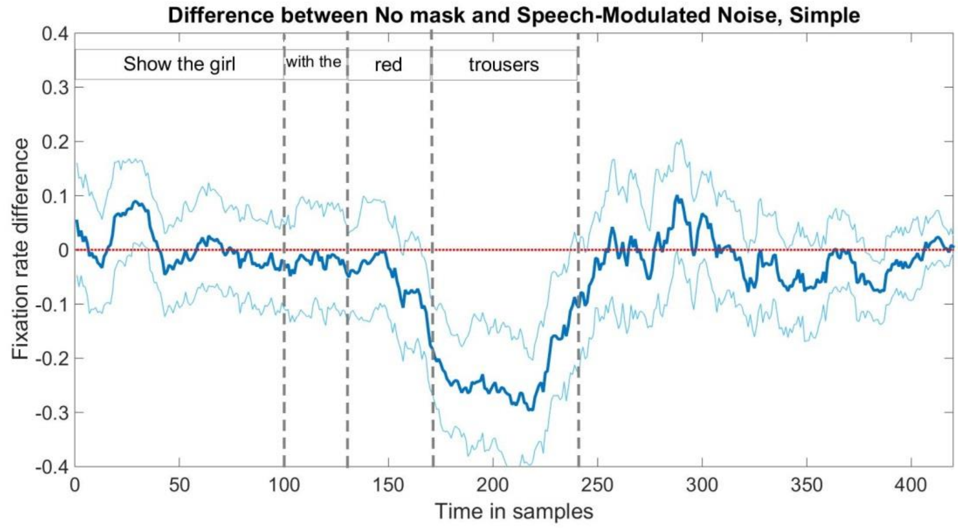


Figure I.12. Experiment 5. Average fixation rates and *t*-statistics for the speech-modulated noise condition with object relative sentences.

I.2 Pairwise comparisons between masks for Experiment 5







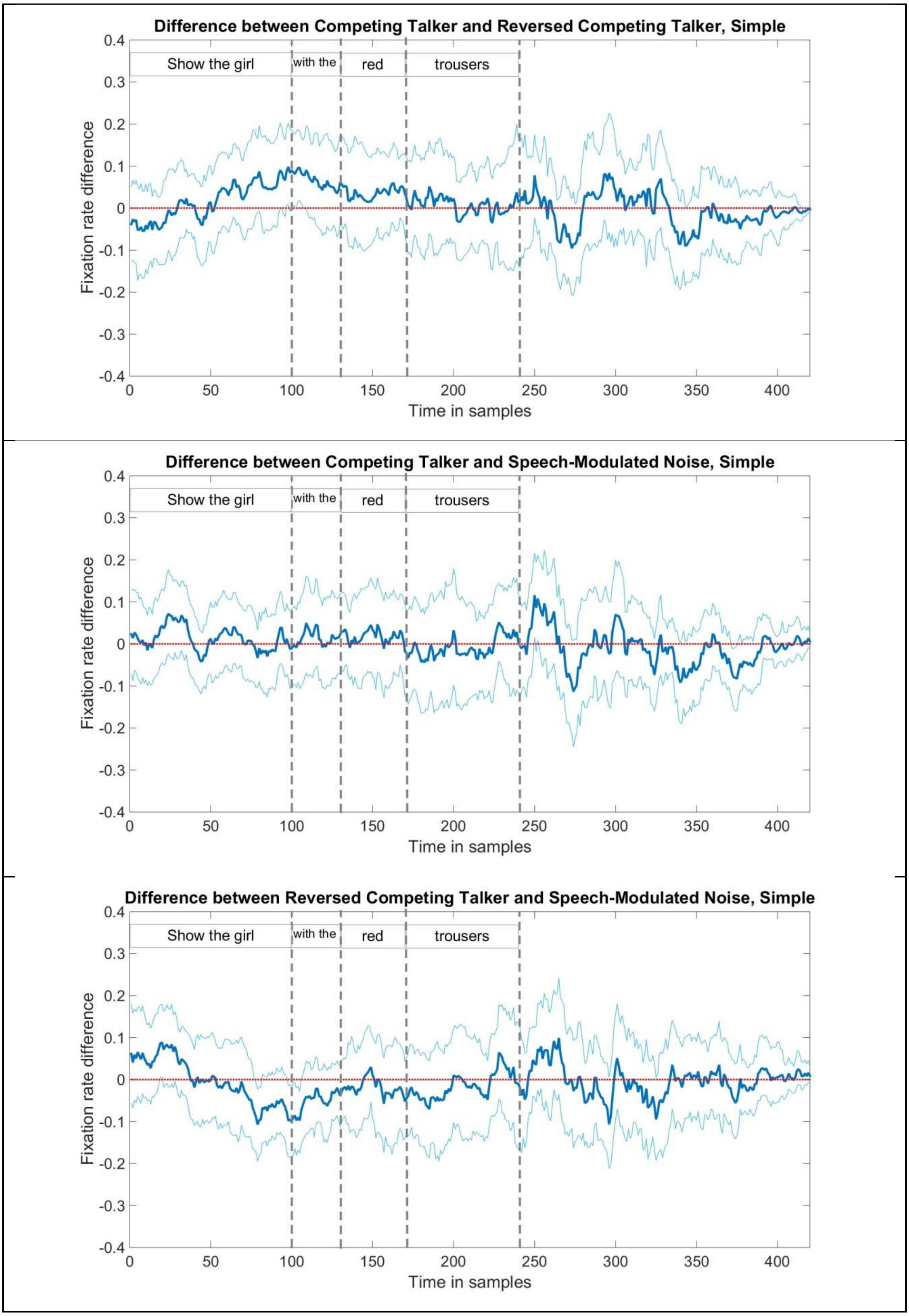
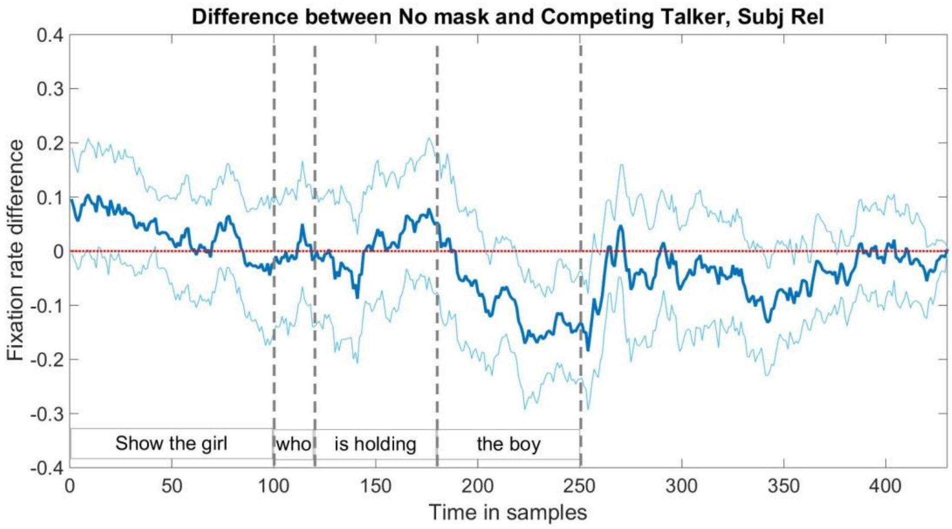
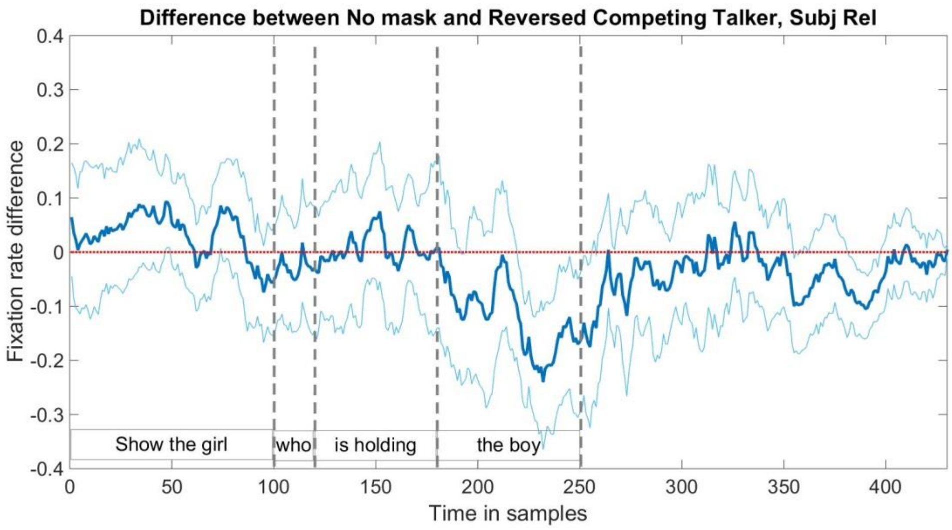
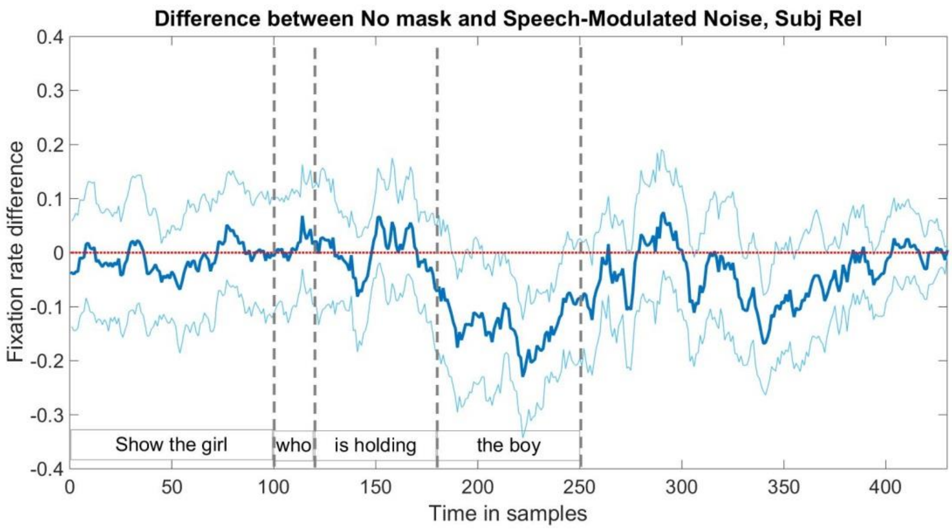


Figure I.13. Experiment 5. Differences between each of the mask conditions for simple sentences, with 99.17% confidence intervals to correct for multiple comparisons.


Difference between No mask and Competing Talker, Subj Rel
Fixation rate difference
Time in samples
Show the girl
who
is holding
the boy
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
50
100
150
200
250
300
350
400


Difference between No mask and Reversed Competing Talker, Subj Rel
Fixation rate difference
Time in samples
Show the girl
who
is holding
the boy
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
50
100
150
200
250
300
350
400


Difference between No mask and Speech-Modulated Noise, Subj Rel
Fixation rate difference
Time in samples
Show the girl
who
is holding
the boy
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
50
100
150
200
250
300
350
400

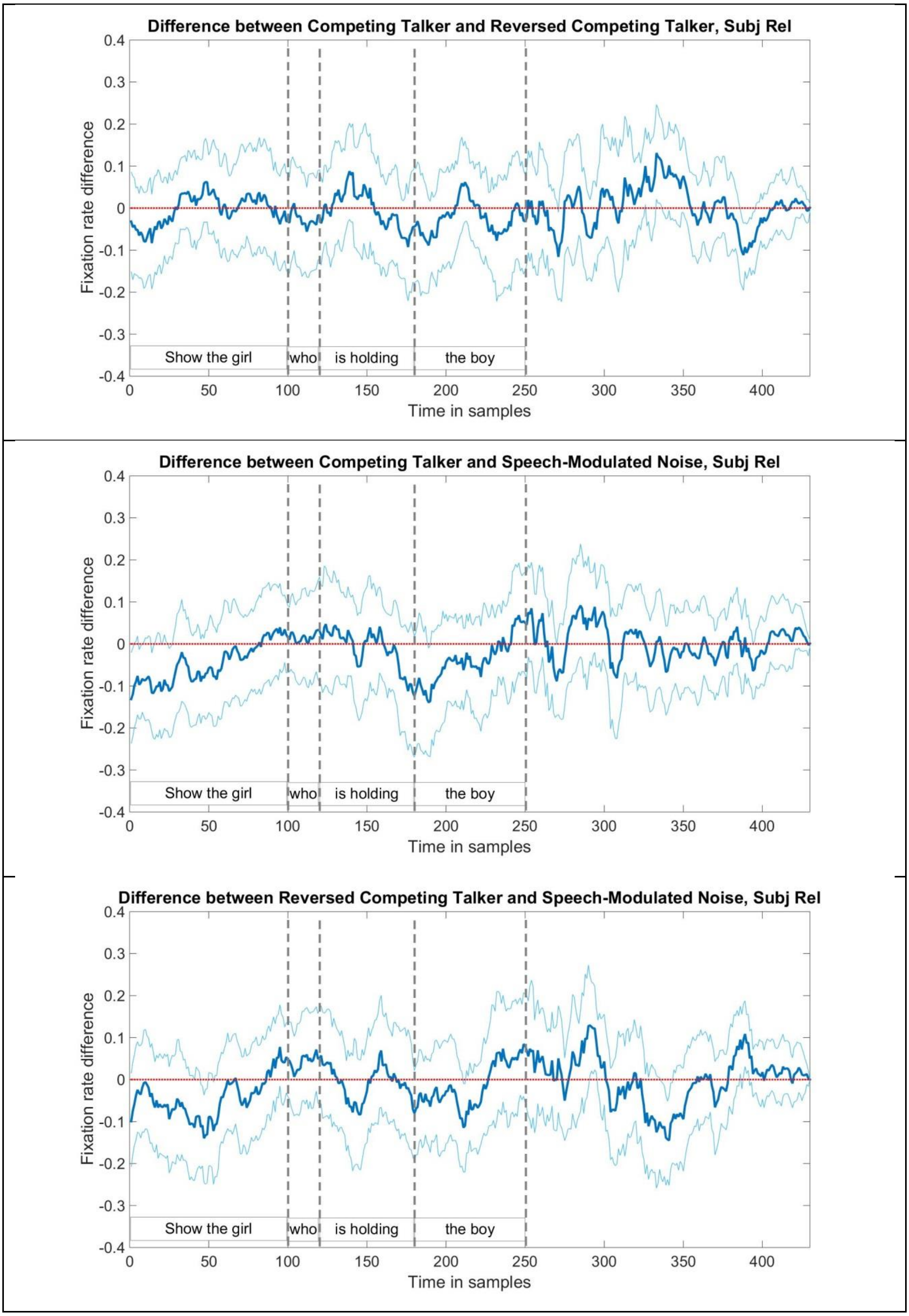
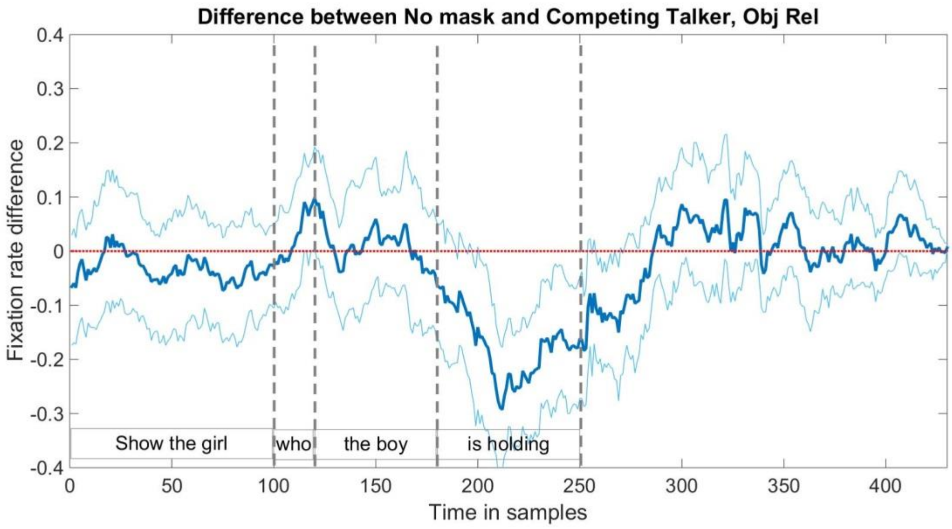
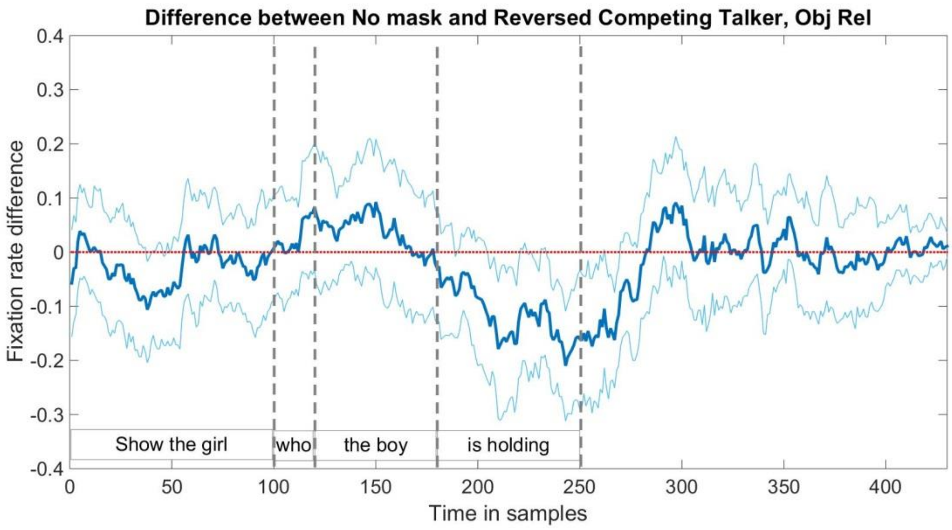


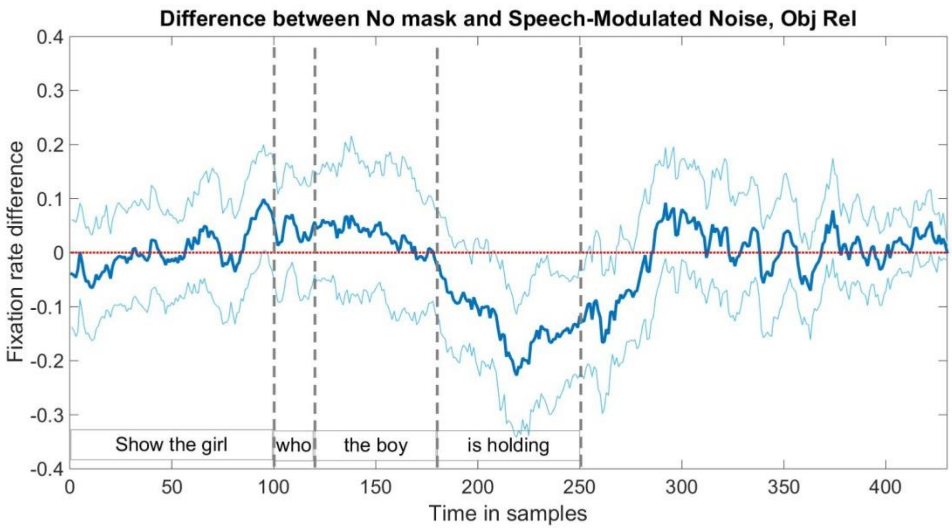
Figure I.14. Experiment 5. Differences between each of the mask conditions for subject relative sentences, with 99.17% confidence intervals to correct for multiple comparisons.



Difference between No mask and Competing Talker, Obj Rel
Fixation rate difference
Time in samples
Show the girl
who
the boy
is holding



Difference between No mask and Reversed Competing Talker, Obj Rel
Fixation rate difference
Time in samples
Show the girl
who
the boy
is holding



Difference between No mask and Speech-Modulated Noise, Obj Rel
Fixation rate difference
Time in samples
Show the girl
who
the boy
is holding

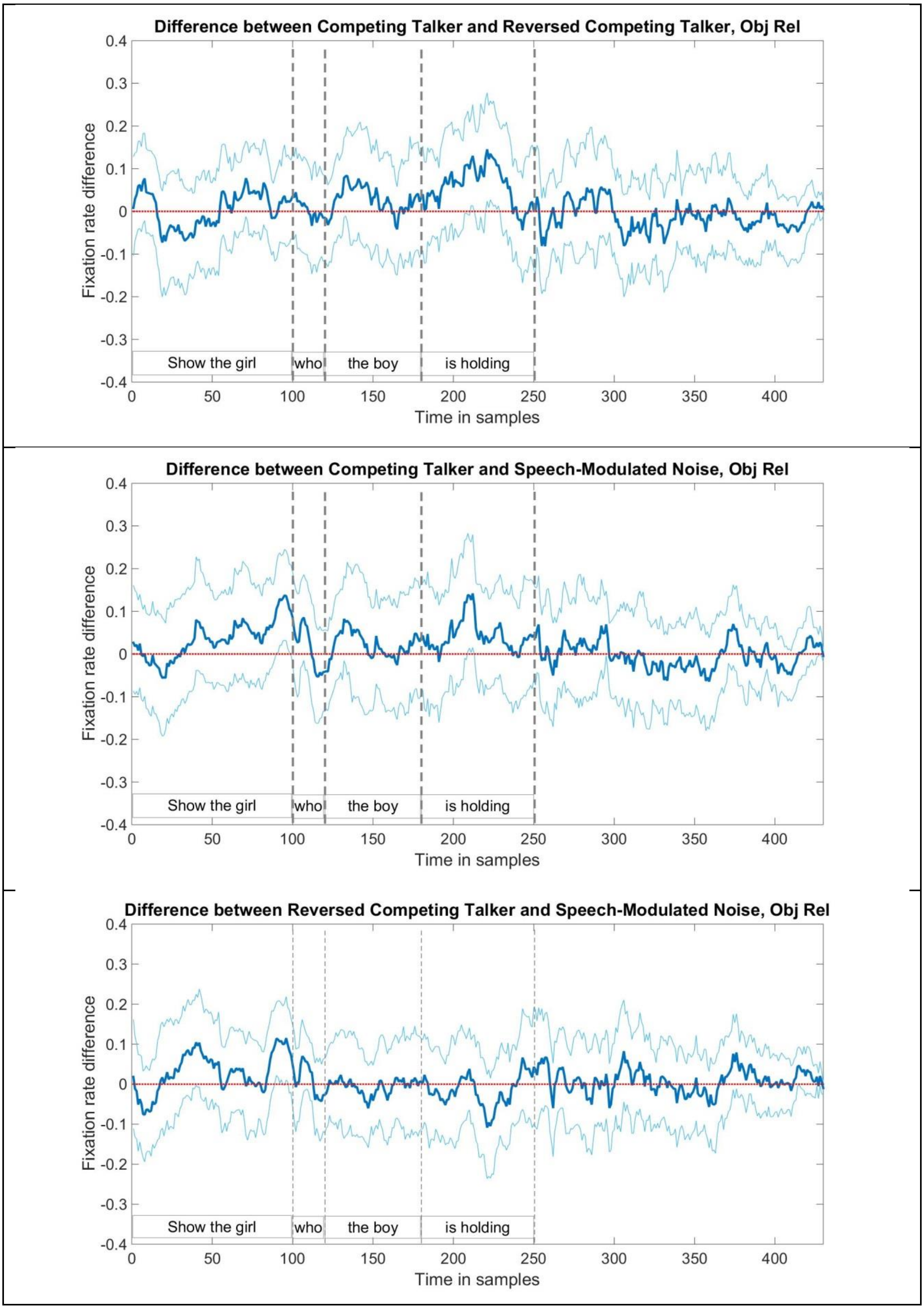


Figure I.15. Experiment 5. Differences between each of the mask conditions for object relative sentences, with 99.17% confidence intervals to correct for multiple comparisons.

Appendix J Individual results for cognitive measures in Experiment 5

J.1 Short-term and working memory scores for Experiment 5

Participant	Non-word repetition	Listening Recall	Processing speed
1	91	119	117
2	99	80	81
3	109	87	85
4	73	129	129
5	99	119	119
6	95	112	117
7	95	80	81
8	113	80	81
9	77	101	99
10	82	84	82
11	127	112	117
12	77	91	91
13	99	122	122
14	99	80	79
15	95	94	96
16	82	80	81
17	92	102	99
18	95	94	99
19	91	80	81
20	77	98	102
21	109	108	113
22	98	120	130
23	95	119	117
24	99	119	118
25	95	91	95
26	91	77	77
27	87	109	114
28	87	117	114
29	99	101	99
30	95	94	98
31	73	94	98
32	118	126	126
33	68	84	83
34	92	95	101
35	109	87	85
36	77	84	83

Table J.1. Experiment 5. Standardised scores for the non-word recall and listening recall (including processing speed) AWMA subtests.

J.2 Selective attention: visual flanker task reaction times for Experiment 5

Participant	Consistent	Inconsistent	Neutral	Inconsistent-Consistent
1	412	476	412	64
2	467	476	448	9
3	431	508	439	77
4	382	416	391	34
5	431	489	427	58
6	466	527	470	61
7	383	452	394	70
8	418	483	415	65
9	393	440	400	48
10	427	447	405	19
11	387	492	422	104
12	432	474	430	41
13	410	463	435	52
14	375	410	385	35
15	376	473	402	97
16	403	460	431	57
17	390	426	364	36
18	417	476	428	59
19	447	452	439	5
20	403	443	420	40
21	391	437	414	46
22	451	518	460	67
23	371	425	390	54
24	411	466	420	54
25	446	497	470	51
26	432	506	428	73
27	393	413	403	20
28	351	395	361	45
29	438	475	446	37
30	398	459	415	61
31	400	427	410	26
32	386	444	399	58
33	377	423	392	46
34	382	469	378	87
35	421	518	414	97
36	411	429	385	18

Table J.2. Experiment 5. Reaction times in ms per participant for each of the flanker task conditions. RTs include only accurate responses.

Appendix K Individual accuracy and reaction time data for the sentence comprehension task in Experiment 5

Participant	Masked - Unmasked	CT – (RCT + SMN)	OR - SR
1	11%	3%	0%
2	39%	-4%	2%
3	20%	-5%	8%
4	18%	-13%	7%
5	34%	-9%	12%
6	34%	-4%	15%
7	24%	1%	-2%
8	53%	-3%	-5%
9	17%	-8%	-7%
10	27%	-3%	20%
11	26%	-7%	8%
12	18%	-13%	-20%
13	20%	-5%	-3%
14	17%	-4%	-3%
15	27%	-15%	8%
16	31%	-6%	-7%
17	34%	-8%	25%
18	11%	-10%	8%
19	26%	-1%	12%
20	30%	-13%	12%
21	28%	3%	8%
22	14%	-7%	8%
23	28%	-3%	-3%
24	23%	-10%	-5%
25	37%	-3%	18%
26	29%	-12%	12%
27	23%	-11%	2%
28	32%	-13%	8%
29	10%	3%	-5%
30	16%	-9%	-2%
31	29%	-16%	2%
32	24%	7%	12%
33	43%	3%	27%
34	24%	-7%	22%
35	40%	3%	-3%
36	43%	-3%	22%

Table K.1. Experiment 5. Individual values for the difference in percent accurate responses between the masked (CT, SMN, RCT) and unmasked condition, between the CT condition and the average of the energetic mask conditions (RCT, SMN), and between the OR and the SR conditions.

Participant	Masked - Unmasked	CT – (RCT + SMN)	OR - SR
1	229	-39	300
2	248	-105	139
3	48	21	363
4	315	86	54
5	270	11	233
6	173	69	197
7	229	29	411
8	241	-129	272
9	217	-158	381
10	202	-110	171
11	205	-19	261
12	406	-131	222
13	283	49	221
14	311	3	-23
15	205	-27	256
16	251	-124	89
17	299	-74	309
18	314	-67	96
19	380	-177	390
20	289	-39	223
21	165	16	185
22	232	73	290
23	227	-14	218
24	408	12	94
25	154	12	173
26	272	-53	206
27	235	123	141
28	137	-46	171
29	238	6	-13
30	299	-2	73
31	142	15	219
32	304	-10	89
33	181	102	176
34	308	47	188
35	555	-22	429
36	392	167	294

Table K.2. Experiment 5. Individual values for the difference in reaction times between the masked (CT, SMN, RCT) and unmasked condition, between the CT condition and the average of the energetic mask conditions (RCT, SMN), and between the OR and the SR conditions.

Appendix L Eye fixation graphs for Experiment 6

L.1 Average fixation rates and *t*-statistics between target and competitor for Experiment 6

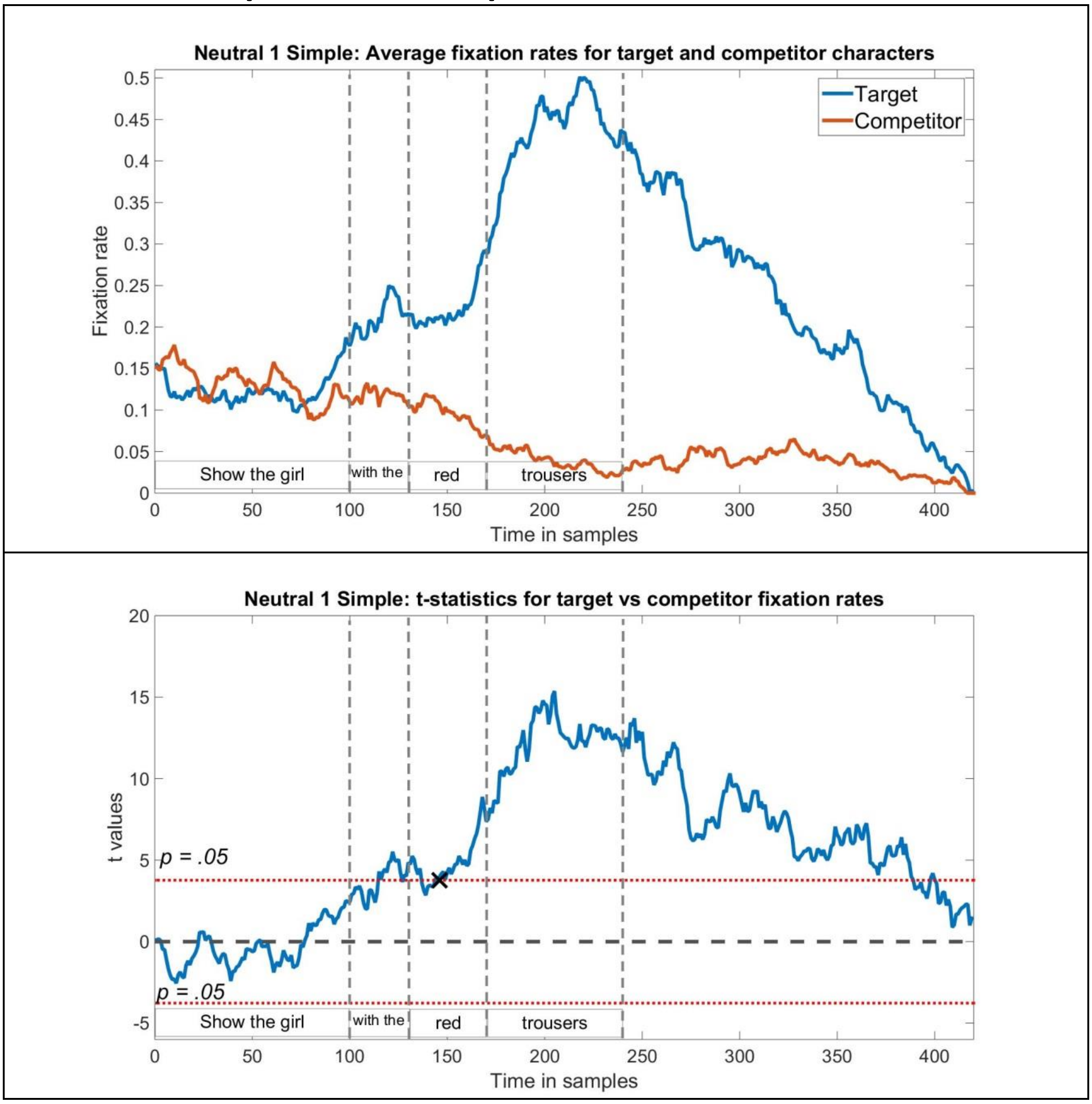


Figure L.1. Experiment 6. Average fixation rates and *t*-statistics for the neutral 1 condition with simple sentences.

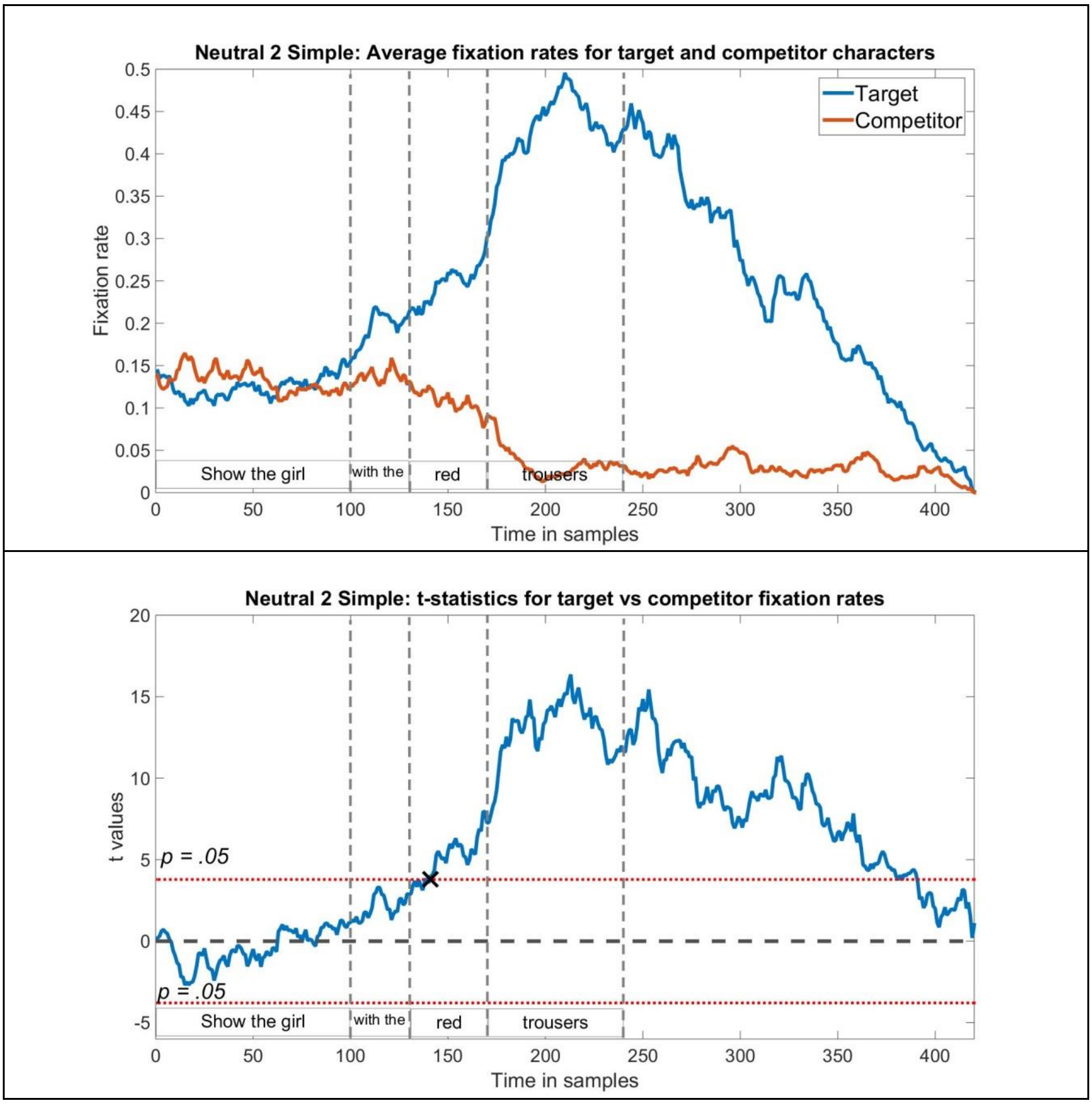


Figure L.2. Experiment 6. Average fixation rates and *t*-statistics for the neutral 2 condition with simple sentences.

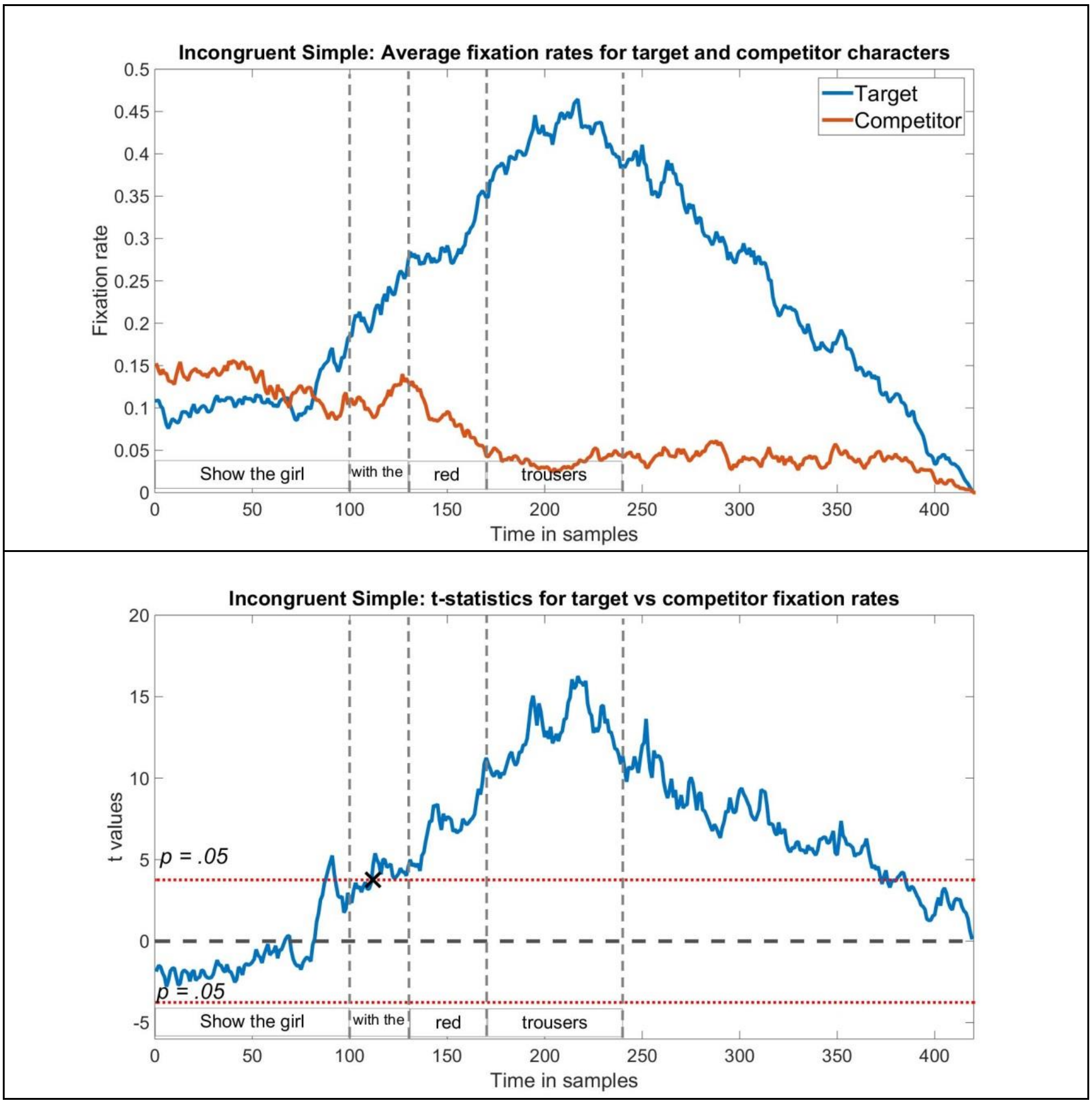


Figure L.3. Experiment 6. Average fixation rates and *t*-statistics for the incongruent condition with simple sentences.[33]

[33] For the incongruent condition, the *t*-value exceeds the critical *t* for a duration of 5 samples at the end of the first segment, before dipping down again for 19 samples.

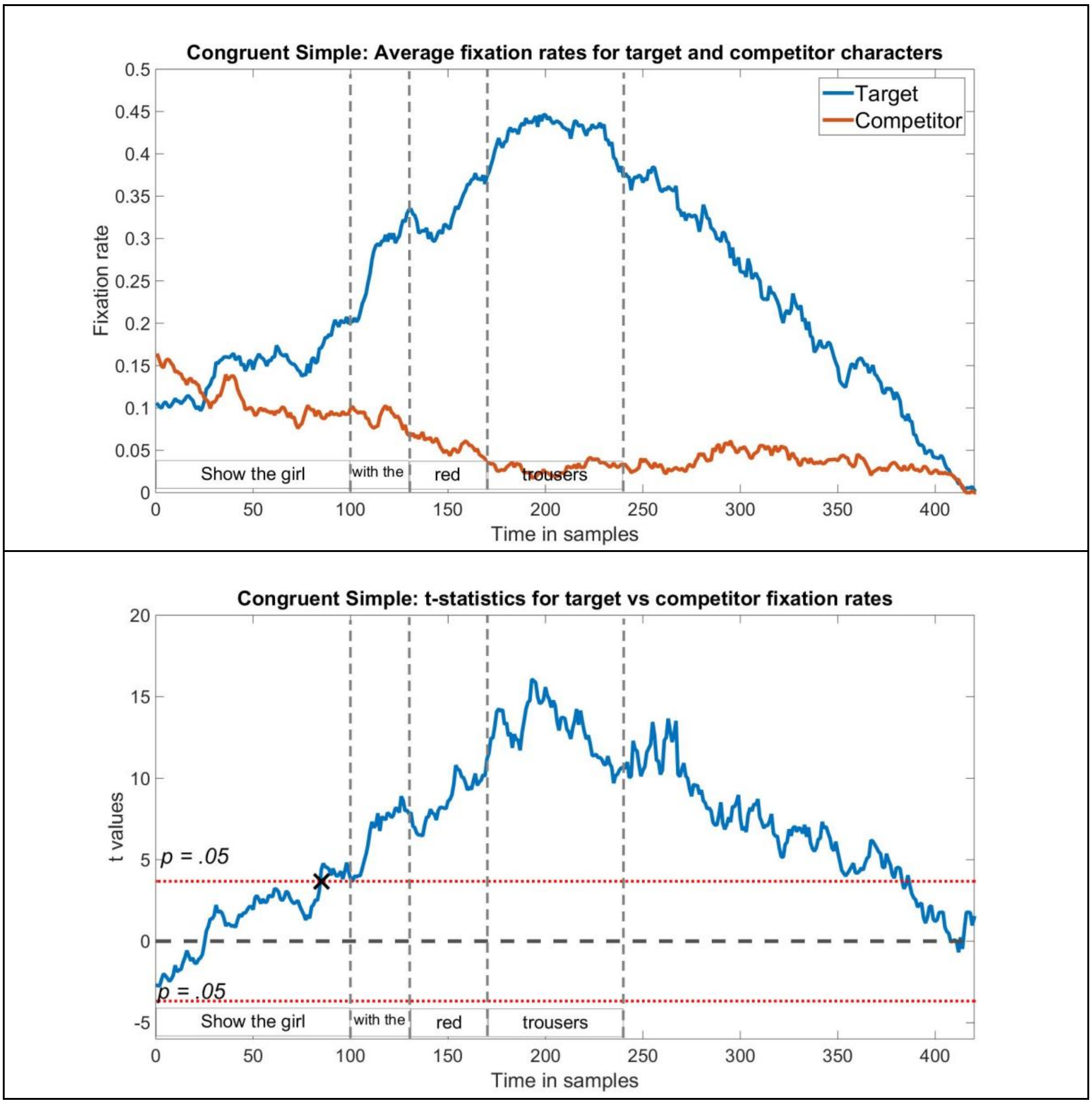


Figure L.4. Experiment 6. Average fixation rates and *t*-statistics for the congruent condition with simple sentences.[34]

[34] For the congruent condition, $p < .05$ from the 85th sample (end of the first segment of the sentence), but at the 101st sample (very beginning of the second segment), $p = .051$ and $t = 3.666$ (critical $t = 3.675$) for just one sample, before going back to $p < .05$ from the 102nd sample.

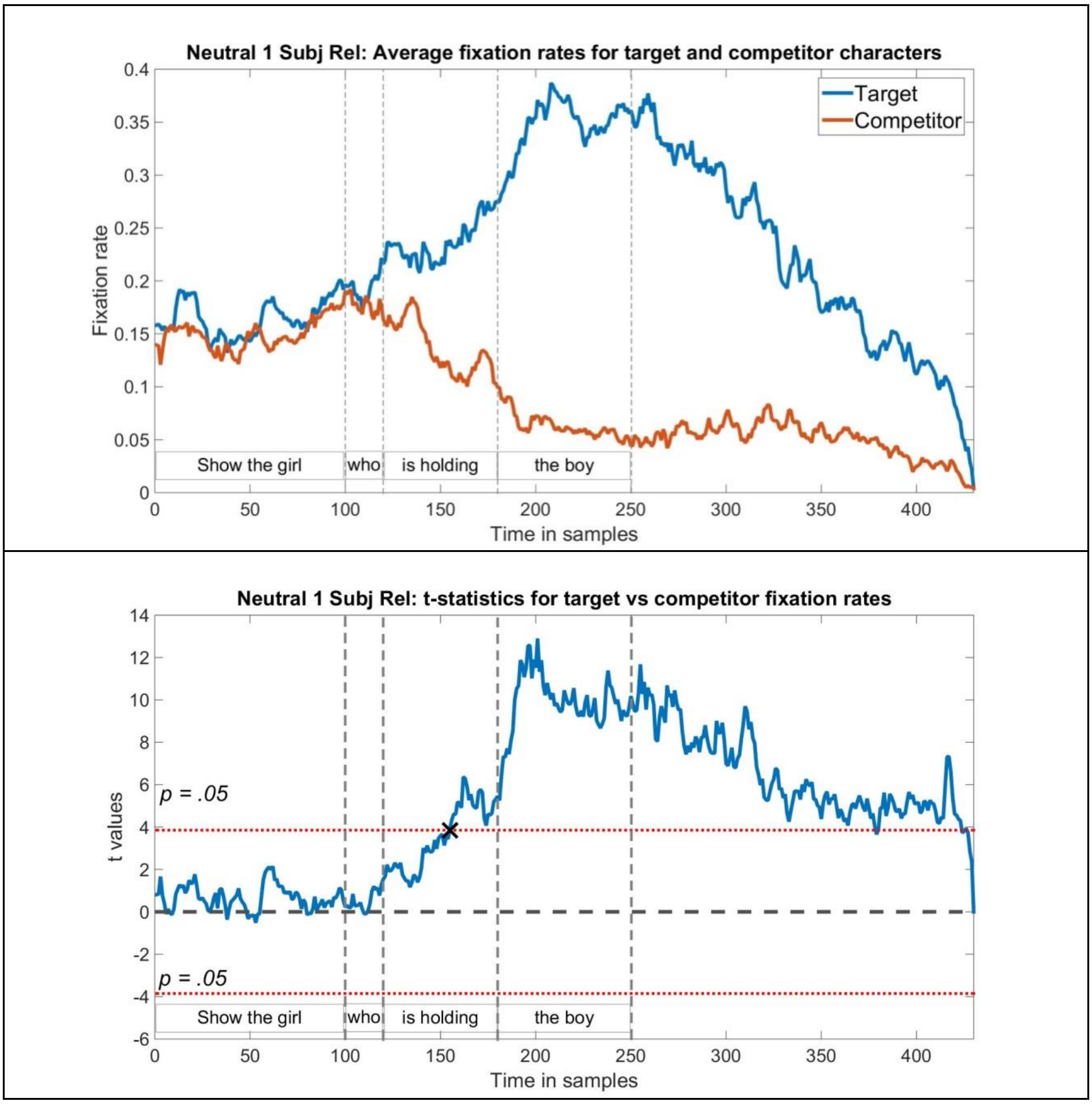


Figure L.5. Experiment 6. Average fixation rates and t-statistics for the neutral 1 condition with subject relative sentences.

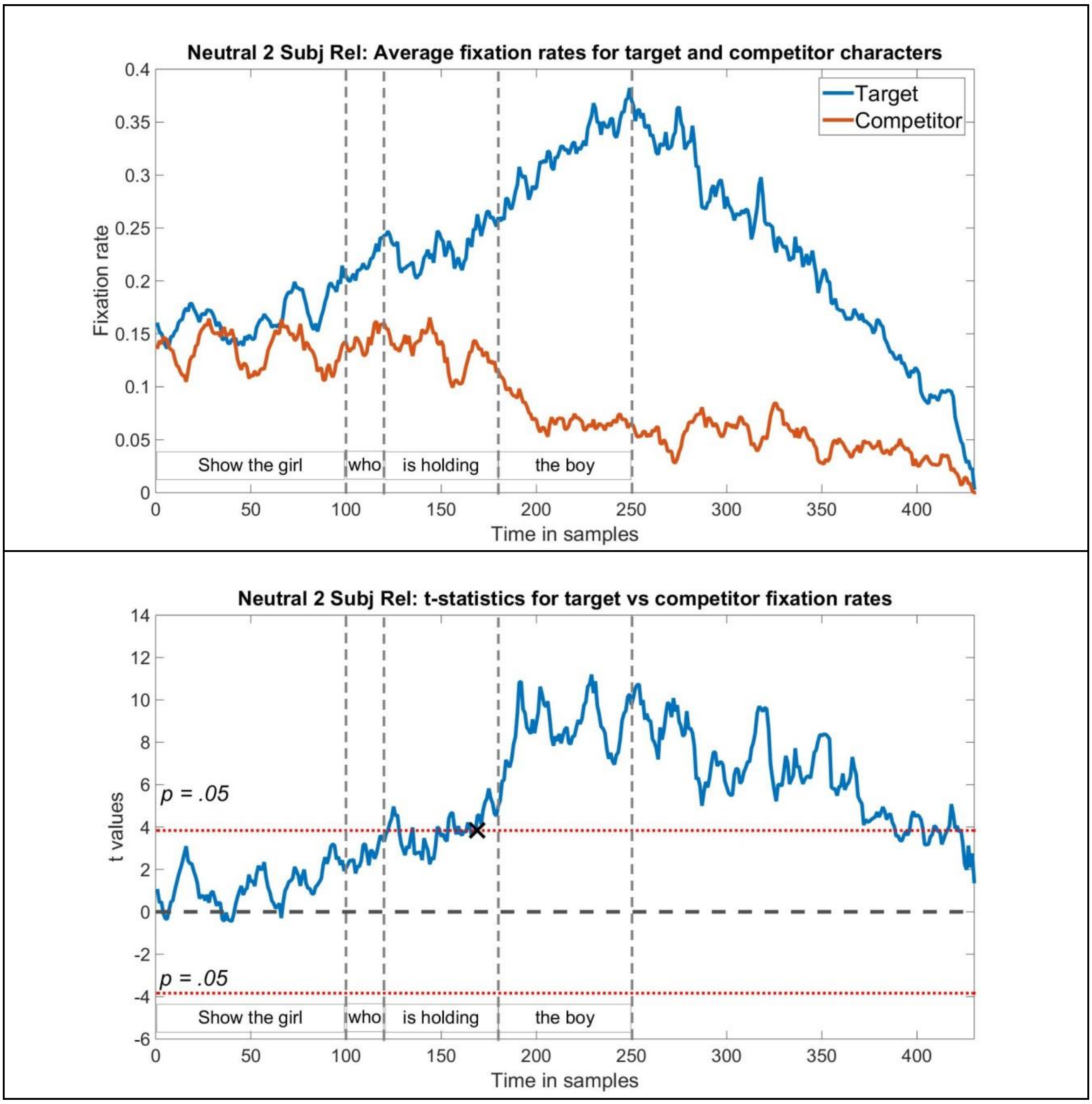


Figure L.6. Experiment 6. Average fixation rates and t-statistics for the neutral 2 condition with subject relative sentences.

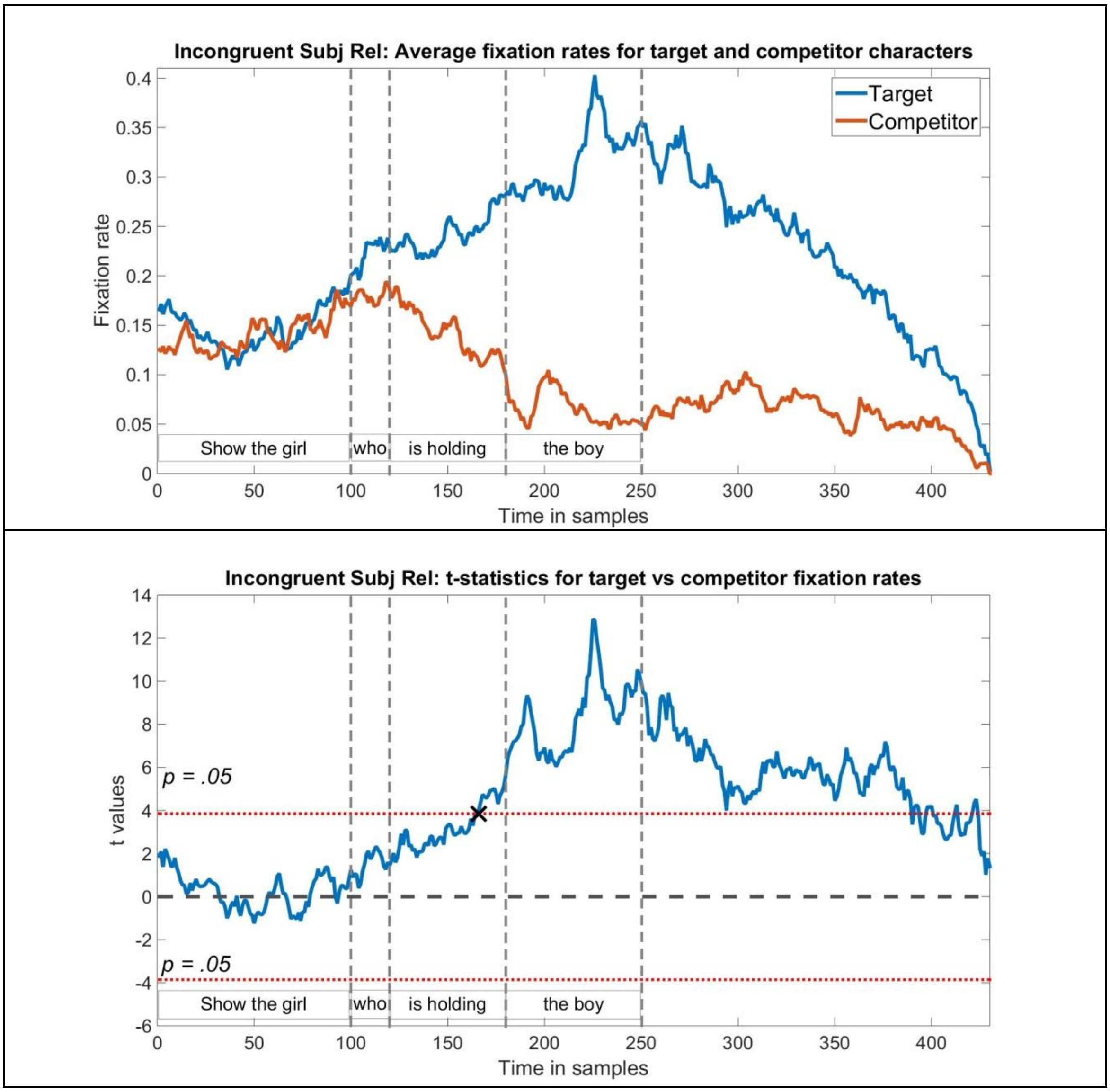


Figure L.7. Experiment 6. Average fixation rates and t-statistics for the incongruent condition with subject relative sentences.

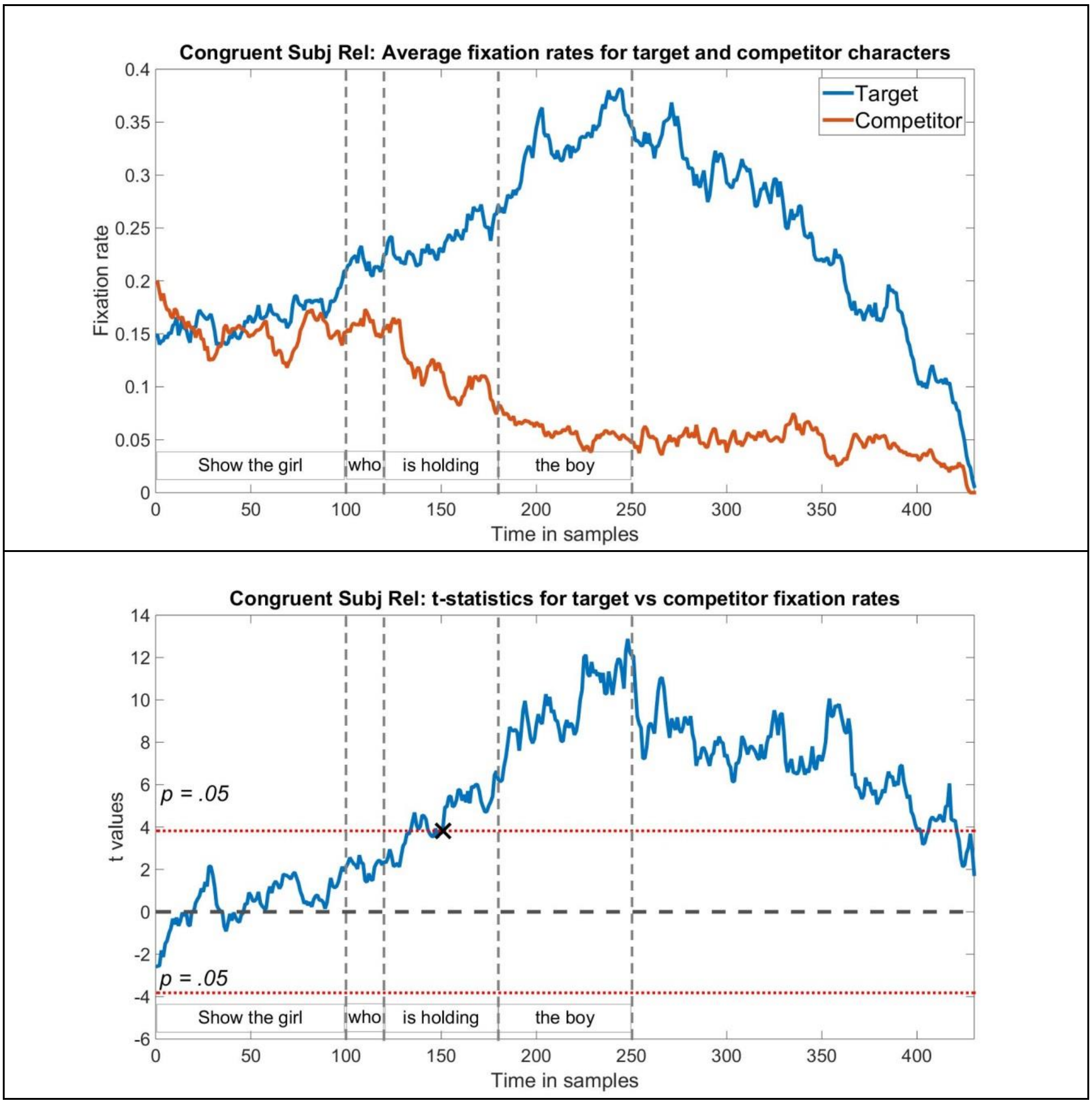


Figure L.8. Experiment 6. Average fixation rates and t-statistics for the congruent condition with subject relative sentences.

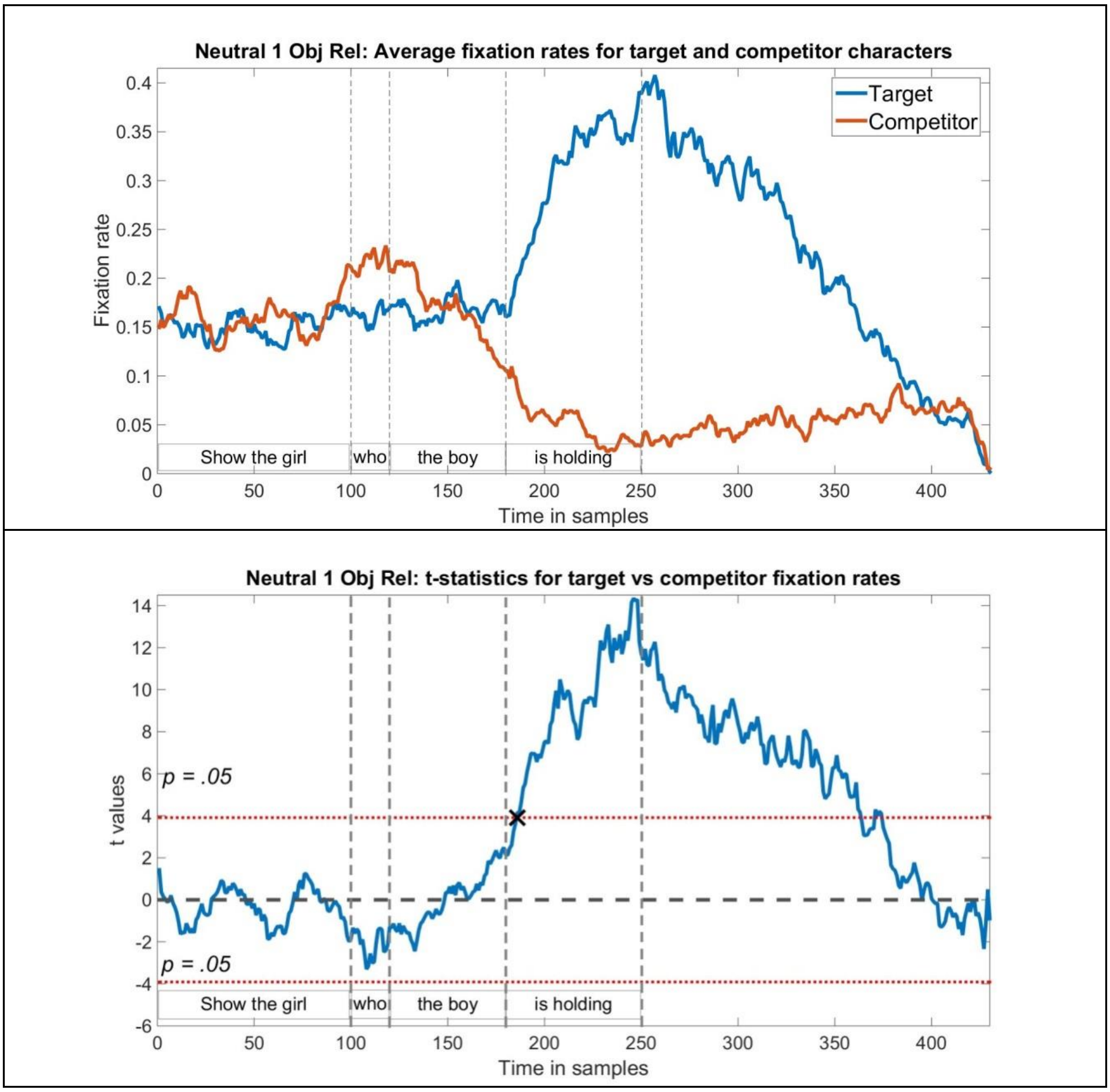


Figure L.9. Experiment 6. Average fixation rates and t-statistics for the neutral 1 condition with object relative sentences.

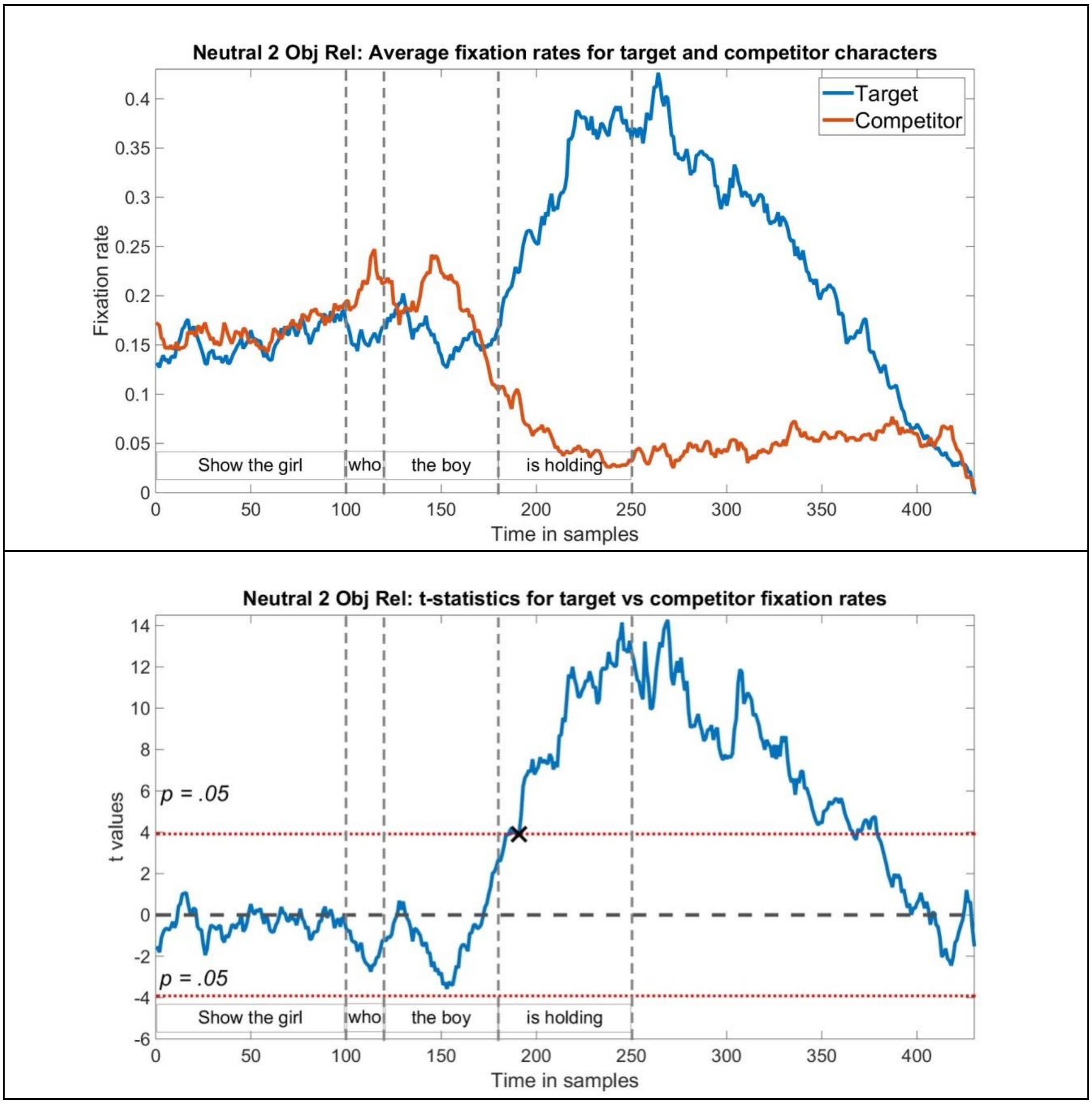


Figure L.10. Experiment 6. Average fixation rates and t-statistics for the neutral 2 condition with object relative sentences.

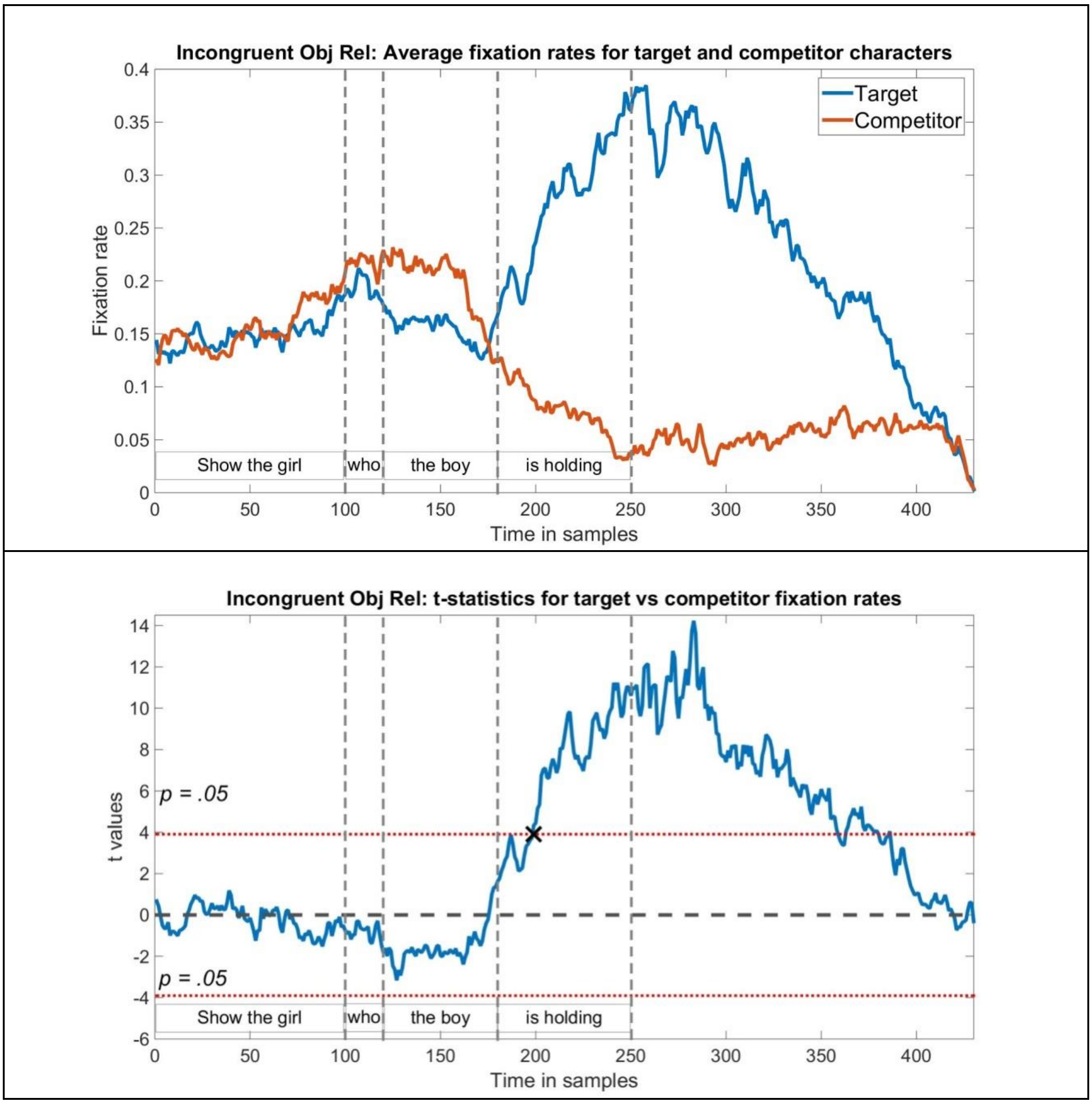


Figure L.11. Experiment 6. Average fixation rates and t-statistics for the incongruent condition with object relative sentences.

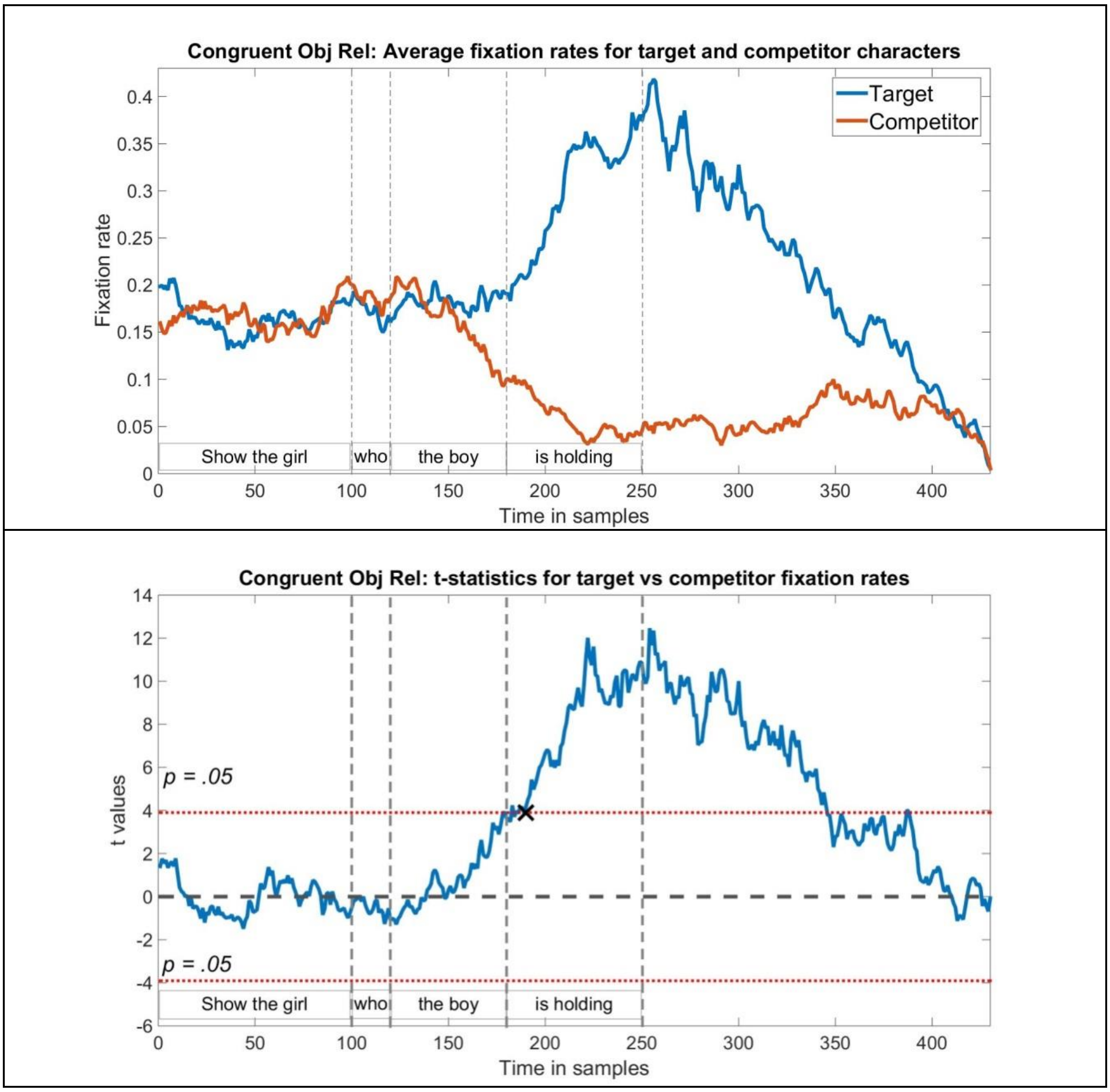
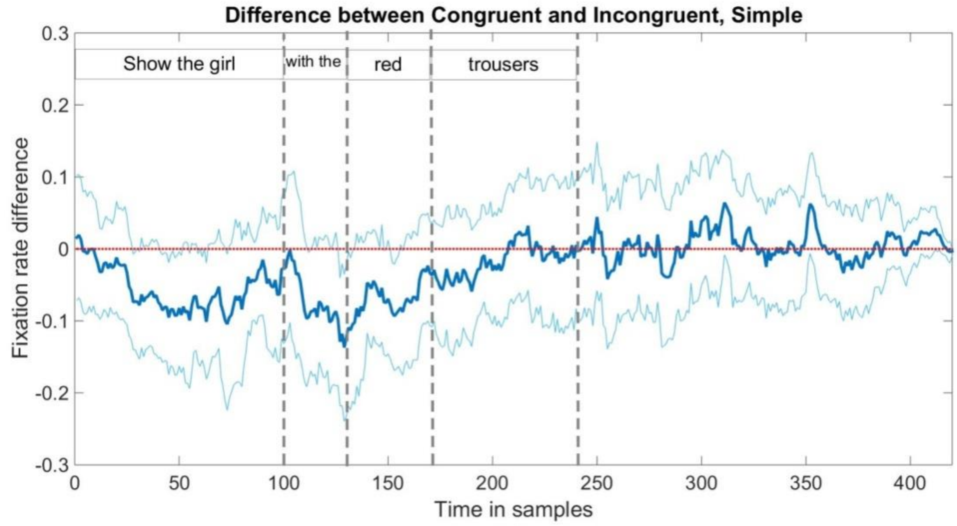
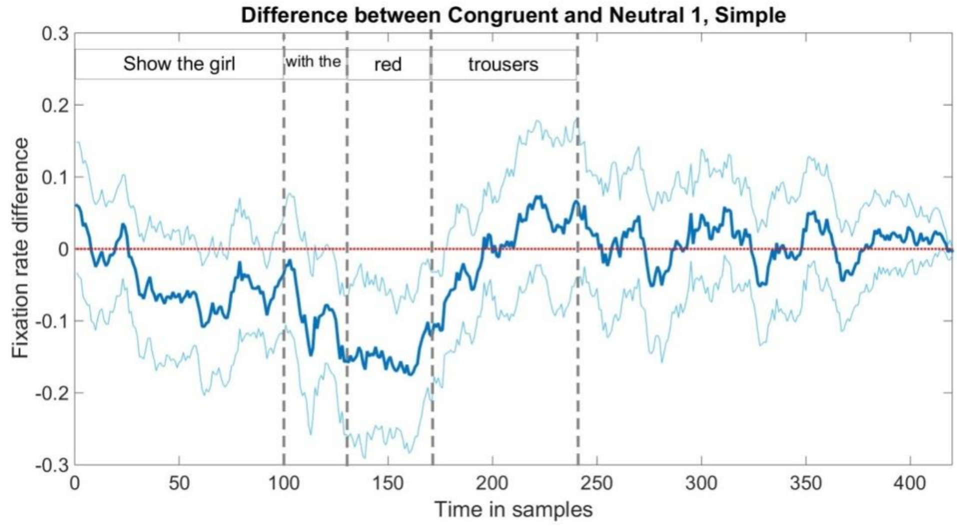
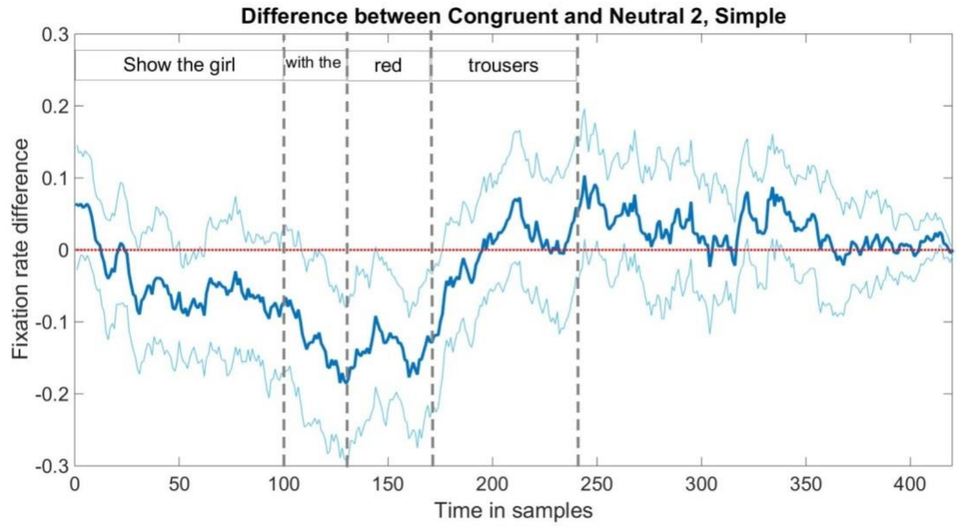


Figure L.12. Experiment 6. Average fixation rates and t-statistics for the congruent condition with object relative sentences.

L.2 Pairwise comparisons between masks, Experiment 6







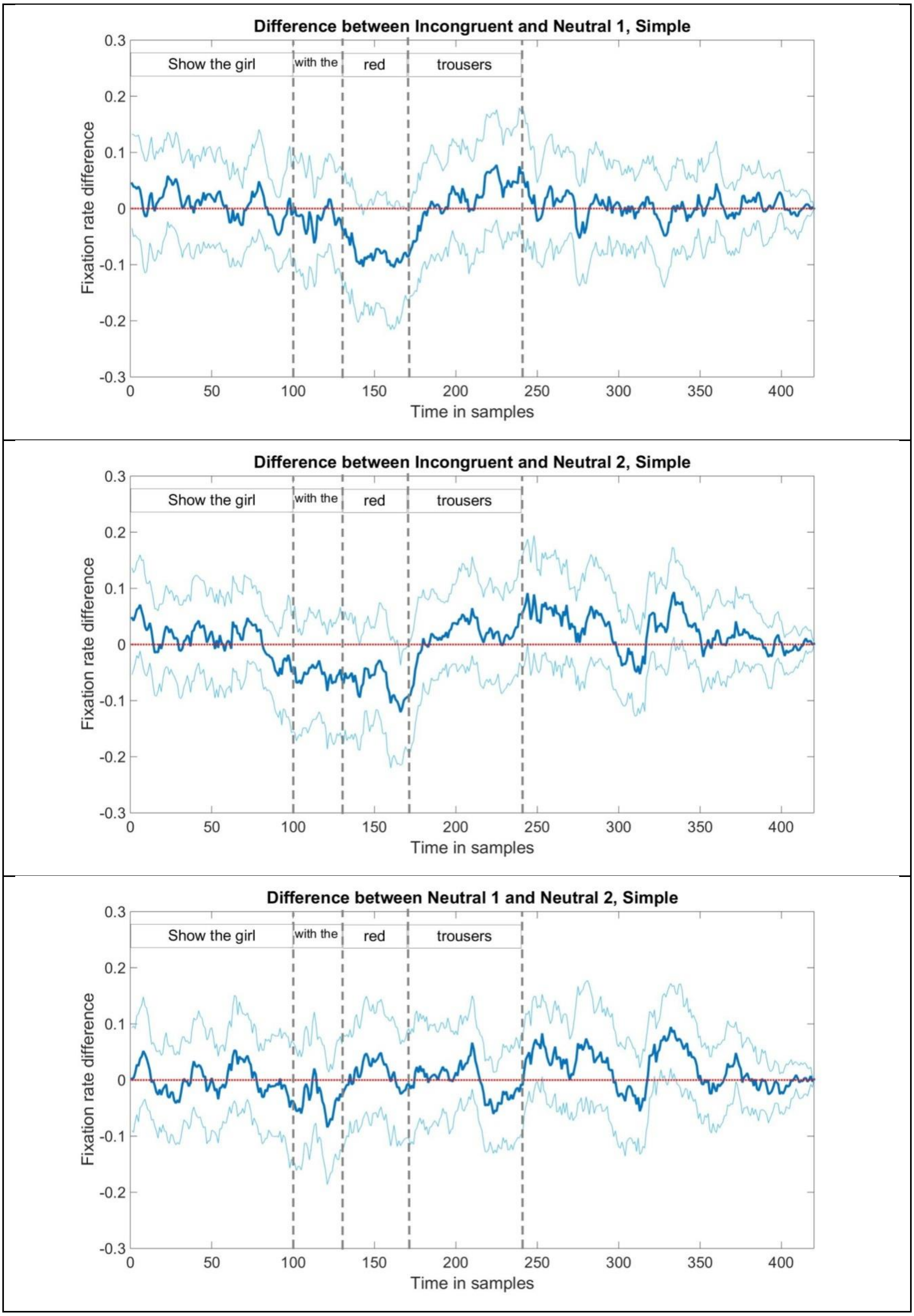
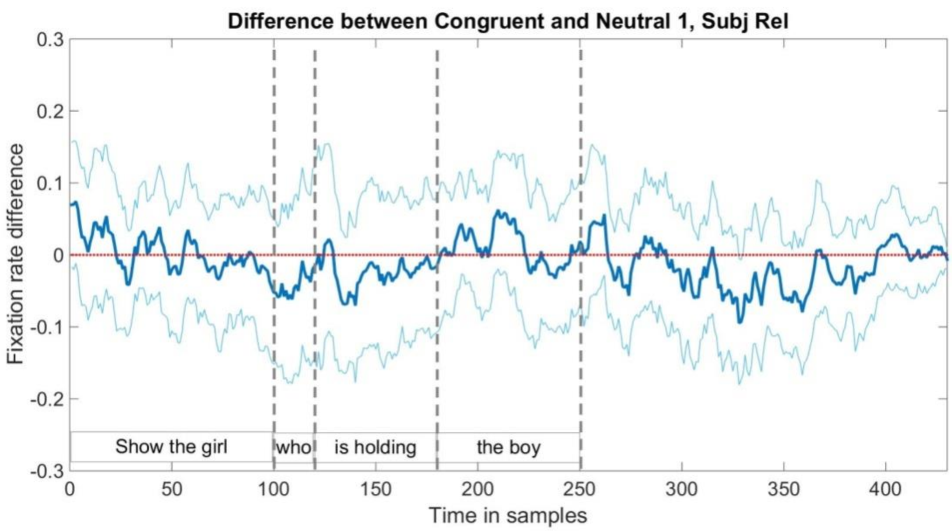
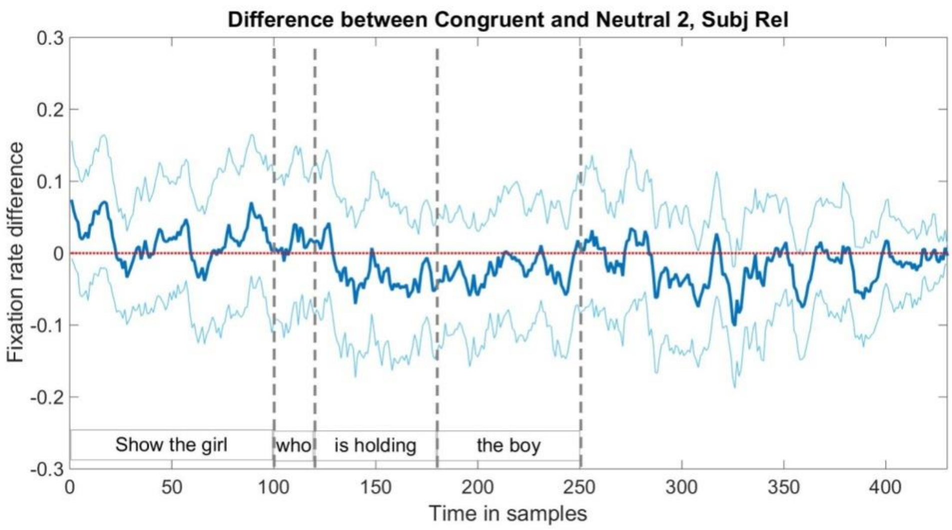
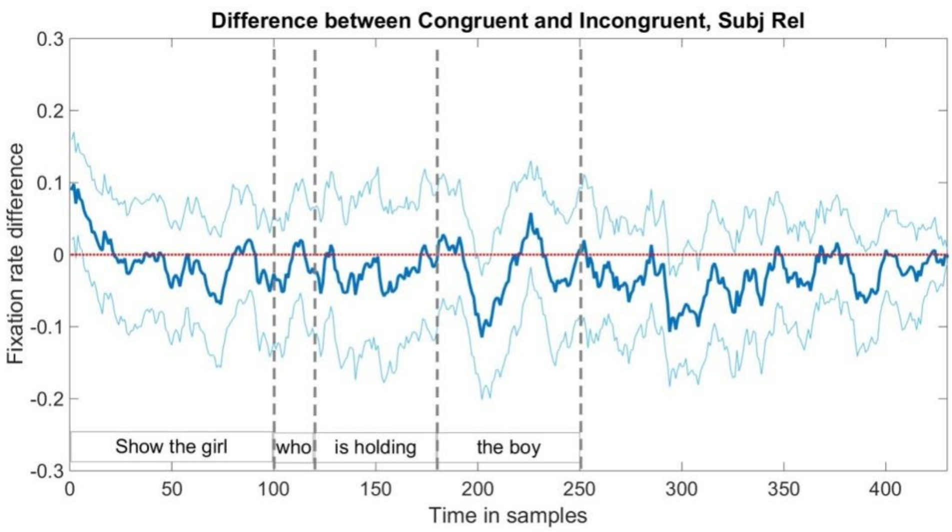


Figure L.13. Experiment 6. Differences between each of the mask conditions for simple sentences, with 99.17% confidence intervals to correct for multiple comparisons.


Difference between Congruent and Neutral 1, Subj Rel
Fixation rate difference
Time in samples
Show the girl
who
is holding
the boy


Difference between Congruent and Neutral 2, Subj Rel
Fixation rate difference
Time in samples
Show the girl
who
is holding
the boy


Difference between Congruent and Incongruent, Subj Rel
Fixation rate difference
Time in samples
Show the girl
who
is holding
the boy

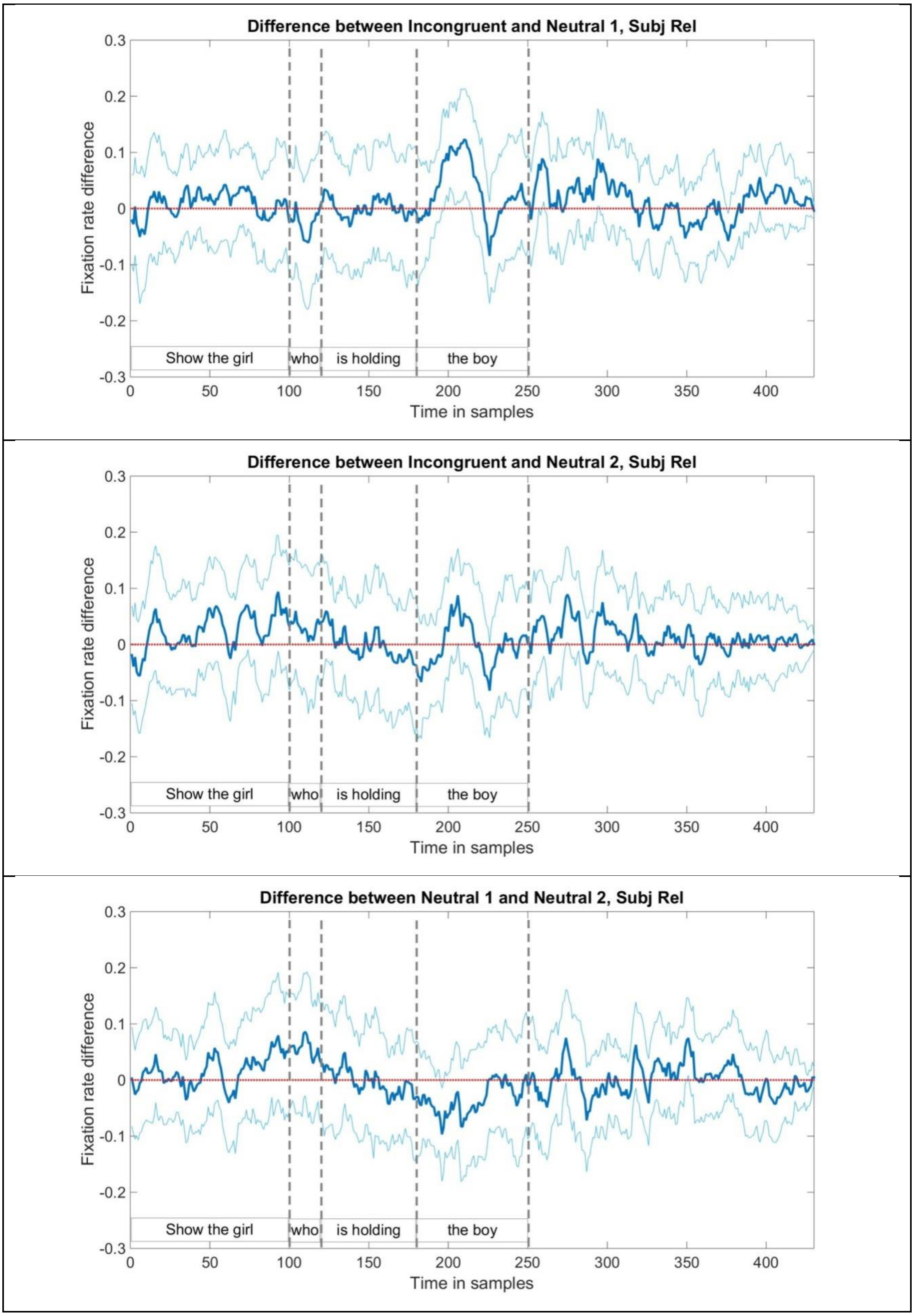
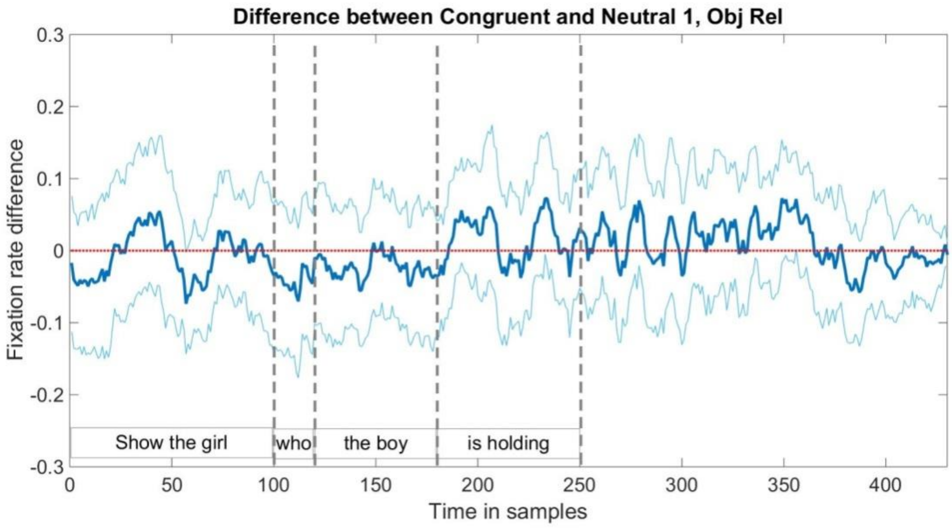
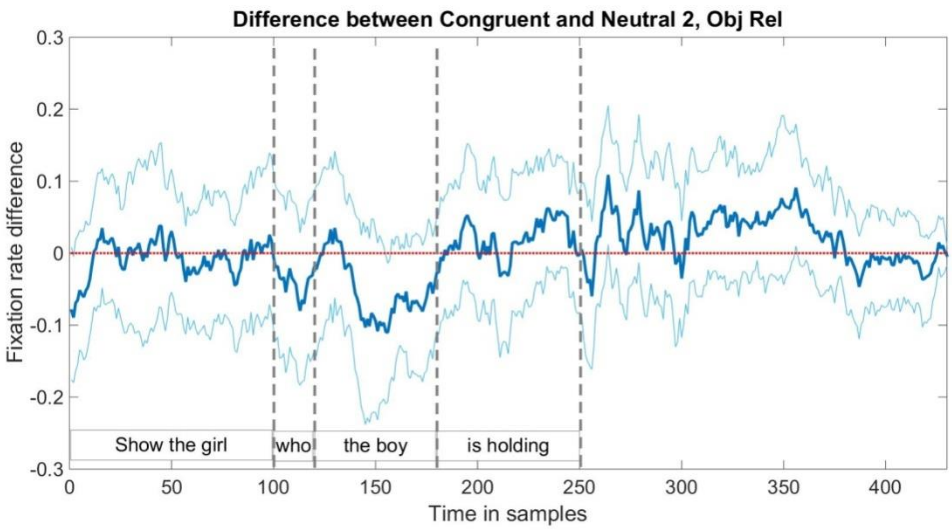
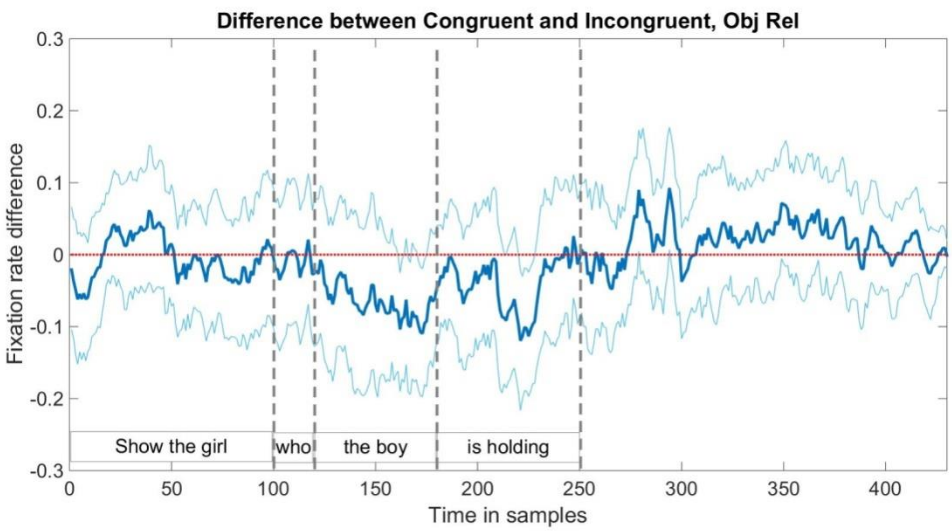


Figure L.14. Experiment 6. Differences between each of the mask conditions for subject relative sentences, with 99.17% confidence intervals to correct for multiple comparisons.


Difference between Congruent and Neutral 1, Obj Rel
Fixation rate difference
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Show the girl
who
the boy
is holding
0
50
100
150
200
250
300
350
400
Time in samples


Difference between Congruent and Neutral 2, Obj Rel
Fixation rate difference
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Show the girl
who
the boy
is holding
0
50
100
150
200
250
300
350
400
Time in samples


Difference between Congruent and Incongruent, Obj Rel
Fixation rate difference
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Show the girl
who
the boy
is holding
0
50
100
150
200
250
300
350
400
Time in samples

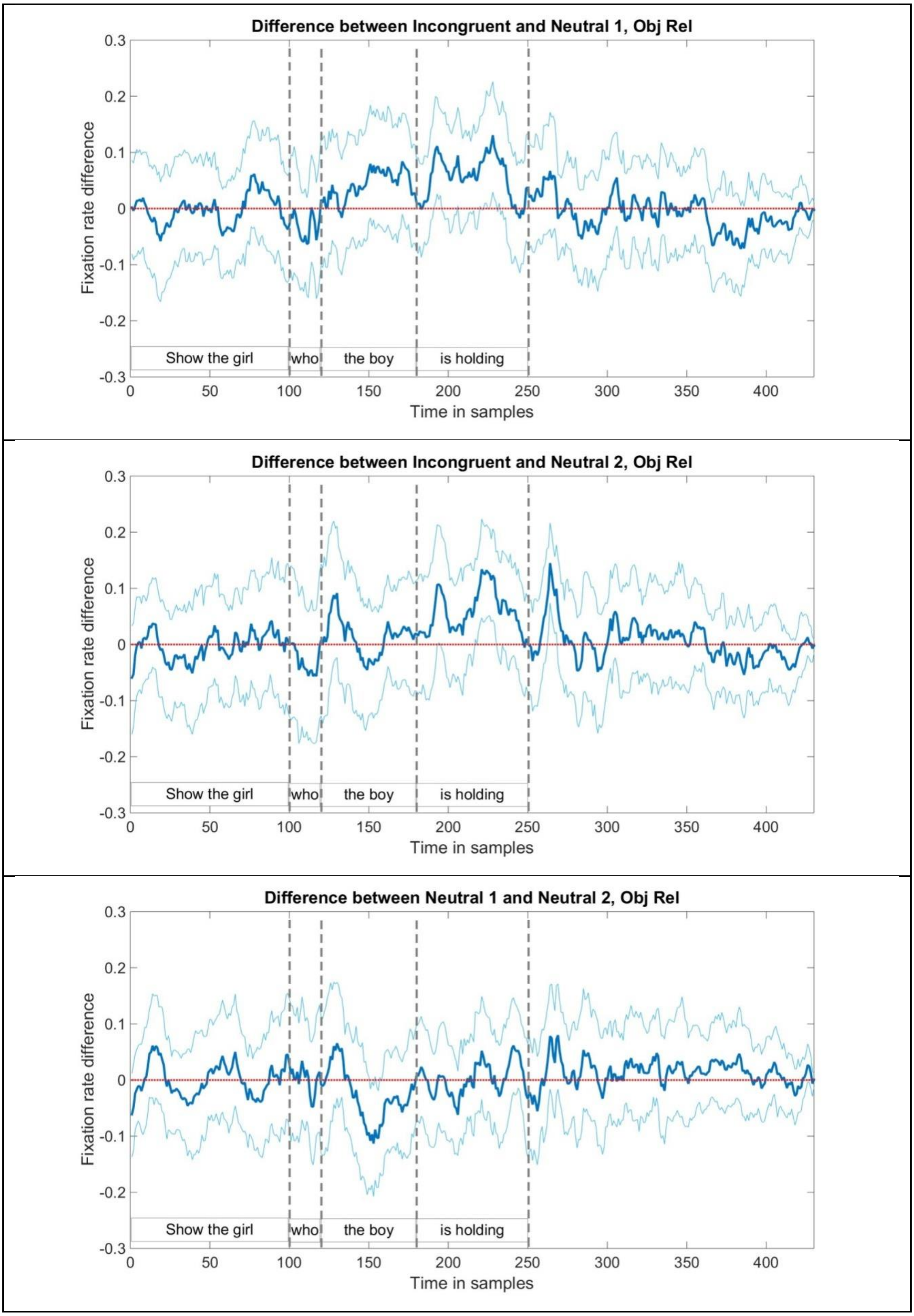


Figure L.15. Experiment 6. Differences between each of the mask conditions for object relative sentences, with 99.17% confidence intervals to correct for multiple comparisons.

Appendix M Individual accuracy and reaction time data for Experiment 6

M.1 Short-term and working memory spans for Experiment 6

Participant	Non-word repetition span	Forward digit span	Backward digit span
1	-	6	4
2	-	6	5
3	-	8	4
4	-	5	4
5	-	5	4
6	-	7	4
7	-	7	3
8	3	4	5
9	4	7	6
10	2	8	7
11	4	6	7
12	4	6	4
13	2	7	7
14	3	6	4
15	3	7	6
16	-	6	3
17	3	6	2
18	3	6	4
19	3	5	3
20	2	6	4
21	2	5	3
22	3	8	4
23	4	7	5
24	3	8	5
25	3	7	6
26	3	6	4
27	3	5	4
28	3	8	4
29	3	6	5
30	3	8	6
31	2	7	5
32	2	4	4
33	3	8	4
34	3	8	7
35	1	3	3
36	3	6	6

Table M.1. Experiment 6. Spans for the non-word recall and listening recall (including processing speed) AWMA subtests.

M.2 Selective attention: visual flanker task reaction times for Experiment 6

Participant	Neutral	Consistent	Inconsistent	Inconsistent-Consistent
1	397	418	473	55
2	381	373	431	58
3	375	390	400	10
4	363	348	387	39
5	397	399	444	45
6	395	393	407	14
7	396	385	448	63
8	405	437	428	-9
9	309	330	358	28
10	402	405	491	86
11	433	444	438	-6
12	428	426	461	35
13	398	396	492	96
14	392	407	462	55
15	362	332	416	84
16	432	397	433	37
17	378	380	428	48
18	404	372	462	90
19	450	439	500	61
20	347	349	403	54
21	380	390	412	22
22	335	366	416	50
23	351	350	373	23
24	376	391	450	59
25	417	436	520	84
26	350	322	408	86
27	346	341	392	51
28	399	385	464	78
29	420	445	470	26
30	456	459	492	33
31	393	408	445	37
32	441	441	478	36
33	328	339	364	25
34	360	360	402	43
35	434	436	489	53
36	341	345	372	27

Table M.2. Experiment 6. Reaction times in ms per participant for each of the flanker task conditions. RTs include only accurate responses.

Appendix N Individual accuracy and reaction time data for the sentence comprehension task in Experiment 6

Participant	Neutral 1 - Congruent	Neutral 2 - Congruent	Incongruent - Congruent	OR - SR
1	8%	5%	-3%	-2%
2	-3%	-2%	-2%	2%
3	0%	2%	-3%	-5%
4	-2%	-2%	-3%	-3%
5	-5%	3%	3%	-3%
6	-2%	-3%	-2%	0%
7	-8%	-10%	5%	-3%
8	3%	3%	-3%	-7%
9	3%	8%	-2%	3%
10	2%	2%	-7%	-7%
11	-2%	3%	2%	-2%
12	-3%	-3%	2%	-2%
13	5%	-3%	-2%	-5%
14	0%	-2%	-3%	2%
15	-2%	-3%	-5%	0%
16	2%	2%	2%	-2%
17	0%	0%	-3%	-3%
18	3%	2%	3%	-7%
19	5%	-10%	3%	2%
20	0%	5%	3%	-12%
21	0%	-3%	-5%	-2%
22	0%	0%	0%	-5%
23	3%	-3%	-3%	-2%
24	2%	-2%	3%	-5%
25	2%	-5%	-5%	-5%
26	5%	-3%	-7%	-5%
27	-2%	-15%	0%	-8%
28	2%	8%	5%	2%
29	2%	0%	0%	-2%
30	2%	-3%	0%	0%
31	-2%	-2%	2%	-3%
32	-7%	-5%	-7%	-15%
33	2%	0%	2%	0%
34	-2%	-2%	-2%	-2%
35	-10%	5%	8%	2%
36	-3%	-3%	-2%	-7%

Table N.1. Experiment 6 Individual values for the difference in percent accurate responses between the congruent condition and each of the other masks, and between the object relative and the subject relative conditions.

Participant	Neutral 1 - Congruent	Neutral 2 - Congruent	Incongruent - Congruent	OR - SR
1	-89	-10	49	202
2	85	124	157	158
3	86	-52	6	106
4	124	194	-99	302
5	14	79	9	197
6	134	129	174	152
7	161	186	270	111
8	197	181	-145	300
9	182	184	-53	221
10	38	-36	31	206
11	-64	132	37	227
12	91	60	277	167
13	67	8	-174	224
14	-87	103	42	138
15	119	103	169	-64
16	318	454	275	167
17	162	-76	10	216
18	-40	82	294	200
19	-59	-73	-57	251
20	-150	-239	-165	254
21	52	-120	60	103
22	138	211	109	206
23	-258	-140	-59	180
24	18	-196	-47	166
25	145	81	52	287
26	122	21	205	96
27	137	227	29	168
28	132	242	86	193
29	286	19	54	221
30	90	65	118	83
31	70	-41	-21	95
32	46	-124	-73	351
33	178	142	-8	265
34	81	-42	-63	319
35	237	-93	121	51
36	76	162	102	293

Table N.2. Experiment 6. Individual values for the difference in reaction times between the congruent condition and each of the other masks, and between the object relative and the subject relative conditions.

References

Adani, F. (2011). Rethinking the acquisition of relative clauses in Italian: towards a grammatically based account. *Journal of Child Language*, *38*(1), 141–65. http://doi.org/10.1017/S0305000909990250

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47 Suppl 2*, S53-71. http://doi.org/10.1080/14992020802301142

Alcántara, J. I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *45*(6), 1107–14. http://doi.org/10.1111/j.1469-7610.2004.t01-1-00303.x

Alloway, T. P. (2007). Automated Working Memory Assessment. Test, London: Pearson Assessment.

Arlinger, S. D., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scandinavian Journal of Psychology*, *50*(5), 371–84. http://doi.org/10.1111/j.1467-9450.2009.00753.x

Arosio, F., Adani, F., & Guasti, M. T. (2009). Grammatical features in the comprehension of Italian relative clauses by children. *Merging Features: Computation, …*, (1), 1–29. http://doi.org/10.1093/acprof:oso/9780199553266.003.0008

Assmann, P. F., & Summerfield, A. Q. (2004). The perception of speech under adverse conditions. *Speech Processing in the Auditory System*, 231–308.

Baayen, H., Piepenbrock, R., & Rijn, H. van. (1993). The CELEX database on CD-ROM. *Linguistic Data Consortium. Philadelpha, PA*. Retrieved from http://celex.mpi.nl/

Baddeley, A. D. (1992). Working Memory. *Science*, *255*(5044), 556–559. http://doi.org/10.1126/science.1736359

Baddeley, A. D. (2000). The episodic buffer : a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual Review* *of Psychology*, *63*, 1–29. http://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review*, *105*(1), 158–173. http://doi.org/10.1037/0033-295X.105.1.158

Baird, R., & Koslick, J. D. (1974). Recall of grammatical relations within clause-containing sentences. *Journal of Psycholinguistic Research*, *3*(2), 165–171. http://doi.org/10.1007/BF01067574

Baldwin, C. L., & Galinsky, A. M. (1999). Pure-Tone Threshold Shifts During Moderate Workload Conditions. *Automation Technology and Human Performance: Current Research and Trends*, 296–300.

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*(3), 108–12.

Bernstein, J. G. W., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *125*(5), 3358–3372. http://doi.org/10.1121/1.3110132

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–50. http://doi.org/10.1016/j.tics.2012.03.001

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*, *27*(March), 32–48. http://doi.org/10.1044/jshr.2701.32

Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, *30*(5), 518–524. http://doi.org/10.1111/j.1469-8986.1993.tb02075.x

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America*, *137*(1), 378–387. http://doi.org/10.1121/1.4904537

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer. Retrieved from http://www.praat.org

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*(2), 1065–1066. http://doi.org/10.1121/1.428288

Broadbent, D. E. (1952a). Failures of attention in selective listening. *Journal of Experimental Psychology*, *44*(6), 428–433. http://doi.org/10.1037/h0057163

Broadbent, D. E. (1952b). Listening to one of two synchronous messages. *Journal of Experimental Psychology*, *44*(1), 51–55. http://doi.org/10.1037/h0056491

Broadbent, D. E. (1958a). *Perception and communication*. Pergamon Press.

Broadbent, D. E. (1958b). Retrospect and Prospect. In *Perception and Communication* (pp. 297–316). Pergamon Press.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 2527–38. http://doi.org/10.1121/1.1345696

Brungart, D. S., Iyer, N., Thompson, E. R., Simpson, B. D., Gordon-Salant, S., Schurman, J., … Grant, K. (2013). Interactions between listening effort and masker type on the energetic and informational masking of speech stimuli. In *Proceedings of Meetings on Acoustics* (Vol. 19, p. 9). http://doi.org/10.1121/1.4800033

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527–38. http://doi.org/10.1121/1.1408946

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (2002). Immature Frontal Lobe Contributions to Cognitive Control in Children. *Neuron*, *33*(2), 301–311. http://doi.org/10.1016/S0896-6273(01)00583-9

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, *128*(2), 860–9. http://doi.org/10.1121/1.3458857

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457–469.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *The Behavioral and Brain Sciences*, *22*(1), 77–126.

Carreiras, M., Duñabeitia, J. A., Vergara, M., de la Cruz-Pavía, I., & Laka, I. (2010). Subject relative clauses are not universally easier to process: Evidence from Basque. *Cognition*, *115*(1), 79–92. http://doi.org/10.1016/j.cognition.2009.11.012

Carroll, R. (2012). *Effects of Syntactic Complexity and Prosody on Sentence Processing and Comprehension in Noise*. Oldenburg University.

Carroll, R., & Ruigendijk, E. (2013). The effects of syntactic complexity on processing sentences in noise. *Journal of Psycholinguistic Research*, *42*(2), 139–59. http://doi.org/10.1007/s10936-012-9213-7

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–70. http://doi.org/10.1016/j.tics.2006.10.002

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331–5.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Cooke, A., Zurif, E. B., DeVita, C., Alsop, D., Koenig, P., Detre, J., … Grossman, M. (2002). Neural basis for sentence comprehension: Grammatical and short-term memory components. *Human Brain Mapping*, *94*(May), 80–94. http://doi.org/10.1002/hbm.10006

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562. http://doi.org/10.1121/1.2166600

Cooke, M., Lecumberri, M. L. G., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, *123*(1), 414–27. http://doi.org/10.1121/1.2804952

Cool Edit Pro. (2002). Phoenix, Arizona: Syntrillium Software Corporation.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology*, *6*(1), 84–107. http://doi.org/10.1016/0010-0285(74)90005-X

Cowan, N. (1999). An Embedded Processes Model of Working Memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 62–101). Cambridge: Cambridge University Press.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668–3678. http://doi.org/10.1121/1.1810292

Dai, B., Kösem, A., McQueen, J. M., & Hagoort, P. (2016). Pure linguistic interference during comprehension of competing speech signals. In *Speech in Noise Workshop* (Vol. January). Groningen.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90. http://doi.org/10.1037/h0039515

Dillon, L. M. (1995, May). *The effect of noise and syntactic complexity on listening comprehension*. The University of British Columbia.

Dirks, D. D., & Bower, D. R. (1969). Masking Effects of Speech Competing Messages. *Journal of Speech, Language, and Hearing Research*, *12*(2), 229–245. http://doi.org/10.1044/jshr.1202.229

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *The Journal of the Acoustical Society of America*, *74*(3), 739–43.

Durlach, N. I. (2006). Auditory masking: Need for improved conceptual structure. *The Journal of the Acoustical Society of America*, *120*(4), 1787. http://doi.org/10.1121/1.2335426

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. http://doi.org/10.3758/BF03203267

Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the Cocktail Party Started: Masking Effects in Speech Perception. *Journal of Cognitive Neuroscience*, *28*(3), 483–500. http://doi.org/10.1162/jocn_a_00913

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). New York: Harper & Row.

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–36.

Filippi, R., Leech, R., Thomas, M. S., Green, D. W., & Dick, F. (2012). A bilingual advantage in controlling language interference during sentence comprehension. *Bilingualism: Language and Cognition*, *15*(4), 858–872. http://doi.org/10.1017/S1366728911000708

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 203–218. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022537183901561

Forster, K. I., & Forster, J. C. (2003, February). *DMDX: A Windows display program with millisecond accuracy*. *Behavior Research Methods, Instruments, & Computers*.

Francart, T., van Wieringen, A., & Wouters, J. (2011). Comparison of fluctuating maskers for speech recognition tests. *International Journal of Audiology*, *50*(1), 2–13. http://doi.org/10.3109/14992027.2010.505582

Francis, A. L. (2010). Improved segregation of simultaneous talkers differentially affects perceptual and cognitive capacity demands for recognizing speech in competing speech. *Attention, Perception & Psychophysics*, *72*(2), 501–516. http://doi.org/10.3758/APP.72.2.501

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, *71*(6), 1360–1374. http://doi.org/10.3758/APP

French, N. R., & Steinberg, J. C. (1947). Factors Governing the Intelligibility of Speech Sounds. *The Journal of the Acoustical Society of America*, *19*(1), 90. http://doi.org/10.1121/1.1916407

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *109*(5), 2112–2122. http://doi.org/10.1121/1.1354984

Füllgrabe, C., & Rosen, S. (2016). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Frontiers in Psychology*, *7*(1268). http://doi.org/10.3389/fpsyg.2016.01268

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*(3), 336–360. http://doi.org/10.1016/0749-596X(90)90004-J

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: a longitudinal study. *Developmental Psychology*, *28*(5), 887–898.

Gautreau, A., Hoen, M., & Meunier, F. (2012). Masques acoustiques et masques linguistiques de différentes langues sur la reconnaissance de mots en français. In *Actes de la conférence conjointe JEP-TALN-RECITAL* (Vol. 1, pp. 755–762). ATALA & AFCP.

Gennari, S. P., & Macdonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*(4), 161–187. http://doi.org/10.1016/j.jml.2007.07.004

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, *51*(1), 97–114. http://doi.org/10.1016/j.jml.2004.02.003

Gordon, P., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1–13. http://doi.org/10.1037//0278-7393.27.6.0

Gratton, G., Coles, M. G. H., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 331–344.

Green, T., Rosen, S., Faulkner, A., & Paterson, R. (2013). Adaptation to spectrally-rotated speech. *The Journal of the Acoustical Society of America*, *134*(2), 1369. http://doi.org/10.1121/1.4812759

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, *29*(2), 261–90. http://doi.org/10.1207/s15516709cog0000_7

Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing Subject-Object Ambiguities in the L2: A Self-Paced Reading Study With German L2 Learners of Dutch. *Language Learning*, *59*(1), 73–112. http://doi.org/10.1111/j.1467-9922.2009.00501.x

Hazeltine, E., Poldrack, R., & Gabrieli, J. D. (2000). Neural activation during response competition. *Journal of Cognitive Neuroscience*, *12 Suppl 2*, 118–29. http://doi.org/10.1162/089892900563984

Helfer, K. S., & Freyman, R. L. (2009). Lexical and indexical cues in masking by competing speech. *The Journal of the Acoustical Society of America*, *125*(1), 447–56. http://doi.org/10.1121/1.3035837

Helfer, K. S., & Freyman, R. L. (2014). Stimulus and listener factors affecting age-related changes in competing speech perception. *The Journal of the Acoustical Society of America*, *136*(2), 748–59. http://doi.org/10.1121/1.4887463

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 417–430.

Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology*, *49*(12), 891–903. http://doi.org/10.3109/14992027.2010.506889

Howard-Jones, P. A., & Rosen, S. (1993a). The Perception of Speech in Fluctuating Noise. *Acustica*, *78*(5), 258–272(15).

Howard-Jones, P. A., & Rosen, S. (1993b). Uncomodulated glimpsing in "checkerboard" noise. *The Journal of the Acoustical Society of America*, *93*(5), 2915–2922. http://doi.org/10.1121/1.405811

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*(1), 3–27. http://doi.org/10.1016/S0010-0277(03)00124-0

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23-32.

http://doi.org/10.1016/j.cognition.2004.10.003

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica*, *137*(2), 151–71. http://doi.org/10.1016/j.actpsy.2010.11.003

Hygge, S., Rönnberg, J., Larsby, B., Arlinger, S. D., & Rönnberg, J. (1992). Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds. *Journal of Speech and Hearing Research*, *35*(1), 208–15. http://doi.org/10.1044/jshr.3501.208

Iyer, N., Brungart, D. S., & Simpson, B. D. (2010). Effects of target-masker contextual similarity on the multimasker penalty in a three-talker diotic listening task. *The Journal of the Acoustical Society of America*, *128*(5), 2998–2910. http://doi.org/10.1121/1.3479547

Jensen De López, K., Sundahl Olsen, L., & Chondrogianni, V. (2014). Annoying Danish relatives: comprehension and production of relative clauses by Danish children with and without SLI. *Journal of Child Language*, *41*(1), 51–83. http://doi.org/10.1017/S0305000912000517

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, New Jersey: Prentice-Hall.

Kidd, G., Best, V., & Mason, C. R. (2008). Listening to every other word: examining the strength of linkage variables in forming streams of speech. *The Journal of the Acoustical Society of America*, *124*(6), 3793–3802. http://doi.org/10.1121/1.2998980

Kidd, G., Mason, C. R., & Best, V. (2014). The role of syntax in maintaining the integrity of streams of speech. *The Journal of the Acoustical Society of America*, *135*(2), 766–777. http://doi.org/10.1121/1.4861354

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2007). Informational Masking. In *Auditory Perception of Sound Sources* (pp. 143–189). Springer.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602. http://doi.org/10.1016/0749-596X(91)90027-H

Klatte, M., Lachmann, T., & Meis, M. (2010). Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. *Noise & Health*, *12*(49), 270–82. http://doi.org/10.4103/1463-1741.70506

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear and Hearing*, *33*(2), 291–300. http://doi.org/10.1097/AUD.0b013e3182310019

Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2014). The influence of informational masking on speech perception and pupil response in adults with hearing impairment. *The Journal of the Acoustical Society of America*, *135*(3), 1596. http://doi.org/10.1121/1.4863198

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82. http://doi.org/10.1016/j.tics.2004.12.004

Lavie, N., & Dalton, P. (2014). Load Theory of Attention and Cognitive Control. *The Oxford Handbook of Attention*, (June), 56–75. http://doi.org/10.1093/oxfordhb/9780199675111.013.003

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197. http://doi.org/10.3758/BF03213897

Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445. http://doi.org/10.1121/1.2180210

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11–12), 864–886. http://doi.org/10.1016/j.specom.2010.08.014

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–43. http://doi.org/10.3758/s13428-011-0146-0

Lew, H., & Jerger, J. (1991). Effect of linguistic interference on sentence identification. *Ear and Hearing*, *12*(5), 365–367. http://doi.org/10.1097/00003446-199110000-00013

Lewis, D. E., Manninen, C. M., Valente, D. L., & Smith, N. A. (2014). Children’s Understanding of

Instructions Presented in Noise and Reverberation. *American Journal of Audiology*. http://doi.org/10.1044/2014_AJA-14-0020

Lewis, R. L., Vasishth, S., & Van Dyke, J. a. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–54. http://doi.org/10.1016/j.tics.2006.08.007

Macdonald, J. S. P., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception & Psychophysics*, *73*(6), 1780–1789. http://doi.org/10.3758/s13414-011-0144-4

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in Self-Ratings of Second Language Proficiency: The Role of Language Anxiety. *Language Learning*, *47*(2), 265–287. http://doi.org/10.1111/0023-8333.81997008

Mair, K. R. (2013). *Speech Perception in Autism Spectrum Disorder : Susceptibility to Masking and Interference*. University College London.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The Influence of Animacy on Relative Clause Processing. *Journal of Memory and Language*, *47*(1), 50–68. http://doi.org/10.1006/jmla.2001.2837

MATLAB. (2010). Natick, Massachussetts: The MathWorks, Inc.

Mattys, S. L., Carroll, L. M., Li, C. K. W. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, *52*(11), 887–899.

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

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "white paper." *International Journal of Audiology*. http://doi.org/10.3109/14992027.2014.890296

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071. http://doi.org/10.3758/PBR.15.6.1064

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin*, *44*(2), 105–129.

Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, *2*(3), 217–228. http://doi.org/10.1016/S0022-5371(63)80087-0

Mishra, S., Lunner, T., Stenfelt, S., Rönnberg, J., & Rudner, M. (2013). Visual information can hinder working memory processing of speech. *Journal of Speech, Language, and Hearing Research : JSLHR*, *56*(4), 1120–32. http://doi.org/10.1044/1092-4388(2012/12-0033)

Moore, B. C. J. (2013). *An Introduction to the Psychology of Hearing* (6th ed.). Bingley, UK: Emerald.

Moore, T. J. (1981). Voice communication jamming research. In *AGARD Conference Proceedings 311: Aural Communication in Aviation* (p. 2:1-2:6). Neuilly-sur-Seine.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. http://doi.org/10.1080/17470215908416289

Murphy, S., Fraenkel, N., & Dalton, P. (2013). Perceptual load does not modulate auditory distractor processing. *Cognition*, *129*(2), 345–355. http://doi.org/10.1016/j.cognition.2013.07.014

Nairne, J. S. (1990). A feature model of immediate memory. *Memory and Cognition*, *18*(3), 251–269.

Nilsson, M., Soli, S. D., & Sullivan, J. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise The BKB sentence materials. *Journal of the Acoustical Society of America*, *95*(June 1993), 1085–1099.

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neuroscience and Biobehavioral Reviews*, *36*(2), 836–54. http://doi.org/10.1016/j.neubiorev.2011.11.008

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*(3), 331–347.

Plomp, R., & Mimpen, A. (1979). Improving the reliability of testing the speech reception threshold for sentences. *International Journal of Audiology*.

Pollack, I. (1975). Auditory informational masking. In *89th Meeting of the Acoustical Society of America* (Vol. 57).

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, *114*(1), 446. http://doi.org/10.1121/1.1579009

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. *Attention, Perception, & Psychophysics*, *77*(2), 483–492. http://doi.org/10.3758/s13414-014-0776-2

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *The Journal of the Acoustical Society of America*, *118*(3 Pt 1), 1274–1277. http://doi.org/10.1121/1.2000751

Rönnberg, J., Lunner, T., Zekveld, A. A., Sörqvist, P., Danielsson, H., Lyxell, B., … Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(July), 2–17. http://doi.org/10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47 Suppl 2*, S99-105. http://doi.org/10.1080/14992020802301167

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London*, *336*(1278), 367–373. http://doi.org/10.1098/rstb.1992.0070

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

Rudner, M., Ng, E. H. N., Rönnberg, N., Mishra, S., Rönnberg, J., Lunner, T., & Stenfelt, S. (2011). Cognitive spare capacity as a measure of listening effort. *Journal of Hearing …*, *1*(2), 47–49.

Salamé, P., & Baddeley, A. D. (1987). Noise, unattended speech and short-term memory. *Ergonomics*, *30*(8), 1185–94. http://doi.org/10.1080/00140138708966007

Schneider, B. A., Trehub, S. E., Morrongiello, B. A., & Thorpe, L. A. (1989). Developmental changes in chickens' masked thresholds. *The Journal of the Acoustical Society of America*, *86*(5), 1733–1742. http://doi.org/10.1002/dev.420260803

Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America*, *115*(2), 813. http://doi.org/10.1121/1.1639336

Shah, P., & Miyake, A. (1999). Models of Working Memory: an Introduction. In A. Miyake & S. Priti (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 1–28). Cambridge.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–6. http://doi.org/10.1016/j.tics.2008.02.003

Sörqvist, P., Ljungberg, J. K., & Ljung, R. (2010). A sub-process view of working memory capacity: evidence from effects of speech on prose memory. *Memory (Hove, England)*, *18*(3), 310–26. http://doi.org/10.1080/09658211003601530

Sörqvist, P., & Rönnberg, J. (2012). Episodic long-term memory of spoken discourse masked by speech: what is the role for working memory capacity? *Journal of Speech, Language and Hearing Research*, *55*, 210–218.

Speaks, C., & Jerger, J. (1965). Method for Measurement of Speech Identification. In *The Journal of the Acoustical Society of America* (Vol. 37, p. 1205). http://doi.org/10.1121/1.1939554

Steinmetzger, K., & Rosen, S. (2015). The role of periodicity in perceiving speech in quiet and in background noise. *The Journal of the Acoustical Society of America*, *138*(6), 3586–3599. http://doi.org/10.1121/1.4936945

Stone, M. A., Füllgrabe, C., & Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America*, *132*(1), 317–26. http://doi.org/10.1121/1.4725766

Stone, M. A., & Moore, B. C. J. (2014). On the near non-existence of "pure" energetic masking release for speech. *The Journal of the Acoustical Society of America*, *135*(4), 1967–1977. http://doi.org/10.1121/1.4868392

Sullivan, J. R., Osman, H., & Schafer, E. C. (2015). The Effect of Noise on the Relationship Between Auditory Working Memory and Comprehension in School-Age Children. *Journal of Speech, Language, and Hearing Research*, *58*(June), 1043–1051.

Trammell, J. L., & Speaks, C. (1970). On the distracting properties of competing speech. *Journal of Speech Language and Hearing Research*, *13*(2), 442. http://doi.org/10.1044/jshr.1302.442

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing Subject and Object Relative Clauses: Evidence from Eye Movements. *Journal of Memory and Language*, *47*(1), 69–90. http://doi.org/10.1006/jmla.2001.2836

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299. http://doi.org/10.1037/h0027242

Tun, P. A., Benichov, J., & Wingfield, A. (2010). Response latencies in auditory sentence comprehension: effects of linguistic versus perceptual challenge. *Psychology and Aging*, *25*(3), 730–5. http://doi.org/10.1037/a0019300

Uslar, V. N., Carroll, R., Hanke, M., Hamann, C., Ruigendijk, E., Brand, T., & Kollmeier, B. (2013). Development and evaluation of a linguistically and audiologically controlled sentence intelligibility test. *The Journal of the Acoustical Society of America*, *134*(4), 3039–56. http://doi.org/10.1121/1.4818760

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native-and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of …*, *121*(October 2005), 519–526.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science*, *21*(3), 384–90. http://doi.org/10.1177/0956797610361340

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *The Journal of the Acoustical Society of America*, *107*(3), 1671. http://doi.org/10.1121/1.428451

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: critique of Just and Carpenter (1992). *Psychological Review*, *103*(4), 761–72.

Watson, C., Kelly, W., & Wroton, H. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *The Journal of the Acoustical Society of America*, *60*(5), 1176–1186.

Wendt, D., Brand, T., & Kollmeier, B. (2014). An Eye-Tracking Paradigm for Analyzing the Processing Time of Sentences with Different Linguistic Complexities. *PloS One*, *9*(6), 1–13. http://doi.org/10.1371/journal.pone.0100186

Wendt, D., Dau, T., & Hjortkjær, J. (2016). Impact of Background Noise and Sentence Complexity on Processing Demands during Sentence Comprehension. *Frontiers in Psychology*, *7*(March), 1–12. http://doi.org/10.3389/fpsyg.2016.00345

Wendt, D., Kollmeier, B., & Brand, T. (2015). How Hearing Impairment Affects Sentence Comprehension: Using Eye Fixations to Investigate the Duration of Speech Processing. *Trends in Hearing*, *19*(0), 1–18. http://doi.org/10.1177/2331216515584149

Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, *118*(5), 3164. http://doi.org/10.1121/1.2082567

Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & Cox, C. L. (2006). Effects of Adult Aging and Hearing Loss on Comprehension of Rapid Speech Varying in Syntactic Complexity. *Journal of the American Academy of Audiology*, *17*(7), 487–497. http://doi.org/10.3766/jaaa.17.7.4

Wood, N., & Cowan, N. (1995). The Cocktail Party Phenomenon Revisited: How Frequent Are Attention Shifts to One's Name in an Irrelevant Auditory Channel? *Journal of Experimental Psychology*, *21*(1), 255–260.

Yampolsky, S., Waters, G. S., Caplan, D., Matthies, M., & Chiu, P. (2002). Effects of acoustic degradation on syntactic processing: implications for the nature of the resource system used in language processing. *Brain and Cognition*, *48*(2–3), 617–25. http://doi.org/10.1006/brcg.2001.1428

Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition*, *14*(3), 412–422. http://doi.org/10.1017/S1366728910000611

Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2013). Task difficulty differentially affects two measures of processing load: the pupil response during sentence processing and delayed cued recall of the sentences. *Journal of Speech, Language, and Hearing Research : JSLHR*, *56*(4), 1156–65. http://doi.org/10.1044/1092-4388(2012/12-0058)

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277–84. http://doi.org/10.1111/psyp.12151

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, *12*(5), 732–45. http://doi.org/10.1111/j.1467-7687.2009.00817.x

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2011). Noise on, voicing off: Speech perception deficits in children with specific language impairment. *Journal of Experimental Child Psychology*, *110*(3), 362–72. http://doi.org/10.1016/j.jecp.2011.05.001

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. http://doi.org/10.1016/j.neuron.2012.12.037