Electrophysiological markers of predictive coding in multisensory integration and autism spectrum disorder

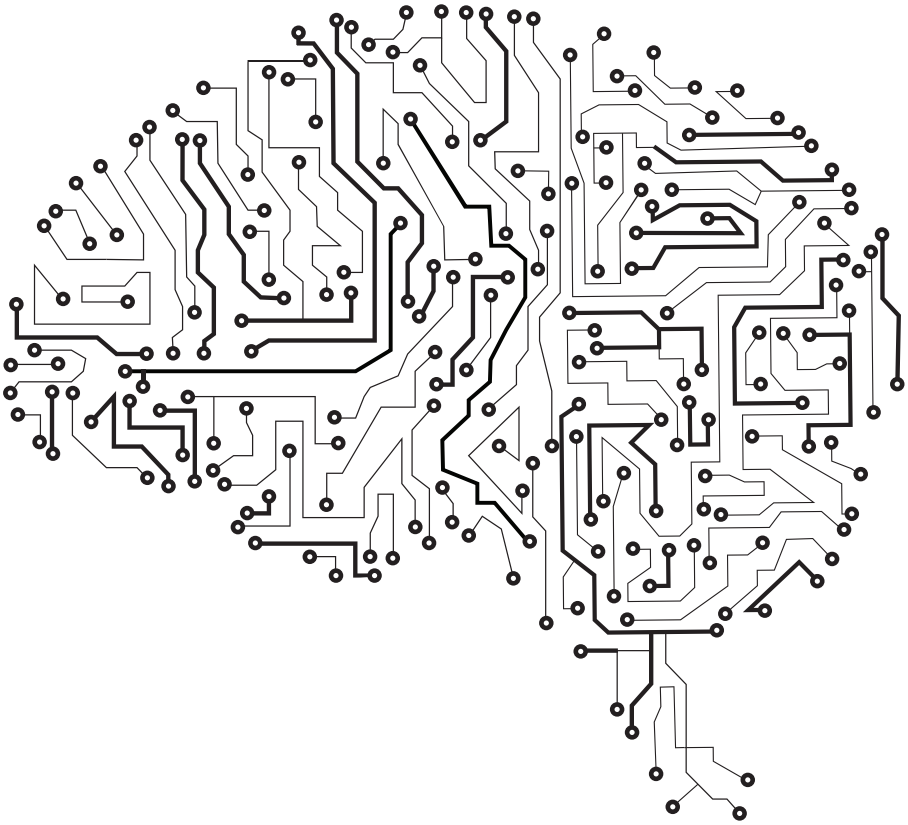
van Laarhoven, T.J.T.M.

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

ELECTROPHYSIOLOGICAL MARKERS OF PREDICTIVE CODING

IN MULTISENSORY INTEGRATION AND AUTISM SPECTRUM DISORDER



THIJS VAN LAARHOVEN

Electrophysiological markers of predictive coding in multisensory integration and autism spectrum disorder

Thijs van Laarhoven

Electrophysiological markers of predictive coding
in multisensory integration and autism spectrum disorder

Cover design and layout: Thijs van Laarhoven
Printing: Ridderprint, www.ridderprint.nl
ISBN: 978-94-6416-531-9

Electrophysiological markers of predictive coding in multisensory integration and autism spectrum disorder

Proefschrift

ter verkrijging van de graad van doctor
aan Tilburg University
op gezag van de rector magnificus
prof. dr. W.B.H.J. van de Donk

in het openbaar te verdedigen ten overstaan van een
door het college voor promoties aangewezen commissie
in de Aula van de Universiteit op vrijdag 23 april 2021 om 13.30 uur

door

Thijs Johannes Theodorus Maria van Laarhoven
geboren op 31 januari 1988
te Oudenbosch

Promotor
Prof. dr. J.H.M. Vroomen (Tilburg University)

Copromotor
Dr. J.J. Stekelenburg (Tilburg University)

Promotiecommissie
Prof. dr. H.M. Geurts (Universiteit van Amsterdam)
Prof. dr. D. Talsma (Ghent University)
Prof. dr. J. Wagemans (KU Leuven)
Prof. dr. M.G.J. Swerts (Tilburg University)

Table of contents

Chapter 1

General introduction

This introductory chapter outlines the main concepts that are essential to fully appreciate the scope of this dissertation. The concept of multisensory integration is briefly discussed first. Next, it is explained how sensory input and prior expectations shape our perception through predictive processing of sensory information, and how these predictive coding mechanisms can be examined with electrophysiological markers. The chapter continues with a brief outline of the alterations in sensory processing and multisensory integration associated with autism spectrum disorder, and describes how these alterations may be manifestations of an underlying impairment in predictive coding. The chapter concludes with an outline of the primary aims and research questions of this dissertation.

1.1 MAIN CONCEPTS

Multisensory integration

As we go about our daily lives, our brain is constantly exposed to sensory information. Despite the enormous variety and complexity of these sensory signals, the brain is somehow able to adequately process and contextualize all this information. While being in a crowded bar, for example, our brain receives a vast amount of input from numerous sources via our various sensory systems. Entering a bar might feel overwhelming at first, but after a short while, we grow accustomed to the changes in our sensory environment. Despite all the distractions and background noise, we are still able to understand what our friends are saying by 'tuning-in' to the sound of their voice and observing their visual articulatory movements. Without this lipread information, however, it is very difficult to have a conversation. Conversely, it is even harder to discriminate words, let alone sentences, while only viewing lip movements without listening to the speech sounds. Binding auditory and visual speech signals into unified percepts thus greatly enhances the ability to comprehend speech, especially under suboptimal listening conditions (Macleod & Summerfield, 1987; Sumby & Pollack, 1954). Combining information from multiple senses may lead to a variety of perceptual and behavioral benefits, including increased discrimination accuracy, faster reaction time, and lower stimulus detection thresholds (for review, see Stevenson, Ghose, et al., 2014). The perceptual and behavioral benefits of combining multimodal sensory information typically increase as unisensory signal quality degrades, a principle known as inverse effectiveness (Meredith & Stein, 1986). The process of combining unisensory information from various senses into unified percepts is commonly referred to as *multisensory integration* (Stein et al., 2010).

To benefit from multisensory integration, the brain must assess whether the incoming sensory information should be integrated into a unified percept. One of the most important cues indicating that multisensory input should be bound together is temporal proximity (Vroomen & Keetels, 2010). Signals occurring close in time are more likely to originate from the same sensory event, so integrating those signals is usually more efficient than processing each signal separately. Being able to perceive the relative timing of incoming sensory signals from multiple modalities is vital to properly integrate multisensory information. To be integrated into a single multisensory percept, sensory signals need to occur within a certain temporal proximity to each other, a construct commonly referred to as the *temporal binding window*.

Predictive coding in multisensory integration

Our perception not only relies on sensory input, but is also affected by prior experience and exposure to the outside world. The ability to integrate multisensory signals starts to develop during infancy (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017; Lewkowicz, Minar, Tift, & Brandon, 2015; Patterson & Werker, 2003), and continues to improve into late childhood (Ross et al., 2011; Tremblay et al., 2007). Prior experience may influence the extent to which multimodal information is integrated. For example, the temporal binding window for audiovisual sensory information is assumed to be symmetric at birth, but typically develops asymmetrically into adulthood (Hillock-Dunn & Wallace, 2012; Hillock, Powers, & Wallace, 2011; van Wassenhove, Grant, & Poeppel, 2007; Zampini, Guest, Shore, & Spence, 2005). Visual input usually arrives at the retina prior to auditory information reaches the cochlea as a result of the difference in speed between light and sound. Hence, audiovisual events in which the visual signal precedes the auditory signal (i.e. a natural situation) are more likely to be integrated than audiovisual events where the auditory signal precedes the visual signal (i.e. an unnatural situation). This asymmetry of the multisensory temporal binding window explains, for example, why lip-sync errors on television are usually more noticeable when the sound precedes the visual lip movements than vice versa.

Although our understanding of the behavioral and neural mechanisms involved in the integration of multisensory information has been greatly increased over the last few decades (for review, see Wallace, Woynaroski, & Stevenson, 2020), it is not yet fully understood how prior experience influences multisensory perception. A contemporary theoretical framework that describes the processing and integration of prior knowledge and sensory information, postulates that our brain continuously generates an internal predictive model of our environment based on sensory input and previous experiences (Friston, 2005). A key assumption of this theoretical framework based on empirical Bayes, commonly referred to as *predictive coding*, is that the brain continuously attempts to infer the probabilistic structure of sensory events. Predictive coding assumes that the brain processes new sensory experiences with a predictive model without specific prior expectations. Uninformative or, in Bayesian terms, 'flat' priors may bias perception towards sensory input. However, most of our sensory experiences share some similarities with prior experiences, so in reality, *our perception of the world is typically a combination of sensory input and prior expectations*.

The internal predictive model can be thought of as a probabilistic map that contains predictions about the current state of our environment that is used to contextualize and inform our perception (Baum, Stevenson, & Wallace, 2015; Lawson, Rees, & Friston, 2014). These predictions are assumed to be sent from higher to lower cortical areas (i.e. top-down), where they are contrasted with incoming sensory information. Any discrepancy between the sensory input and prior expectations results in an error signal that is sent from lower to higher cortical areas (i.e. bottom-up). *Prediction errors* inform our perception about unexpected or otherwise informative information, and indicate that our current internal prediction model needs to be adjusted and updated to minimize the occurrence of similar error signals in the future (Arnal & Giraud, 2012; Friston, 2005).

Predictive coding mechanisms allow us to anticipate upcoming sensory events, distinguish between expected and unexpected sensory events, and discern self-initiated from externally initiated sensory stimulation. Being able to predict the sensory consequences of upcoming sensory events significantly improves the efficiency of processing and contextualizing sensory information, and ensures that cognitive resources are primarily allocated to novel or otherwise newsworthy information. The predictive processing of sensory information enables us to 'make sense' of the world around us, and exhibit appropriate behavioral responses crucial for effective engagement and (social) interaction with our natural environment.

Implementations of predictive coding schemes have been discussed extensively at the theoretical level in both unisensory and multisensory processing and perception (see, e.g., Schröger, Marzecová, & Sanmiguel, 2015; Summerfield & Egner, 2009; Talsma, 2015; van Wassenhove, 2013). In short, it is assumed that the impact of prior expectations on perception increases as the effectiveness of sensory input decreases. The more ambiguous the sensory input, the more informative the prior expectations, and the more perception is affected by previous experience. Conversely, the more ambiguous or less informative the prior expectations, the more perception is driven by sensory input.

Electrophysiological markers of predictive coding in multisensory integration

A frequently applied approach to study predictive coding mechanisms in multisensory integration is by comparing electrophysiological brain responses evoked by predictable and unpredictable sounds. Auditory events can be made predictable in several ways, including stimulus repetition (Budd, Barry, Gordon, Rennie, & Michie, 1998), self-initiated motor actions (Baess, Horváth, Jacobsen, & Schröger, 2011; Baess, Jacobsen, & Schröger, 2008; Bendixen, SanMiguel, & Schröger, 2012; Martikainen, Kaneko, & Hari, 2005), lip-read speech (Besle, Fort, Delpuech, & Giard, 2004; Klucharev, Mottonen, Sams, Möttönen, & Sams, 2003; Van Wassenhove, Grant, & Poeppel, 2005), and anticipatory visual motion (Stekelenburg & Vroomen, 2007, 2012; Vroomen & Stekelenburg, 2010).

Numerous studies have shown that the auditory N1 component of the event-related potential (ERP) is typically suppressed and speeded up for predictable sounds, compared to the N1 evoked by unpredictable sounds with identical temporal and acoustic features (for review, see Bendixen et al., 2012). The amplitude and latency of the auditory N1 are assumed to be modulated by the precision of the internal predictive model (Arnal & Giraud, 2012; Friston, 2005). For predictable sounds, predictions are precise and incoming sounds likely match the prior expectations, so the auditory N1 is suppressed and often speeded up. For unpredictable sounds, prediction precision is low, and so the auditory N1 is not attenuated or speeded up. From a predictive coding perspective, suppression of the auditory N1 can be explained as an indication of the internal prediction model correctly anticipating the upcoming auditory stimulation. The *auditory N1 suppression effect* for sounds that are predictable by a self-initiated motor action or visual anticipatory motion is thus considered an early electrophysiological *marker of fulfilled auditory prediction* driven by multisensory integration (Bendixen et al., 2012).

Predictive coding mechanisms of fulfilled predictions are typically examined in paradigms designed to match the internal prediction (e.g. by self-initiation or anticipatory visual motion). A different approach to examine predictive coding mechanisms is to study the electrophysiological brain responses to *prediction violations*. An extreme case of prediction violation is a situation in which a highly anticipated stimulus is not occurring at all. Unexpected omissions of predictable sounds typically evoke an early negative omission response in the electroencephalography (EEG) during the period of silence where the sound was expected to be heard (SanMiguel, Saupe, & Schröger, 2013; SanMiguel, Widmann, Bendixen, Trujillo-Barreto, & Schröger, 2013; Stekelenburg & Vroomen, 2015). This early omission response resembles the auditory N1 that is typically evoked by regular auditory stimulation, and is therefore commonly referred to as the

auditory *omission N1* (oN1). The amplitude of the auditory oN1 is hypothesized to be modulated by the prediction and prediction error (Arnal & Giraud, 2012; Friston, 2005). For sounds that are highly predictable, precise auditory predictions are formed. Incoming stimuli that do not match (but violate) this precise prior, such as during unexpected auditory omissions, induce large prediction errors, and thus the oN1 is enlarged. If no clear predictions can be formed about an upcoming sound, the prediction is less likely to be violated, and so the oN1 is attenuated or absent during unexpected omissions of unpredictable sounds. Several studies have indeed shown that the oN1 is only elicited by unexpected omissions of predictable sounds, and not by omissions of unpredictable sounds or auditory omissions per se (SanMiguel, Widmann, et al., 2013; Stekelenburg & Vroomen, 2015). Hence, the *auditory oN1* evoked by unexpected omissions of sounds that are predictable by a motor act or anticipatory visual information is considered an early electrophysiological *marker of auditory prediction error* driven by multisensory integration (Bendixen et al., 2012).

While the auditory N1 suppression effect and auditory oN1 response are both early electrophysiological markers of predictive coding in multisensory integration in audition (Bendixen et al., 2012), it is not yet fully understood what characteristics of the anticipated sound drive these effects. Multisensory information and perception are shaped by prior expectations of the world. When we see someone moving their hands together, for example, we typically expect to hear the familiar sound of a handclap when the palms of their hands collide. The internal predictive model thus forms both *temporal* ('when') and *identity* ('what') predictions of upcoming sensory events. However, it is not yet fully understood whether predictive coding mechanisms in audition are primarily driven by the temporal characteristics, or by the identity features of anticipated sounds. Hence, the first aim of this dissertation is to examine how temporal and identity predictability affect the auditory N1 suppression effect and elicitation of the auditory oN1 response.

Multisensory integration in autism spectrum disorder

While the integration of multisensory information is generally considered an automatic process that is frequently utilized in everyday life, some individuals struggle to adequately acquire this fundamental ability. Of particular interest are autistic individuals. Autism spectrum disorder (ASD) is a pervasive neurodevelopmental disorder that affects between 1% and 2% of all children and adolescents, with an estimated worldwide population prevalence of 1% (Baxter et al., 2015; Lai, Lombardo, & Baron-Cohen, 2014; Lord et al., 2020; Maenner et al., 2020). ASD is characterized by deficits in social communication and social interaction and restricted, repetitive patterns of behavior, interests or activities (American Psychiatric Association, 2013; Robertson & Baron-Cohen, 2017), and has been linked to a range of perceptual alterations, including hypo- and hyperresponsiveness to sensory stimulation (Baranek et al., 2013; Robertson & Baron-Cohen, 2017).

Previous research has shown that autistic individuals have alterations in multisensory integration (Baum et al., 2015; Brandwein et al., 2015; Feldman et al., 2018; Russo et al., 2010; Stevenson, Siemann, et al., 2014). For example, autistic individuals tend to benefit less from visual articulatory cues when listening to speech in background noise, indicating that the integration of audiovisual speech signals may be impaired in ASD (Foxe et al., 2015; Stevenson, Baum, et al., 2017). The neural mechanisms underlying these impairments in multisensory integration are not yet fully understood.

As a spectrum disorder, symptoms of ASD are found in varying degrees in the general population (Ruzich et al., 2015). However, the impact of autistic symptoms in the general population on sensory processing is still relatively understudied. There is some evidence that sub-clinical levels of autistic traits may be related to alterations in multisensory integration of artificial audiovisual stimuli (Donohue, Darling, & Mitroff, 2012; Stevenson, Toulmin, et al., 2017; Ujiie, Asai, & Wakabayashi, 2015), but it is unclear whether these findings extent to more ecologically valid stimuli such as audiovisual speech. Hence, the second aim of this dissertation was to examine the relationship between sub-clinical levels of autistic traits and audiovisual speech processing in a large non-clinical population.

Autism as a disorder of predictive coding

Recently, it has been proposed that the alterations in sensory processing and multisensory integration associated with ASD may be manifestations of a decreased ability to anticipate upcoming sensory stimulation (Lawson et al., 2014; Pellicano & Burr, 2012; Sinha et al., 2014; van de Cruys et al., 2014).

If predictive coding of sensory information is indeed altered in autistic individuals, perception in ASD could be less affected by prior expectations and more driven by sensory input. A predictive model that is biased toward sensory input may generate overly precise predictions that are highly effective, but only in specific contexts. Overweighing sensory input over prior experience might be beneficial in contexts that require increased attention to detail, such as discrimination or identification tasks. Indeed, autistic individuals often 'see the trees, but not the forest', and seem more attuned to perceptual details than global percepts (Robertson & Baron-Cohen, 2017). However, prior expectations that are overly precise may not generalize well to other sensory experiences. This "overfitting" of predictions may in turn lead to an increased demand on cognitive resources required for sensory processing, since every sensory experience is seemingly processed afresh, rather than mediated by prior experience.

Failing to efficiently contextualize and generalize sensory information not only affects sensory processing and perception, but may also impact the ability to interact with the natural environment. Without efficient internal predictive coding mechanisms to process sensory information, interaction with the environment becomes increasingly challenging in social situations, which are inherently ambiguous and unpredictable and thus require the adequate processing of prior expectations and sensory input. Understanding the neural basis of the potential impairments in predictive coding in ASD may thus very well be a fundamental part of the explanation of why autistic individuals are more quickly overwhelmed by sensory stimulation, and often struggle with social communication and interaction with their environment.

There is some empirical support for the impaired predictive coding account of altered multisensory processing and perception in ASD. However, most of the evidence is based on behavioral observations or retrospective analysis of neuroimaging data (Sinha et al., 2014). Since the emergence of the predictive coding account of autistic symptomatology, only a few studies have explicitly examined whether predictive coding of sensory information is altered in ASD. At the behavioral level, ASD has been linked to reduced adaptation to stimulus loudness and little or no rapid recalibration to audiovisual

asynchrony (Lawson, Aylward, White, & Rees, 2015; Turi, Karaminis, Pellicano, & Burr, 2016), suggesting that autistic individuals may experience difficulties in inferring the probabilistic structure of sensory events. At the neural level, there is some evidence for reduced sensory suppression in ASD for repeated auditory stimulation (Kolesnik et al., 2019; Seery, Tager-Flusberg, & Nelson, 2014), which suggests that perception in ASD may be less affected by prior experience. Neuroimaging studies in the visual domain have yielded mixed results. While some report reduced suppression effects in visual cortex for repeated visual stimulation (Ewbank et al., 2017), and decreased influence of prior expectations on visual processing of familiar everyday objects in autistic individuals (Sapey-Triomphe et al., 2020), others have reported typical repetition suppression responses in autistic individuals (Utzerath, Schmits, Buitelaar, & de Lange, 2018), and suggest that prior knowledge for perceptual inference in visual cortex is preserved in ASD (Utzerath, Schmits, Kok, Buitelaar, & de Lange, 2019; van de Cruys, Vanmarcke, Van de Put, & Wagemans, 2018).

To date, it is unclear whether predictive coding is indeed altered in ASD. Furthermore, while numerous studies suggest that multisensory integration is impaired in ASD at the behavioral level, particularly in the audiovisual domain (Baum et al., 2015; Brandwein et al., 2015; Feldman et al., 2018; Russo et al., 2010; Stevenson, Siemann, et al., 2014), it is yet to be examined whether these alterations are reflected in the neural correlates of predictive coding in multisensory integration in autistic individuals. The third aim of this dissertation was to address this issue by comparing the electrophysiological markers of predictive coding in auditory prediction by self-initiated motor action and visual anticipatory motion between autistic individuals and age-matched individuals with typical development.

1.2 AIMS AND RESEARCH QUESTIONS

This dissertation has three primary aims: (1) to examine the impact of temporal and identity predictability on electrophysiological markers of predictive coding in multisensory integration in the typical brain, (2) to examine the extent to which sub-clinical levels of autistic symptoms in the general population are related to alterations in multisensory integration, and (3) to examine whether the neural correlates of predictive coding in multisensory integration are altered in autism spectrum disorder.

The aims of this dissertation are discussed by answering the following research questions:

1. Are the electrophysiological markers of fulfilled prediction (i.e. the N1 suppression effect) and prediction error (i.e. the oN1 response) in auditory prediction by vision affected by temporal and identity predictability of the anticipated sound? (Chapters 2 and 3)
2. Are increased sub-clinical levels of autistic traits associated with alterations in multisensory integration of audiovisual speech? (Chapter 4)
3. Are the electrophysiological markers of fulfilled motor-auditory prediction (i.e. the N1 suppression effect) and visual-auditory prediction error (i.e. the oN1 response) altered in autistic individuals? (Chapter 5 and 6)

1.3 THESIS OUTLINE

This dissertation consists of 7 chapters. The present chapter has outlined the main concepts that are essential to this dissertation. Chapters 2 and 3 examine how temporal and identity predictability affect the auditory N1 suppression effect and elicitation of the auditory oN1 response in ecologically valid audiovisual events. Chapter 4 examines the relationship between sub-clinical levels of autistic traits and audiovisual speech processing in a large non-clinical population using a battery of experimental tasks assessing audiovisual perceptual binding, visual enhancement of speech embedded in noise, and audiovisual temporal processing. In chapters 5 and 6, the electrophysiological markers of fulfilled motor-auditory prediction and visual-auditory prediction error are examined in autistic individuals, and compared with age-matched individuals with typical development. Chapter 7 concludes the dissertation with a summary and general discussion of the main findings, suggestions for future research, and overall conclusions.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology*, *48*(9), 1276–1283. https://doi.org/10.1111/j.1469-8986.2011.01196.x

Baess, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology*, *70*(2), 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Baranek, G. T., Watson, L. R., Boyd, B. A., Poe, M. D., David, F. J., & McGuire, L. (2013). Hyporesponsiveness to social and nonsocial sensory stimuli in children with autism, children with developmental delays, and typically developing children. *Development and Psychopathology*, *25*(2), 307–320. https://doi.org/10.1017/S0954579412001071

Baum, S. H., Stevenson, R. A., & Wallace, M. T. (2015). Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Progress in Neurobiology*, *134*, 140–160. https://doi.org/10.1016/j.pneurobio.2015.09.007

Baxter, A. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine, 45*(3), 601-613. https://doi.org/10.1017/S003329171400172X

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology*, *83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*(8), 2225–2234. https://doi.org/10.1111/j.1460-9568.2004.03670.x

Brandwein, A. B., Foxe, J. J., Butler, J. S., Frey, H. P., Bates, J. C., Shulman, L. H., & Molholm, S. (2015). Neurophysiological Indices of Atypical Auditory Processing and Multisensory Integration are Associated with Symptom Severity in Autism. *Journal of Autism and Developmental Disorders*, *45*(1), 230–244. https://doi.org/10.1007/s10803-014-2212-9

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology, 31*(1), 51–68. https://doi.org/10.1016/S0167-8760(98)00040-3

Donohue, S. E., Darling, E. F., & Mitroff, S. R. (2012). Links between multisensory processing and autism. *Experimental Brain Research*, *222*(4), 377–387. https://doi.org/10.1007/s00221-012-3223-4

Ewbank, M. P., Pell, P. J., Powell, T. E., Von Dem Hagen, E. A. H., Baron-Cohen, S., & Calder, A. J. (2017). Repetition suppression and memory for faces is reduced in adults with autism spectrum conditions. *Cerebral Cortex*, *27*(1), 92–103. https://doi.org/10.1093/cercor/bhw373

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *95*, 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H. P., Russo, N. N., Blanco, D., … Ross, L. A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex*, *25*(2), 298–312. https://doi.org/10.1093/cercor/bht213

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, *20*(3). https://doi.org/10.1111/desc.12381

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*, *15*(5), 688–696. https://doi.org/10.1111/j.1467-7687.2012.01171.x

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia*, *49*(3), 461–467. https://doi.org/10.1016/j.neuropsychologia.2010.11.041

Klucharev, V., Mottonen, R., Sams, M., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, *18*(1), 65–75. https://doi.org/10.1016/j.cogbrainres.2003.09.004

Kolesnik, A., Begum Ali, J., Gliga, T., Guiraud, J., Charman, T., Johnson, M. H., & Jones, E. J. H. (2019). Increased cortical reactivity to repeated tones at 8 months in infants with later ASD. *Translational Psychiatry, 9*(46). https://doi.org/10.1038/s41398-019-0393-x

Lai, M. C., Lombardo, M. V., & Baron-Cohen, S. (2014). Autism. *Lancet, 383*(9920), 896–910. https://doi.org/10.1016/S0140-6736(13)61539-1

Lawson, R. P., Aylward, J., White, S., & Rees, G. (2015). A striking reduction of simple loudness adaptation in autism. *Scientific Reports*, *5*, 16157. https://doi.org/10.1038/srep16157

Lawson, R. P., Rees, G., & Friston, K. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, *8*, 302. https://doi.org/10.3389/fnhum.2014.00302

Lewkowicz, D. J., Minar, N. J., Tift, A. H., & Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology*, *130*, 147–162. https://doi.org/10.1016/j.jecp.2014.10.006

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., … Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, *6*(1). https://doi.org/10.1038/s41572-019-0138-4

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131–141. https://doi.org/10.3109/03005368709077786

Maenner, M., Shaw, K., Baio, J., Washington, A., Patrick, M., DiRienzo, M., … Dietz, P. (2020). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR. Surveillance Summaries*, *69*, 1–12. https://doi.org/10.15585/mmwr.ss6904a1

Martikainen, M. H., Kaneko, K. I., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex*, *15*(3), 299–302. https://doi.org/10.1093/cercor/bhh131

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56*(3), 640-662. https://doi.org/10.1152/jn.1986.56.3.640

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196. https://doi.org/10.1111/1467-7687.00271

Pellicano, E., & Burr, D. (2012). When the world becomes “too real”: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, *18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, *33*(12), 2329–2337. https://doi.org/10.1111/j.1460-9568.2011.07685.x

Russo, N., Foxe, J. J., Brandwein, A. B., Altschuler, T., Gomes, H., & Molholm, S. (2010). Multisensory processing in children with autism: High-density electrical mapping of auditory-somatosensory integration. *Autism Research*, *3*(5), 253–267. https://doi.org/10.1002/aur.152

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism*, *6*(1), 1–12. https://doi.org/10.1186/2040-2392-6-2

SanMiguel, I., Saupe, K., & Schröger, E. (2013). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted ”what” but not ”when”. *Frontiers in Human Neuroscience*, *7*, 407. https://doi.org/10.3389/fnhum.2013.00407

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schröger, E. (2013). Hearing silences: Human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience*, *33*(20), 8633–8639. https://doi.org/10.1523/JNEUROSCI.5821-12.2013

Sapey-Triomphe, L. A., Boets, B., Van Eylen, L., Noens, I., Sunaert, S., Steyaert, J., & Wagemans, J. (2020). Ventral stream hierarchy underlying perceptual organization in adolescents with autism. *NeuroImage: Clinical*, *25,* 102197. https://doi.org/10.1016/j.nicl.2020.102197

Schröger, E., Marzecová, A., & Sanmiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*(5), 641–664. https://doi.org/10.1111/ejn.12816

Seery, A., Tager-Flusberg, H., & Nelson, C. A. (2014). Event-related potentials to repeated speech in 9-month-old infants at risk for autism spectrum disorder. *Journal of Neurodevelopmental Disorders, 6*(1), 43. https://doi.org/10.1186/1866-1955-6-43

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., … Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15220–15225. https://doi.org/10.1073/pnas.1416797111

Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Alex Meredith, M., Perrault, T. J., … Lewkowicz, D. J. (2010). Semantic confusion regarding the development of multisensory integration: A practical solution. *European Journal of Neuroscience, 31*(10), 1713–1720. https://doi.org/10.1111/j.1460-9568.2010.07206.x

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*(12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Stekelenburg, J. J., & Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Frontiers in Integrative Neuroscience*, *6*, 26. https://doi.org/10.3389/fnint.2012.00026

Stekelenburg, J. J., & Vroomen, J. (2015). Predictive coding of visual-auditory and motor-auditory events: An electrophysiological study. *Brain Research*, *1626*, 88–96. https://doi.org/10.1016/j.brainres.2015.01.036

Stevenson, R. A., Baum, S. H., Segers, M., Ferber, S., Barense, M. D., & Wallace, M. T. (2017). Multisensory speech perception in autism spectrum disorder: From phoneme to whole-word perception. *Autism Research*, *10*(7), 1280–1290. https://doi.org/10.1002/aur.1776

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., … Wallace, M. T. (2014). Identifying and Quantifying Multisensory Integration: A Tutorial Review. *Brain Topography, 27*. https://doi.org/10.1007/s10548-014-0365-7

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014). Evidence for Diminished Multisensory Integration in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *44*(12), 3161–3167. https://doi.org/10.1007/s10803-014-2179-6

Stevenson, R. A., Toulmin, J. K., Youm, A., Besney, R. M. A., Schulz, S. E., Barense, M. D., & Ferber, S. (2017). Increases in the autistic trait of attention to detail are associated with decreased multisensory temporal adaptation. *Scientific Reports*, *7*(1), 14354. https://doi.org/10.1038/s41598-017-14632-1

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9). https://doi.org/10.1016/j.tics.2009.06.003

Talsma, D. (2015). Predictive coding and multisensory integration: An attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, 9*,* 19 https://doi.org/10.3389/fnint.2015.00019

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and non-speech audio-visual illusions: A developmental study. *PLoS ONE*, *2*(8). https://doi.org/10.1371/journal.pone.0000742

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific Reports*, *6*, 21756. https://doi.org/10.1038/srep21756

Ujiie, Y., Asai, T., & Wakabayashi, A. (2015). The relationship between level of autistic traits and local bias in the context of the McGurk effect. *Frontiers in Psychology*, *6*, 891. https://doi.org/10.3389/fpsyg.2015.00891

Utzerath, C., Schmits, I. C., Buitelaar, J., & de Lange, F. P. (2018). Adolescents with autism show typical fMRI repetition suppression, but atypical surprise response. *Cortex*, *109*, 25–34. https://doi.org/10.1016/j.cortex.2018.08.019

Utzerath, C., Schmits, I. C., Kok, P., Buitelaar, J., & de Lange, F. P. (2019). No evidence for altered up- and downregulation of brain activity in visual cortex during illusory shape perception in autism. *Cortex*, *117*, 247–256. https://doi.org/10.1016/j.cortex.2019.03.011

van de Cruys, S., Evers, K., van der Hallen, R., van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

van de Cruys, S., Vanmarcke, S., Van de Put, I., & Wagemans, J. (2018). The Use of Prior Knowledge for Perceptual Inference Is Preserved in ASD. *Clinical Psychological Science*, *6*(3), 382–393. https://doi.org/10.1177/2167702617740955

van Wassenhove, V. (2013). Speech through ears and eyes: Interfacing the senses with the supramodal brain. *Frontiers in Psychology, 4,* 388. https://doi.org/10.3389/fpsyg.2013.00388

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, and Psychophysics*, *72*(4), 871–884. https://doi.org/10.3758/APP.72.4.871

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*, *22*(7), 1583–1596. https://doi.org/10.1162/jocn.2009.21308

Wallace, M. T., Woynaroski, T. G., & Stevenson, R. A. (2020). Multisensory Integration as a Window into Orderly and Disrupted Cognition and Communication. *Annual Review of Psychology*, *71*(1), 193–219. https://doi.org/10.1146/annurev-psych-010419-051112

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception and Psychophysics*, *67*(3), 531–544. https://doi.org/10.3758/BF03193329

Chapter 2

Suppression of the auditory N1 by visual anticipatory motion is modulated by temporal and identity predictability

This chapter is published in a slightly modified version as:

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2020). Suppression of the auditory N1 by visual anticipatory motion is modulated by temporal and identity predictability. *Psychophysiology, 58*(3), e13749. https://doi.org/10.1111/psyp.13749

ABSTRACT

The amplitude of the auditory N1 component of the event-related potential (ERP) is typically suppressed when a sound is accompanied by visual anticipatory information that reliably predicts the timing and identity of the sound. While this visually-induced suppression of the auditory N1 is considered an early electrophysiological marker of fulfilled prediction, it is not yet fully understood whether this internal *predictive coding* mechanism is primarily driven by the *temporal* characteristics, or by the *identity* features of the anticipated sound. The current study examined the impact of temporal and identity predictability on suppression of the auditory N1 by visual anticipatory motion in an ecologically valid audiovisual event (a video of a handclap). Predictability of auditory timing and identity was manipulated in three different conditions in which sounds were either played in isolation, or in conjunction with a video that either reliably predicted the timing of the sound, the identity of the sound, or both the timing and identity. The results showed that N1 suppression was *largest* when the video reliably predicted both the timing and identity of the sound, and *reduced* when either the timing or identity of the sound was unpredictable. The current results indicate that predictions of timing and identity are both essential elements for predictive coding in audition.

INTRODUCTION

Our brain is constantly exposed to sensory information that must be adequately processed and contextualized in order to facilitate appropriate responses and interactions with our environment. Being able to predict the sensory consequences of our own actions and those of others greatly improves the efficiency of this process, and enables us to allocate cognitive resources to novel or otherwise newsworthy information. A contemporary theoretical framework that describes the *predictive coding* of sensory information, postulates that our brain continuously generates an internal predictive model of our environment based on previous experiences (Friston, 2005). This internal model enables the brain to form both *temporal* ('when') and *identity* ('what') predictions of upcoming sensory events (Arnal & Giraud, 2012).

A frequently applied approach to study predictive coding mechanisms is by comparing electrophysiological brain responses evoked by predictable and unpredictable sounds. Numerous studies have shown that the auditory N1 is typically suppressed and speeded up for sounds that are initiated by motor actions (e.g. a key-press), compared to the N1 evoked by sounds with identical temporal and acoustic features that are triggered externally (for review, see Bendixen, SanMiguel, & Schröger, 2012). N1 suppression effects have also been reported in the visual-auditory domain. Lip-read speech, for example, consistently suppresses and often speeds up the N1 (for review, see Baart, 2016). Other studies have shown that the N1 induced by sounds that are accompanied by anticipatory visual motion (e.g. seeing someone performing a handclap) is typically suppressed and speeded up compared to the same sounds played in isolation (Stekelenburg & Vroomen, 2007, 2012; Vroomen & Stekelenburg, 2010).

The amplitude and latency of the auditory N1 is assumed to be modulated by the precision of our internal predictive model (Arnal & Giraud, 2012; Friston, 2005). When precision is high, an incoming sound likely matches the prediction, and the auditory N1 is attenuated and often – but not always (Baart, 2016) – speeded up. For unpredictable sounds, the precision of the predictive model is low, and so the auditory N1 is not suppressed or speeded up. Given that the timing and identity of sounds that are accompanied by visual anticipatory information are usually quite predictable, the precision of the internal prediction model is typically higher for such sounds than sounds played in isolation or externally-initiated sounds. From a predictive coding perspective, the N1 suppression effect for sounds that are predictable by visual motion can thus be explained as an indication of the internal prediction model correctly anticipating the upcoming auditory stimulation.

While the available data agree that the N1 suppression effect is an early electrophysiological marker of fulfilled prediction (for review, see Bendixen et al., 2012), it is not yet fully understood whether predictive coding mechanisms in audition are primarily driven by the *temporal* characteristics, or by the *identity* features of the anticipated sound. In the visual-auditory domain, there is some evidence that N1 suppression by visual motion may be less affected by identity prediction, and more reliant on temporal prediction (Huhn, Szirtes, Lorincz, & Csépe, 2009; Klucharev, Mottonen, Sams, Möttönen, & Sams, 2003; Stekelenburg & Vroomen, 2007; Van Wassenhove, Grant, & Poeppel, 2005; Vroomen & Stekelenburg, 2010). Several studies have shown that synchronous presentation of speech sounds and visual articulatory movements (i.e. lip-read speech) suppresses the auditory N1, even if the visual information is ambiguous or incongruent with the speech sounds (Klucharev et al., 2003; Stekelenburg & Vroomen, 2007; Van Wassenhove et al., 2005), while suppression of the N1 is reduced if visual articulatory movements and speech sounds are presented asynchronous (Huhn et al., 2009). Similarly, N1 suppression in artificial audiovisual events is reduced when the timing of the auditory signal, relative to the visual signal, is inconsistent (Vroomen & Stekelenburg, 2010). Another study on N1 suppression by visual motion in ecological valid audiovisual stimuli has shown that N1 suppression was not affected by audiovisual congruency (i.e. the N1 was similarly suppressed by a video of a handclap paired with the actual sound of a handclap, as by a video of the handclap paired with the sound of a spoon tapping on a cup), but only occurs when the visual motion precedes the auditory signal and reliably predicts the timing (and not necessarily the identity) of the anticipated sound (Stekelenburg & Vroomen, 2007). These findings suggest that visually-induced N1 suppression appears to be mostly driven by temporal prediction. However, in the studies that examined the impact of identity prediction on visually-induced N1 suppression through manipulation of audiovisual congruency, only two (Stekelenburg & Vroomen, 2007; Van Wassenhove et al., 2005) or four (Klucharev et al., 2003) incongruent audiovisual stimulus pairings were included that were repeated several times. Hence, participants may have learned to expect these few incongruent pairings and may have incorporated both the incongruent and congruent stimulus pairings in their internal predictive model - which in turn may have led to an overall suppression of the auditory N1 induced by both congruent and incongruent audiovisual stimulus pairings. Furthermore, temporal prediction was not manipulated in these studies, so the impact of temporal prediction on N1 suppression by visual motion in ecologically valid audiovisual events is yet to be examined.

Previous studies examining the impact of temporal and identity predictions on N1 suppression induced by visual motion either manipulated audiovisual timing in speech or artificial stimuli (Huhn et al., 2009; Vroomen & Stekelenburg, 2010), or audiovisual congruency in speech and ecologically valid stimuli (Klucharev et al., 2003; Stekelenburg & Vroomen, 2007; Van Wassenhove et al., 2005). To our knowledge, the impact of both temporal and identity prediction on N1 suppression by visual motion has not been formally investigated in ecologically valid audiovisual events.

The visually-induced N1 suppression effect is often accompanied by a suppression of the P2 (for review, see Baart, 2016). Although the impact of temporal and identity prediction on P2 suppression is still unclear, some studies suggest that P2 suppression by visual motion may be less affected by temporal predictability, and more driven by identity prediction (Stekelenburg & Vroomen, 2007; Vroomen & Stekelenburg, 2010). In a previous study using artificial audiovisual stimuli, no effect of temporal predictability on P2 suppression by visual motion was found (Vroomen & Stekelenburg, 2010), whereas a study using ecologically valid audiovisual stimuli suggests that P2 suppression may be modulated by audiovisual congruency (Stekelenburg & Vroomen, 2007).

The current study examined the impact of temporal and identity predictability on suppression of the auditory N1 and P2 by visual anticipatory motion in an ecologically valid audiovisual stimulus. Predictability of auditory onset (relative to visual onset) and sound identity was manipulated in three conditions (Table 2.1). The first condition was adapted from a previous study showing robust and consistent N1-P2 suppression effects (Stekelenburg & Vroomen, 2007). In this natural condition, a video of a handclap was presented synchronously with the sound of the actual handclap. In the other two conditions, either the onset (random-timing condition) or the identity (random-identity condition) of the sound was unpredictable. In the random-timing condition, the sound and video of the handclap were always presented asynchronous. The magnitude of asynchrony varied on a trial-to-trial basis in order to prevent adaptation to temporal asynchrony (Vroomen, Keetels, De Gelder, & Bertelson, 2004). In the random-identity condition, the sound was randomly selected out of a pool of 100 stimuli on a trial-to-trial basis - thereby rendering the video an unreliable predictor for sound identity, while sound onset was always synced to the video. We tested for the presence of visually-induced N1 and P2 suppression effects by presenting randomly intermixed audiovisual (AV), auditory-only (A), and visual-only (V) trials in each condition (natural, random-timing, random-identity). In accord with previous research on early electrophysiological correlates of audiovisual interactions (Besle, Fort, & Giard, 2004; Teder-Salejarvi, McDonald, Di Russo,

& Hillyard, 2002), visual activity was eliminated from ERPs evoked by AV trials by subtracting ERPs evoked by V trials. The additive model (A = AV–V) assumes that the neural activity evoked by AV stimuli is equal to the sum of activities of A and V if the unimodal signals are processed independently. This assumption is valid for extracellular media and is based on the law of superposition of electric fields (Barth, Goldberg, Brett, & Di, 1995). Differences in amplitude and latency of the N1 and P2 between the A and AV–V modality were interpreted as neural correlates of visual-auditory predictive coding.

Suppression of the N1 induced by visual anticipatory motion was expected to be most pronounced in the natural condition, where the video reliably predicted both the timing and identity of the sound. Based on the previous finding that temporal predictability is important for N1 suppression in artificial audiovisual events (Vroomen & Stekelenburg, 2010), suppression of the auditory N1 was expected to be reduced in the random-timing condition. Assuming that identity of the sound is also of importance in the visual-auditory domain, we expected that N1 suppression was also reduced in the random-identity condition. Although the impact of temporal and identity predictability on visually-induced P2 suppression is relatively under-examined, there is some evidence for increased P2 suppression for incongruent audiovisual stimuli (Stekelenburg & Vroomen, 2007). Hence, suppression of the P2 was expected to be increased in the random-identity condition compared to the natural and random-timing condition.

METHODS

All experimental procedures were approved by the Ethics Review Board of the School of Social and Behavioral Sciences of Tilburg University (EC-2016.48), and conducted in accordance with the ethical standards of the Declaration of Helsinki.

Participants

Twenty-nine undergraduate students from Tilburg University participated in this study (23 female, mean age 19.72 years, *SD* = 1.74). Written informed consent was obtained from each participant prior to participation. All participants reported normal hearing and normal or corrected-to-normal vision. None were diagnosed with a neurological disorder and none reported use of medication. All participants were reimbursed with course credits as part of a curricular requirement.

Stimuli

Stimulus materials were adapted from a previous study showing robust and consistent N1 suppression effects (Stekelenburg & Vroomen, 2007). Visual stimuli consisted of a

video recording portraying the visual motion of a single handclap on a black background (Figure 2.1). The video started with a 200 ms fade-in followed by a still image with a randomly jittered duration from 200 to 800 ms showing the hands separated. Subsequently, the hands moved to each other and struck together 500 ms after motion onset. After impact, the hands returned to their original starting position and a 200 ms fade-out was shown. The intertrial interval (ITI) was randomly jittered from 250 to 1750 ms, during which a black screen was displayed. The video was presented on a 19-inch CRT monitor (Iiyama Vision Master Pro 454) at a frame rate of 25 frames/s, a refresh rate of 100 Hz, a resolution of 640 × 480 pixels (14° horizontal and 12° vertical visual angle), and at a viewing distance of approximately 70 cm. Auditory stimuli consisted of an audio recording of the handclap portrayed in the video, and audio recordings of 100 different environmental sounds (e.g. doorbell, dog bark, car horn) adapted from a previous study (Otte et al., 2013). All sounds were recorded at a sampling rate of 44.1 kHz with a duration of 200 ms duration (including 10 ms rise and fall times), and with matched root-mean-square (RMS) amplitudes. Sounds were presented over JAMO S100 stereo speakers, located directly on the left and right side of the monitor, at approximately 61 dB(A) sound pressure level. Stimulus presentation was controlled using E-Prime 1.2 (Psychology Software Tools Inc., Sharpsburg, PA, USA).

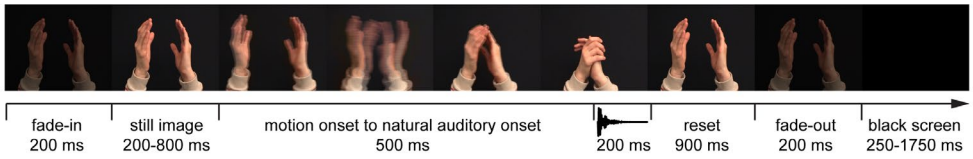


Figure 2.1. Time-course of the video presented in audiovisual and visual trials.

Procedure

Participants were individually tested in a sound attenuated and dimly lit room, and were seated in front of the monitor positioned at eye-level. They were instructed to carefully listen to the presented sounds and to maintain their gaze on the center of the screen. Gaze direction was monitored through CCTV.

Three experimental conditions were included and presented in separate blocks: a *natural* condition, a *random-timing* condition, and a *random-identity* condition (Table 2.1). Three trial types were included in each block: audiovisual (AV), auditory (A), and visual (V). During AV trials in the natural condition, the video of a handclap was presented synchronously with the sound of the actual handclap. During A and V trials in the natural condition, only the sound or video of the handclap was presented, respectively. During AV

trials in the random-timing condition, the sound of the handclap could either precede or follow the visual collision of the two hands at an unpredictable stimulus onset asynchrony (SOA). Based on the results of a simultaneity judgment (SJ) task from a previous study using the same stimuli (van Laarhoven, Stekelenburg, & Vroomen, 2017), SOAs of -250, -230, -210, -190, -170, 210, 240, 260, 290 and 320 were included (all values in ms, negative and positive values indicate sound leading and lagging the onset of the sound in the natural condition, respectively) to ensure that the timing of the sound relative to the video was highly unpredictable and never perceived as synchronous. During A trials in the random-timing condition, the sound of the handclap was presented at the same unpredictable variable onset, but without the video. V trials in the random-timing condition were identical to V trials in the natural condition. During AV trials in the random-identity condition, the video of the handclap was presented synchronously with an environmental sound that was randomly selected out of the pool of 100 sounds. During A trials in the random-identity condition, the same randomly selected environmental sounds were presented. V trials in the random-identity condition were identical to V trials in the natural and random-timing condition.

Table 2.1. Experimental conditions

Condition	Sound timing	Sound identity
Natural	Synchronized with video	Handclap
Random-timing	Random[b]	Handclap
Random-identity	Synchronized with video	Random[a]

[a] The identity of the sound was randomly selected in every trial out of 100 different environmental sounds (e.g. doorbell, dog bark, car horn) with equal rise and fall times, equal length and matched amplitudes
[b] The sound could either precede or follow the visual collision moment of the two hands at a randomly selected SOA of -250, -230, -210, -190, -170, 210, 240, 260, 290, or 320 (all values in ms, negative and positive values indicate sound leading and following the natural synchrony point, respectively)

For each condition, a total of 270 randomly intermixed AV, A, and V trials (90 for each modality) were presented across three blocks of 90 trials. An intermixed design - as opposed to a blocked design, in which each modality is presented in a separate block - was implemented to limit the impact of potential attentional demand differences between modalities on the N1 response (Horváth & Winkler, 2010; Lange, Rösler, & Röder, 2003). Block order was quasi-randomized across participants with the restriction that natural, random-timing, and random-identity blocks were never repeated successively.

EEG acquisition and processing

The EEG was sampled at 512 Hz from 64 locations using active Ag-AgCl electrodes (BioSemi, Amsterdam, the Netherlands) mounted in an elastic cap and two mastoid electrodes. Electrodes were placed in accordance with the extended International 10-20 system. Two additional electrodes served as reference (Common Mode Sense active electrode) and ground (Driven Right Leg passive electrode). Horizontal electrooculogram (EOG) was recorded using two electrodes placed at the outer canthi of the left and right eye. Vertical EOG was recorded from two electrodes placed above and below the right eye. BrainVision Analyzer 2.0 (Brain Products, Gilching, Germany) was used for ERP analyses. EEG was referenced offline to an average of left and right mastoids and band-pass filtered (0.01-30 Hz, 24 dB/octave). The (residual) 50 Hz interference was removed by a 50-Hz notch filter. Raw data were segmented into epochs of 1000 ms, including a 200-ms pre-stimulus baseline period. Epochs were time-locked to the sound onset in the AV and A trials, and to the corresponding timestamp in the V trials separately for all conditions. After EOG correction (Gratton, Coles, & Donchin, 1983), epochs with an amplitude change exceeding ±150 μV at any EEG channel were rejected and subsequently averaged and baseline corrected for each condition separately. On average 7.75 percent (*SD* = 7.24) of the trials were rejected. There were no significant differences in rejected trials between conditions or trial types (natural: AV 8.00%, A 8.25%, V 7.72%; random-timing: AV 7.59%, A 9.27%, V 7.66%; random-identity AV 5.71%, A 7.28%, V 6.51%). To facilitate a direct comparison between the AV and A trials within and between each condition, V ERPs were subtracted from the AV ERPs for each condition to nullify the contribution of visual activity (Besle et al., 2004; Stekelenburg & Vroomen, 2007; Teder-Salejarvi et al., 2002; Vroomen & Stekelenburg, 2010).

RESULTS

Time windows and regions of interest

The auditory (A) and audiovisual minus visual (AV–V) ERPs showed clearly identifiable N1 and P2 responses in all three conditions (Figure 2.2). Statistical analysis focused on whether visual motion suppressed and speeded up these auditory-evoked responses equally in all three conditions by comparing the peak amplitude and peak latency of the N1 and P2 components of the A and AV–V ERPs within and between conditions.

In accordance with previous research on audiovisual processing (Baart, 2016; Stekelenburg & Vroomen, 2007, 2012; Vroomen & Stekelenburg, 2010), the N1 and P2 had a fronto-central scalp distribution in all conditions (Figure 2.3). Hence, analyses for both the N1 and P2 component were conducted at a fronto-central (C1, Cz, C2, FCz) region of interest (ROI). Peak amplitude and peak latency for N1 and P2 were determined in a

window of 50-150 ms and 120-300 ms, respectively. The presence of statistically significant suppression and latency effects was tested by conducting separate repeated measures MANOVAs on the peak amplitude and peak latency for the N1 and P2 time windows, with the within-subjects variables Electrode (C1, Cz, C2, FCz), Condition (natural, random-timing, random-identity), and Modality (A, AV–V). Significant interaction effects were further examined with post hoc paired samples t-test. The multiple comparisons problem was addressed with the Benjamini-Hochberg procedure with a false discovery rate of 0.05 (Benjamini & Hochberg, 1995).

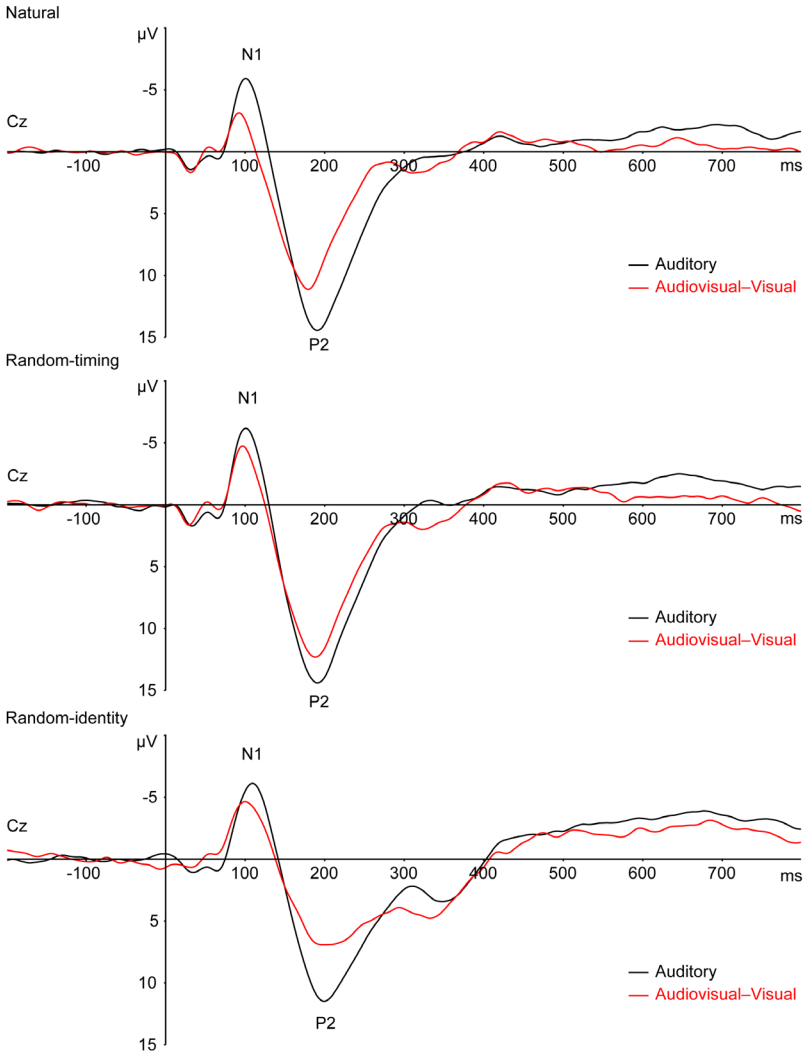


Figure 2.2. Grand averaged auditory and audiovisual event-related potential (ERP) waveforms for the natural, random-timing, and random-identity condition. Audiovisual ERPs were corrected for visual activity via subtraction of the visual waveform.

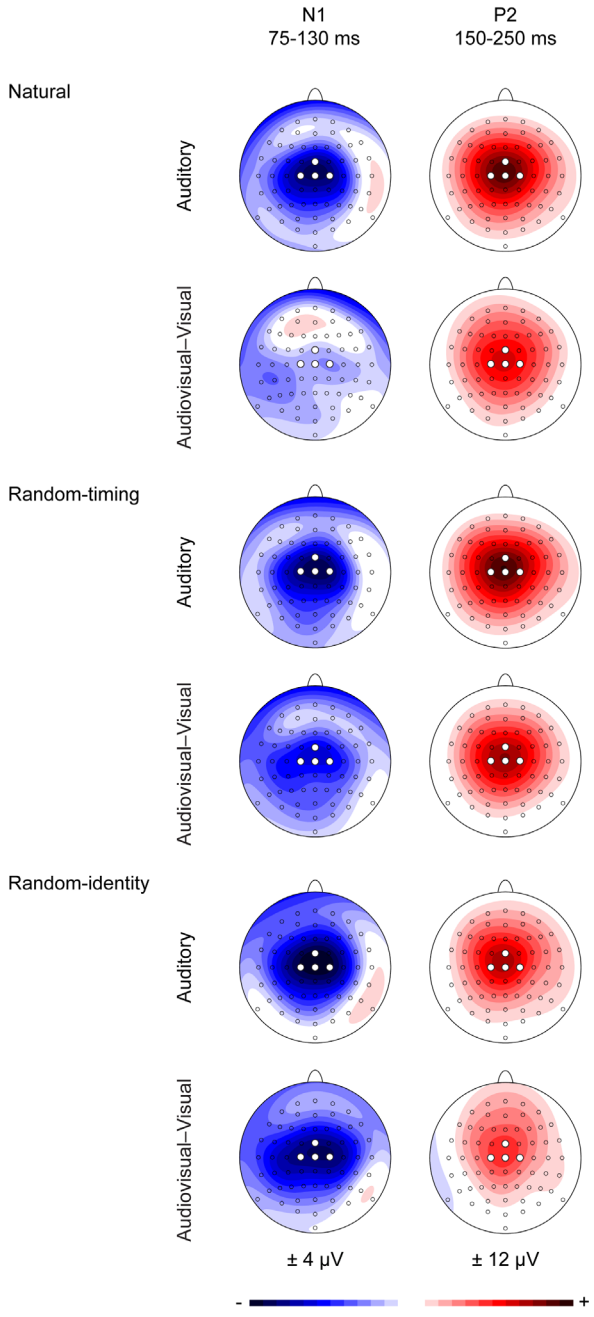


Figure 2.3. Scalp potential maps of the grand averaged auditory and audiovisual–visual N1 and P2 responses for the natural, random-timing, and random-identity condition. Based on these scalp distributions, a fronto-central region of interest (C1, Cz, C2, FCz) was selected for both the N1 and P2 component.

N1 amplitude suppression

The overall peak amplitude in the N1 time-window differed from pre-stimulus baseline levels, $F(1, 28) = 138.58$, $p < .001$, $\eta_p^2 = .83$. There were main effects of Electrode, $F(3, 26) = 10.32$, $p < .001$, $\eta_p^2 = .54$, Condition, $F(2, 27) = 5.90$, $p < .01$, $\eta_p^2 = .30$, and Modality, $F(1, 28) = 23.11$, $p < .001$, $\eta_p^2 = .45$. There was a significant Electrode × Modality interaction, $F(3, 26) = 3.66$, $p = .03$, $\eta_p^2 = .30$. Post hoc paired samples t-tests indicated that N1 peak amplitude (averaged across conditions) was attenuated in the AV–V modality compared to the A modality at all electrodes (all t values < -4.28, and p values < .001). Most importantly, the MANOVA for N1 peak amplitude showed there was a significant Condition × Modality interaction, $F(2, 27) = 3.90$, $p = .03$, $\eta_p^2 = .22$. Post hoc paired samples t-tests indicated that N1 peak amplitude in the fronto-central ROI was attenuated in the AV–V modality compared to the A modality in all conditions (all t values < -2.78, and p values < .01). N1 peak amplitude in the AV–V modality was lower in the natural condition compared to the random-timing ($t(28) = -3.60$, $p = .001$) and random-identity condition ($t(28) = -2.67$, $p = .01$). There were no significant differences in N1 peak amplitude between the three conditions in the A modality (all t values < 1.07, and p values > .30), which indicates that the extent to which the N1 was suppressed by visual motion in each condition could likely not be attributed to differences in unisensory auditory processing between conditions. A MANOVA on the AV–V–A N1 peak amplitude differences (averaged across all electrodes in the fronto-central ROI) showed a main effect of Condition, $F(2, 27) = 3.90$, $p = .03$, $\eta_p^2 = .22$. Post hoc t-tests showed that N1 suppression was larger in the natural condition compared to the random-timing ($t(28) = 2.49$, $p = .02$) and random-identity condition ($t(28) = 2.36$, $p = .03$), and did not significantly differ between the random-timing and random-identity condition ($t(28) = 0.03$, $p = .98$). Averaged across all electrodes in the fronto-central ROI, as shown in Figure 2.4a, visual information suppressed the N1 by 2.60 μV in the natural condition, 1.38 μV in the random-timing condition, and 1.38 μV in the random-identity condition (natural: A -6.46 μV, AV–V -3.86 μV; random-timing: A -6.85 μV, AV–V -5.47 μV, random-identity: A -6.80 μV, AV–V -5.42 μV).

N1 latency facilitation

The MANOVA for N1 peak latency showed a main effect of Modality $F(1, 28) = 11.70$, $p = .002$, $\eta_p^2 = .30$. The N1 was speeded up by 7.91 ms in the AV–V modality compared to the A modality (AV–V: 96.78 ms, A:104.68 ms). There was a main effect of Condition $F(2, 27) = 8.75$, $p = .001$, $\eta_p^2 = .39$. Post hoc paired samples t-tests indicated that (averaged across Electrode and Modality), N1 latency peaked ~8 ms later in the random-identity condition compared to the natural ($t(28) = 3.61$, $p < .001$) and random-timing ($t(28) = 3.56$, $p < .001$) condition (natural: 98.60 ms, random-timing: 97.51 ms, random-identity: 106.08 ms).

There was no main effect of Electrode, and no Condition × Modality, Condition × Electrode or Condition × Modality × Electrode interactions (all F values < 1.65, and p values > .18), indicating that the N1 was similarly speeded up by visual motion in all conditions (see Figure 2.4b).

P2 amplitude suppression

The overall peak amplitude in the P2 time-window differed from pre-stimulus baseline levels, $F(1, 28) = 319.22$, $p < .001$, $\eta_p^2 = .92$. There were main effects of Electrode, $F(3, 26) = 28.85$, $p < .001$, $\eta_p^2 = .77$, Condition, $F(2, 27) = 26.65$, $p < .001$, $\eta_p^2 = .66$, and Modality, $F(1, 28) = 60.39$, $p < .001$, $\eta_p^2 = .68$. There was a significant Electrode × Modality interaction, $F(3, 26) = 5.00$, $p < .01$, $\eta_p^2 = .37$. Post hoc paired samples t-tests indicated that P2 peak amplitude (averaged across conditions) was attenuated in the AV–V modality compared to the A modality at all electrodes (all t values < -6.61, and p values < .001). In addition, there was a significant Condition × Modality interaction, $F(2, 27) = 3.56$, $p = .04$, $\eta_p^2 = .21$. Post hoc paired samples t-tests indicated that P2 peak amplitude in the fronto-central ROI was attenuated in the AV–V modality compared to the A modality in all conditions (all t values < -4.57, and p values < .001). A post hoc MANOVA and subsequent t-tests on the AV–V–A P2 peak amplitude differences averaged across all electrodes in the fronto-central ROI (see Figure 2.4c), showed a main effect of Condition, $F(2, 27) = 3.56$, $p = .04$, $\eta_p^2 = .21$, with larger P2 suppression in the natural condition compared to the random-timing condition ($t(28) = 2.42$, $p = .02$). However, this difference did not remain significant after correction for multiple comparisons. There was no significant difference in P2 suppression between the natural and random-identity condition ($t(28) = 0.09$, $p = .93$), and between the random-timing and random-identity condition ($t(28) = 1.57$, $p = .13$). Averaged across all conditions and electrodes in the fronto-central ROI, visual information suppressed the P2 by 2.78 µV (natural: A 14.48 µV, AV–V 11.26 µV; random-timing: A 14.73 µV, AV–V 12.76 µV, random-identity: A 11.54 µV, AV–V 8.39 µV).

P2 latency facilitation

The MANOVA for P2 peak latency showed a main effect of Condition $F(2, 27) = 27.18$, $p < .001$, $\eta_p^2 = .67$. There was a significant Electrode × Condition interaction, $F(3, 26) = 2.75$, $p = .04$, $\eta_p^2 = .42$. Post hoc paired samples t-tests indicated that P2 peak latency (averaged across modalities) was different between conditions at all electrodes (all t values > 2.59, and p values < .02), but did not differ between electrodes within each condition (all t values < 1.81, and p values > .08). Crucially, there was a significant Condition × Modality interaction, $F(2, 27) = 9.66$, $p = .001$, $\eta_p^2 = .42$. Post hoc paired samples t-tests indicated that P2 peak latency was speeded up in the AV–V modality compared to the A modality in

the natural condition ($t(28) = -4.69$, $p < .001$), but not in the random-timing condition ($t(28) = -0.57$, $p = .58$). P2 peak latency appeared to be slower in the AV–V modality compared to the A modality in the random-identity condition ($t(28) = 2.11$, $p = .04$), but this difference did not remain significant after correction for multiple comparisons. A MANOVA and subsequent t-tests on the AV–V–A P2 peak latency differences (averaged across all electrodes in the fronto-central ROI) showed a main effect of Condition, $F(2, 27) = 9.66$, $p = .001$, $\eta_p^2 = .42$, with a larger P2 latency effect in the natural condition compared to the random-timing ($t(28) = 2.89$, $p < .01$) and random-identity condition ($t(28) = 4.21$, $p < .001$), and no difference in P2 latency effect between the random-timing and random-identity condition ($t(28) = 2.08$, $p = .05$). Averaged across all electrodes in the fronto-central ROI, as shown in Figure 2.4d, visual information speeded up the P2 by 17.34 ms in the natural condition, while the P2 was not significantly speeded up in the random-timing and random-identity condition (natural: A 195.43 ms, AV–V 178.09 ms; random-timing: A 192.94 ms, AV–V 190.95 ms, random-identity: A 203.04 ms, AV–V 216.38 ms).

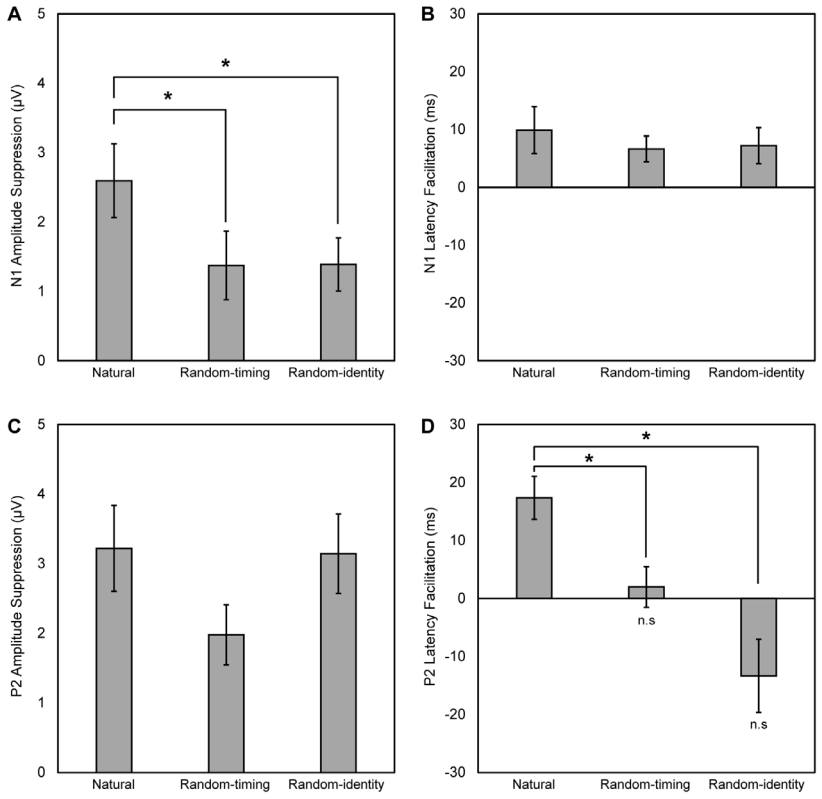


Figure 2.4. Mean amplitude suppression and latency facilitation for the N1 (A, B) and P2 (C, D) responses for the natural, random-timing and random-identity condition, averaged across all electrodes in the fronto-central region of interest (C1, Cz, C2, FCz). Error bars represent ±one standard error of the mean.

DISCUSSION

The current study examined the impact of temporal and identity predictability on suppression of the auditory N1 by visual motion. In three conditions, a video of a handclap either reliably predicted both the timing and identity of the sound (the natural condition), or only the identity (random-timing condition) or timing (random-identity condition). Compared to the N1 evoked by sounds played in isolation, the N1 was suppressed and speeded up when the same sounds were accompanied by visual motion, regardless of whether the video reliably predicted the timing or identity of the sound. The largest suppression of the N1 occurred when both the timing and identity were predictable. N1 suppression was similarly reduced when either the timing or identity were unpredictable. It is unlikely that these differences in N1 suppression between conditions can be attributed to differences in unisensory auditory processing. The current results thus demonstrate that visually-induced temporal and identity predictions both contribute to suppression of the N1 in the visual-auditory domain.

Previous studies have shown that N1 suppression by visual motion is reduced in audiovisual speech and artificial audiovisual stimuli when the timing of the auditory and visual signal varies from trial to trial (Huhn et al., 2009; Vroomen & Stekelenburg, 2010). The current results regarding the random-timing condition are in line with these findings, and demonstrate that visually-induced N1 suppression is also reduced in ecologically valid audiovisual events when the visual signal does not reliably predict the timing of the anticipated sound, but only the identity. The current findings regarding the random-identity condition, however, are not in agreement with those of previous studies using audiovisual speech stimuli (Klucharev et al., 2003; Stekelenburg & Vroomen, 2007; Van Wassenhove et al., 2005), and ecologically valid audiovisual stimuli (Stekelenburg & Vroomen, 2007), in which no effect of audiovisual congruency on N1 suppression was found. A plausible explanation for these different results could be due to the fact that only two (Stekelenburg & Vroomen, 2007; Van Wassenhove et al., 2005) or four (Klucharev et al., 2003) different sounds were used in these studies to manipulate identity predictability - whereas in the random-identity condition of the current study, sound identity was more unpredictable as the sound was randomly selected out of a pool of 100 stimuli on a trial-to-trial basis. Hence, participants in the aforementioned studies may have still been able to generate identity predictions to some degree, whereas in the current study this was virtually impossible and hence identity predictions were likely much less precise - as indicated by the reduced N1 suppression effect for sounds with unpredictable identity. In a future study, it would be interesting to examine whether the extent to which the N1 is suppressed by visual motion is indeed affected by the number of distinct sounds that are

paired with the visual signal by including different conditions with varying degrees of identity predictability. In addition, it has been hypothesized that the signaling of temporal predictions predominantly involves slow delta-theta oscillations, while identity predictions seem to be mediated on beta and gamma frequencies (Arnal & Giraud, 2012). It would therefore be interesting to see if a future study incorporating a time-frequency analysis might be able to examine the underlying oscillatory patterns of early electrophysiological markers of temporal and identity prediction in the visual-auditory domain.

The current results are in accordance with a previous study in the motor-auditory domain (Baess, Jacobsen, & Schröger, 2008), which showed that the N1 is suppressed for self-initiated sounds compared to externally-initiated sounds, even when the onset and identity of the sound was unpredictable. Similar to the current results, this study found that suppression was largest when both sound frequency and onset were predictable, and reduced when the sound frequency could not be predicted (Baess et al., 2008). Unlike in the study by Baess et al. (2008), suppression of the N1 was not larger in the current study when the onset of the sound was unpredictable when compared to a context with unpredictable sound identity. A plausible explanation for the different results is that, compared to auditory prediction by a self-generated motor act, auditory prediction by vision might be more affected by the temporal association between the auditory and visual stimulus. While strict intersensory temporal associations are not necessarily involved in the act of a button press in daily life, perceiving a video of a natural audiovisual event may induce relatively strong temporal auditory expectations based on lifelong experience. When operating electronic or mechanical devices, a button press may induce sensory stimulation with a great variety of delays. In the perception of audiovisual events, however, intersensory timing is much more confined by the naturally occurring lag in arrival time of the sensory information streams. Any disruption of this natural order of events thus likely results in less accurate predictions and decreased sensory attenuation - as indicated by the reduced suppression effect for sounds with unpredictable timing in the current study.

In line with previous studies in the visual-auditory domain (Baart, 2016; Stekelenburg & Vroomen, 2007, 2012; Vroomen & Stekelenburg, 2010), the N1 suppression effect was accompanied by a suppression of the P2. The P2 was equally suppressed by visual motion in all three conditions, regardless of whether the video reliably predicted the timing and identity of the sound. The current results are in line with previous research using artificial audiovisual stimuli (Vroomen & Stekelenburg, 2010), in which no effect of temporal

predictability on P2 suppression by visual motion was observed. The current findings regarding suppression of the P2 are not in accord with a previous study using similar audiovisual stimuli as used in the current study (Stekelenburg & Vroomen, 2007). In this study, an increased P2 suppression was found for incongruent audiovisual stimulus pairings. However, only two different incongruent audiovisual stimuli were used in this study to manipulate identity predictability, and incongruent and congruent stimulus pairings were administered in the same block. It could therefore be speculated that the incongruent audiovisual stimulus pairings were considered as a deviant stimulus category, and that the observed increase in P2 suppression in the study by Stekelenburg and Vroomen (2007) was in part caused by an overlapping mismatch negativity (MMN) ERP component related to the detection of a deviant stimulus or conflict between the visually anticipated sound and presented sound (Horváth, Schilberg, & Thomson, 2013; Näätänen, Gaillard, & Mäntysalo, 1978; Näätänen, Paavilainen, Rinne, & Alho, 2007). In the random-identity condition of the current study, the P2 suppression was likely not affected by an overlapping MMN-like component, since all audiovisual stimulus pairings in this condition were incongruent, and hence equally salient.

Suppression of the P2 for sounds with predictable timing and identity has also been reported in the motor-auditory domain (Horváth, Maess, Baess, & Tóth, 2012; Knolle, Schröger, Baess, & Kotz, 2012; Knolle, Schröger, & Kotz, 2013b). Unfortunately, the few studies that manipulated temporal or identity prediction in the motor-auditory domain did not examine the P2 component (Baess et al., 2008; Horváth, 2013). To our knowledge, only one study has reported an effect of identity prediction on P2 suppression in the motor-auditory domain (Hughes, Desantis, & Waszak, 2013). However, the manipulation of identity prediction in this study was very limited (i.e. only two different sounds were used), so the validity of these results is unclear and further research is warranted in order to determine the impact of identity and temporal prediction on P2 suppression in the motor-auditory domain. Nevertheless, the current results suggest that suppression of the auditory P2 may be less affected by temporal and identity predictability.

Although the functional interpretation of the auditory P2 is poorly understood (for review, see Crowley & Colrain, 2004), the current findings suggest that different processing mechanisms may underly N1 and P2 suppression effects. Functional dissociation of N1 and P2 suppression has been reported in the motor-auditory domain as well (Horváth et al., 2012; Knolle, Schröger, & Kotz, 2013a; Knolle et al., 2013b). It has been speculated that auditory P2 suppression may reflect the conscious post-hoc detection that a sound is initiated by a particular action (e.g. a self-initiated button press, or action performed by

someone else) - as opposed to a suppressed N1 response, which reflects the effect of an automatic prospective internal forward prediction mechanism (Knolle et al., 2012, 2013b, 2013a). In the current study, the suppressed P2 may reflect the conscious realization that the sounds were initiated by someone else, namely the actor portraying the handclap in the video. Given that the actor in the video was equal throughout the experiment, this might explain why there was no difference in visually-induced P2 suppression between conditions.

The N1 was equally speeded up by visual motion in all conditions. The dissociation between N1 suppression and latency effects in the current data suggests that different processing mechanisms may underly N1 amplitude and latency facilitation. It has been suggested that, unlike N1 and P2 suppression, facilitation of the N1 and P2 is affected by attentional task demands (Alsius, Möttönen, Sams, Soto-Faraco, & Tiippana, 2014). Given that latency facilitation of the N1 was equal in all conditions, it can thus be speculated that attentional load was equal in all three conditions. However, an argument against this interpretation of the current results is that the P2 was only speeded up in the natural condition, and not in the random-timing and random-identity condition. Another explanation of the current results is that the extent to which the N1 and P2 are speeded up depends on the degree to which the visual motion predicts the auditory signal, as previously proposed (Van Wassenhove et al., 2005). The current results may thus suggest that latency facilitation of both N1 *and* P2 only occurs when precision of the internal prediction model is optimally enhanced by visual motion (i.e. when both timing and identity are predictable). It should be noted, however, that visually-induced N1 and P2 latency effects are less robust than N1-P2 suppression effects (for review, see Baart, 2016), so the current results regarding N1-P2 latency facilitation require further validation and replication. Still, the current data support the predictive coding account for sensory processing, and demonstrate that suppression and facilitation of early electrophysiological indicators for predictive processing in audition depend on both temporal and identity precision of the internal prediction model.

CONCLUSIONS

The current study examined the impact of temporal and identity prediction on suppression of the auditory N1 by visual motion. Predictability of the timing and identity of auditory stimulation was manipulated in a fully crossed design using three different conditions in which sounds were either played in isolation or in conjunction with a video that either reliably predicted both the timing and identity of the sound, or only the timing or identity. The extent to which the N1 was suppressed by visual motion was affected by the temporal and identity predictability of the sound. N1 suppression was *largest* when the video reliably predicted the timing and identity of the sound, and *reduced* when either the timing or identity of the sound was unpredictable. Several studies in the auditory and motor-auditory domain have shown that both temporal and identity expectations may lead to suppression of the auditory N1 (for review, see Bendixen et al., 2012). The current results extend the existing literature by demonstrating that temporal and identity predictions contribute to auditory N1 suppression in the visual-auditory domain as well. Taken together, these findings indicate that predictions of timing and identity are both essential elements for predictive coding in audition. Future studies should examine if the current findings apply to other intersensory domains as well.

ACKNOWLEDGEMENTS

The authors thank Maria-Elise van der Sluys for her help with the data collection.

REFERENCES

Alsius, A., Möttönen, R., Sams, M. E., Soto-Faraco, S., & Tiippana, K. (2014). Effect of attentional load on audiovisual speech perception: Evidence from ERPs. *Frontiers in Psychology, 5,* 727. https://doi.org/10.3389/fpsyg.2014.00727

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Baart, M. (2016). Quantifying lip-read-induced suppression and facilitation of the auditory N1 and P2 reveals peak enhancements and delays. *Psychophysiology, 53*(9), 1295–1306. https://doi.org/10.1111/psyp.12683

Baess, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology*, *70*(2), 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Barth, D. S., Goldberg, N., Brett, B., & Di, S. (1995). The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Research, 678*(1-2), 177–190. https://doi.org/10.1016/0006-8993(95)00182-P

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology*, *83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Besle, J., Fort, A., & Giard, M.-H. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing, 5,* 189–192. https://doi.org/10.1007/s10339-004-0026-y

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Horváth, J. (2013). Attenuation of auditory ERPs to action-sound coincidences is not explained by voluntary allocation of attention. *Psychophysiology*, *50*(3), 266–273. https://doi.org/10.1111/psyp.12009

Horváth, J., Maess, B., Baess, P., & Tóth, A. (2012). Action-sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. *Journal of Cognitive Neuroscience, 24*(9), 1919–1931. https://doi.org/10.1162/jocn_a_00215

Horváth, J., Schilberg, L., & Thomson, J. (2013). Does sight predominate sound? Electrophysiological evidence for multisensory mismatch negativity correlation. *Neurophysiology*, *45*(5–6), 459–467. https://doi.org/10.1007/s11062-013-9394-1

Horváth, J., & Winkler, I. (2010). Distraction in a continuous-stimulation detection task. *Biological Psychology*, *83*(3), 229–238. https://doi.org/10.1016/j.biopsycho.2010.01.004

Hughes, G., Desantis, A., & Waszak, F. (2013). Attenuation of auditory N1 results from identity-specific action-effect prediction. *European Journal of Neuroscience, 37*(7), 1152–1158. https://doi.org/10.1111/ejn.12120

Huhn, Z., Szirtes, G., Lorincz, A., & Csépe, V. (2009). Perception based method for the investigation of audiovisual integration of speech. *Neuroscience Letters, 465*(3), 204–209. https://doi.org/10.1016/j.neulet.2009.08.077

Klucharev, V., Mottonen, R., Sams, M., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, *18*(1), 65–75. https://doi.org/10.1016/j.cogbrainres.2003.09.004

Knolle, F., Schröger, E., Baess, P., & Kotz, S. A. (2012). The Cerebellum Generates Motor-to-Auditory Predictions: ERP Lesion Evidence. *Journal of Cognitive Neuroscience*, *24*(3), 698–706. https://doi.org/10.1162/jocn_a_00167

Knolle, F., Schröger, E., & Kotz, S. A. (2013a). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, *49*(9), 2449–2461. https://doi.org/10.1016/j.cortex.2012.12.012

Knolle, F., Schröger, E., & Kotz, S. A. (2013b). Prediction errors in self- and externally-generated deviants. *Biological Psychology*, *92*(2), 410–416. https://doi.org/10.1016/j.biopsycho.2012.11.017

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. *Psychophysiology*, *40*(5), 806–817. https://doi.org/10.1111/1469-8986.00081

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Otte, R. A., Winkler, I., Braeken, M. A. K. A., Stekelenburg, J. J., van der Stelt, O., & Van den Bergh, B. R. H. (2013). Detecting violations of temporal regularities in waking and sleeping two-month-old infants. *Biological Psychology, 92*(2), 315–322. https://doi.org/10.1016/j.biopsycho.2012.09.009

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*(12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Stekelenburg, J. J., & Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Frontiers in Integrative Neuroscience*, *6*, 26. https://doi.org/10.3389/fnint.2012.00026

Teder-Salejarvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*, *14*(1), 106–114. https://doi.org/10.1016/S0926-6410(02)00065-4

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2017). Temporal and identity prediction in visual-auditory events: Electrophysiological evidence from stimulus omissions. *Brain Research*, *1661*, 79–87. https://doi.org/10.1016/j.brainres.2017.02.014

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*(1), 32–35. https://doi.org/10.1016/j.cogbrainres.2004.07.003

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*, *22*(7), 1583–1596. https://doi.org/10.1162/jocn.2009.21308

Chapter 3

Temporal and identity prediction in visual-auditory events: electrophysiological evidence from stimulus omissions

This chapter is published as:

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2017). Temporal and identity prediction in visual-auditory events: Electrophysiological evidence from stimulus omissions. *Brain Research*, 1661, 79–87. https://doi.org/10.1016/j.brainres.2017.02.014

ABSTRACT

A rare omission of a sound that is predictable by anticipatory visual information induces an early negative omission response (oN1) in the EEG during the period of silence where the sound was expected. It was previously suggested that the oN1 was primarily driven by the *identity* of the anticipated sound. Here, we examined the role of *temporal* prediction in conjunction with identity prediction of the anticipated sound in the evocation of the auditory oN1. With incongruent audiovisual stimuli (a video of a handclap that is consistently combined with the sound of a car horn) we demonstrate in Experiment 1 that a natural match in identity between the visual and auditory stimulus is *not* required for inducing the oN1, and that the perceptual system can adapt predictions to unnatural stimulus events. In Experiment 2 we varied either the auditory onset (relative to the visual onset) or the identity of the sound across trials in order to hamper temporal and identity predictions. Relative to the natural stimulus with correct auditory timing and matching audiovisual identity, the oN1 was abolished when either the timing or the identity of the sound could not be predicted reliably from the video. Our study demonstrates the flexibility of the perceptual system in predictive processing (Experiment 1) and also shows that precise predictions of timing and content are both essential elements for inducing an oN1 (Experiment 2).

INTRODUCTION

One of the main and arguably most basal functions of the human brain is to 'make sense' of our environment. Understanding which events in the outside world caused activation of specific sensory systems is what is generally considered to be the essence of perception (Lochmann & Deneve, 2011). This notion is central to the predictive coding theory, in which perceiving is considered a process of inferring the most probable causes explaining sensory signals (Friston, 2005). A key element of predictive coding is the assumption that the brain generates internal templates of the world in higher cortical areas (Mumford, 1992). These templates supposedly contain specific activation patterns of sensory systems that an occurring stimulus would normally elicit. The generated templates are presumed to be sent from higher to lower cortical processing areas (top-down), where they induce a predicted pattern of activation (Friston, 2005). If the bottom-up activation pattern induced by a stimulus matches the prediction, recognition of the stimulus occurs. Any violation of the predicted patterns by the sensory input is sent from lower sensory levels to higher cortical processing areas, reflecting the prediction error (Arnal & Giraud, 2012; Wacongne, Changeux, & Dehaene, 2012).

An approach that has been applied recently to explore the neurophysiological mechanisms of sensory prediction relies on the electrophysiological responses to infrequent unexpected stimulus *omissions*. According to the predictive coding framework, early sensory responses reflect the difference between the prediction and sensory input (Friston, 2005; Wacongne et al., 2012). During stimulus omissions there is no sensory input and the neural response to stimulus omissions is thus hypothesized to represent the neural code of top-down prediction devoid of stimulus-evoked sensory processing (Arnal & Giraud, 2012; SanMiguel, Widmann, Bendixen, Trujillo-Barreto, & Schroger, 2013b). An auditory event can be made predictable either by a motor act or anticipatory visual information regarding the onset and identity of the sound (SanMiguel et al., 2013b; Stekelenburg & Vroomen, 2015). An occasional unexpected omission of the sound evokes an early negative omission response (oN1), likely originating in the auditory cortex, suggesting that both motor and visual predictions are able to activate a sensory template of an expected auditory stimulus in the auditory cortex.

While the available data agree that the oN1 response is an electrophysiological indicator of automatic predictive processing, it is not yet fully understood whether auditory prediction is primarily driven by *temporal* information (timing) or by the *identity* of the anticipated sound. In the motor-auditory (MA) domain, a study of SanMiguel, Saupe and Schröger (2013a) suggests that auditory omission responses are primarily driven by

identity prediction, with only a modulatory effect of temporal prediction. In their study either a single sound or a random sound was presented after a self-paced button press. Prediction-related auditory omission responses were only observed in the single sound condition, suggesting that the sensory system, even with exact foreknowledge of the stimulus onset, does not formulate predictions if the identity of the predicted stimulus cannot be anticipated (SanMiguel, Saupe, & Schroger, 2013a). However, the timing of the sound was not specifically manipulated in their study, which calls upon further investigation of the role of temporal prediction using a stimulus omission paradigm.
The present study investigated the neural mechanisms of temporal and identity auditory predictions in the visual-auditory (VA) domain by using infrequent auditory stimulus omissions. We conducted two separate experiments. In both experiments, we used a video of an actor performing a single handclap (Figure 3.1) as a visual stimulus containing anticipatory information about sound identity and sound onset (Stekelenburg & Vroomen, 2007, 2015).

In the first experiment, we examined whether visual-to-auditory predictions (reflected in the omission response) are flexible and adapt, in short-term, to unnatural VA incongruences, or rather depend on long-term established associations. Compared to auditory prediction by a self-generated motor act, prediction of a sound by vision might be more affected by the informational association between the visual and auditory stimulus. While strict informational associations are not necessarily involved in the act of a button press – as a button press can elicit various different sounds in daily practice – a video of a natural visual event may induce relatively strong auditory associations based on lifelong experience. Furthermore, although previous studies have shown that unnatural VA pairings may lead to enhancements in auditory processing (Fort, Delpuech, Pernier, & Giard, 2002; Giard & Peronnet, 1999; Thorne & Debener, 2008), it is unclear whether auditory omission responses are affected by VA congruency of identity or not. Hence, the first experiment was conducted to examine the influence of VA congruency of identity on prediction-related auditory omission responses. VA congruency was manipulated block-wise in two separate conditions. The video of the handclap was presented synchronously with either the sound of the actual handclap (natural condition) or the sound of a car horn (incongruent condition). The timing of the incongruent sound matched the timing of the natural sound. The sound of a car horn was specifically chosen to obtain a high level of VA incongruence with respect to real-world situations. VA trials were interspersed with unpredictable omissions of the sound in 12% of the trials in both conditions, c.f. SanMiguel et al. (2013a) and Stekelenburg and Vroomen (2015). Based on previous findings (SanMiguel et al., 2013b; Stekelenburg & Vroomen, 2015), three distinct

omission ERP components – elicited by rare omissions of the expected sound – were expected for the natural condition: an initial negative deflection at around 50-100 ms after the expected sound onset (oN1), reflecting prediction error, followed by a second negative response at around 200 ms (oN2), and finally a more broadly distributed positive response at 300 ms (oP3), presumably reflecting higher-order error evaluation, attention orienting and subsequent updating of the forward model (Baldi & Itti, 2010; Polich, 2007). A statistically significant difference between the omission responses of the natural and incongruent conditions would suggest that the omission response depends on long-term learned VA associations.

In the second experiment, we examined the separate contributions of temporal and identity information on VA omission responses by randomizing (on a trial-to-trial basis) either auditory onset relative to visual onset or sound identity. Three experimental conditions were included: a *natural* condition, a *random-timing* condition and a *random-identity* condition (Table 3.1). The natural condition was identical to the natural condition of Experiment 1. In the other two conditions, either the onset (random-timing condition) or the identity (random-identity condition) of the sound was unpredictable. Temporal prediction was disrupted in the random-timing condition by presenting VA stimuli (88% of total number of trials) for which sound and vision were always asynchronous. The magnitude of asynchrony varied on a trial-to-trial basis in order to prevent adaptation to temporal asynchrony (Vroomen, Keetels, de Gelder, & Bertelson, 2004). In the random-identity condition the identity of the sound was different for each trial (c.f. the random-sound condition in SanMiguel et al. (2013a)). Based on previous findings in the MA domain, prediction-related neural activity induced by auditory omissions was expected to be most evident in the natural condition (SanMiguel et al., 2013a; Stekelenburg & Vroomen, 2015), and to be diminished in the random-identity condition (SanMiguel et al., 2013a). Assuming that timing of the sound is also of importance in the VA domain (Vroomen & Stekelenburg, 2010), we expected that the omission responses would also be diminished in the random-timing condition.

METHODS

The study was conducted in accordance with the Declaration of Helsinki. All experiments were undertaken with the understanding and written consent of each participant. The Ethics Review Board of the School of Social and Behavioral Sciences of Tilburg University approved all experimental procedures (EC-2016.48). All participants received course credits. None were diagnosed with a neurological condition and none reported use of medication. All participants reported normal hearing and normal or corrected-to-normal vision.

Experiment 1

Participants. Seventeen students of Tilburg University (11 female, all right-handed) with a mean age of 20.82 years (*SD* = 2.92) participated in the study.

4.1.2 Stimuli. Auditory stimuli consisted of recordings of a handclap and a car horn of equal length (200 ms) and sampling rate (44.1 kHz), with matched amplitudes based on the root mean square method. Audio files were presented over JAMO S100 stereo speakers, located directly on the left and right side of the monitor, at approximately 61dB(A) sound pressure level. Visual stimuli consisted of a video recording portraying the motion of a single handclap (Figure 3.1). The video started with the hands separated. Subsequently, the hands moved to each other and after collision returned to their original starting position. Total duration of the video was 1300 ms. The video was presented at a frame rate of 25 frames/s on a 19-inch Iiyama Vision Master Pro 454 CRT monitor at a refresh rate of 100 Hz, a resolution of 640 x 480 pixels (14° horizontal and 12° vertical visual angle) and at a viewing distance of approximately 70 cm.

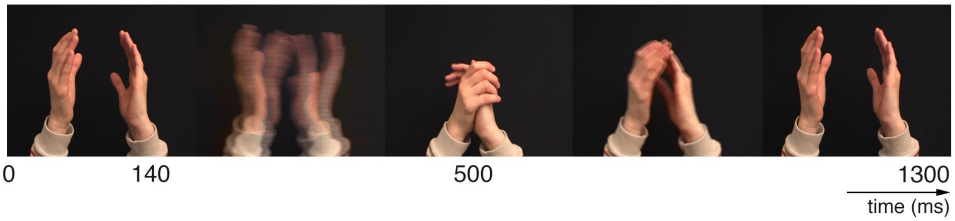


Figure 3.1. Time-course of the video used in all experimental conditions of Experiment 1 and Experiment 2.

Procedure. Participants were individually tested in a sound attenuated and dimly lit booth. They were instructed to carefully listen to the presented audio files and to maintain their focus on the center of the screen. VA congruency was manipulated block-wise in two separate conditions: a natural condition and an incongruent condition (Table 3.1). During VA trials in the natural condition, a video of a handclap was presented synchronously with the audio recording of the actual handclap. For the incongruent condition the handclap sound was replaced by the sound of a car horn. This sound was specifically chosen to obtain a high level of VA incongruence with respect to real-world situations. In both conditions, the sound occurred 500 ms after video onset and 360 ms after the start of the hand movement, while the inter-stimulus interval (from auditory onset) was 1300 ms (Figure 3.1). VA trials were interspersed with unpredictable omissions of the sound in 12% of the trials in both conditions, c.f. SanMiguel et al. (2013a) and Stekelenburg and Vroomen (2015). These omission trials were randomly intermixed with VA trials with the restrictions that the first five trials of each block and the two trials immediately following an omission trial were always VA trials. Each condition was presented in seven blocks of 200 trials (with a short break between blocks). This resulted in a total of 1400 stimulus presentations in each condition, including 168 auditory stimulus omissions. Block order was varied quasi-randomly. After every fourth block a short block of 100 visual-only trials was presented (i.e. three visual-only blocks for each participant), during which only the visual recording of a handclap was presented. The visual-only (V) condition was introduced to correct for visual activity in the auditory omission trials (see 'EEG recording'). An auditory-only condition was not included, since a previous study using the same VA stimuli and a similar inter-stimulus interval demonstrated that unexpected omissions of the sound *as such* do not evoke a significant neural response (Stekelenburg & Vroomen, 2015). To ensure that participants watched the visual stimuli, 8% of all VA and V trials consisted of catch trials. Participants were required to respond with a button press after onset of a catch stimulus (i.e. a small white square superimposed on the handclap video, presented at the center of the screen, measuring 1° horizontal and 1° vertical visual angle). To prevent possible interference of (delayed) motor responses, these catch trials never preceded an omission trial. Participants were unaware of the total amount of catch trials presented in each block. After each block, the percentage of missed catch trials and false alarms was displayed at the center of the screen. Average percentage of detected catch trials across conditions was high ($M = 98.76$, $SD = 1.76$) and did not differ between conditions or subjects and there was no condition × subject interaction effect, indicating that the participants attentively watched the video in all conditions.

Table 3.1. Experimental conditions included in Experiment 1 and Experiment 2

Condition	Sound timing	Sound identity
Natural [Exp1, Exp2]	Synchronized with video	Handclap
Incongruent [Exp1]	Synchronized with video	Car horn
Random-identity [Exp2]	Synchronized with video	Random[a]
Random-timing [Exp2]	Random[b]	Handclap

[a] The identity of the sound was randomly selected in every trial out of 100 different environmental sounds (e.g. a doorbell, barking dog or a car horn) of equal length and matched amplitudes
[b] The sound could either precede or follow the visual collision moment of the two hands at a randomly selected SOA of -250, -230, -210, -190, -170, 210, 240, 260, 290, or 320 (all values in ms, negative and positive values indicate sound leading and following the natural synchrony point, respectively)

EEG recording. The EEG was sampled at 512 Hz from 64 locations using active Ag-AgCl electrodes (BioSemi, Amsterdam, the Netherlands) mounted in an elastic cap and two mastoid electrodes. Electrodes were placed in accordance with the extended International 10-20 system. Two additional electrodes served as reference (Common Mode Sense active electrode) and ground (Driven Right Leg passive electrode). EEG was referenced offline to an average of left and right mastoids and band-pass filtered (1-30 Hz, 24 dB/octave). To facilitate a more direct comparison between the current data and the results of the previous auditory omission study in the MA domain, the same high-pass filter settings were applied as in SanMiguel et al. (2013a). Furthermore, the relatively long interval between visual and auditory stimulus onset might elicit anticipatory slow waves that may contaminate early ERP components (Teder-Salejarvi, McDonald, Di Russo, & Hillyard, 2002). To prevent this anticipatory activity from contaminating or simply obscuring the oN1 component, a relatively high high-pass filter of 1 Hz was applied. The (residual) 50 Hz interference was removed by a 50 Hz notch filter. Raw data were segmented into epochs of 1000 ms, including a 200-ms pre-stimulus baseline period. Epochs were time-locked to the expected sound onset in the natural and incongruent conditions, and to the corresponding timestamp in the V condition. After EOG correction (Gratton, Coles, & Donchin, 1983), epochs with an amplitude change exceeding ±120 μV at any EEG channel were rejected and subsequently averaged and baseline corrected for each condition separately. On average 6.34 percent (*SD* = 6.84) of the omission trials were rejected. There was no significant difference in rejected omission trials between conditions. The ERP of the V condition was subtracted from the VA omission ERPs of the natural and incongruent conditions to correct for the contribution of visual activity to the omission ERPs (Figure 3.6). Consequently, the VA-V difference waves reflect prediction related activity – induced by unexpected auditory omissions – devoid of visual related activity (Stekelenburg & Vroomen, 2015).

Experiment 2

Participants. Twenty-seven students of Tilburg University (23 female, 4 left-handed) with a mean age of 19.93 years (*SD* = 2.40) participated after given written informed consent. None of them participated in Experiment 1. None reported use of prescription drugs or were diagnosed with a neurological disorder. All participants reported normal hearing and normal or corrected-to-normal vision and received credits in hours as part of a curricular requirement.

Stimuli and procedure. Three experimental conditions were included: a *natural* condition, a *random-timing* condition and a *random-identity* condition (Table 3.1). The natural condition was identical to the natural condition of Experiment 1. In the random-timing condition, the sound could either precede or follow the visual collision of the two hands at an unpredictable stimulus onset asynchrony (SOA). Based on the results of a simultaneity judgment (SJ) task ran prior to the EEG experiment (Figure 3.7), SOAs of -250, -230, -210, -190, -170, 210, 240, 260, 290 and 320 were chosen (all values in ms, negative and positive values indicate sound leading and following the natural synchrony point, respectively). In the random-identity condition, 100 different environmental sounds (e.g. a doorbell, barking dog or a car horn) of equal length (200 ms) and matched amplitudes were used. The video showed the same handclap as before and was presented synchronously with an environmental sound that was randomly selected in every trial out of the pool of 100 sounds (c.f. the random-sound condition in SanMiguel et al. (2013a)). The experimental design of the three conditions was identical to Experiment 1 (a total of 1400 trials per condition; 12% auditory omission trials; 8% catch trials). Each condition was presented in seven blocks of 200 trials, while quasi-random block sequences were allocated to each participant using a counterbalanced measures design. After every sixth block a short block of 100 V trials was presented. Average percentage of detected catch trials across conditions was high (*M* = 98.48, *SD* = 2.51). There were no main effects of condition and subject and no condition × subject interaction effect, indicating that all participants attentively watched the video in each condition.

EEG recording. EEG recording and filtering was equivalent to Experiment 1. Epochs of 1000 ms (including a 200-ms pre-stimulus baseline period) were time-locked to the expected sound onset in the natural and random-identity conditions, and to the corresponding timestamp in the random-timing and V condition. All omission trials not rejected due to artifacts were included in the average visual-corrected omission-ERP for

each condition. On average, 5.51% (*SD* = 5.74) of all omission trials were rejected. There were no significant differences in rejected omission trials between conditions.

RESULTS

Experiment 1

Three distinct deflections in the omission ERP were observed for both the natural and incongruent condition (Figure 3.2). The first negative component peaked in a time-window of 45-80 ms and is denoted as oN1. A second negative component reached its maximum at 120-240 ms (oN2). The two negative components were followed by a broadly distributed positive deflection in a window of 240-500 ms (oP3). The oN1 deflection showed a bilateral scalp distribution with a right preponderance in both conditions, while the oN2 and oP3 components had a bilateral scalp distribution with no clear preponderance towards either hemisphere (Figure 3.3). Based on these scalp distributions, a left fronto-temporal (F7, F5, FT7, FC5) and right temporal (FC6, FT8, C6, T8) ROI were selected for the oN1 time-window. A frontal (F1, Fz, F2) and frontal-central (FC1, FCz, FC2) ROI was selected for the oN2 and oP3 time-window respectively. Mean amplitudes were calculated for each time-window. The presence of statistically significant omission responses was tested with separate repeated measures ANOVAs for each time-window with the within-subjects variables condition, electrode and ROI for the oN1 time-window and condition and electrode for the oN2 and oP3 time-windows.

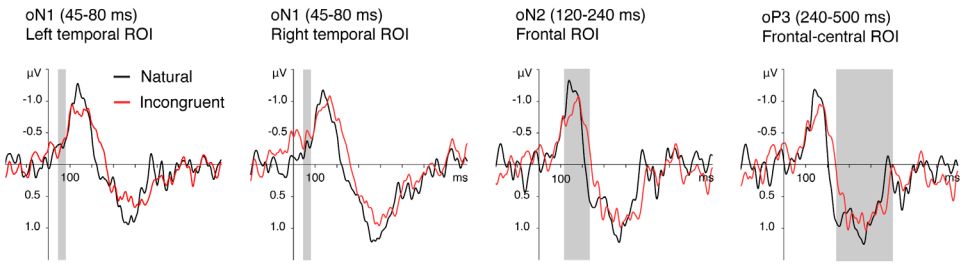


Figure 3.2. Direct comparison of the grand average omission-ERPs between the natural and incongruent condition. Omission responses were corrected for visual activity via subtraction of the visual-only waveform and collapsed over electrodes in each region of interest (ROI). The first negative component peaked in a time-window of 45–80 ms (oN1). A second negative component reached its maximum in 120–240 ms (oN2). The two negative components were followed by late positive potentials in a time-window of 240–500 ms (oP3).

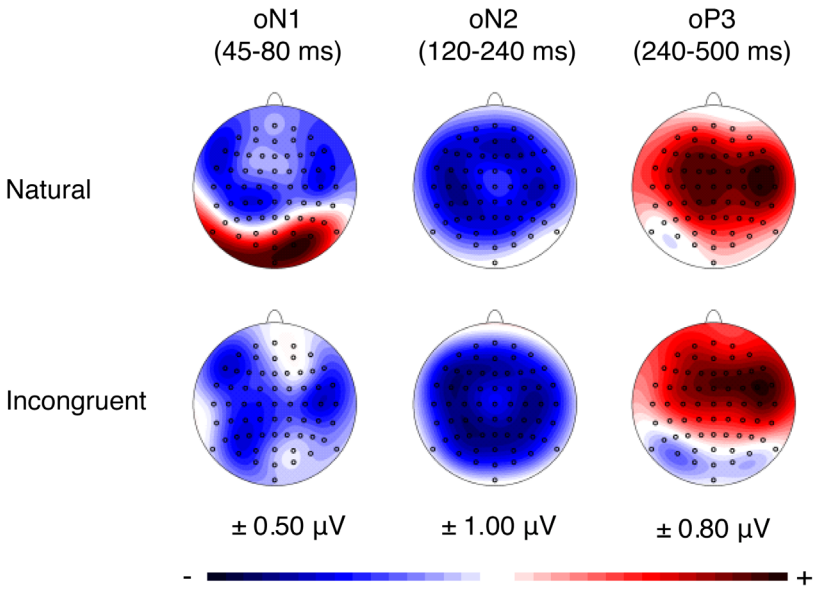


Figure 3.3. Scalp potential maps of the grand average visual-corrected omission responses for the natural and incongruent condition in the denoted oN1 (45-80 ms), oN2 (120-240 ms) and oP3 (240-500 ms) time-windows.

The mean activity in the oN1, oN2 and oP3 time-windows differed from pre-stimulus baseline levels (oN1: F (1, 16) = 5.97, p < .05, η_p^2 = .27, oN2: F (1, 16) = 20.76, p < .001, η_p^2 = .57, oP3: F (1, 16) = 5.33, p = .05, η_p^2 = .25). Most importantly, there were no significant main effects of condition, ROI and electrode, and no interaction effects (all p values > .05), indicating that the omission responses for the natural and incongruent conditions were alike. Of note, upon visual inspection of the omission ERPs shown in Figure 3.2, it appears there was a difference in amplitude between the two conditions around the expected sound onset. However, statistical analysis of the mean activity recorded at the electrodes showing maximal activity in a time-window of -20-40 ms – using the same repeated measures ANOVA as used for the oN1 time-window – showed no significant main effects of condition, ROI, and electrode, and no interaction effects (all p values > .12). Figure 3.2 also suggests a latency difference between the two conditions in the oN2 time-window. We tested peak latency of the oN2 response in both conditions – using the same repeated measures ANOVA used for the mean activity in the oN2 time-window – and found no significant main effects of condition and electrode and no interaction effect (all p values > .19).

Experiment 2

Three distinct deflections were observed in the omission ERP of the natural condition: oN1 peaking in a temporal window of 45-100 ms; oN2 at 120-230 ms and oP3 at 240-550 ms (Figure 3.4). The oN1 component for the natural condition had a bilateral scalp distribution with a left preponderance, while the oN1 components for the random-timing and the random-identity condition showed a more lateralized distribution toward the left hemisphere. The oN2 and oP3 deflections had bilateral scalp topographies for all conditions. Based on these scalp potential maps (Figure 3.5), a left temporal (FT7, FC5, T7, C5) and right temporal (FC6, FT8, C6, T8) ROI were selected for the oN1 time-window. A frontal (F1, Fz, F2) and frontal-central (FC1, FCz, FC2) ROI was selected for the oN2 and oP3 time-window, respectively. After calculation of the mean amplitudes in each time-window, the presence of statistically significant omission responses in the oN1 time-window was tested with a repeated measures ANOVA with within-subjects variables condition, ROI and electrode. The oN2 and oP3 responses were tested with repeated measures ANOVAs with condition and electrode as within-subjects variables.

The overall mean activity in the oN1 time-window differed from pre-stimulus baseline levels, $F(1, 26) = 10.03$, $p < .01$, $\eta_p^2 = .28$. There was a main effect of condition, $F(2, 25) = 5.41$, $p < .05$, $\eta_p^2 = .30$. Post hoc paired samples t-tests (Holm-Bonferroni corrected) showed that the mean activity in the oN1 time-window was significantly more negative in the natural condition than in the random-timing condition and random-identity condition (both p values < .05). Mean activity in the oN1 time-window did not differ between the random-timing and random-identity condition. To further examine whether the oN1 differed from pre-stimulus baseline levels within each condition, we tested the mean activity in the oN1 time-window for each condition with separate repeated measures ANOVAs with within-subjects variables ROI and electrode. This analysis revealed that the mean activity in the oN1 time-window only differed from zero in the natural condition, $F(1, 26) = 20.51$, $p < .001$, $\eta_p^2 = .44$. There were no main effects of ROI and electrode, but the ROI × electrode interaction was statistically significant, $F(1, 24) = 10.03$, $p < .01$, $\eta_p^2 = .43$. Simple effect tests examining the effect of electrode within each ROI showed no main effect of electrode in the right temporal ROI, whereas a significant main effect of electrode was revealed in the left temporal ROI, $F(3, 24) = 3.44$, $p < .05$, $\eta_p^2 = .30$. Post hoc paired samples t-tests indicated that the mean activity in the oN1 time-window was more negative at C5 than at FC5 and T7 (all p values < .05). There were no other interaction effects.

The overall mean activity in the oN2 time-window differed from pre-stimulus baseline levels, $F(1, 26) = 15.85$, $p < .001$, $\eta_p^2 = .38$. There was a main effect of condition, $F(2, 25) = 4.21$, $p < .05$, $\eta_p^2 = .25$. The mean activity in the oN2 time-window was more negative in the natural condition than in the random-timing and random-identity condition (both p values < .05). There was no difference in mean activity between the random-timing and random-identity condition. Further examination of the oN2 activity for each condition with separate repeated measures ANOVAs (with electrode as within-subjects variable) showed that the mean amplitude only differed from zero in the natural condition, $F(1, 26) = 32.08$, $p < .001$, $\eta_p^2 = .55$. There were no other main or interaction effects.

The overall mean amplitude in the oP3 time-window differed from pre-stimulus baseline levels, $F(1, 26) = 16.53$, $p < .001$, $\eta_p^2 = .39$. There was a main effect of condition, $F(2, 25) = 4.77$, $p < .05$, $\eta_p^2 = .28$. The mean activity in the oP3 time-window was more positive in the natural condition than in the random-timing and random-identity condition (both p values < .03). There was no difference in mean activity between the random-timing and random-identity condition. Testing of the oP3 activity for each condition separately – following the same procedure used on the oN2 activity – showed that the mean amplitude in the oP3 time-window only differed from zero in the natural condition, $F(1, 26) = 16.53$, $p < .001$, $\eta_p^2 = .39$. There were no other main or interaction effects.

In sum, auditory omissions induced three distinct deflections in the natural condition: oN1 (45-100 ms), oN2 (120-230 ms) and oP3 (240-550 ms). Statistical analysis indicated that the mean activity in all time-windows differed between the natural and random-timing condition, and the natural and random-identity condition. Further examination revealed that the mean amplitude in the oN1, oN2 and oP3 time-windows, tested in the selected ROIs, only differed from pre-stimulus baseline levels in de natural condition.

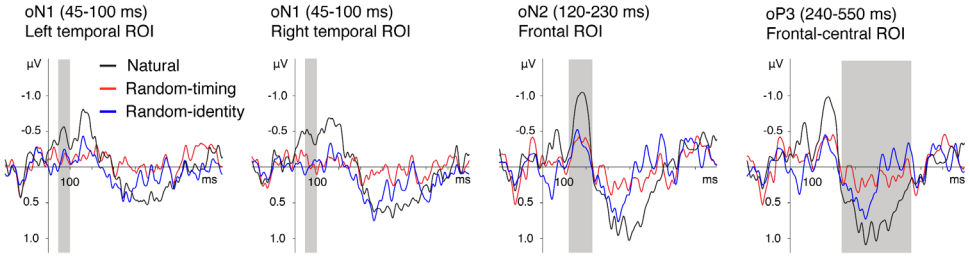


Figure 3.4. Direct comparison of the grand average omission-ERPs between the natural, random-timing and random-identity condition. Omission responses were corrected for visual activity via subtraction of the visual-only waveform and collapsed over electrodes in each region of interest (ROI). The first negative component peaked in a time-window of 45-100 ms (oN1). A second negative component was observed in a time-window of 120-230ms (oN2). The two negative components were followed by a positive deflection that reached its maximum at 240-550 ms (oP3).

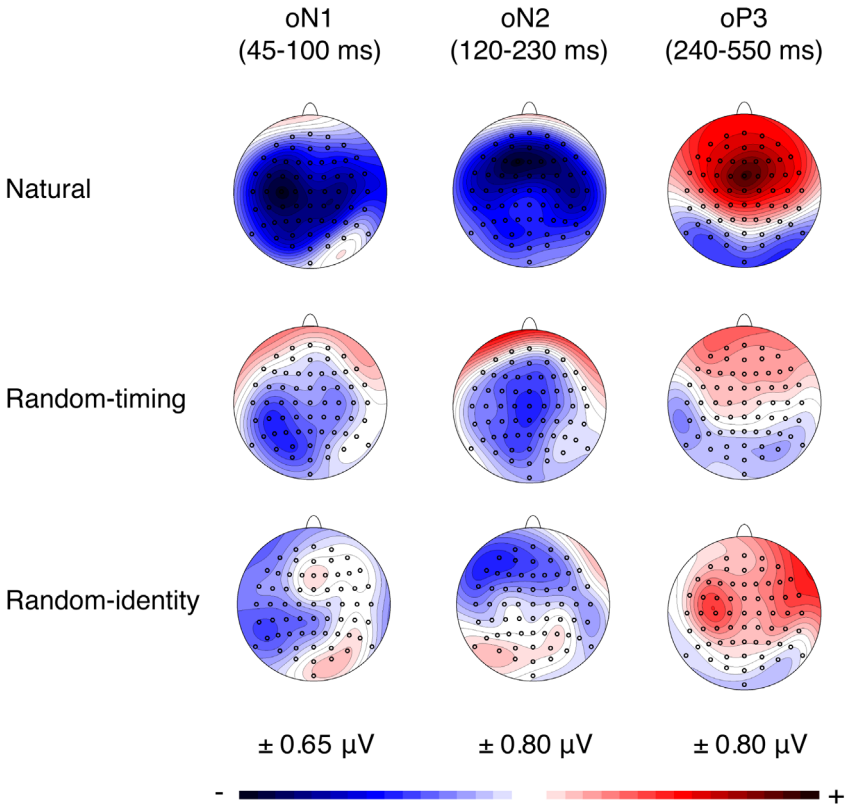


Figure 3.5. Scalp potential maps of the grand average visual-corrected omission responses for the natural, random-timing and random-identity condition in the denoted oN1 (45-100 ms), oN2 (120-230 ms) and oP3 (240-550 ms) time-windows.

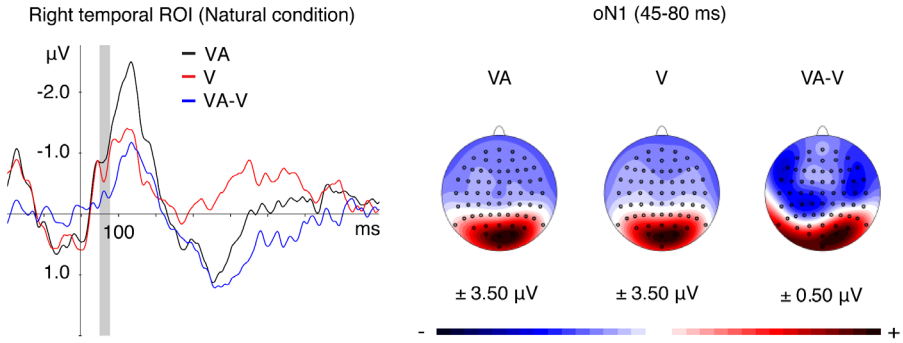


Figure 3.6. Comparison between visual-uncorrected and visual-corrected omission ERPs and oN1 scalp potential maps in the natural condition. The ERP of the visual-only (V) condition was subtracted from the visual-auditory (VA) omission ERP to correct for the contribution of visual activity to the omission ERPs. Consequently, the VA-V difference wave reflects prediction related activity devoid of visual activity.

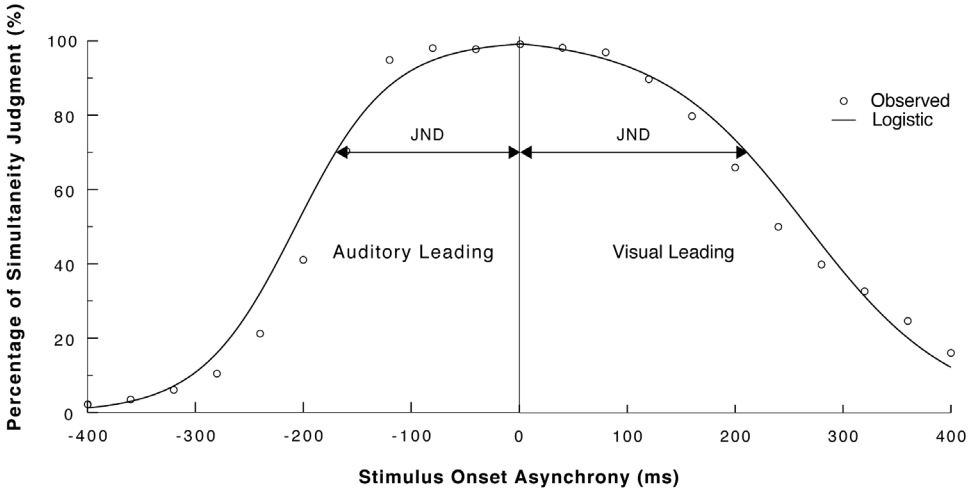


Figure 3.7. Grand average percentages of simultaneity judgment (SJ) as a function of stimulus onset asynchrony (SOA). Two logistic curves were fitted to the data - one on the auditory-leading side and one on the visual-leading side. The just noticeable difference (JND) of 70% was calculated for both logistic curves and used as a reference for the smallest SOAs included in the random-timing condition of Experiment 2. This ensured that none of the visual-auditory (VA) trials in this condition would be perceived as synchronous VA events. The remaining SOAs were obtained from both curves by calculating the intersections at 60, 50, 40 and 30 percent of simultaneity judgment. Consequently, the following SOAs were included: -250, -230, -210, -190, -170, 210, 240, 260, 290 and 320 (all values in ms).

DISCUSSION

The current study examined the neural correlates of auditory prediction by vision using a stimulus omission paradigm. In Experiment 1 we examined whether the identity of the sound should match the natural identity of the visual information in order for the oN1 to occur, or whether an incongruent sound can also elicit the oN1, provided that the sound remains consistent across trials and synchronized with the visual event. The results of Experiment 1 showed that occasional auditory omissions in otherwise natural (video and sound of a handclap) and unnatural (video of a handclap combined with a car horn) VA combinations induced prediction-related ERP components (oN1, oN2 and oP3) of similar amplitude. This indicates that a match in identity between sound and vision of a natural event is not required per se for auditory prediction by vision. Presumably, given that the stimulus was highly predictable in both content and timing, the perceptual system learned to expect an incongruent sound, which suggests that sensory predictions adapt to unnatural stimulus events when presented repeatedly. These findings are in line with previous studies showing that unnatural VA pairings of artificial stimuli are integrated by the perceptual system in a seemingly automatic fashion (Fort et al., 2002; Giard & Peronnet, 1999; Thorne & Debener, 2008). More importantly, the current data show that visual-to-auditory predictions are not bound to long-term established VA associations – as reflected in the highly similar omission responses in the natural and incongruent condition – but are able to adapt to unnatural VA incongruences. This ability may be crucial in order to deal with the inherent imprecision of visual to auditory predictions in real life situations.

In Experiment 2, the relative contribution of temporal and identity prediction to omission responses was explored by varying either the relative timing of the sound (while keeping sound identity constant) or the identity of the sound (while keeping relative timing constant). We found that only in the natural situation − where sound onset and identity were highly predictable from visual context – the oN1 and subsequent mid- and late latency responses (oN2, oP3) occurred. No omission responses were observed if either temporal or identity prediction was disrupted. This thus suggests that VA prediction is dependent on both *timing* and *identity*.

The results of Experiment 2 are partly consistent with studies on stimulus prediction as measured by the attenuation of the auditory N1. The amplitude of the auditory N1 is hypothesized to be a reflection of the prediction error (Arnal & Giraud, 2012; Friston, 2005). As an example, when an incoming sound matches the predicted stimulus, the amplitude of the auditory N1 is attenuated, while the neural response is enlarged when

the prediction error is large. Several studies have indeed demonstrated that the amplitude of the auditory N1 is significantly attenuated when sounds are self-initiated compared to sounds triggered externally (Baess, Jacobsen, & Schroger, 2008; Martikainen, Kaneko, & Hari, 2005), or when a sound is preceded by a visual stimulus that reliably predicts the onset of the sound (Stekelenburg & Vroomen, 2007; van Wassenhove, Grant, & Poeppel, 2005). Our results support a study on predictive processing in the MA domain (Baess et al., 2008), which showed that attenuation of the auditory N1 depended on both the identity and timing of the auditory stimuli − with less attenuation when the auditory stimuli varied randomly in pitch and timing relative to the motor act. In the VA domain, randomization of VA asynchrony also abolished the attenuation of the auditory N1 (Stekelenburg & Vroomen, 2007). Interestingly, though, VA congruency of identity had no effect on N1-suppression (Klucharev, Mottonen, & Sams, 2003; Stekelenburg & Vroomen, 2007). How can the latter results be reconciled with the current data showing an effect of identity on predictive processing? It could be reasoned that attenuation of the auditory N1 and the elicitation of the oN1 reflect different processes. However, an argument against this view is that the neural source of the oN1 and the attenuation of the auditory N1 induced by the same visual stimulus − the handclap video − appear to be similar (Stekelenburg & Vroomen, 2015; Vroomen & Stekelenburg, 2010), despite obvious limitations in spatial resolution of EEG. Assuming that both the oN1 and attenuation of the N1 reflect corresponding stages in predictive processing, the issue remains that different experimental paradigms produce different outcomes regarding identity predictions in the VA domain. However, a solution to this contradiction may lie in the manipulation of congruency of identity. In studies showing an effect of identity on early prediction related potentials, the incongruent trials consisted of many different incongruent VA pairings (Baess et al., 2008; SanMiguel et al., 2013a), whereas studies showing no effect of identity used only a limited number (2 to 4) of different incongruent VA pairings (Klucharev et al., 2003; Stekelenburg & Vroomen, 2007). Considering the results of Experiment 1 of the current study, we speculate that in these latter studies participants could adapt to the violations of visual prediction of identity because a limited number of different incongruent VA pairings was repeated several times. Therefore, participants may have learned to expect either of the few different sounds. This expectation is presumably incorporated in the predictive model. Considering the numerous different incongruent pairings included in the random-identity condition of Experiment 2, no predictive model of identity could be constructed here, and hence no omission responses were elicited. In a future study it would therefore be interesting to test whether in the VA domain N1-suppresion is diminished or abolished if multiple incongruent VA pairings were presented as in the random-identity condition of

Experiment 2. Likewise, it would be of interest to examine if an omission response is induced if natural VA stimuli (allowing precise prediction of timing and identity) are presented in the context of a larger and more varied stimulus set.

The results regarding the natural and random-identity conditions of Experiment 2 are in accordance with the auditory omission study in the MA domain (SanMiguel et al., 2013a). In both studies, auditory omission responses were elicited in the natural condition but not in the random-identity condition. The new finding – besides the fact that we now tested the antecedents of predictive coding in the VA domain instead of the MA domain – is that no omission responses were elicited when a temporal prediction could not be formulated. The studies that explicitly varied auditory onset relative to visual or motor onset all agree with Experiment 2 on the importance of the timing of the to be predicted stimulus (Baess et al., 2008; Vroomen & Stekelenburg, 2010). Based on their results, SanMiguel et al. (2013a), however, did not ascribe a critical role to temporal prediction. Although the role of temporal prediction was not specifically examined in their omission study, the fact that no omission responses were observed when timing – but not identity – of the sound was predictable, led SanMiguel et al. (2013a) to conclude that motor-to-auditory prediction is primarily based on identity, with only a modulatory role for timing. The data of SanMiguel et al. (2013a) and the random-identity condition of the current study indeed suggest that identity is a prominent factor in stimulus prediction. However, if temporal prediction is indeed only of secondary importance, one would expect similar omission responses for both natural and random-timing conditions on the basis of intact identity predictions. In our opinion, the results of Experiment 2 therefore demonstrate that timing *does* play an important role in stimulus prediction, since no omission responses were observed when the timing of the sound was unpredictable. Visual-to-auditory prediction is thus greatly hampered if the auditory onset cannot be predicted from the visual context.

The critical role of timing in predictive models fits within a theory of stimulus prediction in which the brain generates predictions of "when" parallel to "what" (Arnal & Giraud, 2012). Predictive timing ("*when*") and predictive coding ("*what*") are thought of as integral parts of a common framework, although with different functions and underlying neural bases in terms of neural rhythms. The alleged function of predictive timing is to facilitate sensory processing – taking less into account the validity of the prediction – by alignment of low frequency oscillations relative to incoming stimuli. Predictive coding concerns content-specific predictions driven by a combined role of gamma and beta oscillations. The common framework of predictive timing and coding assumes that only when an

event falls inside the expected temporal window the anticipated stimulus is compared to the actual input. Our data concur with this notion of a common framework of timing and identity predictions and demonstrate that reliable prediction of the timing of the anticipated stimulus may serve as a precondition for identity prediction. Experiment 2 demonstrated that intact prediction of either solely timing or identity was insufficient to elicit prediction-related activity, thus indicating that only when the auditory stimulus is correctly timed to its anticipated onset, stimulus-specific predictions can be made. Future studies might investigate the contribution of temporal- versus identity prediction to omission responses in the MA domain by contrasting similar experimental conditions as used in the current study. It should be noted that our study cannot rule out the possibility that, conversely, intact identity prediction is necessary for temporal prediction. However, other electrophysiological studies on intersensory prediction show that visual to auditory prediction – reflected in the suppression of the auditory N1 – does not depend on audiovisual congruency of identity (Klucharev et al., 2003; Stekelenburg & Vroomen, 2007), but is abolished when sound onset could not be accurately predicted from the visual signal (Vroomen & Stekelenburg, 2010). This is in line with our initial interpretation of our data and favors the notion that identity prediction is more dependent on timing prediction than vice versa.

An alternative account for the results of the random-timing condition of Experiment 2 we have to consider, is that stimulus-specific predictions did remain intact, but due to the random onset of the sound the omission responses were jittered over time and smoothed out across the omission ERP. Similarly, it might also be argued that in the random-timing condition, the sensory system develops a set of predictions that corresponds to the ranges of SOAs the participant has been confronted with. In this view, participants thus may expect sounds to occur either too early (i.e., between -250 and -170 ms) or too late (i.e., between +210 and +320 ms). If confronted with an auditory omission, participants may predict by the time that the natural sound would have occurred (typically at 0 ms) that the forthcoming sound will be late – the well-known foreperiod effect, for a review see (Niemi & Naatanen, 1981). Following this reasoning, one expects the oN1 to be elicited starting at approximately 320 ms after the natural sound onset (i.e. the last possible time-point at which a sound may have occurred in the random-timing condition). However, inspection of Figure 3.4 shows that there was no negative deflection in the omission ERP of the random-timing condition within 200 ms after this time-point, which makes the probability of a time-jittered prediction less likely, although future studies might examine this more carefully.

The oN2 and oP3 followed the oN1 in the natural condition, but were absent when the oN1 was not elicited in the random-identity and random-timing conditions. These results mirror those of SanMiguel et al. (2013a, 2013b) in the MA domain, who also report a strict coupling between the oN1 and oN2-oP3. The strong coupling of N1, N2 and P3 components is often found in oddball paradigms (Escera & Corral, 2007), but there is also evidence that a P3 response can be elicited without a concurrent N1-N2 response (Horvath, Winkler, & Bendixen, 2008). The oN2 is thought to reflect a higher-order error evaluation associated with stimulus conflict – in this case a conflict between the visually anticipated sound and the omitted sound. The oP3 probably reflects attention orienting triggered by the unexpected omission of the sound, and the subsequent updating of the internal forward model to minimize future error (Baldi & Itti, 2010; Polich, 2007). Assuming that the oN2 and oP3 are manifestations of processing of stimulus deviancy, the question is why they were not elicited in the random-identity and random-timing conditions? It could be reasoned that, despite the fact that the timing and identity of the sound could not be predicted in the random-timing and random-identity condition, participants still might have perceived the auditory omissions in these conditions as deviant events. Still, because of the severe disruption of the predictive model, the sensory system likely did not assign any significance to stimulus deviancy and failed to see the need for updating of the forward model because there was no viable model to be updated. The dissociation between the oN1 and oN2-oP3 in the current data suggests that stimulus prediction and stimulus deviancy are not processed in parallel, but rather points to a serial organization of different processing stages in which deviant events are only processed in depth if both timing and identity prediction can be formulated (SanMiguel et al., 2013b).

CONCLUSIONS

Auditory omission responses adapted to unnatural VA incongruences – such that they were highly similar to natural VA auditory omission responses – and they abolished if either the timing or the identity of the sound could not be predicted from visual context. Taken together, these findings suggest that predictions of timing and content are both essential elements for stimulus prediction in the VA domain.

REFERENCES

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16*(7), 390-398. https://doi.org/10.1016/j.tics.2012.05.003

Baess, P., Jacobsen, T., & Schroger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology, 70*(2), 137-143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks, 23*(5), 649-666. https://doi.org/10.1016/j.neunet.2009.12.007

Escera, C., & Corral, M. J. (2007). Role of mismatch negativity and novelty-P3 in involuntary auditory attention. *Journal of Psychophysiology, 21*(3-4), 251-264. https://doi.org/10.1027/0269-8803.21.34.251

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research, 14*(1), 20-30. https://doi.org/10.1016/s0926-6410(02)00058-7

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815-836. https://doi.org/10.1098/rstb.2005.1622

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473-490. https://doi.org/10.1162/089892999563544

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468-484. https://doi.org/10.1016/0013-4694(83)90135-9

Horvath, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology, 79*(2), 139-147. https://doi.org/10.1016/j.biopsycho.2008.04.001

Klucharev, V., Mottonen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research, 18*(1), 65-75. https://doi.org/10.1016/j.cogbrainres.2003.09.004

Lochmann, T., & Deneve, S. (2011). Neural processing as causal inference. *Current Opinion in Neurobiology, 21*(5), 774-781. https://doi.org/10.1016/j.conb.2011.05.018

Martikainen, M. H., Kaneko, K., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex, 15*(3), 299-302. https://doi.org/10.1093/cercor/bhh131

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics, 66*(3), 241-251. https://doi.org/10.1007/BF00202389

Niemi, P., & Naatanen, R. (1981). Foreperiod and Simple Reaction-Time. *Psychological Bulletin, 89*(1), 133-162. https://doi.org/10.1037/0033-2909.89.1.133

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

SanMiguel, I., Saupe, K., & Schroger, E. (2013a). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when". *Frontiers in Human Neuroscience, 7*, 407. https://doi.org/10.3389/fnhum.2013.00407

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schroger, E. (2013b). Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience, 33*(20), 8633-8639. https://doi.org/10.1523/JNEUROSCI.5821-12.2013

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*(12), 1964-1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Stekelenburg, J. J., & Vroomen, J. (2015). Predictive coding of visual-auditory and motor-auditory events: An electrophysiological study. *Brain Research, 1626*, 88-96. https://doi.org/10.1016/j.brainres.2015.01.036

Teder-Salejarvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research, 14*(1), 106-114. https://doi.org/10.1016/S0926-6410(02)00065-4

Thorne, J. D., & Debener, S. (2008). Irrelevant visual stimuli improve auditory task performance. *Neuroreport, 19*(5), 553-557. doi: 10.1097/WNR.0b013e3282f8b1b6

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the USA, 102*(4), 1181-1186. https://doi.org/10.1073/pnas.0408949102

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research, 22*(1), 32-35. https://doi.org/10.1016/j.cogbrainres.2004.07.003

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience, 22*(7), 1583-1596. https://doi.org/10.1162/jocn.2009.21308

Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience, 32*(11), 3665-3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Chapter 4

Increased sub-clinical levels of autistic traits are associated with reduced multisensory integration of audiovisual speech

This chapter is published as:

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2019). Increased sub-clinical levels of autistic traits are associated with reduced multisensory integration of audiovisual speech. *Scientific Reports, 9*(1), 9535. https://doi.org/10.1038/s41598-019-46084-0

ABSTRACT

Recent studies suggest that sub-clinical levels of autistic symptoms may be related to reduced processing of artificial audiovisual stimuli. It is unclear whether these findings extent to more natural stimuli such as audiovisual speech. The current study examined the relationship between autistic traits measured by the Autism spectrum Quotient and audiovisual speech processing in a large non-clinical population using a battery of experimental tasks assessing audiovisual perceptual binding, visual enhancement of speech embedded in noise and audiovisual temporal processing. Several associations were found between autistic traits and audiovisual speech processing. Increased autistic-like imagination was related to reduced perceptual binding measured by the McGurk illusion. Increased overall autistic symptomatology was associated with reduced visual enhancement of speech intelligibility in noise. Participants reporting increased levels of rigid and restricted behaviour were more likely to bind audiovisual speech stimuli over longer temporal intervals, while an increased tendency to focus on local aspects of sensory inputs was related to a more narrow temporal binding window. These findings demonstrate that increased levels of autistic traits may be related to alterations in audiovisual speech processing, and are consistent with the notion of a spectrum of autistic traits that extends to the general population.

INTRODUCTION

Autism Spectrum Disorder (ASD) is a pervasive neurodevelopmental disorder characterized by restricted interests, repetitive behaviour and deficits in social communication (American Psychiatric Association, 2013; Robertson & Baron-Cohen, 2017). Although widely reported in ASD (Leekam, Nieto, Libby, Wing, & Gould, 2007), atypical sensory processing was only recently included as a core diagnostic criteria for ASD with the introduction of the DSM-5 (American Psychiatric Association, 2013). Emerging evidence suggests that many of the atypical sensory experiences seen in ASD may stem from a general inability to properly integrate sensory information from multiple modalities into accurate and meaningful percepts (Baum, Stevenson, & Wallace, 2015; Beker, Foxe, & Molholm, 2018; Marco, Hinkley, Hill, & Nagarajan, 2011).

Evidence supporting this notion has been widely – but not exclusively (Brandwein et al., 2015; Russo et al., 2010; Stevenson, Siemann, Woynaroski, et al., 2014b) – reported in studies examining audiovisual speech perception in ASD (Feldman et al., 2018). In typically developing (TD) individuals, integration of multimodal inputs allows the brain to process sensory information more efficiently and provides significant perceptual benefits. Lip-reading under noisy listening conditions, for instance, significantly improves speech intelligibility (Macleod & Summerfield, 1987; Sumby & Pollack, 1954). Compared to TD controls, individuals with ASD tend to benefit less from visual articulatory cues when listening to noise-masked speech, indicating that they show alterations in multisensory integration (MSI) of audiovisual speech (Foxe et al., 2015; J. R. Irwin, Tornatore, Brancazio, & Whalen, 2011; Smith & Bennetto, 2007; Stevenson, Baum, et al., 2017). While visual cues are especially useful under suboptimal listening conditions where the auditory signal is degraded, visual input may also affect auditory perception of clearly audible speech. A prime example of this is the McGurk illusion (Mcgurk & Macdonald, 1976), in which the presentation of an incongruent audiovisual stimulus pairing (e.g. auditory /ba/ visual /ga/) typically induces an illusory percept (e.g. /da/). Previous research shows that individuals with ASD are less susceptible to the McGurk illusion compared to TD controls (Bebko, Schroeder, & Weiss, 2014; de Gelder, Vroomen, & Van der Heide, 1991; Iarocci, Rombough, Yager, Weeks, & Chua, 2010; J. R. Irwin & Brancazio, 2014; Mongillo et al., 2008; Stevenson, Siemann, Woynaroski, et al., 2014a). This reduced perceptual binding suggests that speech perception in ASD is less affected by visual input, and hence more biased towards the auditory modality. A possible underlying cause of these alterations in MSI is that individuals with ASD have impaired temporal binding abilities for multisensory speech signals (Stevenson et al., 2018). To benefit from lip-read cues in audiovisual speech perception, one must be able to assess whether the incoming

auditory and visual information should be integrated into a unified percept. One of the most important cues indicating that multisensory input should be bound together is temporal proximity (Vroomen & Keetels, 2010). Being able to perceive the relative timing of incoming sensory signals from multiple modalities is thus vital to properly integrate audiovisual speech. Several studies have shown that individuals with ASD have reduced temporal acuity and a wider temporal binding window (TBW) for speech stimuli compared to TD controls (de Boer-Schellekens, Eussen, & Vroomen, 2013; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011a; Stevenson et al., 2016). Evidence for an explicit link between multisensory temporal processing and audiovisual perceptual binding is found in TD (Stevenson, Zemtsov, & Wallace, 2012; van Wassenhove, Grant, & Poeppel, 2007) and ASD populations (Stevenson et al., 2018; Stevenson, Siemann, Schneider, et al., 2014), suggesting that the atypical patterns of MSI observed in ASD might indeed be linked to alterations in multisensory temporal processing.

The current study aims to investigate whether autistic traits in the general population are related to MSI. As a spectrum disorder, symptoms of ASD are found to varying degrees in the general population (Ruzich et al., 2015). Given the presumed relationship between MSI and ASD in clinical populations, one might expect that, in the general population, MSI and sub-clinical autistic symptoms are associated as well. However, it is unclear whether there is a steady decrease of MSI with increasing severity of ASD (across subclinical and clinical groups), or if atypical patterns of MSI may only emerge when a certain (clinical) threshold of severity of ASD is exceeded. To our knowledge, only four studies have examined the impact of sub-clinical levels of autistic traits on MSI (Donohue, Darling, & Mitroff, 2012; Stevenson, Toulmin, et al., 2017; Ujiie, Asai, Tanaka, & Wakabayashi, 2015; Ujiie, Asai, & Wakabayashi, 2015). One study examined the relationship between autistic traits and susceptibility to the McGurk illusion in a Japanese sample of 46 TD individuals (Ujiie, Asai, & Wakabayashi, 2015). Autistic traits were assessed via the Adult Autism Spectrum Quotient (AQ) self-report questionnaire. The AQ is a widely used screening instrument for ASD that assesses five subdomains associated with autistic symptomatology: *social skill, attention switching*, *attention to detail*, *communication* and *imagination* (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Wakabayashi et al., 2007). The experiment included auditory, audiovisual congruent and audiovisual incongruent stimulus presentations of the utterances /pa/, /ta/ and /ka/. The results showed that in the incongruent condition, AQ score was *negatively* correlated with the rate of fused (McGurk) responses (e.g. /ta/ in response to auditory /pa/ visual /ka/), but positively correlated with auditory responses (e.g. /pa/ in response to auditory /pa/ visual /ka/). This suggests that – similar to clinical ASD populations – speech perception in TD

individuals with higher AQ scores is less affected by visual input, and more reliant on the auditory modality. However, another study using a similar experimental design but with the addition of background noise found that AQ score was *positively* correlated with fused responses for McGurk stimuli (Ujiie, Asai, Tanaka, et al., 2015). These inconsistencies have not been addressed to date, so it is unclear if these mixed findings are caused by differences in participant populations or experiment-related factors.

Another study examined the relationship between multisensory temporal processing and autistic traits using a simultaneity judgement (SJ) task wherein 101 TD participants reported whether an auditory beep and a visual flash occurred at the same time or not (Donohue et al., 2012). The results showed that the point of subjective simultaneity (PSS) – the stimulus onset asynchrony (SOA) at which a participant most likely perceived the auditory and visual stimuli as occurring simultaneously – was related to autistic traits assessed via the AQ, with the PSS shifting toward an auditory-leading bias as autistic symptoms increased. More specifically, individuals with higher AQ scores and increased difficulties in the ability to switch attention had a stronger tendency to report auditory-leading stimulus presentations as occurring simultaneously. One interpretation of this shift toward an ecologically less valid point is that individuals in the general population with higher levels of autistic traits prioritize auditory information over visual information; which is in line with the presumed over-reliance on the auditory modality observed in ASD (Iarocci et al., 2010; J. R. Irwin & Brancazio, 2014; Stevenson, Siemann, Woynaroski, et al., 2014a). Another explanation for this finding is that individuals with more ASD traits have a decreased ability to infer the probabilistic structure of sensory events. Without a precise internal probabilistic representation of the environment, their perception may be less affected by prior experience and more driven by sensory input (Lawson, Rees, & Friston, 2014; Pellicano & Burr, 2012; van Boxtel & Lu, 2013; van de Cruys et al., 2014). Evidence for this interpretation is found in another study that examined how multisensory temporal adaptation is related to sub-clinical symptoms of ASD measured by the AQ (Stevenson, Toulmin, et al., 2017). Using a statistical learning paradigm including visual flashes and beeps, 60 TD participants were exposed to three-minute adaptation sessions of synchronous, auditory-leading and visual-leading audiovisual stimulus presentations. After exposure to visual-leading stimulus pairings, the participants' perception of synchrony shifted towards visual-leading presentations, as was reported before in TD (Vroomen, Keetels, De Gelder, & Bertelson, 2004). The strength of this temporal recalibration effect was significantly related to the level of autistic traits that participants exhibited. Specifically, an increased tendency to focus on local details of sensory input was related to weaker temporal recalibration. This suggests that individuals with

increased levels of autistic traits are indeed less able to utilize temporal regularities in the environment, and that their perception may thus be less affected by prior expectations and more driven by sensory input.

Taken together, the results of these studies (Donohue et al., 2012; Stevenson, Toulmin, et al., 2017; Ujiie, Asai, Tanaka, et al., 2015; Ujiie, Asai, & Wakabayashi, 2015) suggest that sub-clinical levels of autistic traits are indeed related to alterations in MSI. However, the studies to date that examined the impact of autistic traits on audiovisual temporal processing only used artificial stimuli (i.e. beeps and flashes) (Donohue et al., 2012; Stevenson, Toulmin, et al., 2017), so it is unclear whether the results of these studies extent to more natural stimuli with higher ecological validity such as audiovisual speech. The studies that did use (elementary components) of speech to examine the relationship between autistic traits and MSI yielded inconsistent results (Ujiie, Asai, Tanaka, et al., 2015; Ujiie, Asai, & Wakabayashi, 2015), and it remains to be elucidated whether these mixed results are caused by differences in participant populations or experiment-related factors. Furthermore, AQ subdomain scores were not reported in these studies (Ujiie, Asai, Tanaka, et al., 2015; Ujiie, Asai, & Wakabayashi, 2015), so it is unclear how autistic traits in the general population within each subdomain of autistic symptomatology relate to MSI of audiovisual speech.

Here, we examined the relationship between sub-clinical levels of autistic traits and MSI of audiovisual speech in a large population of TD individuals (N = 101). ASD traits were assessed via the AQ (Baron-Cohen et al., 2001). The primary measures of MSI were audiovisual perceptual binding (assessed with a McGurk task), visual enhancement of speech intelligibility (assessed with a speech-in-noise task) and audiovisual temporal processing (assessed with an SJ task). To our knowledge, this is the first study to link sub-clinical autistic traits to MSI of audiovisual speech using several assessments in a within-subjects design. This approach allowed for a direct comparison of results across paradigms and aimed to further determine the specificity of the relation between autistic symptomatology and MSI in the general population.

METHODS

Participants

A total of 101 undergraduate students (86 female, mean age 20.10 years, *SD* = 2.45 from Tilburg University participated in this study. All participants reported normal hearing and normal or corrected-to-normal vision. None were diagnosed with a neurological disorder and none reported use of medication. All participants received course credits as part of a curricular requirement. Written informed consent was obtained from each participant prior to participation. The study was conducted in accordance with the Declaration of Helsinki. All experimental procedures were approved by the Ethics Review Board of the School of Social and Behavioral Sciences of Tilburg University (EC-2016.48).

ASD trait assessment

Self-reported levels of autistic traits were assessed with the AQ (Baron-Cohen et al., 2001). The AQ is a widely used sensitive and reliable screening measure for autistic traits in the general population(Baron-Cohen et al., 2001; Woodbury-Smith, Robinson, Wheelwright, & Baron-Cohen, 2005), and comprises 50 statements examining five subdomains associated with ASD: *social skill* ("I find it easy to work out what someone is thinking or feeling just by looking at their face"), *attention switching* ("If there is an interruption, I can switch back to what I was doing very quickly"), *attention to detail* ("I often notice small sounds when others do not"), *communication* ("Other people frequently tell me that what I've said is impolite, even though I think it is polite"), and *imagination* ("If I try to imagine something, I find it very easy to create a picture in my mind"). Each scale is represented by 10 statements. Participants were instructed to read each statement very carefully and rate how strongly they agreed or disagreed with it on a 4-point Likert scale (definitely agree, slightly agree, slightly disagree, definitely disagree). Scores for each subscale can range from 0 to 10 and the total score on the questionnaire can range from 0 to 50, with higher scores indicating more symptoms of ASD. Poor social skill, poor communication skill, poor imagination, exceptional attention to detail and problems with attention switching (i.e. exhibiting more rigid and restricted patterns of behaviour) are associated with autistic-like behaviour. A total AQ score ≥ 32 is indicative of ASD (Baron-Cohen et al., 2001). During a subsequent visit to the laboratory on a different day, participants completed a McGurk task, a speech-in-noise task, and an SJ task. Administering the AQ questionnaire and conducting the experimental procedures on two separate occasions ensured that participants were unaware of the fact that their AQ scores were correlated with their performance on the experimental tasks.

Stimuli and experimental procedures

Participants were individually tested in a dimly lit and sound attenuated room and were seated in front of a 25-inch LCD monitor (BenQ Zowie XL2540) positioned at eye-level at a viewing distance of approximately 70 cm. Visual stimuli were presented on the 25-inch LCD monitor at a resolution of 1920 × 1080 pixels and a refresh rate of 240 Hz. Auditory stimuli were recorded at a sampling rate of 44.1 kHz and presented over two loudspeakers (JAMO S100) located directly to the left and the right of the monitor. Stimulus presentation was controlled using E-Prime 3.0 (Psychology Software Tools Inc., Sharpsburg, PA).

McGurk task. Stimuli for the McGurk task were adapted from a previous study on perception of intersensory synchrony in audiovisual speech (Vroomen & Stekelenburg, 2011), and consisted of audiovisual recordings of the pseudowords /tabi/ and /tagi/ pronounced by a male speaker. The entire face was visible on a neutral background and subtended approximately 9.80° horizontal (ear to ear) and 14.65° vertical (hairline to chin) visual angle. Videos were presented at a frame rate of 25 frames/s. Speech sounds were presented at a fixed level of 50 dB sound pressure level (SPL) at ear-level. Trials included audiovisual congruent (auditory /tabi/ visual /tabi/; auditory /tagi/ visual /tagi/) and audiovisual incongruent (auditory /tabi/ visual /tagi/ [fused]; auditory /tagi/ visual /tabi/ [combination]) stimulus pairings. All stimulus pairings were temporally synchronous and had a duration of 2000 ms. Each pairing was presented 15 times in random order (60 trials in total). Participants were instructed to carefully listen to the sounds and attentively watch the speakers lip movements on the monitor. After each trial participants reported what the speaker said by pressing one of four keys, “b,” “d,” “bg” or “g”. Task duration was approximately five minutes.

Speech-in-noise task. Stimulus materials and experimental design were adapted from two previous studies on audiovisual speech perception (van Der Zande, Jesse, & Cutler, 2014; van Laarhoven, Keetels, Schakel, & Vroomen, 2018). Stimuli consisted of audiovisual recordings of 120 different simple mono- and disyllabic nouns pronounced by a male speaker (e.g., tree, room, sugar). The entire face of the speaker was visible on a neutral background and measured approximately 9.80° horizontally (ear to ear) and 14.65° vertically (hairline to chin). Videos were presented at a frame rate of 25 frames/s. Speech sounds were presented at a fixed level of 50 dB SPL at ear-level. Speech sounds were embedded in four levels of pink noise presented at 50, 54, 58 and 62 dB SPL, resulting in SNRs of 0, −4, −8 and −12 dB SPL. Noise onset was synchronized with video onset. The

length of the videos (4 s) and auditory onset (1.5 s after video onset) were identical across all nouns.

Two conditions were included in the speech-in-noise task. In the audiovisual (AV) condition, nouns were presented in conjunction with the corresponding video of the speaker articulating the noun. In the auditory (A) condition, nouns were presented in conjunction with a still image of the speaker's face (with closed mouth). To ensure that participants were less likely to anticipate the experimental condition prior to auditory onset, different still images were created for each noun by extracting still frames from the corresponding videos. A visual-only condition was not included since previous work using the same stimuli reported very low identification scores in unimodal lip-read word recognition(van Der Zande et al., 2014).

Eight of the 120 nouns included in the stimulus set were selected for a practice session that participants completed prior to the main experiment. The remaining 112 nouns were divided into eight subsets of equal size and difficulty. Subset difficulty was based on average viseme overlap (van Der Zande et al., 2014) and proportion of disyllabic versus monosyllabic nouns. Each condition (AV, A) × SNR (0, −4, −8, −12) combination was assigned to one of the eight subsets. The resulting 14 trials for each combination were presented in random order. To reduce possible item-specific effects, eight different stimulus lists were generated and counterbalanced across participants such that each condition × SNR combination was assigned equally to all subsets.

Participants were instructed to attentively listen to the speech sounds and watch the speaker's face, and were informed that a real noun would be presented during each trial. After each trial, participants reported the word they heard using a computer keyboard. Participants were able to correct their answer. Pressing the return key confirmed an answer and initiated the next trial. Participants were allowed to take a short break after every 14th trial to minimize potential fatigue effects. Total duration of the experiment was approximately 20 minutes. After the experiment, a list of all the nouns presented during the experiment was presented. To identify participants with insufficient vocabulary knowledge, participants were instructed to encircle the nouns they did not know the meaning of. Eight pseudowords were intermixed with the regular nouns in the list to control for the possible tendency of participants to report all words as known or not report unknown words as a social desirable response. The average percentage of unknown words was low (M = 3.42, SD = 3.79). No participants were excluded due to insufficient vocabulary knowledge.

Simultaneity judgment task. Stimuli for the simultaneity judgment (SJ) task were adapted from a previous study on perception of intersensory synchrony in audiovisual speech (Vroomen & Stekelenburg, 2011), and consisted of audiovisual recordings of the pseudoword /tabi/ pronounced by a male speaker. The entire face of the speaker was visible on a neutral background and measured approximately 9.80° horizontally (ear to ear) and 14.65° vertically (hairline to chin). SOAs were set relative to the visual onset moment of the speech. A total of 21 SOAs were included: -400, -360, -320, -280, -240, -200, -160, -120, -80, -40, 0, 40, 80, 120, 160, 200, 240, 280, 320, 360, 400 (all values in ms, negative values mean auditory-leading). Fifteen trials were presented for each SOA. The entire task included 315 randomly intermixed trials. After each trial, participants performed a two-alternative forced-choice task in which they indicated whether or not they perceived the presented sound and video as synchronous events. Total duration of the SJ-task was approximately 15 minutes.

RESULTS

Autism Spectrum Quotient

Descriptive statistics of the average total AQ and subscale scores are presented in Table 4.1. Total AQ score ranged from 8 to 32 with a mean of 17.33 (*SD* = 4.88), which is in line with the expected average total AQ score of a non-clinical population(Ruzich et al., 2015).

Table 4.1. Descriptive statistics of the average total AQ and subscale scores

	Mean (SD)	Range
Total AQ (0-50)	17.33 (4.88)	8-32
Social skill (0-10)	2.30 (2.03)	0-9
Imagination (0-10)	2.04 (1.64)	0-7
Attention to detail (0-10)	5.38 (2.29)	1-10
Attention switching (0-10)	4.96 (1.89)	1-9
Communication (0-10)	2.65 (1.92)	0-8

AQ: Autism Spectrum Quotient

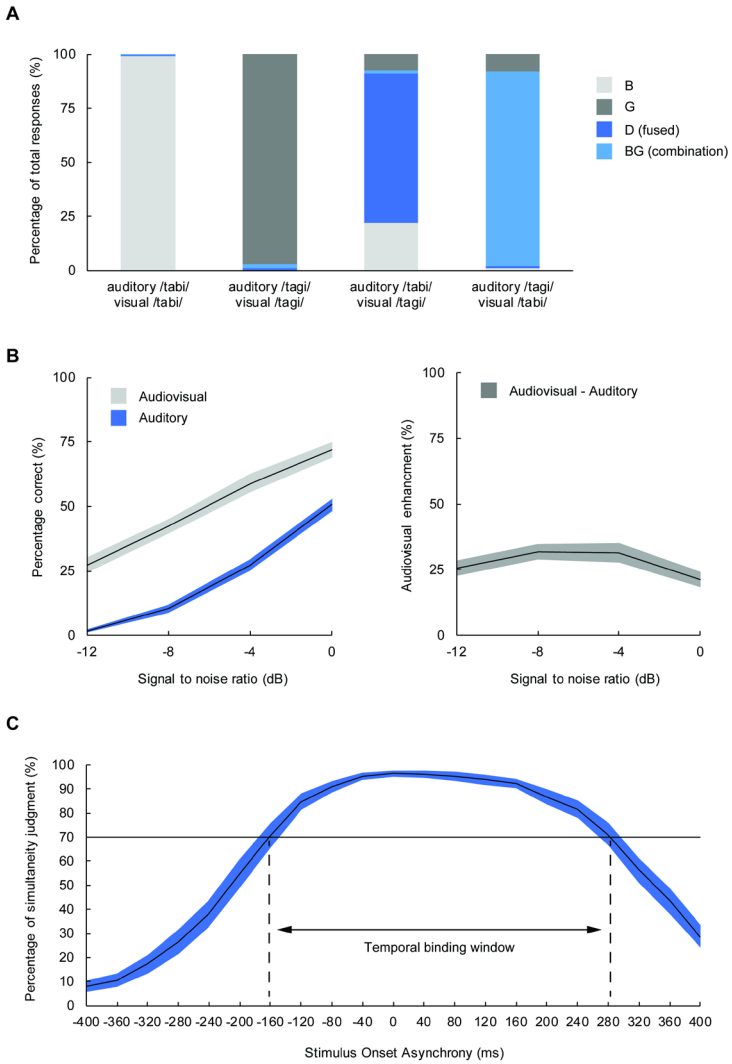


Figure 4.1. Overview of the behavioural data from each experimental task. **A**: Grand average response rates for each audiovisual stimulus pairing presented in the McGurk task. Trials included audiovisual congruent (auditory /tabi/ visual /tabi/; auditory /tagi/ visual /tagi/) and audiovisual incongruent (auditory /tabi/ visual /tagi/ [fused]; auditory /tagi/ visual /tabi/ [combination]) stimulus pairings. Possible responses to the stimuli were auditory (B or G), visual (B or G), fused (D) or combination (BG). **B**: Grand average word recognition performance for each condition (auditory, audiovisual) included in the speech-in-noise task and audiovisual enhancement (audiovisual – auditory performance) as a function of signal-to-noise ratio (SNR). Shaded areas represent 95% confidence intervals. **C**: Grand average percentages of simultaneity judgment (SJ) for each stimulus onset asynchrony (SOA) included in the SJ task. Shaded area represents the 95% confidence interval. For each participant, percentages perceived as synchronous were calculated for each SOA. Two separate logistic curves were fitted on the negative (auditory-leading) and positive (visual-leading) SOAs, respectively. The TBW was calculated for each participant as the difference in ms between the SOAs at which the y-value of the logistic curves equalled 70% (Vroomen & Keetels, 2010).

McGurk task

Mean response rates were calculated for each of the four conditions (Figure 4.1A). Mean percentages auditory responses were submitted to a repeated measures MANOVA with the within subjects factor Stimulus (auditory /tabi/ visual /tabi/; auditory /tagi/ visual /tagi/; auditory /tabi/ visual /tagi/; auditory /tagi/ visual /tabi/). Average percentages auditory responses to the congruent stimulus pairings were high (99% and 97% for auditory /tabi/ visual /tabi/ and auditory /tagi/ visual /tagi/, respectively), indicating that participants were able to correctly identify the syllables for natural stimulus pairings. The MANOVA revealed a main effect of Stimulus $F(3, 98) = 1077.64$, $p < 0.001$, $\eta_p^2 = 0.97$. Post hoc paired samples t-tests (Bonferroni corrected) showed that there was no difference in correct responses between the two congruent stimulus pairings, and that the average percentage auditory responses was lower for incongruent than congruent stimulus pairings (all p values < .001) – indicating the occurrence of the McGurk illusion. Furthermore, the average percentage of auditory responses to the incongruent stimulus pairing auditory /tabi/ visual /tagi/ was higher (22%) than the average percentage of auditory responses to the incongruent stimulus pairing auditory /tagi/ visual /tabi/ (8%); $t(100) = 4.34$, $p < .001$, $d = 0.53$).

To examine the associations between autistic traits and perceptual binding, the percentages of fused and combination responses to the incongruent stimulus pairings (i.e. /tadi/ responses to the stimulus pairing auditory /tabi/ visual /tagi/, and /tabgi/ responses to the stimulus pairing auditory /tagi/ visual /tabi/) were calculated for each participant. In addition, individual percentages of auditory responses to each incongruent stimulus pairing were calculated to examine a potential bias towards the auditory modality. These indices of perceptual binding were then correlated with individual total AQ and subscale scores, and indices of visual enhancement of speech-in-noise and temporal processing (see below).

Speech-in-noise task

Responses were checked for typographical errors and scored as either correct or incorrect. For each participant, percentages of correctly recognized words were calculated for each signal-to-noise ratio (SNR) in both the audiovisual (AV) and auditory (A) condition. Grand average percentages of correct responses for each condition as a function of SNR are shown in Figure 4.1B. A two-way repeated measures MANOVA including the within subjects factors condition (AV, A) and SNR (0, -4, -8, -12 dB) revealed a two-way interaction between these factors $F(3, 98) = 11.22$, $p < 0.001$, $\eta_p^2 = 0.26$. Post hoc paired samples t-tests (Bonferroni corrected) showed that the average percentage

correctly identified words was on average 28% (*SD* = 10.28) higher in the AV condition compared to the A condition at all SNRs (all *p* values < .001), thereby replicating numerous studies showing that observing a speaker's articulatory movements can substantially enhance speech comprehension under suboptimal listening conditions(Macleod & Summerfield, 1987; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954). In accordance with previous research on audiovisual speech perception in noise in ASD(Foxe et al., 2015), visual enhancement of speech intelligibility (AV gain) was indexed for each participant as the difference in percentage correctly recognized words between the AV and A condition (AV − A) averaged across all four SNRs.

Simultaneity judgment task

For each participant, percentages perceived as synchronous were calculated for each SOA. Two separate logistic curves were fitted on the negative (auditory-leading) and positive (visual-leading) SOAs, respectively. The TBW was calculated for each participant as the difference in ms between the SOAs at which the y-value of the logistic curves equalled 70% (Vroomen & Keetels, 2010). Data from 14 participants were excluded from the analyses because their calculated temporal binding window exceeded the boundaries of the SOAs included in the task, indicating that they did not adhere to the task instructions or were unable to perform the task correctly. Simultaneity judgment percentages for each SOA averaged across the remaining 87 participants are shown in Figure 4.1C. The average TBW width was 502.35 ms (*SD* = 138.98), which is similar to previous research on the TBW for audiovisual speech (van Wassenhove et al., 2007).

Correlation effects

Pearson product-moment correlation coefficients (bivariate) were calculated to determine the relationships between the total AQ and subscale scores, AV gain and TBW. Spearman's rank-order correlation coefficients were computed to examine the relationships between the AQ and subscale scores, AV gain, TBW, and perceptual binding indices since the average percentages of reported fused, combination and auditory responses in the McGurk task were not normally distributed. The multiple comparisons problem was addressed with the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) with a false discovery rate of 0.05. Non-significant correlations were further examined with Bayesian correlation tests using a default prior width of 1 (JASP version 0.9.2, https://jasp-stats.org/) to determine if the data support the null hypothesis over the alternative hypothesis. A Bayes Factor (BF_{01}) larger than 1 indicates that the data support the null hypothesis, while a BF_{01} smaller than 1 indicates that the data support the alternative hypothesis. Data were interpreted as anecdotal, moderate, or strong evidence

in favour of the null hypothesis if the BF_{01} was between 1-3, 3-10, and 10-30, respectively (Lee & Wagenmakers, 2013).

Autistic traits. We first examined the correlations between the different AQ subscales. There was a significant relationship between the subscales social skill and attention switching (r = .31, p = .002), and between the subscales social skill and communication (r = .31, p = .002). Since the subscales social skill and communication were not significantly related to any of the measures of MSI, these correlations will not be further discussed here. Correlations between the subscales imagination, attention switching and attention to detail were all non-significant (all p values > .39). Further examination of these non-significant correlations using Bayesian correlation tests provided moderate evidence for the null hypothesis (all BF_{01} between 5.60 and 7.93), indicating that these subscales likely assessed different subdomains of autistic symptomatology.

Correlations between measures of MSI. There was a negative correlation between audiovisual enhancement and percentage of auditory responses to the incongruent McGurk stimulus pairing auditory /tabi/ visual /tagi/ (r_s = -.22, p = .03), but this relationship did not remain significant after adjustment for multiple comparisons using the FDR controlling procedure. A Bayesian correlation test provided only anecdotal evidence for the alternative hypothesis (BF_{01} = 0.67), which suggests that, although the current results could be indicative of a relationship between audiovisual enhancement and auditory responses to incongruent McGurk stimuli, the current data are insensitive to detect a correlation between these indices. There were no other significant correlations between the indices of perceptual binding, audiovisual enhancement, and audiovisual temporal processing (all p values > .05). Bayesian correlation tests provided anecdotal evidence for the null hypothesis for correlations between the indices of perceptual binding (other than auditory responses to incongruent McGurk stimuli, see above) and audiovisual enhancement (all BF_{01} between 1.39 and 3.06.), indicating that the current data were insensitive and therefore unable to provide support for the lack of a relationship between these indices. Bayesian correlation tests did provide moderate evidence for the null hypothesis for correlations between audiovisual temporal processing and audiovisual enhancement, and audiovisual temporal processing and perceptual binding (all BF_{01} between 3.07 and 7.25) – indicating that the temporal processing paradigm likely tapped into different processes of MSI than the paradigms used to assess audiovisual enhancement and perceptual binding, and, importantly, that significant associations between autistic traits and these measures of MSI were likely not interdependent.

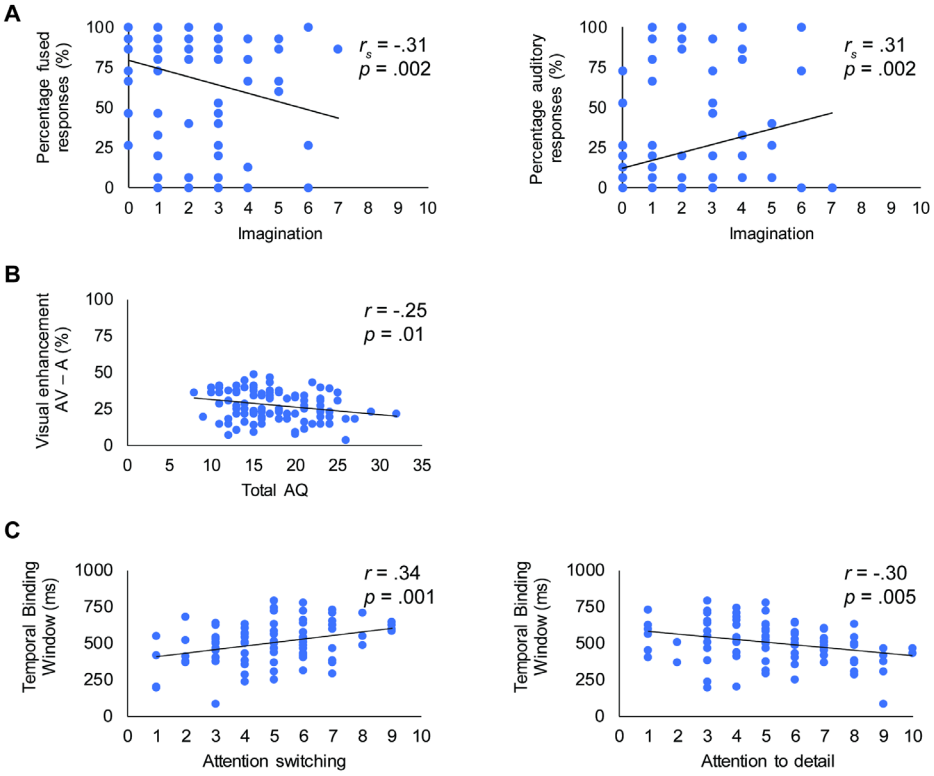


Figure 4.2. Several significant correlations were found between specific subdomains of autistic traits and audiovisual speech processing. **A**: The subscale *imagination* was significantly related to audiovisual perceptual binding of incongruent McGurk stimuli. Individuals reporting higher (more autistic-like) scores on the subscale imagination reported fewer fused responses, but more auditory responses to the incongruent stimulus pairing auditory /tabi/ visual /tagi/ compared to individuals with lower scores on this subscale. **B**: Increased overall autistic-like behaviour (indexed by *total AQ* score) was associated with reduced visual enhancement of speech intelligibility in noise (indexed by audiovisual (AV) – auditory (A) performance). **C**: Participants experiencing increased difficulties with *attention switching* (and exhibiting more rigid and restricted behaviour) were more likely to bind audiovisual speech stimuli over longer temporal intervals, while increased *attention to detail* (i.e. the tendency to focus on local aspects of sensory inputs) was related to a more narrow temporal binding window.

McGurk task. There was no significant correlation between total AQ and indices of perceptual binding assessed by the McGurk task (all *p* values > .05, all BF_{01} between 6.53 and 7.93). However, the subscale *imagination* was significantly related to audiovisual perceptual binding of incongruent McGurk stimuli. Individuals reporting higher (more autistic-like) scores on the subscale imagination reported fewer fused responses (r_s = -.31, *p* = .002), but more auditory responses (r_s = .31, *p* = .002) to the incongruent stimulus pairing auditory /tabi/ visual /tagi/ compared to individuals with lower scores on this subscale (Figure 4.2A). Further examination revealed a negative correlation between the subscale imagination and percentage of combination responses to the incongruent stimulus pairing auditory /tagi/ visual /tabi/ (r_s = -.20, *p* = .04), but this relationship did not remain significant after adjustment for multiple comparisons using the FDR controlling procedure. Bayesian correlation tests provided moderate evidence in favour of the null hypothesis (BF_{01} = 6.03), indicating that the relationship between these indices was indeed non-significant. There were no significant correlations between the other AQ subscales and indices of perceptual binding (all *p* values > .07, all BF_{01} between 5.46 and 8.03).

Speech-in-noise task. *Total AQ* was significantly correlated with visual enhancement of speech embedded in noise (*r* = -.25, *p* = .01). Participants with a higher total AQ showed less AV gain (i.e. AV – A) from lip-read information in the speech-in-noise task (Figure 4.2B). There were no significant correlations between any of the AQ subscales and AV-gain (all *p* values > .05, all BF_{01} between 1.56 and 6.93).

Simultaneity judgment task. There was no significant correlation between total AQ and audiovisual temporal processing indexed by the TBW (*p* = .42, BF_{01} = 5.40). There was, however, a significant relationship between the subscale *attention switching* and TBW (*r* = .34, *p* = .001). Participants experiencing more difficulties with attention switching exhibited a wider TBW. The subscale *attention to detail* was negatively correlated with TBW (*r* = -.30, *p* = .005). Participants with a stronger tendency to focus on small details of sensory input (at the expense of more coherent perceptions) exhibited a more narrow TBW (Figure 4.2C). There was a positive correlation between the subscale social skill and TBW (*r* = .24, *p* = .03), but this relationship did not remain significant after adjustment for multiple comparisons using the FDR controlling procedure. A Bayesian correlation test provided anecdotal evidence for the alternative hypothesis (BF_{01} = 0.66), which suggests that, although the current results could be indicative of a relationship between the subscale social skill and TBW, the current data are insensitive to detect a correlation between these indices. There were no significant correlations between the subscale

imagination and TBW (p = .35, BF_{01} = 4.89), and between the subscale communication and TBW (p = .33, BF_{01} = 4.64).

DISCUSSION

Altered perception of audiovisual speech has been widely reported in ASD, including differences in perceptual binding and temporal processing, and impaired perception of noise-masked audiovisual speech (Feldman et al., 2018). To our knowledge, this study is the first to demonstrate that sub-clinical autistic traits are related to reduced audiovisual speech processing performance across multiple experimental paradigms assessing MSI. Associations between autistic traits and MSI were specific for the subscale *imagination* (reduced perceptual binding of incongruent McGurk stimuli), total *AQ* score (reduced audiovisual gain) and the subscales *attention switching* and *attention to detail* (wider and narrower TBW, respectively). There was no relationship between the subscales imagination, attention switching and attention to detail, or between any of the measures of MSI. The current results therefore demonstrate that autistic traits in TD individuals do not necessarily co-occur in every subdomain within the same individual, which is in line with the notion of a heterogeneous spectrum of ASD symptoms that extends to the general population. Importantly, the current results suggest that each subdomain of autistic traits may affect audiovisual speech processing abilities in a specific way.

Perceptual binding and imagination

Reduced audiovisual perceptual binding − characterized by reduced fused responses to incongruent McGurk stimuli − has been widely reported in ASD (Bebko et al., 2014; de Gelder et al., 1991; Iarocci et al., 2010; J. R. Irwin & Brancazio, 2014; Mongillo et al., 2008; Stevenson, Siemann, Woynaroski, et al., 2014a). Studies on the relationship between autistic traits and susceptibility to the McGurk illusion in the general population have yielded inconsistent results. Some have linked increased levels of autistic traits to reduced fused responses (Ujiie, Asai, & Wakabayashi, 2015), while others showed stronger fused responses for McGurk stimuli embedded in background noise(Ujiie, Asai, Tanaka, et al., 2015). The current study is in accord with previous work relating autistic traits to reduced audiovisual integration of incongruent McGurk stimuli (Ujiie, Asai, & Wakabayashi, 2015), and extends the existing literature by demonstrating that perceptual binding of incongruent audiovisual speech may be related to an individuals' imagination abilities.

In the current study, individuals reporting a more limited (autistic-like) capacity to imagine reported fewer fused responses, but more auditory responses to the incongruent

McGurk stimuli. This reduced perceptual binding behaviour is also found in clinical ASD populations (Bebko et al., 2014; de Gelder et al., 1991; Iarocci et al., 2010; J. R. Irwin & Brancazio, 2014; Mongillo et al., 2008; Stevenson, Siemann, Woynaroski, et al., 2014a), and suggests that audiovisual speech perception in individuals with diminished (autistic-like) imagination abilities may be less affected by visual input, and more reliant on the auditory modality. Another explanation for the observed relationship between reduced susceptibility to the McGurk illusion and autistic-like imagination is that individuals with reduced imagination abilities may have a more literal perception of the world that is less affected by prior experiences, but more reliant on the sensory input (Pellicano & Burr, 2012). Perceptual binding of incongruent audiovisual (i.e. McGurk) stimuli is primarily based on the prior expectation that auditory and visual stimuli that are presented in close spatial and temporal proximity are more likely to originate from the same external event, and should therefore be processed as a single unified percept (Hillock-Dunn & Wallace, 2012; Lewkowicz & Flom, 2014; Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006). Underweighting this prior expectation − 'hypo-priors', in Bayesian terms (Pellicano & Burr, 2012) − could lead to a decreased tendency to automatically bind incongruent audiovisual speech inputs, which in turn may result in a more literal perception of the world in which individual components of audiovisual speech inputs are more likely to be perceived than the unified percept. Given that the auditory component of the McGurk stimuli in the current study was less ambiguous than the visual component, the hypo-priors account might be a plausible explanation of the relationship between autistic-like imagination and increased auditory responses to McGurk stimuli − at the expense of unified (i.e. fused) responses − found in the current study. Indirect evidence for this explanation is reported in a recent study examining recognition accuracy of low-pass filtered and thresholded grayscale images, so-called Mooney images (Mooney, 1957), in relation to autistic traits (van de Cruys, Vanmarcke, Van de Put, & Wagemans, 2018). It was found that individuals with higher scores on the AQ subscale imagination were less likely to recognize Mooney images than those with lower scores, even after exposure to the original source images. This suggests that perception in individuals with more autistic-like imagination is indeed more literal and less susceptible to perceptual change. It should be noted that in one study, autistic traits were linked to *increased* perceptual binding of McGurk stimuli embedded in background noise (Ujiie, Asai, Tanaka, et al., 2015). Further research is therefore needed to examine the underlying mechanisms of the potential link between imagination abilities and perceptual flexibility, and the role of background noise. Still, the current results suggest that increased levels of autistic-like imagination may affect MSI of incongruent audiovisual speech.

Visual enhancement of speech intelligibility in noise and ASD traits

Increased total AQ score was related to reduced visual enhancement of speech intelligibility in noise. Participants with increased levels of autistic-like traits showed less gain from lip-read information when perceiving noise-masked speech. Impaired audiovisual perception of noise-masked speech has been widely reported in ASD (Foxe et al., 2015; J. R. Irwin et al., 2011; Smith & Bennetto, 2007; Stevenson, Baum, et al., 2017). To our knowledge, this study is the first to report a relation between autistic traits and audiovisual speech-in-noise perception in a population of TD individuals, thereby demonstrating that alterations in audiovisual perception may be observed across a spectrum of ASD symptoms that extends to the general population. There was no specific link between audiovisual enhancement and any of the subdomains assessed by the AQ. This suggests that, in the current sample, individual differences in autistic traits within each subdomain may have been too subtle to impact audiovisual speech perception in noise, even though the cumulative impact of autistic traits across subdomains was significant.

Audiovisual temporal processing, attention switching and attention to detail

The current results revealed a relationship between difficulties with attention switching and temporal processing suggesting that individuals with a stronger tendency to show rigid and restricted patterns of behaviour may bind audiovisual speech stimuli over longer temporal intervals. Analogous findings have been reported in a previous study showing that TD individuals with higher total AQ scores and increased difficulties with attention switching were more likely to perceive artificial audiovisual stimuli (i.e. beeps and flashes) as simultaneous when performing an SJ task than individuals with lower total AQ scores and less restricted patterns of behaviour, specifically for auditory-leading stimuli (Donohue et al., 2012). The current results are also in accordance with previous studies in clinical populations demonstrating wider TBWs for audiovisual speech stimuli in individuals with ASD (de Boer-Schellekens et al., 2013; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011b; Stevenson et al., 2016). Taken together, these findings suggest that individuals with increased levels of autistic traits associated with inflexible behaviour tend to have a *wider* TBW for audiovisual stimuli.

Autistic traits in the subdomain attention to detail were also related to temporal processing. Specifically, an increased tendency to focus on local aspects of sensory inputs (at the expense of global information) was associated with a more *narrow* TBW. This positive relationship between ASD traits and temporal precision may sound somewhat counter-intuitive, given that an *enlarged* TBW is generally assumed to reflect *decreased*

temporal acuity (Stevenson et al., 2016) – as it may result in the perceptual binding of stimuli that *should not* be bound together – while a narrow TBW, on the other hand, is assumed to reflect *increased* precision of multisensory temporal processing (Wallace & Stevenson, 2014). These seemingly contradictory results can be reconciled if we consider that *overly precise* temporal processing (i.e. temporal hyperacuity) may lead to the separate processing of stimuli that *should* be bound together. Evidence for this interpretation is reported in previous research on audiovisual temporal recalibration in TD individuals (Stevenson, Toulmin, et al., 2017), which demonstrated that the extent to which the visual-leading side of the TBW is malleable to temporal recalibration is related to the level of autistic symptoms exhibited in the attention to detail subdomain. Given that a certain degree of 'tolerance' to asynchronous (visual-leading) sensory input is required in temporal adaptation, defaulting towards a more narrow TBW may limit the range of temporal recalibration effects. Impairments in temporal recalibration have also been reported in clinical populations of individuals with ASD (Noel, De Niear, Stevenson, Alais, & Wallace, 2017; Turi, Karaminis, Pellicano, & Burr, 2016), although it remains to be elucidated whether these impairments are specifically related to increased attention to detail.

The temporal binding window typically develops asymmetrically to reflect the statistics of the natural environment, where visual input typically arrives at the retina prior to auditory information reaches the cochlea. This results in a steeper slope of the left (auditory-leading) side compared to the right (visual-leading) side (Hillock-Dunn & Wallace, 2012; Hillock, Powers, & Wallace, 2011; Zampini, Guest, Shore, & Spence, 2005). A TBW that is either too wide (diffuse) or too narrow (restricted) poorly reflects the temporal statistics of the environment, and may thus significantly impair an individuals' ability to properly bind multisensory input. The current findings indicate that increased levels of autistic traits may be related to such alterations of the TBW. It has been hypothesized that these alterations of the TBW may have potential cascading effects on the perception of multisensory input such as audiovisual speech (Wallace & Stevenson, 2014). Evidence supporting this notion has been found in recent work in a clinical ASD population (Stevenson, Siemann, Schneider, et al., 2014), were a correlational pathway was found between multisensory temporal processing (indexed by the TBW) and audiovisual perceptual binding (indexed by the McGurk illusion). No such link was observed in the current study. In the aforementioned study (Stevenson, Siemann, Schneider, et al., 2014), no link between temporal processing and perceptual binding was found in a neurotypical control group. The participants included in the current study can be considered to be neurotypical, and, thus, the absence of a direct link between the

indices of MSI in the current study is in line with previous work (Stevenson et al., 2018). This shows that – as mentioned in the introduction – experimental observations in clinical ASD populations do not necessarily translate to individuals in the general population with subclinical autistic traits. For the speech in noise task it should be noted that the stimuli used in the SJ and speech-in-noise task in the current study were of different complexity (i.e. phonemes and nouns, respectively). The absence of a link between temporal processing and visual enhancement of speech intelligibility in noise could thus (in part) be explained by a difference between phoneme and whole-word perception. The lack of a direct link between visual enhancement of speech intelligibility in noise and susceptibility to the McGurk illusion is in line with a recent study showing no relationship between audiovisual sentence recognition in noise and susceptibility to the McGurk illusion (van Engen, Xie, & Chandrasekaran, 2017) – although the failure to find a direct link in the current study should not be considered as evidence that there is no relationship. Nevertheless, the current findings suggest that a cascading pathway of alterations in MSI from impaired temporal processing, through reduced perceptual binding, to impaired speech-in-noise perception is only found in clinical populations of individuals with ASD (Stevenson et al., 2018). Hence, it could be speculated that the associations between the different subdomains of autistic traits and indices of MSI may be reliant on particular thresholds of overall autistic symptomatology. Further research is needed to unravel the various patterns of associations between autistic traits and MSI of audiovisual speech observed in the current study.

The current data suggest that potential subgroups characterized by a particular range of autistic-like behaviours and multisensory functioning may exist in the general population. An interesting avenue of research to pursue would therefore be to examine if similar subgroups can also be identified in clinical populations. Identifying potential subgroups may have important implications for conceptualisations of MSI in ASD. If specific alterations in MSI are indeed linked to distinct subdomains of autistic traits, the impact of these alterations might be reduced by explicit interventions. Previous research has demonstrated that in TD individuals, audiovisual speech-in-noise perception (Song, Skoe, Banai, & Kraus, 2012) and temporal processing (Powers, Hillock, & Wallace, 2009; Stevenson, Wilson, Powers, & Wallace, 2013) can be enhanced with training. However, the impact of training on audiovisual speech perception in ASD is still largely unknown. A recent study demonstrated that individuals with ASD exhibit typical rapid audiovisual temporal recalibration effects for phonemes (Noel et al., 2017), which suggests that the TBW for audiovisual speech in ASD is malleable – although the longer-term effects are still unclear. Still, these findings suggest that audiovisual temporal acuity in ASD may be

susceptible to perceptual training protocols. Another study showed that speech-in-noise performance in children with ASD may improve after extensive app-based audiovisual training (J. Irwin, Preston, Brancazio, D'angelo, & Turcios, 2015). However, the sample size of this study was very small (N = 4), and an untrained control group was not included, so further research is needed to corroborate these results. Still, MSI training in individuals with ASD seems to offer a promising avenue of research, that may ultimately reduce the impact of alterations in MSI on daily life of individuals with ASD.

Study limitations

A limitation of the current study is that a visual-only condition was not included to control for potential individual differences in lip-reading abilities. It may therefore be questioned whether the observed associations between autistic traits and indices of MSI can partly be explained by higher or lower lip-reading abilities. To our knowledge, no study to date has related sub-clinical autistic traits to lip-reading performance. The literature on lip-reading in clinical ASD populations is inconclusive; while some studies have reported reduced lip-reading in ASD (Foxe et al., 2015; Irwin et al., 2011), others have found that lip-reading is intact in ASD and comparable to neurotypical controls (Bebko et al., 2014; de Gelder et al., 1991). Still, this alternative account cannot be ruled out entirely. However, variability in lip-reading abilities likely would have had little effect on the observed links between temporal processing and autistic traits, since lip-reading is not essential for executing audiovisual simultaneity judgments. It is also unlikely that variability in lip-reading ability is solely responsible for the observed association between overall autistic symptomatology and audiovisual enhancement, as previous work has demonstrated that lip-reading abilities are not the driving factor in audiovisual enhancement of speech-in-noise perception in adolescents aged 13-15 years with and without ASD (Foxe et al., 2015). When extrapolating these findings to the current study, we may argue that for the even slightly older participants in the current study, lip-reading abilities are unlikely to explain differences in audiovisual enhancement indexed by the speech-in-noise task.

CONCLUSIONS

The current study replicates previous findings demonstrating that autistic traits are found in varying degrees in the general population (Ruzich et al., 2015). Importantly, this study reports a relationship between autistic traits and multiple indices of MSI of audiovisual speech in a non-clinical population. These findings demonstrate that increased autistic symptomatology may underlie alterations in audiovisual speech processing, not only in clinical populations of individuals with ASD, but also in TD individuals.

REFERENCES

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Baum, S. H., Stevenson, R. A., & Wallace, M. T. (2015). Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Progress in Neurobiology, 134*, 140–160. https://doi.org/10.1016/j.pneurobio.2015.09.007

Bebko, J. M., Schroeder, J. H., & Weiss, J. A. (2014). The McGurk effect in children with autism and asperger syndrome. *Autism Research, 7*(1), 50–59. https://doi.org/10.1002/aur.1343

Beker, S., Foxe, J. J., & Molholm, S. (2018). Ripe for solution: Delayed development of multisensory processing in autism and its remediation. *Neuroscience and Biobehavioral Reviews, 84*, 182–192. https://doi.org/10.1016/j.neubiorev.2017.11.008

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289–300. https://doi.org/10.2307/2346101

Brandwein, A. B., Foxe, J. J., Butler, J. S., Frey, H. P., Bates, J. C., Shulman, L. H., & Molholm, S. (2015). Neurophysiological Indices of Atypical Auditory Processing and Multisensory Integration are Associated with Symptom Severity in Autism. J*ournal of Autism and Developmental Disorders, 45*(1), 230–244. https://doi.org/10.1007/s10803-014-2212-9

de Boer-Schellekens, L., Eussen, M. L. J. M., & Vroomen, J. (2013). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Frontiers in Integrative Neuroscience, 7*, 8. https://doi.org/10.3389/fnint.2013.00008

de Gelder, B., Vroomen, J., & Van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology, 3*(1), 69–86. https://doi.org/10.1080/09541449108406220

Donohue, S. E., Darling, E. F., & Mitroff, S. R. (2012). Links between multisensory processing and autism. *Experimental Brain Research, 222*(4), 377–387. https://doi.org/10.1007/s00221-012-3223-4

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews, 95*, 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H. P., Russo, N. N., Blanco, D., … Ross, L. A. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex, 25*(2), 298–312. https://doi.org/10.1093/cercor/bht213

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science, 15*(5), 688–696. https://doi.org/10.1111/j.1467-7687.2012.01171.x

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia, 49*(3), 461–467. https://doi.org/10.1016/j.neuropsychologia.2010.11.041

Iarocci, G., Rombough, A., Yager, J., Weeks, D. J., & Chua, R. (2010). Visual influences on speech perception in children with autism. *Autism, 14*(4), 305–320. https://doi.org/10.1177/1362361309353615

Irwin, J., Preston, J., Brancazio, L., D'angelo, M., & Turcios, J. (2015). Development of an audiovisual speech perception app for children with autism spectrum disorders. *Clinical Linguistics and Phonetics, 29*(1), 76–83. https://doi.org/10.3109/02699206.2014.966395

Irwin, J. R., & Brancazio, L. (2014). Seeing to hear? Patterns of gaze to speaking faces in children with autism spectrum disorders. *Frontiers in Psychology, 5*, 397. https://doi.org/10.3389/fpsyg.2014.00397

Irwin, J. R., Tornatore, L. A., Brancazio, L., & Whalen, D. H. (2011). Can children with autism spectrum disorders "hear" a speaking face? *Child Development, 82*(5), 1397–1403. https://doi.org/10.1111/j.1467-8624.2011.01619.x

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011a). Altered Auditory and Multisensory Temporal Processing in Autism Spectrum Disorders. *Frontiers in Integrative Neuroscience, 4*. https://doi.org/10.3389/fnint.2010.00129

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011b). Altered Auditory and Multisensory Temporal Processing in Autism Spectrum Disorders. *Frontiers in Integrative Neuroscience, 4*, 129. https://doi.org/10.3389/fnint.2010.00129

Lawson, R. P., Rees, G., & Friston, K. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience, 8*, 302. https://doi.org/10.3389/fnhum.2014.00302

Lee, M. D., & Wagenmakers, E. J. (2013). Bayesian cognitive modeling: A practical course. Bayesian Cognitive Modeling: A Practical Course. https://doi.org/10.1017/CBO9781139087759

Leekam, S. R., Nieto, C., Libby, S. J., Wing, L., & Gould, J. (2007). Describing the sensory abnormalities of children and adults with autism. *Journal of Autism and Developmental Disorders, 37*(5), 894–910. https://doi.org/10.1007/s10803-006-0218-7

Lewkowicz, D. J., & Flom, R. (2014). The Audiovisual Temporal Binding Window Narrows in Early Childhood. *Child Development, 85*(2), 685–694. https://doi.org/10.1111/cdev.12142

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131–141. https://doi.org/10.3109/03005368709077786

Marco, E. J., Hinkley, L. B. N., Hill, S. S., & Nagarajan, S. S. (2011). Sensory processing in autism: A review of neurophysiologic findings. *Pediatric Research, 69*(5), 48R–54R. https://doi.org/10.1203/PDR.0b013e3182130c54

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. https://doi.org/10.1038/264746a0

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(7), 1349–1358. https://doi.org/10.1007/s10803-007-0521-y

Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology, 11*(4), 219–226. https://doi.org/10.1037/h0083717

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science, 9*(5), 454–464. https://doi.org/10.1111/j.1467-7687.2006.00512.x

Noel, J. P., De Niear, M. A., Stevenson, R., Alais, D., & Wallace, M. T. (2017). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Research, 10*(1), 121–129. https://doi.org/10.1002/aur.1633

Pellicano, E., & Burr, D. (2012). When the world becomes “too real”: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual Training Narrows the Temporal Window of Multisensory Binding. *Journal of Neuroscience, 29*(39), 12265–12274. https://doi.org/10.1523/jneurosci.3501-09.2009

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience, 18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*(5), 1147–1153. https://doi.org/10.1093/cercor/bhl024

Russo, N., Foxe, J. J., Brandwein, A. B., Altschuler, T., Gomes, H., & Molholm, S. (2010). Multisensory processing in children with autism: High-density electrical mapping of auditory-somatosensory integration. *Autism Research, 3*(5), 253–267. https://doi.org/10.1002/aur.152

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism, 6*(1), 1–12. https://doi.org/10.1186/2040-2392-6-2

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 48*(8), 813–821. https://doi.org/10.1111/j.1469-7610.2007.01766.x

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to improve hearing speech in noise: Biological mechanisms. *Cerebral Cortex, 22*(5), 1180–1190. https://doi.org/10.1093/cercor/bhr196

Stevenson, R. A., Baum, S. H., Segers, M., Ferber, S., Barense, M. D., & Wallace, M. T. (2017). Multisensory speech perception in autism spectrum disorder: From phoneme to whole-word perception. *Autism Research, 10*(7), 1280–1290. https://doi.org/10.1002/aur.1776

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., Camarata, S., & Wallace, M. T. (2016). Keeping time in the brain: Autism spectrum disorder and audiovisual temporal processing. *Autism Research, 9*(7), 720–738. https://doi.org/10.1002/aur.1566

Stevenson, R. A., Segers, M., Ncube, B. L., Black, K. R., Bebko, J. M., Ferber, S., & Barense, M. D. (2018). The cascading influence of multisensory processing on speech perception in autism. *Autism, 22*(5), 609–624. https://doi.org/10.1177/1362361317704413

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory Temporal Integration in Autism Spectrum Disorders. *Journal of Neuroscience, 34*(3), 691–697. https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014a). Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(6), 1470–1477. https://doi.org/10.1007/s10803-013-1992-7

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014b). Evidence for Diminished Multisensory Integration in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 44*(12), 3161–3167. https://doi.org/10.1007/s10803-014-2179-6

Stevenson, R. A., Toulmin, J. K., Youm, A., Besney, R. M. A., Schulz, S. E., Barense, M. D., & Ferber, S. (2017). Increases in the autistic trait of attention to detail are associated with decreased multisensory temporal adaptation. *Scientific Reports, 7*(1), 14354. https://doi.org/10.1038/s41598-017-14632-1

Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Experimental Brain Research, 225*(4), 479–489. https://doi.org/10.1007/s00221-012-3387-y

Stevenson, R. A., Zemtsov, R., & Wallace, M. T. (2012). Individual Differences in the Multisensory Temporal Binding Window Predict Susceptibility to Audiovisual Illusions. Journal of Experimental Psychology: *Human Perception and Performance, 38*(6), 1517–1529. https://doi.org/10.1037/a0027339

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America, 26*(2), 212–215. https://doi.org/10.1121/1.1907309

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific Reports, 6*, 21756. https://doi.org/10.1038/srep21756

Ujiie, Y., Asai, T., Tanaka, A., & Wakabayashi, A. (2015). The McGurk effect and autistic traits; an analogue perspective. *Letters on Evolutionary Behavioral Science, 6*(2), 9–12. https://doi.org/10.5178/lebs.2015.34

Ujiie, Y., Asai, T., & Wakabayashi, A. (2015). The relationship between level of autistic traits and local bias in the context of the McGurk effect. *Frontiers in Psychology, 6*, 891. https://doi.org/10.3389/fpsyg.2015.00891

van Boxtel, J. J. A., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology, 4*, 19. https://doi.org/10.3389/fpsyg.2013.00019

van de Cruys, S., Evers, K., van der Hallen, R., van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review, 121*(4), 649–675. https://doi.org/10.1037/a0037665

van de Cruys, S., Vanmarcke, S., Van de Put, I., & Wagemans, J. (2018). The Use of Prior Knowledge for Perceptual Inference Is Preserved in ASD. *Clinical Psychological Science, 6*(3), 382–393. https://doi.org/10.1177/2167702617740955

van Der Zande, P., Jesse, A., & Cutler, A. (2014). Hearing words helps seeing words: A cross-modal word repetition effect. *Speech Communication, 59*, 31–43. https://doi.org/10.1016/j.specom.2014.01.001

van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention, Perception, and Psychophysics, 79*(2), 396–403. https://doi.org/10.3758/s13414-016-1238-9

van Laarhoven, T., Keetels, M., Schakel, L., & Vroomen, J. (2018). Audio-visual speech in noise perception in dyslexia. *Developmental Science, 21*(1), e12504. https://doi.org/10.1111/desc.12504

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. Attention, *Perception, and Psychophysics, 72*(4), 871–884. https://doi.org/10.3758/APP.72.4.871

Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research, 22*(1), 32–35. https://doi.org/10.1016/j.cogbrainres.2004.07.003

Vroomen, J., & Stekelenburg, J. J. (2011). Perception of intersensory synchrony in audiovisual speech: Not that special. *Cognition, 118*(1), 78–86. https://doi.org/10.1016/j.cognition.2010.10.002

Wakabayashi, A., Baron-Cohen, S., Uchiyama, T., Yoshida, Y., Tojo, Y., Kuroda, M., & Wheelwright, S. (2007). The Autism-Spectrum Quotient (AQ) children's version in Japan: A cross-cultural comparison. *Journal of Autism and Developmental Disorders, 37*(3), 491–500. https://doi.org/10.1007/s10803-006-0181-3

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64,* 105–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger Syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, *35*(3), 331–335. https://doi.org/10.1007/s10803-005-3300-7

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception and Psychophysics*, *67*(3), 531–544. https://doi.org/10.3758/BF03193329

Chapter 5

Electrophysiological alterations in motor-auditory predictive coding in autism spectrum disorder

This chapter is published as:

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L. J. M., & Vroomen, J. (2019). Electrophysiological alterations in motor-auditory predictive coding in autism spectrum disorder. *Autism Research, 12*(4), 589–599. https://doi.org/10.1002/aur.20870

ABSTRACT

The amplitude of the auditory N1 component of the event-related potential (ERP) is typically attenuated for self-initiated sounds, compared to sounds with identical acoustic and temporal features that are triggered externally. This effect has been ascribed to internal forward models predicting the sensory consequences of one's own motor actions. The predictive coding account of autistic symptomatology states that individuals with autism spectrum disorder (ASD) have difficulties anticipating upcoming sensory stimulation due to a decreased ability to infer the probabilistic structure of their environment. Without precise internal forward prediction models to rely on, perception in ASD could be less affected by prior expectations and more driven by sensory input. Following this reasoning, one would expect diminished attenuation of the auditory N1 due to self-initiation in individuals with ASD. Here, we tested this hypothesis by comparing the neural response to self- versus externally-initiated tones between a group of individuals with ASD and a group of age matched neurotypical controls. ERPs evoked by tones initiated via button-presses were compared with ERPs evoked by the same tones replayed at identical pace. Significant N1 attenuation effects were only found in the TD group. Self-initiation of the tones did not attenuate the auditory N1 in the ASD group, indicating that they may be unable to anticipate the auditory sensory consequences of their own motor actions. These results show that individuals with ASD have alterations in sensory attenuation of self-initiated sounds, and support the notion of impaired predictive coding as a core deficit underlying autistic symptomatology.

INTRODUCTION

Autism Spectrum Disorder (ASD) is a pervasive neurodevelopmental disorder characterized by deficits in social communication and social interaction and restricted, repetitive patterns of behavior, interests or activities (American Psychiatric Association, 2013; Robertson & Baron-Cohen, 2017). ASD has been linked to a range of sensory processing atypicalities, including atypical processing of faces and emotional stimuli (Eussen et al., 2015; Harms, Martin, & Wallace, 2010; Pellicano, Jeffery, Burr, & Rhodes, 2007; Uljarevic & Hamilton, 2013) and hyper- and hyposensitivity to perceptual stimuli (Baranek et al., 2013; Robertson & Baron-Cohen, 2017). Emerging evidence suggests that many of these atypical sensory experiences reported in ASD may stem from a more general inability to properly integrate sensory information from different sensory sources into accurate and meaningful percepts (Baum, Stevenson, & Wallace, 2015; Beker, Foxe, & Molholm, 2018; Marco, Hinkley, Hill, & Nagarajan, 2011). Given that sensory cues play a central role in human perception and social interaction, understanding the basis of the atypicalities in sensory processing seen in ASD may very well be a fundamental part of the explanation of why individuals with ASD often struggle with social communication and interaction with their environment.

A recently proposed theory that attempts to account for these symptoms, posits that individuals with ASD have impaired predictive coding abilities (Lawson, Rees, & Friston, 2014; Pellicano & Burr, 2012; van Boxtel & Lu, 2013; Van de Cruys et al., 2014). A key element of the predictive coding theory is the assumption that our brain is constantly generating predictions about the current state of our environment based on previous sensory experience. Collectively, these predictions – or prior expectations, in Bayesian terms – form our internal representation of the world (Friston, 2005; Mumford, 1992). This internal forward model can be thought of as a probabilistic map that is used to contextualize and inform our perception (Baum et al., 2015; Lawson et al., 2014). Sensory input is continuously contrasted with our internal predictions. The discrepancy between the sensory input and predictions is reflected in the prediction error (Friston, 2005). Any unexpected or otherwise informative information is stored in this prediction error, which is then passed up to higher cortical areas, where it is used to readjust and improve the forward model to minimize prediction errors in the future. These predictive mechanisms allow us to anticipate upcoming sensory stimulation and distinguish between expected and unexpected events. The predictive coding account of ASD states that individuals with ASD have a decreased ability to infer the probabilistic structure of their environment (Lawson et al., 2014; Pellicano & Burr, 2012; van Boxtel & Lu, 2013; Van de Cruys et al., 2014). As a result, they do not possess a precise internal predictive representation of the

world around them and may therefore fail to contextualize sensory information in an optimal fashion. Given that statistical learning is vital for acquisition of sensory associations and multisensory integration (Mitchel, Christiansen, & Weiss, 2014; Mitchel & Weiss, 2011; Seitz, Kim, Van Wassenhove, & Shams, 2007), impairments in this process will likely have cascading effects on sensory processing, perception and social interaction.

One of the most rudimentary predictive coding mechanisms is the ability to distinguish between self-initiated and external sensory events. This ability is crucial for effective and efficient perceptual organization and interaction with the environment, and has been ascribed to an efference copy/corollary discharge mechanism that enables us to anticipate the sensory consequences of our own motor actions (for review, see Crapse & Sommer, 2008). A frequently applied approach to examine this predictive mechanism is by recording auditory potentials in a motor-sensory prediction paradigm. Several studies have shown that the amplitude of the auditory N1 is typically attenuated for self-initiated sounds, compared to sounds with identical acoustic and temporal features that are triggered externally (Baess, Horváth, Jacobsen, & Schröger, 2011; Baess, Jacobsen, & Schröger, 2008; Bendixen, SanMiguel, & Schröger, 2012; Martikainen, Kaneko, & Hari, 2005). Within the predictive coding framework, the amplitude of the auditory N1 is assumed to be modulated by the prediction error (Arnal & Giraud, 2012; Friston, 2005). When an incoming sound matches the prediction, the prediction error is small and thus the amplitude of the auditory N1 is attenuated. For unexpected sounds the prediction error is more pronounced and so the amplitude of the auditory N1 is enlarged. Since self-initiated sounds are typically experienced as more predictable than externally-initiated sounds, the prediction error, and hence the N1, for such sounds is typically smaller. From a predictive coding perspective, the N1 attenuation effect for self-initiated sounds can thus be explained as an attenuation of the prediction error caused by the internal forward model correctly predicting the auditory consequences of one's own motor actions (Martikainen et al., 2005).

If predictive coding is truly impaired in ASD, and individuals with ASD do indeed lack a precise internal forward model to rely on, then perception in ASD is presumably less affected by prior expectations and more driven by sensory input. Following this reasoning, one would expect diminished or absent attenuation of early auditory neural responses by motor-to-auditory prediction mechanisms in individuals with ASD. To our knowledge, this has never been formally tested. Hence, the current study examined the neural response to self- versus externally-initiated sounds in individuals with ASD. An experimental paradigm was applied that was similar to those used in previous studies showing robust and consistent motor-to-auditory N1 attenuation effects in neurotypical

individuals (Baess et al., 2008; Martikainen et al., 2005). EEG was recorded in a group of older adolescents and young adults with a clinical diagnosis of ASD and in a group of age matched controls with typical development (TD). Motor-to-auditory N1 attenuation was examined by comparing event-related potentials (ERPs) evoked by tones initiated via button-presses with ERPs evoked by the same tones replayed at an identical pace. Differences between ERPs evoked by self- versus externally-initiated tones were interpreted as top-down prediction effects (Baess et al., 2011, 2008; Martikainen et al., 2005). Diminished or absent N1 attenuation, as a neural marker for motor-sensory predictions, was considered as evidence for impaired predictive coding mechanisms.

METHODS

Participants

Thirty individuals with ASD (8 female, mean age 18.55 years, SD = 2.13) and 30 individuals with TD (6 female, mean age 18.83 years, SD = 1.32) participated in this study.

Inclusion criteria for participants in both groups were: between 15 and 25 years of age, Full Scale IQ (FSIQ) >=80, normal or corrected-to-normal vision and hearing, absence of physical disabilities and no active use of sedatives two days prior to the experiment. Additional inclusion criteria for the ASD group were: a clinical DSM-IV TR classification of ASD (American Psychiatric Association, 2000) and absence of severe comorbid neurological disorders (e.g. epilepsy). Additional inclusion criteria for the TD group were: absence of any neurological or neuropsychiatric disorder (e.g. ASD, ADHD, epilepsy).

Participants with ASD were recruited at a mental health institution for ASD (de Steiger, Yulius Mental Health, Dordrecht, the Netherlands). At the time of the experiment, all participants in the ASD group were receiving clinical treatment at this mental health institution due to severe mental problems and impaired functioning in activities of daily living linked to ASD. Participants with TD were recruited at Tilburg University and a high school located in the city of Tilburg.

Table 5.1. Demographics for the Autism Spectrum Disorder (ASD) and Typically Developing (TD) Group

	ASD	TD
Gender $^{n.s.}$	22 male, 8 female	24 male, 6 female
Age $^{n.s.}$	18.55 (2.13)	18.83 (1.32)
Full Scale IQ *	103.00 (16.47)	111.97 (11.49)
ADOS	10.11 (5.04) $n = 18$	-
SRS	72.91 (9.68) $n = 22$	-

$^{n.s.}$ non-significant $^{*}p < .05$ values within parenthesis represent standard deviations

For all participants in the ASD group the clinical DSM-IV TR classification of ASD was confirmed by two independent clinicians. Additional diagnostic information was retrieved when available, including Autism Diagnostic Observation Schedule (ADOS) scores (Lord et al., 2012) and Social Responsiveness Scale (SRS) scores (Constantino & Gruber, 2012). FSIQ was measured with the Dutch versions of the Wechsler Adult Intelligence Scale (WAIS-IV-NL) in participants ≥ 18 years, and the Wechsler Intelligence Scale for Children (WISC-III-NL) in participants < 18 years. Demographic details of the ASD group and the TD control group are shown in Table 5.1. There were no differences in age and gender but the average FSIQ score was higher for the TD group (mean FSIQ 111.97, *SD* = 11.49) compared to the ASD group (mean FSIQ 103.00, *SD* = 16.47), $t(58) = 2.45$, $p = .02$. All procedures were undertaken with the understanding and written consent of each participant and – for participants under the age of 18 – a parent or another legally authorized representative. Participants with ASD and TD participants that were recruited at the high school were reimbursed with 25 EUR for their participation. TD participants recruited at Tilburg University received course credits as part of a curricular requirement. All experimental procedures were approved by the local medical ethical review board (METC Brabant, protocol ID: NL52250.028.15) and performed in accordance with the ethical standards of the Declaration of Helsinki.

Stimuli and procedure

Participants were individually tested in a dimly lit and sound attenuated room and were seated in front of a 19-inch CRT monitor (Iiyama Vision Master Pro 454) positioned at eye-level at a viewing distance of approximately 70 cm. To ensure that the pace of motor actions was comparable across participants, each participant completed a training session prior to the experiment in which they were trained to adapt their pace to approximately 3000 ms. At the start of the training session, eight 50 ms pure tones of 1000 Hz with an inter stimulus interval of 3000 ms were presented at 70 db(A) through two loudspeakers located directly to the left and the right of the monitor. Previous motor-auditory prediction studies typically use headphones for auditory stimulus presentation; however, in the current study loudspeakers were preferred over headphones because they were less obtrusive for the participants in the ASD group. Participants were required to press the left button of a silent mouse with their right index finger in synchrony with the tones, and to continue to press at the same pace after the end of the tone sequence. After 20 button presses (including the 8 pacing tones) their mean press interval was presented on the monitor. When the mean interval deviated more than 1500 ms from the required 3000 ms pace, participants were encouraged to speed up or slow down their pace accordingly. The training session was repeated twice for each participant.

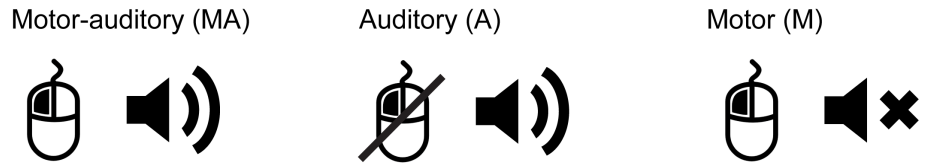


Figure 5.1. Schematic illustration of the three experimental conditions. In the motor-auditory (MA) condition, tones were self-initiated via a button press and the inter-tap-interval was recorded. In the Auditory (A) condition, the tones were presented at the exact pace of the MA condition and no button presses were allowed. In the motor (M) condition, participants were required to press the button at the same pace as in the MA task, but no tones were presented after each button press

Three conditions were included in the experiment: motor-auditory (MA), auditory (A) and motor (M) (Figure 5.1). In the MA condition, participants pressed the left mouse button and were encouraged to maintain the previously trained pace of about 3000 ms. After each button press, a 50 ms pure tone of 1000 Hz was presented. Due to hardware restrictions, the temporal delay between the button press and onset of the sound was ~20 ms, which is below the typical detection threshold of motor-auditory delays (Van Vugt & Tillmann, 2014). The inter-press-interval of the MA condition was recorded to ensure that in the auditory (A) condition, the tones were presented at the exact pace of the MA condition. No button presses were allowed in the A condition and participants were required to refrain from moving their hands, head, fingers or feet in synchrony with the tones. In the motor (M) condition, participants were required to press at the same pace as in the MA condition, but no pure tones were presented after each button press. This condition served as a control condition to rule out the possibility of mere motor activity being a confounder for the expected differences between the A and MA condition (Baess et al., 2008). Each condition consisted of 120 trials divided across 2 blocks of 60 trials. Block order was quasi-randomized across participants with the restrictions that an A block was always preceded by an M and MA block, or an MA and M block. Stimulus presentation and button press performance logging was controlled using E-Prime 1.2 (Psychology Software Tools Inc., Sharpsburg, PA, USA).

To prevent visual EEG activity associated with motor actions, participants were asked to fix their gaze to the monitor and to refrain from looking at the mouse. Participants constantly held their right index finger on the left mouse button and produced mostly isometric muscle contractions without raising their finger before pressing the button to ensure no finger movements were visible in the peripheral visual field. To prevent auditory EEG activity induced by the button presses, we used a mouse specifically designed to produce no clear audible clicks (Nexus SM-9000). Unlike the switches used in

a conventional mouse, the switches used in this mouse lack the typical "click" sound when pressed. In addition, white noise (Hewlett Packard 8057A Precision Noise Generator) was presented during the entire experiment at approximately 60 dB(A) through a single small speaker located at 10 cm behind the mouse, which masked any faint sound originating from the finger movement.

EEG acquisition and processing

The EEG was sampled at 512 Hz from 64 locations using active Ag-AgCl electrodes (BioSemi, Amsterdam, the Netherlands) mounted in an elastic cap and two mastoid electrodes. Electrodes were placed in accordance with the extended International 10-20 system. Two additional electrodes served as reference (Common Mode Sense active electrode) and ground (Driven Right Leg passive electrode). Horizontal electrooculogram (EOG) was recorded using two electrodes placed at the outer canthi of the left and right eye. Vertical EOG was recorded from two electrodes placed above and below the right eye. BrainVision Analyzer 2.0 (Brain Products, Gilching, Germany) and BESA Statistics 2.0 (Brain Electrical Source Analysis, Gräfelfing, Germany) software were used for ERP analyses. EEG was referenced offline to an average of left and right mastoids and band-pass filtered (0.01-30 Hz, 24 dB/octave). The (residual) 50 Hz interference was removed by a 50 Hz notch filter. Raw data were segmented into epochs of 600 ms, including a 200-ms pre-stimulus baseline period. Epochs were time-locked to the sound onset in the MA and A conditions, and to the corresponding timestamp in the M condition. After EOG correction (Gratton, Coles, & Donchin, 1983), epochs with an amplitude change exceeding ±150 μV at any EEG channel were rejected and subsequently averaged and baseline corrected for each condition separately. On average 5.35 percent (*SD* = 7.40) of the trials were rejected. There were no significant differences in rejected trials between groups or conditions (A: TD 4.92, ASD 5.81, MA: TD 3.78, ASD 6.58, M: TD 4.39 ASD 6.61). To facilitate a direct comparison between the A and MA condition, the ERP of the M condition was subtracted from the MA ERP to nullify the contribution of motor activity (Baess et al., 2008; Stekelenburg & Vroomen, 2015).

Time windows and regions of interest

The group-averaged auditory-evoked ERPs showed clearly identifiable N1 and P2 responses in the A and MA – M condition in both groups (Figure 5.2A, 5.2B). Visual inspection of the ERPs showed that only in the TD group, the N1 was attenuated for self-generated tones in the MA condition compared to the same tones replayed in the A condition. The ERPs from both the ASD and TD group showed that the P2 in the MA condition was attenuated and speeded up compared to the A condition.

To test these observations more formally, a cluster-based non-parametric permutation procedure was performed to identify time windows and regions of interest for the N1 and P2 (Maris & Oostenveld, 2007). Difference waveforms reflecting motor-to-auditory prediction effects were computed for each group by subtracting MA – M ERPs from A ERPs (i.e. A – MA – M). The time-course of the difference waveforms of the two groups was compared in the latency range from -200 to 400 ms with a preliminary point-wise independent samples t-test identifying clusters that included data points that fell below the cluster alpha level ($p < .05$). For each identified cluster, a cluster value was calculated by taking the sum of all the t-values of all data points within that cluster. This preliminary clustering procedure was followed by a permutation procedure that randomly interchanged the cluster values 1000 times. For each permutation, new clusters were identified and the according cluster values were derived. Finally, a new distribution of cluster values was established across all permutations. Clusters were considered significant if the probability of observing a larger cluster value in the new distribution was below the significance level of .05.

Clusters revealing significant between group differences in motor-to-auditory prediction effects were further explored by comparing ERPs for each condition (A, MA – M) within each group using cluster based permutation tests with parameters similar to those used to examine the A – MA – M difference waveforms. Regions of interest were defined based on the scalp topographies of the time windows identified by the permutation procedures.

RESULTS

Behavioral performance

The average button press interval in the MA and M conditions was 2987.30 ms (SD = 688.34 ms) and 3133.25 ms (SD = 679.35 ms), respectively. Average press intervals for each group and condition were submitted to a repeated measures MANOVA with the within-subjects variable Condition (MA, A) and between-subjects factor Group (ASD, TD). The MANOVA produced a significant Condition × Group interaction $F(1, 58) = 6.51$, $p = .01$, $\eta_p^2 = .10$. Simple main effects tests revealed that for the ASD group, the average press interval was slightly faster (~265 ms) in the MA condition compared to the M condition $F(1, 29) = 16.15$, $p < .001$, $\eta_p^2 = .22$. However, the average press interval during all conditions was within the required range of 2500 – 3500 ms, indicating that participants were able to maintain the required pressing pace throughout the entire experiment.

Between group differences in motor-to-auditory prediction (A – MA – M)

The cluster-based permutation test revealed a time window of interest for the N1 in the latency range from 110-130 ms showing a significant difference (p = .03) between the ASD and TD group that was most pronounced over fronto-central electrodes (Figure 5.3A). No other time windows of interest were identified, indicating that the difference in mean activity between self- versus externally-initiated tones in the P2 latency range was similar for both groups.

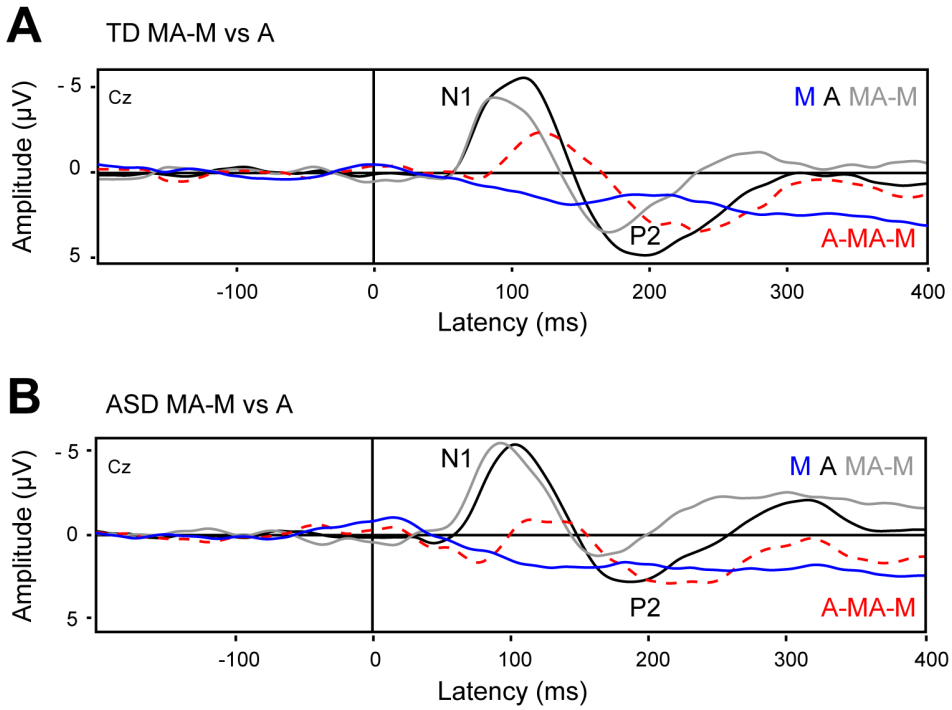


Figure 5.2. Group-averaged auditory-evoked ERPs for the auditory (A) and motor-auditory (MA – M) condition for the TD group (**A**) and ASD group (**B**). Motor-auditory ERPs were corrected for motor activity via subtraction of the motor (M) waveform. ERPs were time-locked to the sound onset in the MA and A conditions, and to the corresponding timestamp in the M condition.

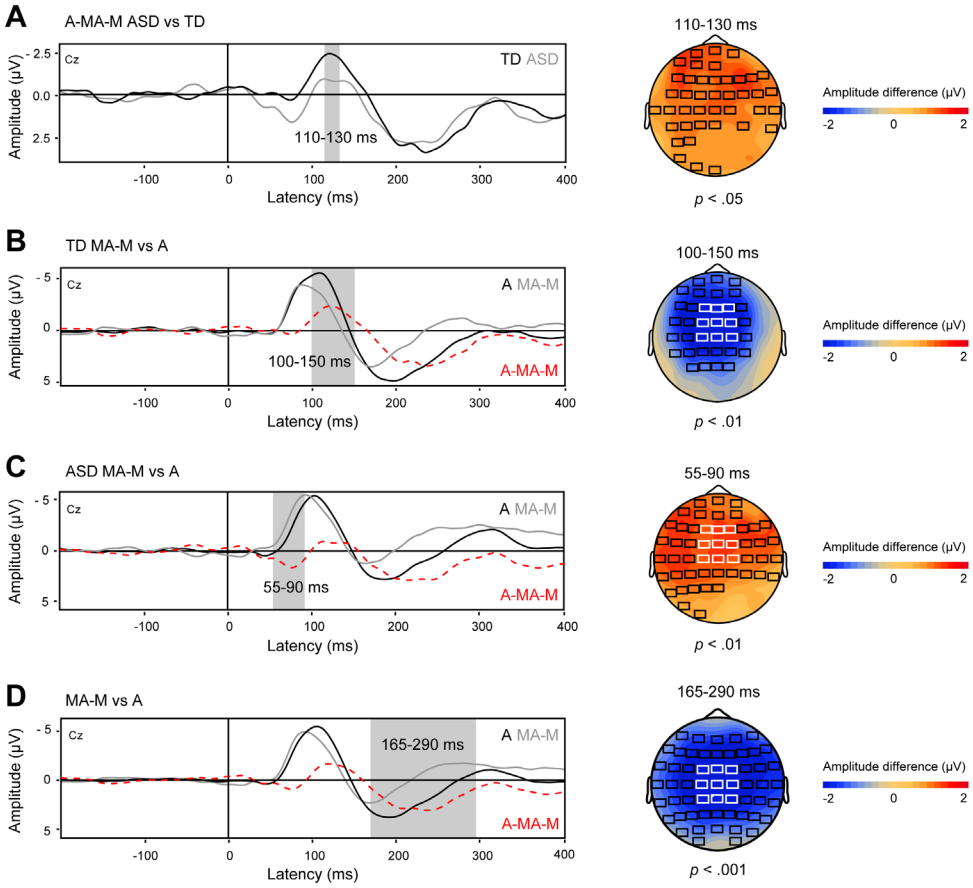


Figure 5.3. Results of the cluster-based permutation tests. **A**: Group-averaged difference waveforms reflecting motor-to-auditory prediction effects were computed for each group by subtracting MA – M ERPs from A ERPs (i.e. A – MA – M). Waveforms were time-locked to the sound onset in the A and MA conditions, and to the corresponding timestamp in the M condition. A time window of interest was identified in the latency range from 110-130 ms showing a significant difference (p = .03) between the ASD and TD group that was most pronounced over fronto-central electrodes. The between-group difference in the 110-130 ms time window was further explored by comparing ERPs for each condition within each group (**B** and **C**). **B**: For the TD group, a time window of interest was identified in the latency range from 100-150 ms indicating a significant difference between the A and MA – M condition that was most pronounced over fronto-central electrodes. **C**: For the ASD group, an earlier time window of interest was identified in the latency range from 55-90 ms indicating a significant difference between conditions that was most pronounced over fronto-central electrodes. **D**: Waveforms reflecting overall neural activity across groups were computed for each condition to examine differences in P2 mean activity between the A and MA – M condition. A time window of interest in the latency range from 165-290 ms was revealed showing a significant difference between the A and MA – M condition that was most pronounced over central electrodes. **Scalp topographies**: Black rectangles indicate electrodes showing a significant difference in motor-to-auditory prediction effects (**A**) or a significant difference in mean activity between the A and MA – M condition (**B**, **C** and **D**). White rectangles depict electrodes included in confirmatory parametric analysis.

N1 responses to self- versus externally-initiated tones

N1 time window. To further explore the between-group difference in the 110-130 ms time window of interest for the N1, ERPs for each Condition (A, MA – M) were compared within each group using cluster based permutation tests similar to those used to examine the A – MA – M difference waveforms. For the TD group, the permutation tests revealed a significant difference between the A and MA – M condition in the latency range from 100-150 ms (Figure 5.3B). Mean activity in this time window was significantly attenuated for self-initiated compared to externally-initiated tones ($p < .01$). Importantly, this time window showed substantial overlap with the previously identified 110-130 ms time window of interest. For the ASD group, there was no significant difference between conditions in the 110-130 ms time window. However, an earlier time window of interest was identified (Figure 5.3C). Mean activity in the latency range from 55-90 ms was significantly *increased* (i.e. more negative) for self-initiated compared to externally-initiated tones ($p < .01$). Given the morphology of the ERPs, this increase in N1 mean activity likely reflects a difference in onset and latency.

To further examine the observed amplitude and latency effects, additional confirmatory parametric testing was carried out on the peak amplitude and peak latency values in the latency range from 55-150 ms. This latency range was selected to include the previously identified time windows of interest for each group (i.e. ASD: 55-90 ms, TD: 100-150 ms). Based on the scalp topographies of the time windows identified by the permutation procedure (Figure 5.3B, 5.3C), a fronto-central region of interest (ROI) including nine electrodes with FCz at its center was defined. Individual N1 peak amplitude and peak latency values within the 55-150 ms time window were calculated for each condition and electrode and submitted to repeated measures MANOVAs with the within-subjects variables Condition (A, MA – M) and Electrode (Cz, C1, C2, FCz, FC1, FC2, Fz, F1, F2) and between-subjects factor Group (ASD, TD).

N1 amplitude. The MANOVA for N1 amplitude produced a significant Condition × Group interaction, $F(1, 58) = 5.70$, $p = .02$, $\eta_p^2 = .09$ and a main effect of Electrode, $F(8, 51) = 18.32$, $p < .001$, $\eta_p^2 = .74$. The main effect of Electrode was further examined with post hoc paired samples t-tests (Bonferroni corrected), which showed that N1 amplitude was less negative at C1, Cz and C2 than at FC1, FCz, FC2, Fz and F2 (all p values < .05), and less negative at F1 than at FCz, Fz and F2 (all p values < .05). The Condition × Group interaction was further explored with simple main effects tests examining the effect of Condition within each Group. For the TD group, there was a main effect of Condition, $F(1, 29) = 8.06$, $p < .01$, $\eta_p^2 = .12$, indicating that the amplitude of the auditory N1 was significantly

attenuated for self-initiated tones in the MA – M condition compared to the same tones replayed in the A condition. There was no main effect of Condition for the ASD group, $F(1, 29) = 0.29$, $p = .59$, $\eta_p^2 = .005$, indicating that self-initiation of the sound did not modulate the amplitude of the auditory N1 (see Figure 5.4 for individual N1 amplitude differences between the A and MA – M condition).

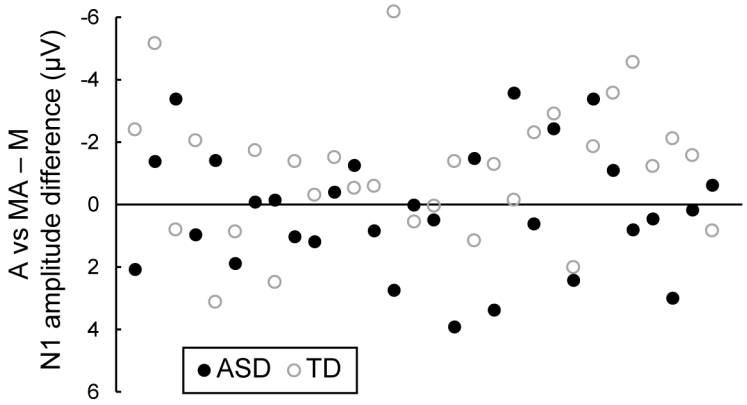


Figure 5.4. Scatter plot showing individual differences in N1 amplitude between the A and MA – M condition in the fronto-central ROI (Cz, C1, C2, FCz, FC1, FC2, Fz, F1, F2). Negative values indicate N1 attenuation.

To ensure that the difference in FSIQ between the ASD and TD group was not a confounding factor for the absent N1 attenuation in the ASD group, a post hoc partial correlation analysis controlling for group membership was conducted correlating individual N1 amplitude difference between the A and MA – M condition in the fronto-central ROI to FSIQ. This analysis revealed that the extent of N1 attenuation was not affected by FSIQ (r = .16, $p = .22$), thereby ruling out FSIQ as a confounding factor for the absent N1 attenuation in the ASD group.

N1 latency. The MANOVA for N1 latency showed a main effect of Condition, $F(1, 58) = 30.21$, $p < .001$, $\eta_p^2 = .34$. The N1 was speeded up by ~9 ms in the MA condition compared to the A condition (A: 105.18 ms MA – M: 96.58 ms). There was no main effect of Group or Condition × Group interaction, indicating that the N1 for self-initiated tones was speeded up similarly in the ASD and TD group. This confirmed that the difference in mean activity between the A and MA – M condition in the 55-90 ms time window for the ASD group (as revealed by the cluster based permutation tests) was indeed due to a temporal shift of the rising flank of the N1 – and not due to a difference in N1 amplitude between modalities.

P2 responses to self- versus externally-initiated tones

P2 time window. The initial permutation test on the A – MA – M difference waveforms revealed no significant differences between the ASD and TD group in the P2 latency range. Visual inspection of the ERPs (Figure 5.2A, 5.2B) suggests that in both groups, the mean activity in the P2 latency range was less positive and speeded up in the MA – M condition compared to the A condition. To verify this observation, neural auditory activity across both groups was computed for each condition and submitted to a cluster-based permutation test. This procedure revealed a time window of interest in the latency range from 165-290 ms showing a significant difference ($p < .001$) between the A and MA – M condition that was most pronounced over central electrodes (Figure 5.3D). Confirmatory parametric testing was carried out on the peak amplitude and peak latency values in this time window in a central ROI including Cz and eight directly surrounding electrodes. Individual P2 peak amplitude and peak latency values within the 165-290 ms time window were calculated for each condition and electrode and submitted to repeated measures MANOVAs with the within-subjects variables Condition (A, MA – M) and Electrode (CPz, CP1, CP2, Cz, C1, C2, FCz, FC1, FC2).

P2 amplitude. The MANOVA for P2 amplitude produced a significant Condition x Electrode interaction, $F(8, 52) = 2.68$, $p = .02$, $\eta_p^2 = .29$. This interaction was further explored with simple main effects tests examining the effect of Condition at each Electrode. In all electrodes, P2 amplitude was significantly attenuated in the MA condition compared to the A condition (all p values < .03, average amplitude difference 1.53 µV).

P2 latency. The MANOVA for P2 latency showed a main effect of Condition, $F(1, 59) = 46.41$, $p < .001$, $\eta_p^2 = .44$, indicating that the P2 was speeded up by ~18 ms in the MA condition compared to the A condition (A: 192.43 ms MA – M: 173.99 ms).

Summary

N1 latency and attenuation effects for self-initiated tones were found in the TD group. In the ASD group, the auditory N1 for self-initiated tones was speeded up but – crucially – not attenuated, whereas the P2 for self-initiated tones was speeded up and attenuated in both groups.

DISCUSSION

The current study tested the predictive coding account for autistic symptomatology by comparing the neural response to self- versus externally-initiated tones in individuals with ASD and TD. The data revealed clear group differences in the neural correlates of internal motor-to-auditory prediction mechanisms. Significant N1 attenuation effects were found in the TD group, indicating that a forward model predicted the auditory consequences of their motor actions. These results are consistent with the literature on typical electrophysiological indicators for predictive processing in audition (Baess et al., 2008; Bendixen et al., 2012). Most importantly, self-initiation of the tones did not attenuate the auditory N1 in the ASD group. The extent of N1 attenuation is presumed to be positively correlated with the accuracy of the prediction of the upcoming stimulus (Arnal & Giraud, 2012; Friston, 2005). The absence of N1 attenuation in the ASD group could thus indicate that, even in a relatively stable context, individuals with ASD experience difficulties in anticipating upcoming sensory events and seemingly process every stimulus afresh – rather than mediated by prior expectation. The current results could be indicative of impaired motor-to-auditory predictions in ASD, and support the impaired predictive coding account of autistic symptomatology (Lawson et al., 2014; Pellicano et al., 2007; van Boxtel & Lu, 2013; Van de Cruys et al., 2014).

Although the N1 was not attenuated for self-initiated tones in the ASD group, it was speeded up similar as in the TD group. Previous studies have shown that N1 latency facilitation only occurs if the preceding stimulus provides reliable predictive information about the identity of the upcoming sound (Arnal, Morillon, Kell, & Giraud, 2009; Paris, Kim, & Davis, 2017). The similar N1 latency facilitation in both the ASD and TD group may thus suggest that predictions regarding the identity of the tones were intact in the ASD group. Yet the absence of N1 attenuation in the ASD group indicates that auditory predictions for self-initiated tones were not enhanced by the cues provided by the preceding motor action. It could be speculated that participants in the ASD group failed to infer the temporal relationship of the tones relative to the button presses. As a result, predictions about the onset of self-initiated tones may have been impaired. This interpretation aligns with recent observations of impaired multisensory temporal acuity in ASD (Noel, De Niear, Stevenson, Alais, & Wallace, 2017; Stevenson et al., 2016). It should be noted, however, that in TD individuals, significant (albeit smaller) auditory N1 attenuation effects have been reported for self-initiated sounds with unpredictable timing and content (Baess et al., 2008; Knolle, Schröger, & Kotz, 2013b). Others have shown that tones triggered by a key-press elicit a smaller N1 than tones following a visual cue with predictable timing (Lange, 2011), suggesting that the attenuated N1 to self-initiated tones is not merely

caused by the fact that self-initiation provides a highly reliably cue for tone onset. Thus, N1 attenuation for self-initiated sounds may in part reflect a more general predictive mechanism (Baess, Widmann, Roye, Schröger, & Jacobsen, 2009; Martikainen et al., 2005; Sanmiguel, Todd, & Schröger, 2013). Based on the current study it cannot be resolved whether the absence of N1 attenuation to self-initiated sounds in the ASD group was caused by impairments in temporal- identity- or general prediction. In a future study it would therefore be interesting to investigate the relative contribution of temporal- and identity predictions in individuals with ASD by contrasting a single sound condition with a random sound condition (cf. Baess et al., 2008).

For both the TD and ASD group, the N1 for self-initiated sounds was followed by an attenuated and speeded up P2 response. Although N1 attention effects are often accompanied by a suppression of the P2 component, the P2 can be functionally dissociated from the N1 (Crowley & Colrain, 2004). While the exact functional interpretation of the auditory P2 component is still unclear, it has been argued that an attenuated P2 response to self- initiated tones may reflect the conscious post-hoc realization that a sound closely following a button press must have been self-initiated - as opposed to an attenuated N1 response, which reflects the effect of an automatic prospective internal forward prediction mechanism (Knolle, Schröger, & Kotz, 2013a). The current data could therefore indicate that, even though individuals with ASD are aware of the fact that auditory stimulation can be self-initiated, they are unable to effectively use the predictive information provided by their own motor actions to anticipate the auditory sensory consequences of those actions.

Previous studies have shown that increasing attention towards an auditory stimulus may result in higher N1 amplitudes (Lange, Rösler, & Röder, 2003), whereas drawing attention away may attenuate the N1 response (Horváth & Winkler, 2010). It could therefore be argued that increased attention to self-initiated sounds – relative to externally-initiated tones – may have resulted in an amplitude increase of the auditory N1 in the ASD group. An argument against this view is that attenuation of the P2 was similar in the ASD and TD group, indicating that a potential difference in allocation of attention between self- and externally-initiated tones was likely similar in both groups. Still, the N1 was significantly attenuated – rather than enlarged – in the TD group, thereby rendering sustained attentional differences between experimental conditions an unlikely account for the absence of N1 attenuation in the ASD group. Furthermore, this attentional account was specifically tested and refuted in a study using a N1 suppression paradigm where self- and externally-initiated sounds were randomly intermixed and presented within the same

block (Baess et al., 2011). Because externally-initiated sounds occurred at unpredictable intervals within the same block as self-initiated sounds, ERP differences between self- and externally-initiated sounds could not stem from a difference in task demands between the experimental conditions. The results showed an even larger N1 attenuation effect for self-initiated sounds than typically observed in a blocked N1 suppression paradigm (as used in the current experiment), indicating that N1 attenuation for self- versus externally-initiated sounds is independent of attention. It can also be argued that the difference in N1 attenuation between the ASD and TD group was due to a difference in allocation of attention between modalities during self-initiation of the tones. Increased attention to the auditory tones – relative to the motor act – may have led to an amplitude increase of the auditory N1 in the ASD group. However, this attentional account was also examined and refuted in a recent study (Timm, SanMiguel, Saupe, & Schröger, 2013). Using a similar mixed N1 suppression paradigm as Baess et al. (2011), allocation of attention was manipulated block-wise to either the sound, the motor act or to a visual stimulus. The results showed similar N1 attenuation effects for self-initiated sounds in all three attention conditions.

Taken together, these findings imply that the lack of N1 attenuation for self-initiated tones in the ASD group cannot be explained by potential differences in allocation of attention, but instead, more likely reflects the activity of an impaired motor-to-auditory prediction mechanism.

Future Directions

If individuals with ASD are indeed unable to anticipate the sensory consequences of their own actions, this raises the question if their ability to predict actions of other individuals is impaired as well. Given that other people's behavior is are arguably more difficult to predict than self-initiated actions, and the fact that individuals with ASD have great difficulty with understanding the thoughts and emotions of their own and those of others (Robertson & Baron-Cohen, 2017), it is reasonable to assume that this might indeed be the case. There is indeed evidence suggesting that individuals with ASD have specific deficits in attributing mental states to others (i.e. mentalizing), whereas processing of lower-level social information is intact (David et al., 2010; Sebanz, Knoblich, Stumpf, & Prinz, 2005; Zwickel, White, Coniston, Senju, & Frith, 2011). Future studies should address if these findings can be linked to electrophysiological alterations. Previous studies have reported that in TD individuals, attenuation effects of auditory potentials are not limited to the motor-auditory domain but are found in other inter-sensory domains as well. For example, seeing someone performing a handclap provides predictive information about

the upcoming sound. Several studies have demonstrated that such anticipatory information attenuates and speeds up the auditory N1 and P2 (Stekelenburg & Vroomen, 2007, 2012; Vroomen & Stekelenburg, 2010). Others have reported that a rare omission of a sound that is predictable by anticipatory visual information typically induces an early negative response in the EEG during the period of silence where the sound was expected (Stekelenburg & Vroomen, 2015; van Laarhoven, Stekelenburg, & Vroomen, 2017). In a future study it would therefore be interesting to investigate if the alterations in motor-to-auditory prediction observed in the current group of individuals with ASD extend to the visual-auditory domain.

CONCLUSIONS

The current results confirm our hypothesis that individuals with ASD show alterations in sensory attenuation of self-initiated sounds. Specifically, predictive cues provided by button presses did not attenuate the auditory N1 in our sample of individuals with ASD. The current data indicate that motor-to-auditory prediction may be impaired in ASD, and support the notion of impaired predictive coding as a core deficit underlying atypical sensory processing in ASD.

ACKNOWLEDGMENTS

We wish to thank the staff of Yulius Mental Health Organization, location de Steiger, in particular Sandra Kint, for their help in recruiting participants with ASD and providing diagnostic assessment information. We also thank Jet Roos, José Hordijk, Justlin van Bruggen for collecting the data.

REFERENCES

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th TR). Washington, DC: Author.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A.-L. (2009). Dual Neural Routing of Visual Facilitation in Speech Processing. *Journal of Neuroscience, 29*(43), 13445–13453. https://doi.org/10.1523/JNEUROSCI.3194-09.2009

Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology, 48*(9), 1276–1283. https://doi.org/10.1111/j.1469-8986.2011.01196.x

Baess, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology, 70*(2), 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Baess, P., Widmann, A., Roye, A., Schröger, E., & Jacobsen, T. (2009). Attenuated human auditory middle latency response and evoked 40-Hz response to self-initiated sounds. *European Journal of Neuroscience, 29*(7), 1514–1521. https://doi.org/10.1111/j.1460-9568.2009.06683.x

Baranek, G. T., Watson, L. R., Boyd, B. A., Poe, M. D., David, F. J., & McGuire, L. (2013). Hyporesponsiveness to social and nonsocial sensory stimuli in children with autism, children with developmental delays, and typically developing children. *Development and Psychopathology, 25(*2), 307–320. https://doi.org/10.1017/S0954579412001071

Baum, S. H., Stevenson, R. A., & Wallace, M. T. (2015). Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Progress in Neurobiology, 134*, 140–160. https://doi.org/10.1016/j.pneurobio.2015.09.007

Beker, S., Foxe, J. J., & Molholm, S. (2018). Ripe for solution: Delayed development of multisensory processing in autism and its remediation. *Neuroscience and Biobehavioral Reviews, 84*, 182–192. https://doi.org/10.1016/j.neubiorev.2017.11.008

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology, 83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Constantino, J. N., & Gruber, C. P. (2012). Social Responsiveness Scale (SRS-2). Torrance, CA: Western Psychological Services.

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge across the animal kingdom. *Nature Reviews Neuroscience, 9*(8), 587–600. https://doi.org/10.1038/nrn2457

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology, 115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

David, N., Aumann, C., Bewernick, B. H., Santos, N. S., Lehnhardt, F. G., & Vogeley, K. (2010). Investigation of mentalizing and visuospatial perspective taking for self and other in asperger syndrome. *Journal of Autism and Developmental Disorders, 40*(3), 290–299. https://doi.org/10.1007/s10803-009-0867-4

Eussen, M. L. J. M., Louwerse, A., Herba, C. M., Van Gool, A. R., Verheij, F., Verhulst, F. C., & Greaves-Lord, K. (2015). Childhood Facial Recognition Predicts Adolescent Symptom Severity in Autism Spectrum Disorder. *Autism Research, 8*(3), 261–271. https://doi.org/10.1002/aur.1443

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Hampson, D. R., & Blatt, G. J. (2015). Autism spectrum disorders and neuropathology of the cerebellum. *Frontiers in Neuroscience, 9*. https://doi.org/10.3389/fnins.2015.00420

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review, 20*(3), 290–322. https://doi.org/10.1007/s11065-010-9138-6

Horváth, J., & Winkler, I. (2010). Distraction in a continuous-stimulation detection task. *Biological Psychology, 83*(3), 229–238. https://doi.org/10.1016/j.biopsycho.2010.01.004

Knolle, F., Schröger, E., Baess, P., & Kotz, S. A. (2012). The Cerebellum Generates Motor-to-Auditory Predictions: ERP Lesion Evidence. *Journal of Cognitive Neuroscience, 24*(3), 698–706. https://doi.org/10.1162/jocn_a_00167

Knolle, F., Schröger, E., & Kotz, S. A. (2013a). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex, 49*(9), 2449–2461. https://doi.org/10.1016/j.cortex.2012.12.012

Knolle, F., Schröger, E., & Kotz, S. A. (2013b). Prediction errors in self- and externally-generated deviants. *Biological Psychology, 92*(2), 410–416. https://doi.org/10.1016/j.biopsycho.2012.11.017

Lange, K. (2011). The reduced N1 to self-generated tones: An effect of temporal predictability? *Psychophysiology, 48*(8), 1088–1095. https://doi.org/10.1111/j.1469-8986.2010.01174.x

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. *Psychophysiology, 40*(5), 806–817. https://doi.org/10.1111/1469-8986.00081

Lawson, R. P., Rees, G., & Friston, K. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience, 8*, 302 https://doi.org/10.3389/fnhum.2014.00302

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). Autism Diagnostic Observation Schedule-2 Manual. Los Angeles, CA: Western Psychological Services.

Marco, E. J., Hinkley, L. B. N., Hill, S. S., & Nagarajan, S. S. (2011). Sensory processing in autism: A review of neurophysiologic findings. *Pediatric research, 69*(5 Pt 2), 48R–54R. https://doi.org/10.1203/PDR.0b013e3182130c54

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. https://doi.org/10.1016/J.JNEUMETH.2007.03.024

Martikainen, M. H., Kaneko, K., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex, 15*(3), 299–302. https://doi.org/10.1093/cercor/bhh131

Mitchel, A. D., Christiansen, M. H., & Weiss, D. J. (2014). Multimodal integration in statistical learning: Evidence from the McGurk illusion. *Frontiers in Psychology, 5,* 407. https://doi.org/10.3389/fpsyg.2014.00407

Mitchel, A. D., & Weiss, D. J. (2011). Learning Across Senses: Cross-Modal Effects in Multisensory Statistical Learning. *Journal of Experimental Psychology: Learning Memory and Cognition, 37*(5), 1081–1091. https://doi.org/10.1037/a0023700

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics, 66*(3), 241–251. https://doi.org/10.1007/BF00202389

Noel, J. P., De Niear, M. A., Stevenson, R., Alais, D., & Wallace, M. T. (2017). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Research, 10*(1), 121–129. https://doi.org/10.1002/aur.1633

Paris, T., Kim, J., & Davis, C. (2017). Visual form predictions facilitate auditory processing at the N1. *Neuroscience, 343*, 157–164. https://doi.org/10.1016/j.neuroscience.2016.09.023

Pellicano, E., & Burr, D. (2012). When the world becomes “too real”: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Pellicano, E., Jeffery, L., Burr, D., & Rhodes, G. (2007). Abnormal Adaptive Face-Coding Mechanisms in Children with Autism Spectrum Disorder. *Current Biology, 17*(17), 1508–1512. https://doi.org/10.1016/j.cub.2007.07.065

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience, 18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Sanmiguel, I., Todd, J., & Schröger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. *Psychophysiology, 50*(4), 334–343. https://doi.org/10.1111/psyp.12024

Sebanz, N., Knoblich, G., Stumpf, L., & Prinz, W. (2005). Far from action-blind: Representation of others' actions in individuals with autism. *Cognitive Neuropsychology, 22*(3–4), 433–454. https://doi.org/10.1080/02643290442000121

Seitz, A. R., Kim, R., Van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception, 36*(10), 1445–1453. https://doi.org/10.1068/p5843

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*(12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Stekelenburg, J. J., & Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Frontiers in Integrative Neuroscience, 6*, 26. https://doi.org/10.3389/fnint.2012.00026

Stekelenburg, J. J., & Vroomen, J. (2015). Predictive coding of visual-auditory and motor-auditory events: An electrophysiological study. *Brain Research, 1626*, 88–96. https://doi.org/10.1016/j.brainres.2015.01.036

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., Camarata, S., & Wallace, M. T. (2016). Keeping time in the brain: Autism spectrum disorder and audiovisual temporal processing. *Autism Research, 9*(7), 720–738. https://doi.org/10.1002/aur.1566

Timm, J., SanMiguel, I., Saupe, K., & Schröger, E. (2013). The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neuroscience, 14*, 2. https://doi.org/10.1186/1471-2202-14-2

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders, 43*(7), 1517–1526. https://doi.org/10.1007/s10803-012-1695-5

van Boxtel, J. J. A., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00019

van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review, 121*(4), 649–675. https://doi.org/10.1037/a0037665

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2017). Temporal and identity prediction in visual-auditory events: Electrophysiological evidence from stimulus omissions. *Brain Research, 1661*, 79–87. https://doi.org/10.1016/j.brainres.2017.02.014

van Vugt, F. T., & Tillmann, B. (2014). Thresholds of auditory-motor coupling measured with a simple task in musicians and non-musicians: Was the sound simultaneous to the key press? *PLoS ONE, 9*(2). https://doi.org/10.1371/journal.pone.0087176

Vroomen, J., & Stekelenburg, J. J. (2010). Visual Anticipatory Information Modulates Multisensory Interactions of Artificial Audiovisual Stimuli. *Journal of Cognitive Neuroscience, 22*(7), 1583–1596. https://doi.org/10.1162/jocn.2009.21308

Zwickel, J., White, S. J., Coniston, D., Senju, A., & Frith, U. (2011). Exploring the building blocks of social cognition: Spontaneous agency perception and visual perspective taking in autism. *Social Cognitive and Affective Neuroscience, 6*(5), 564–571. https://doi.org/10.1093/scan/nsq088

Chapter 6

Atypical visual-auditory predictive coding in autism spectrum disorder

Electrophysiological evidence from stimulus omissions

This chapter is published as:

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L. J. M., & Vroomen, J. (2020). Atypical visual-auditory predictive coding in autism spectrum disorder: Electrophysiological evidence from stimulus omissions. *Autism, 24(7),* 1849-1859. https://doi.org/10.1177/1362361320926061

ABSTRACT

Autism Spectrum Disorder (ASD) is a pervasive neurodevelopmental disorder that has been linked to a range of perceptual processing alterations, including hypo- and hyperresponsiveness to sensory stimulation. A recently proposed theory that attempts to account for these symptoms, states that autistic individuals have a decreased ability to anticipate upcoming sensory stimulation due to overly precise internal prediction models. Here, we tested this hypothesis by comparing the electrophysiological markers of prediction errors in auditory prediction by vision between a group of autistic individuals and a group of age-matched individuals with typical development (TD). Between-group differences in prediction error signaling were assessed by comparing event-related potentials evoked by unexpected auditory omissions in a sequence of audiovisual recordings of a handclap in which the visual motion reliably predicted the onset and content of the sound. Unexpected auditory omissions induced an *increased* early negative omission response (oN1) in the ASD group, indicating that violations of the prediction model produced larger prediction errors in the ASD group compared to the TD group. The current results show that autistic individuals have alterations in visual-auditory predictive coding, and support the notion of impaired predictive coding as a core deficit underlying atypical sensory perception in ASD.

INTRODUCTION

Autism Spectrum Disorder (ASD) is a pervasive neurodevelopmental disorder characterized by impairments in social communication and social interaction, and restricted, repetitive patterns of behavior, interests or activities (American Psychiatric Association, 2013; Robertson & Baron-Cohen, 2017). In addition, ASD has been linked to a range of perceptual processing alterations, including atypical processing of facial emotions (Eussen et al., 2015; Harms, Martin, & Wallace, 2010; Pellicano, Jeffery, Burr, & Rhodes, 2007; Uljarevic & Hamilton, 2013) and hypo- and hyperresponsiveness to sensory stimulation (Baranek et al., 2013; Robertson & Baron-Cohen, 2017).

A recently proposed theory that attempts to account for these symptoms, states that autistic individuals have a decreased ability to anticipate upcoming sensory stimulation (Lawson, Rees, & Friston, 2014; Pellicano & Burr, 2012; van de Cruys et al., 2014). A key element of the predictive coding framework is the assumption that incoming sensory information is continuously contrasted with internal predictions about the current state of our environment based on previous experiences (Friston, 2005). Any discrepancy between the sensory input and prior expectations results in the computation of an error signal. These prediction errors are crucial to adequately contextualize sensory information. They inform our perception about the current state of the world, and indicate that our current internal predictive model is not able to adequately predict upcoming sensory stimulation, and, thus, needs to be updated to resolve similar prediction errors in the future. Given that the world is not static (i.e. two perceptual experiences are never completely alike), prediction errors are always present to some extent. Although prediction errors are typically evoked by unexpected and 'newsworthy' sensory stimulation that ought to increase our attention (e.g. a car ignoring a cross walk), they may sometimes be spurious and uninformative (e.g. someone dropping a glass at a party). Thus, in order to adequately adjust the impact of prediction violations on updates of the predictive model, prediction errors need to be processed with a certain degree of flexibility: some prediction errors should be processed with 'high priority', while others should be ignored and suppressed. Recently, it has been proposed that an inability to flexibly process prediction errors may be the core deficit underlying the socio-communicative impairments in ASD (van de Cruys et al., 2014). Others have posited that an imbalance in the importance ascribed to sensory input – relative to prior expectations – may cause autistic individuals to overweigh the significance of prediction errors (Lawson et al., 2014). It has also been argued that autistic individuals have a decreased ability to infer the probabilistic structure of their environment – resulting in imprecise or attenuated prior expectations or (in Bayesian terms) 'hypo-priors' (Pellicano & Burr,

2012). Although conceptually distinct (for an overview, see Brock, 2012; Friston, Lawson, & Frith, 2013; Lawson et al., 2014; Van de Cruys, De-Wit, Evers, Boets, & Wagemans, 2013), all these theoretical accounts may result in an over-reliance on sensory input. Given that perceptual cues are often noisy and ambiguous, a predictive model that is biased towards sensory input – rather than modulated by prior experience – may generate predictions that are 'overfitted' to specific contexts, but do not generalize well to new experiences where the sensory environment is often volatile. Following this reasoning, new experiences may generate large prediction errors in autistic individuals, since their overfitted prior expectations are likely to be violated by novel sensory input. Failing to contextualize and generalize sensory information in an optimal fashion – based on both current sensory input and prior expectations – may lead to atypical sensitivity to sensory stimulation (including hypo- and hyperresponsiveness), which could ultimately affect sensory processing, perception, and social interaction. Understanding the neural basis of the potential impairments in predictive coding in ASD may thus very well be a fundamental part of the explanation of why autistic individuals often struggle with social communication and interaction with their environment.

Recent evidence suggests that predictive coding might indeed be impaired in autistic individuals (van Laarhoven, Stekelenburg, Eussen, & Vroomen, 2019). In this study, the neural response to self- versus externally-initiated tones was examined in a group of autistic individuals and a group of age-matched individuals with typical development (TD). The amplitude of the auditory N1 component of the event-related potential (ERP) is typically attenuated for self-initiated sounds, compared to sounds with identical acoustic and temporal features that are triggered externally (Baess, Horváth, Jacobsen, & Schröger, 2011; Baess, Jacobsen, & Schröger, 2008; Bendixen, SanMiguel, & Schröger, 2012; Martikainen, Kaneko, & Hari, 2005). This attenuation effect has been ascribed to internal prediction models predicting the sensory consequences of one's own motor actions. The results of this study showed that (unlike in the TD group), self-initiation of the tones did not attenuate the auditory N1 in the ASD group, indicating that they may be unable to fully anticipate the (auditory) sensory consequences of their own motor actions. This raises the question if the ability to predict the actions of other individuals is altered as well in ASD. Given that the behavior of other individuals is arguably more difficult to predict than self-initiated actions, and the fact that autistic individuals have great difficulty with understanding the thoughts and emotions of their own and those of others (Robertson & Baron-Cohen, 2017), this seems plausible.

A growing area of interest and relevance in the study of predictive coding focuses on the electrophysiological responses to unexpected stimulus omissions of predictable sounds (SanMiguel, Widmann, Bendixen, Trujillo-Barreto, & Schröger, 2013; Stekelenburg & Vroomen, 2015). Auditory stimulation can be made predictable either by a motor act or anticipatory visual information (such as in a handclap, in which the movement of the hands precedes the sound) that reliably predicts the timing and content of the sound. Unexpected omissions of predictable sounds typically evoke an early negative omission response (oN1) that peaks between 45 and 100 ms in the EEG during the period of silence where the sound was expected (SanMiguel, Saupe, & Schröger, 2013; SanMiguel, Widmann, et al., 2013; Stekelenburg & Vroomen, 2015; van Laarhoven, Stekelenburg, & Vroomen, 2017). The amplitude of the auditory oN1 is hypothesized to be modulated by the prediction and prediction error (Arnal & Giraud, 2012; Friston, 2005). For sounds that are highly predictable, precise auditory predictions can be formed about the content and timing of the sound. If incoming auditory stimulation does not match (but violates) this prior expectation, such as during unexpected auditory omissions, the prediction error is large, and thus the oN1 is enlarged. If no clear predictions can be formed about an upcoming sound, the prediction is less likely to be violated, and so the oN1 is attenuated or absent during auditory omissions. Several studies have indeed shown that the oN1 is only elicited by unexpected omissions of sounds of which both the timing and content is predictable either by a motor act or anticipatory visual information, and not by omissions of unpredictable sounds or auditory omissions per se (Bendixen et al., 2012; SanMiguel, Saupe, et al., 2013; SanMiguel, Widmann, et al., 2013; van Laarhoven et al., 2017). Hence, the oN1 can be considered as an early marker of auditory prediction error.

In the current study, we used a stimulus omission approach to examine the electrophysiological markers of prediction errors in auditory prediction by vision in autistic individuals to assess their ability to anticipate the sensory consequences of others' actions. An experimental paradigm was applied that was similar to those used in previous studies showing robust and consistent visual-auditory oN1 effects in TD individuals (Stekelenburg & Vroomen, 2015; van Laarhoven et al., 2017). EEG was recorded in a group of older adolescents and young adults with a clinical diagnosis of ASD, and in an age-matched group of individuals with TD. Between group differences in visual-auditory predictive coding were assessed by comparing ERPs evoked by unexpected auditory omissions in a sequence of audiovisual recordings of a handclap, in which the visual motion reliably predicted the timing and content of the sound (Stekelenburg & Vroomen, 2007, 2015). Atypical enlargement of the oN1 response, a neural marker of

prediction error, was considered as evidence for altered visual-auditory predictive coding and a potential indication of over-reliance on sensory input.

METHODS

Participants

Twenty-nine autistic individuals (8 females), and twenty-nine age-matched individuals with TD (6 females) participated in this study (ASD: *M* = 18.64 years, *SD* = 2.11; TD: *M* = 18.93 years, *SD* = 1.22). Inclusion criteria for participants in both groups were: normal or corrected-to-normal vision and hearing, Full Scale IQ (FSIQ) >=80, and no active use of sedatives two days prior to the experiment. Additional inclusion criteria for the ASD group were: a clinical DSM-IV TR classification of ASD (American Psychiatric Association, 2000), and absence of severe comorbid neurological disorders (e.g. epilepsy). Additional inclusion criteria for the TD group were: absence of any neurological or neuropsychiatric disorder (e.g. ASD, ADHD, epilepsy).

Participants with ASD were recruited at a mental health institution for ASD (de Steiger, Yulius Mental Health, Dordrecht, the Netherlands). At the time of the experiment, all participants in the ASD group were receiving clinical treatment at this mental health institution due to severe mental problems and impaired functioning in activities of daily living linked to ASD. Participants with TD were recruited at Tilburg University and a high school located in the city of Tilburg.

For all participants in the ASD group the clinical DSM-IV TR classification of ASD was confirmed by two independent clinicians. Additional diagnostic information was retrieved when available, including Autism Diagnostic Observation Schedule (ADOS) scores (Lord et al., 2012), and Social Responsiveness Scale (SRS) scores (Constantino & Gruber, 2012). FSIQ was measured with the Dutch versions of the Wechsler Adult Intelligence Scale (WAIS-IV-NL) in participants ≥ 18 years, and the Wechsler Intelligence Scale for Children (WISC-III-NL) in participants < 18 years. Demographic details of the ASD group and TD group are shown in Table 6.1. Specific data on socioeconomic status and educational attainment levels were not recorded. There were no significant differences in age, $t(56) = -0.64$, $p = .53$, and gender, $t(56) = 0.61$, $p = .55$), but the average FSIQ score was higher for the TD group (mean FSIQ 112.07, *SD* = 11.68) compared to the ASD group (mean FSIQ 103.03, *SD* = 16.76), $t(56) = 2.38$, $p = .02$.

All procedures were undertaken with the understanding and written consent of each participant and – for participants under the age of 18 – a parent or another legally authorized representative. Participants with ASD and TD participants who were recruited at the high school were reimbursed with 25 EUR for their participation. TD participants recruited at Tilburg University received course credits as part of a curricular requirement. All experimental procedures were approved by the local medical ethical review board (METC Brabant, protocol ID: NL52250.028.15) and performed in accordance with the ethical standards of the Declaration of Helsinki.

Table 6.1. Demographics for the Autism Spectrum Disorder (ASD) and Typically Developing (TD) Group

	ASD	TD
Gender $^{n.s.}$	21 male, 8 female	23 male, 6 female
Age $^{n.s.}$	M = 18.64, SD = 2.11	M = 18.93, SD = 1.22
	range: 15-24	range: 15-20
Full Scale IQ *	M = 103.03, SD = 16.76	M = 112.07, SD = 11.68
	range: 80-134	range: 88-136
ADOS n = 17	M = 10.06, SD = 5.19	-
	range: 4-22	
SRS n = 22	M = 72.91, SD = 9.68	-
	range: 55-92	

$^{n.s.}$ non-significant $^{*}p < .05$, M = mean, SD = standard deviation

Stimuli and procedure

Participants were individually tested in a sound attenuated and dimly lit room, and were seated in front of a 19-inch CRT monitor (Iiyama Vision Master Pro 454) positioned at eye-level at a viewing distance of approximately 70 cm. Stimulus materials were adapted from previous work on visual-auditory predictive coding in TD individuals (Stekelenburg & Vroomen, 2015; van Laarhoven et al., 2017). Visual stimuli consisted of a video recording portraying the visual motion of a single handclap (Figure 6.1). The video started with the hands separated. Subsequently, the hands moved to each other, and after collision returned to their original starting position. Total duration of the video was 1300 ms. The video was presented at a frame rate of 25 frames/s, at a refresh rate of 100 Hz, and a resolution of 640 × 480 pixels (14° horizontal and 12° vertical visual angle). Auditory stimuli consisted of an audio recording (sampling rate 44.1 kHz) of the handclap portrayed in the video, and were presented at approximately 61 dB(A) sound pressure level over JAMO S100 stereo speakers, located directly on the left and right side of the monitor. Stimulus presentation was controlled using E-Prime 1.2 (Psychology Software Tools Inc., Sharpsburg, PA, USA).

Figure 6.1. Time-course of the video used in the visual-auditory (VA) and visual (V) condition.

Three conditions were included in the experiment: visual-auditory (VA), visual (V), and auditory (A). In the VA condition, the video of the handclap was presented synchronously with the audio recording of the handclap. The handclap sound occurred 360 ms after the start of the hand movement. The auditory inter-stimulus interval was 1300 ms. Standard VA trials were interspersed with unpredictable omissions of the handclap sound in 12% of the trials, c.f. (SanMiguel, Saupe, et al., 2013; Stekelenburg & Vroomen, 2015). These omission trials were randomly intermixed with standard VA trials with the restrictions that the first five trials of each block, and the two trials immediately following an omission trial were always standard VA trials. The VA condition was presented in seven blocks of 200 trials, resulting in a total of 1400 stimulus presentations in the VA condition (1232 standard VA trials and 168 auditory stimulus omissions). In the V and A condition, only the video recording or the sound of the handclap was presented, respectively. The V and A conditions were presented in two blocks of 100 trials, resulting in a total of 200 stimulus presentations in the V and A condition. Block order was quasi-randomized across participants such that V and A blocks were never presented successively.

The V condition was included to correct for visual activity in the auditory omission trials of the VA condition (see 'EEG recording'). The auditory oN1 is assumed to be correlated to the amplitude of the N1 that the expected sound would normally elicit (SanMiguel, Widmann, et al., 2013). The A condition was therefore included to test whether potential between-group differences in omission responses could be attributed to differences in sensory processing of the handclap sound itself.

To ensure that participants watched the visual stimuli and remained vigilant, 8% of all VA, V and A trials consisted of catch trials. Participants were required to respond with a button press after onset of a catch stimulus (i.e. a small white square superimposed on the handclap video, presented at the center of the screen, measuring 1° horizontal and 1° vertical visual angle). To prevent possible interference of (delayed) motor responses, these catch trials never preceded an omission trial. Average percentage of detected catch

trials across conditions was high (*M* = 98.30, *SD* = 2.81) and did not differ between conditions or groups, and there was no condition × group interaction effect (all *p* values > .08), indicating that participants in both groups attentively participated in all conditions. Total duration of the experiment was approximately 45 minutes.

EEG acquisition and processing

The EEG was sampled at 512 Hz from 64 locations using active Ag-AgCl electrodes (BioSemi, Amsterdam, the Netherlands) mounted in an elastic cap and two mastoid electrodes. Electrodes were placed in accordance with the extended International 10-20 system. Two additional electrodes served as reference (Common Mode Sense active electrode) and ground (Driven Right Leg passive electrode). Horizontal electrooculogram (EOG) was recorded using two electrodes placed at the outer canthi of the left and right eye. Vertical EOG was recorded from two electrodes placed above and below the right eye. BrainVision Analyzer 2.0 (Brain Products, Gilching, Germany) software was used for ERP analyses. EEG was referenced offline to an average of left and right mastoids and band-pass filtered (0.01-30 Hz, 24 dB/octave). The (residual) 50 Hz interference was removed by a 50 Hz notch filter. Raw data were segmented into epochs of 1000 ms, including a 200-ms pre-stimulus baseline period. Epochs were time-locked to the expected sound onset of auditory omission trials in the VA condition, and to the corresponding timestamp of trials in the V condition and to sound onset in the A condition. After EOG correction (Gratton, Coles, & Donchin, 1983), epochs with an amplitude change exceeding ±150 µV at any EEG channel were rejected, averaged, and baseline corrected for each condition separately. All participants were included in the final analysis. On average, 13.45 (*SD* = 17.02) of the presented 168 auditory omission trials were rejected, corresponding to 7.96 percent (*SD* = 10.13). Percentages of rejected trials were similar for the standard trials in the VA condition (*M* = 8.22, *SD* = 10.01), visual trials in the V condition (*M* = 7.37, *SD* = 8.98), and auditory trials in the A condition (*M* = 11.00, *SD* = 15.65). Across all conditions, 8.64 percent (*SD* = 9.85) of the trials were rejected. There were no significant differences in percentages of rejected trials between groups or conditions, and there was no condition × group interaction effect (all *p* values > .10). The ERP of the V condition was subtracted from the auditory omission ERPs in the VA condition to nullify the contribution of visual activity to the omission ERPs. Consequently, the VA–V difference waves reflect prediction related activity – induced by unexpected auditory omissions – devoid of visual activity (Stekelenburg & Vroomen, 2015; van Laarhoven et al., 2017).

RESULTS

The group-averaged auditory omission ERPs (Figure 6.2) showed two distinct negative deflections in both groups: oN1 (45-100 ms), oN2 (100-200 ms). In accordance with previous research on auditory omission responses (SanMiguel, Widmann, et al., 2013), maximal amplitude of the oN1 and oN2 was measured at electrode FT7. The two negative omission responses were followed by late positive potentials oP3 (300-550 ms), showing maximal amplitudes measured at electrodes Cz.

Visual inspection of the ERPs showed that the oN1 for the ASD group was more pronounced compared to the oN1 for the TD group, while the oN2 and oP3 deflections appeared to be similar for both groups. The oN1 and oN2 deflections showed a bilateral scalp distribution in both groups, while the oP3 components had a central scalp distribution (Figure 6.3). Based on these scalp distributions, a left-temporal (F7, F5, F3, FT7, FC5, FC3, T7, C5, C3) and right-temporal (F4, F6, F8, FC4, FC6, FT8, C4, C6, T8) region of interest (ROI) were selected for the oN1 and oN2 time windows. A central-parietal (C1, Cz, C2, CP1, CPz, CP2) ROI was selected for the oP3 time window. The presence of statistically significant omission responses was tested by conducting separate repeated measures MANOVAs on the mean activity for each time window, with the within-subjects variables Electrode and ROI for the oN1 and oN2 time windows, and Electrode for the oP3 time window, and between-subjects factor Group (ASD, TD) for all time windows.

oN1 time window (45-100 ms)

The overall mean activity in the oN1 time window differed from pre-stimulus baseline levels, $F(1, 57) = 5.73$, $p = .02$, $\eta_p^2 = .09$. There was a main effect of Group, $F(1, 56) = 4.32$, $p = .04$, $\eta_p^2 = .07$, indicating that the mean activity in the oN1 time window (averaged across ROIs and electrodes) was 0.52 µV more negative in the ASD group compared to the TD group (see Figure 6.4A for group medians and interquartile ranges). There were no main effects of ROI, $F(1, 56) = 0.40$, $p = .53$, $\eta_p^2 = .01$, and Electrode, $F(8, 49) = 1.95$, $p = .07$, $\eta_p^2 = .24$, and no significant interaction effects between the factors Electrode, ROI and Group (all p-values > .31).

To examine if the between-group difference in oN1 mean activity could be attributed to differences in sensory processing of the sound and video of the handclap stimulus itself, three separate repeated measures MANOVAs were conducted on the peak amplitude of the N1 evoked by auditory trials in the A condition in a time window of 50-150 ms, the peak amplitude of the N1 evoked by standard trials in the VA–V condition in a time window of 50-150 ms, and the mean activity evoked by visual trials in the V condition in a

time window of 75-175 ms. All analyses were conducted on the electrodes showing maximal activity (A: Cz, CPz; standard VA–V: Cz, CPz, V: O1, Oz, O2). The MANOVA on the peak amplitude of the auditory N1 in the A condition revealed no main effect of Group, $F(1, 56) = 0.19$, $p = .66$, $\eta_p^2 = .003$, and Electrode, $F(1, 56) = 0.05$, $p = .83$, $\eta_p^2 = .001$, and no interaction effect between the factors Group and Electrode, $F(1, 56) = 2.17$, $p = .15$, $\eta_p^2 = .04$. Similarly, the MANOVA on the peak amplitude of the auditory N1 evoked by standard trials in the VA–V condition revealed no main effect of Group, $F(1, 56) = 0.04$, $p = .85$, $\eta_p^2 = .001$, and Electrode, $F(1, 56) = 1.33$, $p = .25$, $\eta_p^2 = .02$, and no interaction effect between the factors Group and Electrode, $F(1, 56) = 0.01$, $p = .91$, $\eta_p^2 < .001$, indicating that the N1 evoked by the handclap sound was similar for both groups.

The MANOVA on the mean activity of the visual N1 in the V condition revealed a main effect of Electrode, $F(2, 55) = 20.83$, $p < .001$, $\eta_p^2 = .43$. Post hoc paired samples t-tests (Bonferroni corrected) showed that the overall mean activity in the visual N1 time window significantly differed between all three electrodes (all p values < .05), such that activity was most negative at Oz, and least negative at O2. More importantly, there was no main effect of Group, $F(1, 56) = 1.73$, $p = .19$, $\eta_p^2 = .03$, and no significant Group × Electrode interaction, $F(2, 55) = 2.71$, $p = .08$, $\eta_p^2 = .09$. Hence, the between-group difference in oN1 response could not be attributed to differences in auditory or visual stimulus processing per se, but instead, more likely reflects a difference in prediction error signaling.
To ensure that the difference in FSIQ between the ASD and TD group was not a confounding factor for the difference in oN1 mean activity, a post hoc partial correlation analysis controlling for group membership was conducted correlating individual oN1 mean activity averaged across the left- and right temporal ROI to FSIQ. This analysis revealed that the oN1 mean activity was not affected by FSIQ ($r = -.03$, $p = .85$), thereby ruling out FSIQ as a confounding factor for the difference in oN1 mean activity between the ASD and TD group.

oN2 time window (100-200 ms)

The overall mean activity in the oN2 time window differed from pre-stimulus baseline levels, $F(1, 57) = 21.27$, $p < .001$, $\eta_p^2 = .28$. There was no main effect of Group, $F(1, 56) = 2.07$, $p = .16$, $\eta_p^2 = .04$, ROI, $F(1, 56) = 0.72$, $p = .40$, $\eta_p^2 = .01$, and Electrode, $F(8, 49) = 1.55$, $p = .20$, $\eta_p^2 = .03$, and no significant interaction effects between the factors Group, ROI, and Electrode (all p-values > .19), indicating that the mean activity in the oN2 time window was similar in both groups (see Figure 6.4B for group medians and interquartile ranges).

oP3 time window (300-550 ms)

The overall mean activity in the oP3 time window differed from pre-stimulus baseline levels, $F(1, 57) = 49.61$, $p < .001$, $\eta_p^2 = .47$. There was no main effect of Group, $F(1, 56) = 6.88$, $p = .72$, $\eta_p^2 = .002$, and Electrode, $F(5, 52) = 2.29$, $p = .06$, $\eta_p^2 = .18$, and no significant Group × Electrode interaction, $F(5, 52) = 0.15$, $p = .98$, $\eta_p^2 = .01$, indicating that the mean activity in the oP3 time window was similar in both groups (see Figure 6.4C for group medians and interquartile ranges).

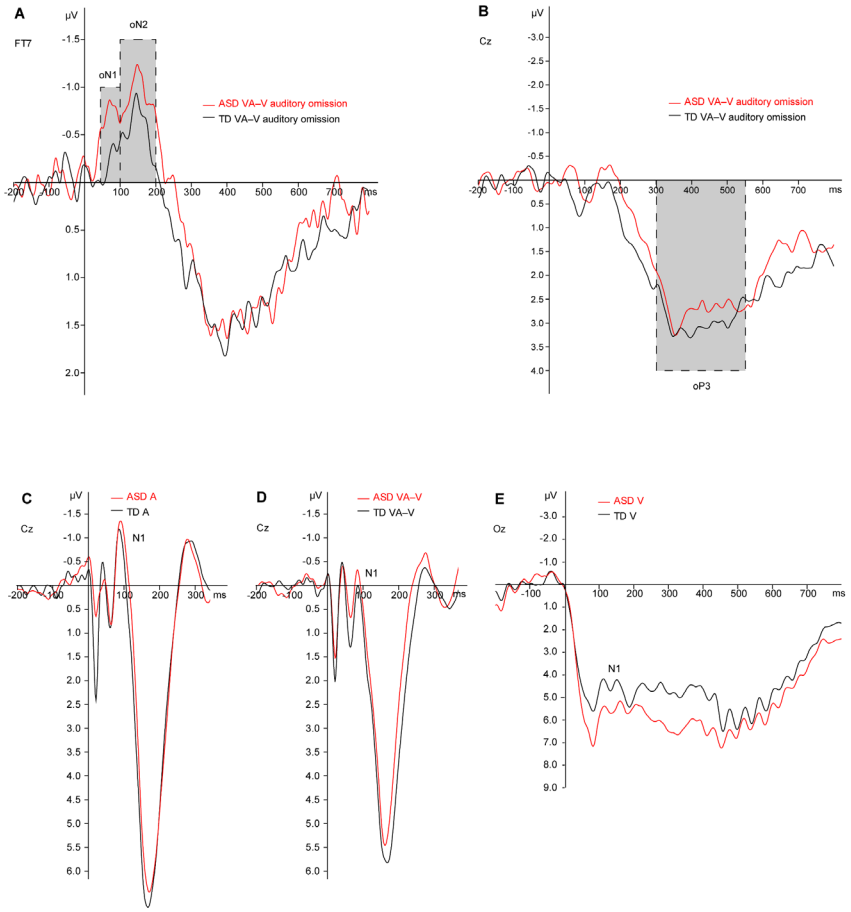


Figure 6.2. Direct comparison of the group-averaged ERPs. Auditory omission ERPs and visual-auditory (VA) ERPs were corrected for visual activity via subtraction of the visual (V) waveform. **A**: The first negative component of the auditory omission ERPs peaked in a time window of 45-100 ms (oN1). A second negative component reached its maximum in 100-200 ms (oN2). Maximal amplitude of the oN1 and oN2 was measured at electrode FT7. **B**: The two negative omission responses were followed by late positive potentials showing maximal amplitudes measured at electrodes Cz in a time window of 300-550 ms (oP3). **C-E**: Group-averaged ERPs for auditory (A), standard visual-auditory (VA–V), and visual (V) stimulation showing maximal amplitudes measured at electrodes Cz (A, standard VA–V), and Oz (V).

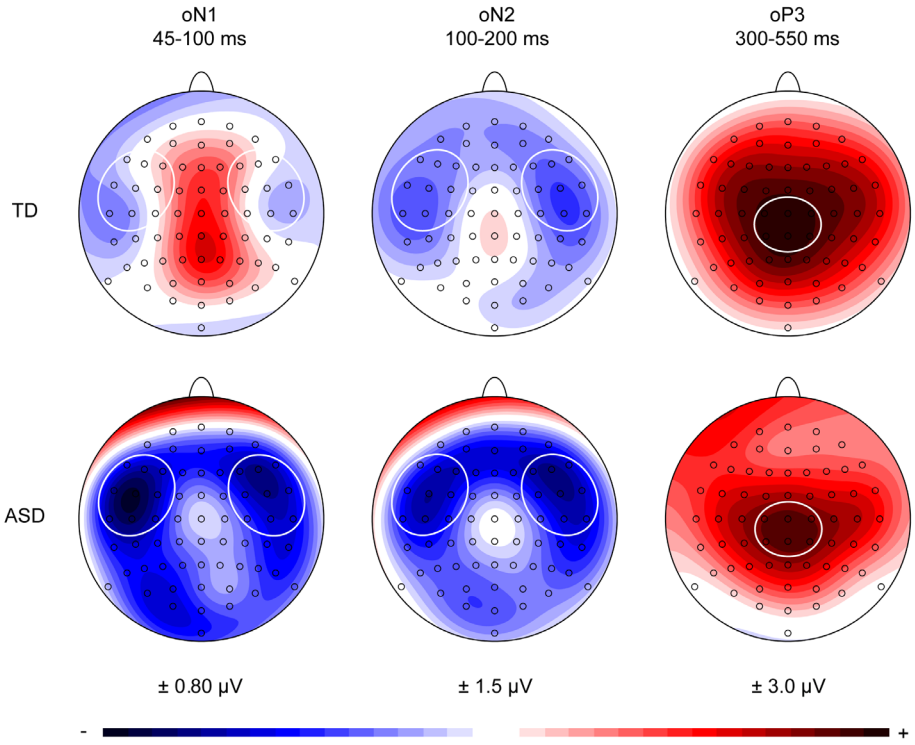


Figure 6.3. Scalp potential maps of the group-averaged visual-corrected auditory omission responses in the denoted oN1 (45-100 ms), oN2 (100-200 ms), and oP3 (300-550 ms) time windows. Based on these scalp distributions, a left-temporal (F7, F5, F3, FT7, FC5, FC3, T7, C5, C3) and right-temporal (F4, F6, F8, FC4, FC6, FT8, C4, C6, T8) region of interest were selected for the oN1 and oN2 time windows. A central-parietal (C1, Cz, C2, CP1, CPz, CP2) region of interest was selected for the oP3 time window.

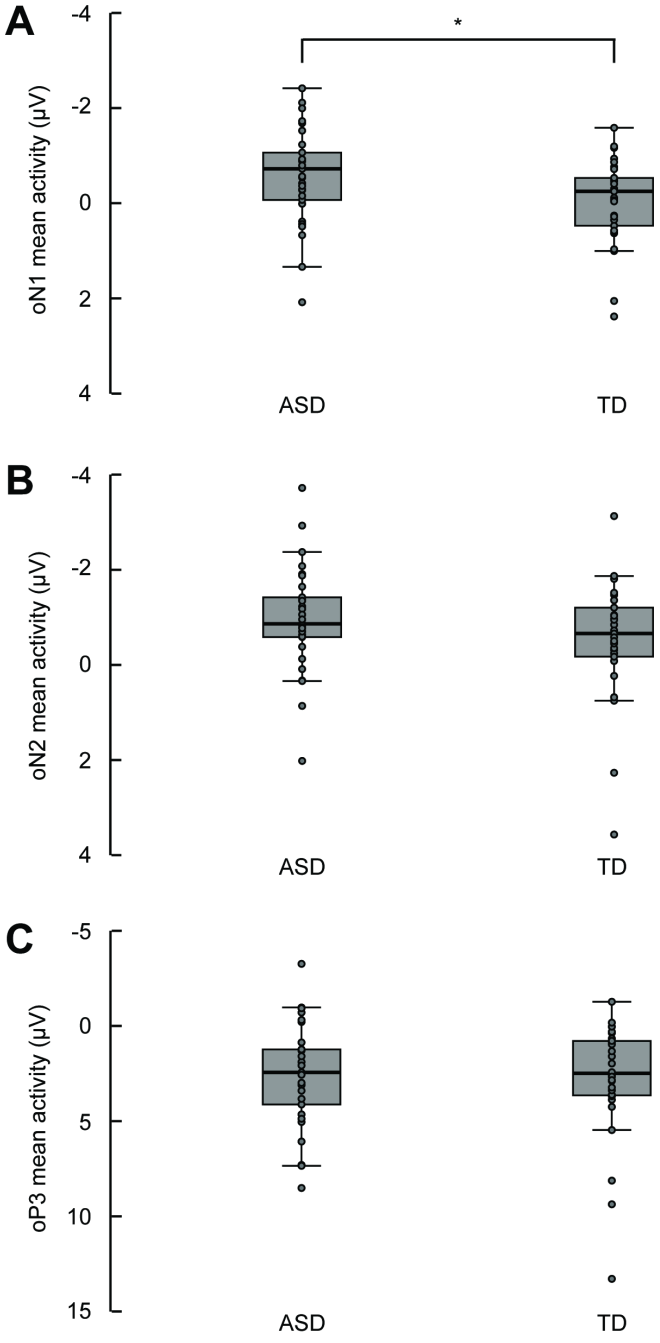


Figure 6.4. Boxplots displaying the group medians and interquartile ranges overlaid with individual data points of the visual-corrected auditory omission responses for the ASD and TD group in the denoted oN1 (45-100 ms), oN2 (100-200 ms), and oP3 (300-550 ms) time windows averaged across regions of interest and electrodes. **A**: The mean activity in the oN1 time window was significantly more negative in the ASD group compared to the TD group. **B** and **C**: The mean activity in the oN2 and oP3 time windows was similar in both groups.

DISCUSSION

The current study tested the hypothesis that predictive coding is impaired in ASD due to overly precise internal prediction models by comparing the neural correlates of visual-auditory prediction errors between autistic individuals and individuals with typical development using a stimulus omission paradigm. The data revealed clear group differences in the early electrophysiological indicators of visual-auditory predictive coding. The oN1, a neural marker of prediction error, was significantly more pronounced in the ASD group, indicating that violations of the visual-auditory predictive model – induced by unexpected auditory omissions – produced larger prediction errors in the ASD group compared to the TD group. Importantly, the increased prediction error signaling in the ASD group could not be explained by between-group differences in the processing of the physical characteristics of the applied stimuli. The current results could thus be indicative of altered visual-auditory predictive coding in ASD.

Previous studies have shown that increasing attention towards an auditory stimulus may increase the amplitude of the N1 response (Lange, Rösler, & Röder, 2003), whereas drawing attention away may result in N1 attenuation (Horváth & Winkler, 2010). Whether attention can affect the oN1 remains to be investigated. But if so, it might be argued that increased attention to the handclap sounds may have resulted in an amplitude increase of the oN1 in the ASD group. An argument against this view is that the N1 for auditory and audiovisual stimulation during standard trials was similar in the ASD and TD group, indicating that sustained attentional differences between groups are an unlikely account for the increased oN1 response in the ASD group.

In both the TD and ASD group, the oN1 was followed by an oN2 and oP3 response. The current results mirror those of previous studies applying motor- and visual-auditory omission paradigms (SanMiguel, Saupe, et al., 2013; SanMiguel, Widmann, et al., 2013; Stekelenburg & Vroomen, 2015; van Laarhoven et al., 2017), in which the oN1 was also followed by an oN2 and oP3 response. The oN2 is assumed to reflect higher-order error evaluation associated with stimulus deviance or the presence of conflict in the context of action monitoring (SanMiguel, Saupe, et al., 2013; SanMiguel, Widmann, et al., 2013; Stekelenburg & Vroomen, 2015; van Laarhoven et al., 2017); in this case a conflict between the visually anticipated sound and the omitted sound. The oP3 likely reflects attention orienting triggered by the unexpected omission of the sound, and the subsequent updating of the internal forward model to minimize future error (Baldi & Itti, 2010; Polich, 2007). Previous research has shown that the oN1 response and oN2-oP3 complex is only elicited by unexpected omissions of sounds of which both the timing and content is

predictable (SanMiguel, Saupe, et al., 2013; SanMiguel, Widmann, et al., 2013; Stekelenburg & Vroomen, 2015; van Laarhoven et al., 2017). The enlarged oN1 response and typical oN2 and oP3 suggest that individuals in the ASD group were able to use the visual motion to predict the upcoming sound during audiovisual stimulation in the standard trials. The current results thus seem to argue against the imprecise or attenuated priors account of ASD (Pellicano & Burr, 2012). When the visual-auditory prediction was not fulfilled, but disrupted by an auditory omission, the ASD group showed an increased error response – as indicated by the atypically large oN1. Given that the amplitude of the oN1 is assumed to be modulated by the precision of the prediction (Arnal, Wyart, & Giraud, 2011; Friston, 2005), the current results suggest that sensory prediction might be overly precise in ASD, as previously hypothesized (van de Cruys et al., 2014). An overly precise predictive model may generate predictions that are overfitted to specific contexts. This overfitting significantly impairs the generalizability of prior expectations to new sensory experiences, which in turn leads to disproportionately large prediction errors in response to unexpected variations in sensory input. The continuous signaling of prediction errors and overfitting of prediction models likely requires an excessive amount of attentional resources – which might explain why autistic individuals are often overwhelmed by sensory stimulation.

In relatively rigid, unambiguous situations, autistic individuals can successfully learn and apply new contingencies (Dawson, Mottron, & Gernsbacher, 2008), and they often excel in detail-focused tasks in which overfitted predictions are advantageous (Robertson & Baron-Cohen, 2017). The experimental paradigm applied in the current study provided a relatively unambiguous context (especially when compared with complex and social interactions). One might therefore expect that, even though the auditory omissions were infrequent and unpredictable, an overly precise predictive model would incorporate the occasional occurrence of an auditory omission after a certain number of iterations to minimize prediction errors in the future. Still, the prediction error – reflected in the oN1 – remained atypically large, which suggests that there was little to no habituation to the auditory omissions in the ASD group. A chronical bias towards sensory input impedes the influence of prior expectations on perception and may cause each unexpected sensory experience to be handled as an error. The current findings may thus be in line with the notion that autistic individuals show alterations in habituation to (unexpected) sensory stimulation because they systematically overweigh the significance of sensory input over prior expectations (Lawson et al., 2014). It should be noted, however, that the signal to noise ratio of the current data does not allow for an analysis of oN1 amplitude over time; so whether habituation to the auditory omissions was indeed absent in the ASD group

remains to be elucidated. Future studies should therefore address if the increased prediction error response in the ASD group can be attributed to overly precise sensory predictions or a lack of habituation to unexpected sensory stimulation. Nevertheless, the current results imply that even in a relatively stable context with little noise, autistic individuals may experience difficulties in anticipating upcoming auditory stimulation.
Recent evidence has shown that self-initiation of tones does not attenuate the auditory N1 in autistic individuals, indicating that autistic individuals may have alterations in anticipating the auditory consequences of their own motor actions (van Laarhoven et al., 2019). The current study extends these findings by demonstrating that the ability to anticipate the sensory consequences of others' actions may be altered in ASD as well. While different predictive mechanisms may underly N1 attenuation (as a marker of fulfilled prediction) and elicitation of the oN1 (as a marker of prediction error), both the absence of N1 attenuation and increased prediction error signaling may indicate that autistic individuals experience difficulties in anticipating upcoming sensory events and seemingly process every new experience afresh rather than mediated by prior expectations. Interaction with the environment becomes especially challenging in social situations, which are inherently noisy and volatile – and thus require flexible and fine-tuned processing of prior expectations, sensory input and prediction errors. A potential consequence of this failure to contextualize sensory information and suppress prediction errors is a constant state of vigilance or sensory alertness – symptoms associated with sensory overload and hyperresponsiveness to sensory stimulation. Indeed, there is evidence that autistic individuals systematically overestimate the volatility of their environment (Lawson, Mathys, & Rees, 2017). Over time, this may lead to frustration, (social) anxiety, repetitive behaviors (e.g. insistence on sameness and stimming as an adaptive coping strategy to control sensory stimulation and attempt to minimize prediction errors), and ultimately, avoidance or hypo-responsiveness to sensory stimulation.

Future studies should focus on when the currently observed alterations in prediction error signaling first emerge throughout development, as the neural response to prediction disruptions may serve as an early marker of autistic symptomatology and potential target for intervention. Ultimately, future work may reveal if and how these alterations in predictive coding can be remediated through clinical applications to improve sensory-perceptual and social functioning of autistic individuals.

CONCLUSIONS

The current results confirm our hypothesis that autistic individuals show alterations in visual-auditory predictive coding. Specifically, unexpected auditory omissions in a sequence of audiovisual recordings in which the visual motion reliably predicted the timing and content of the sound elicited an increased prediction error response in our sample of autistic individuals. The current data suggest that autistic individuals may have impairments in the ability to anticipate the sensory consequences of others' actions, and support the notion of impaired predictive coding as a core deficit underlying atypical sensory perception in ASD.

ACKNOWLEDGMENTS

We wish to thank the staff of Yulius Mental Health Organization, location de Steiger, in particular Sandra Kint, for their assistance with the recruitment of participants with ASD and providing diagnostic assessment information. We also thank Jet Roos, José Hordijk, and Justlin van Bruggen for their help with the data collection.

REFERENCES

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th TR). Washington, DC: Author.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16(*7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Arnal, L. H., Wyart, V., & Giraud, A. L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience, 14*(6), 797–801. https://doi.org/10.1038/nn.2810

Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology, 48*(9), 1276–1283. https://doi.org/10.1111/j.1469-8986.2011.01196.x

Baess, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology, 70*(2), 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks, 23*(5), 649–666. https://doi.org/10.1016/j.neunet.2009.12.007

Baranek, G. T., Watson, L. R., Boyd, B. A., Poe, M. D., David, F. J., & McGuire, L. (2013). Hyporesponsiveness to social and nonsocial sensory stimuli in children with autism, children with developmental delays, and typically developing children. *Development and Psychopathology, 25*(2), 307–320. https://doi.org/10.1017/S0954579412001071

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology, 83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: Comment on Pellicano and Burr. *Trends in Cognitive Sciences, 16*(12), 573–574. https://doi.org/10.1016/j.tics.2012.10.005

Constantino, J. N., & Gruber, C. P. (2012). Social Responsiveness Scale (SRS-2). Torrance, CA: Western Psychological Services.

Dawson, M., Mottron, L., & Gernsbacher, M. A. (2008). Learning in Autism. *In Learning and Memory: A Comprehensive Reference* (pp. 759–772). https://doi.org/10.1016/b978-012370509-9.00152-2

Eussen, M. L. J. M., Louwerse, A., Herba, C. M., Van Gool, A. R., Verheij, F., Verhulst, F. C., & Greaves-Lord, K. (2015). Childhood Facial Recognition Predicts Adolescent Symptom Severity in Autism Spectrum Disorder. *Autism Research, 8*(3), 261–271. https://doi.org/10.1002/aur.1443

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K., Lawson, R., & Frith, C. D. (2013). On hyperpriors and hypopriors: Comment on Pellicano and Burr. *Trends in Cognitive Sciences, 17*(1), 1. https://doi.org/10.1016/j.tics.2012.11.003

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review, 20*(3), 290–322. https://doi.org/10.1007/s11065-010-9138-6

Horváth, J., & Winkler, I. (2010). Distraction in a continuous-stimulation detection task. *Biological Psychology, 83*(3), 229–238. https://doi.org/10.1016/j.biopsycho.2010.01.004

Lange, K., Rösler, F., & Röder, B. (2003). Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: An event-related potential study. *Psychophysiology, 40*(5), 806–817. https://doi.org/10.1111/1469-8986.00081

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience, 20*(9), 1293–1299. https://doi.org/10.1038/nn.4615

Lawson, R. P., Rees, G., & Friston, K. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience, 8*, 302. https://doi.org/10.3389/fnhum.2014.00302

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). Autism Diagnostic Observation Schedule-2 Manual. Los Angeles, CA: Western Psychological Services.

Martikainen, M. H., Kaneko, K. I., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex, 15*(3), 299–302. https://doi.org/10.1093/cercor/bhh131

Pellicano, E., & Burr, D. (2012). When the world becomes “too real”: A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences, 16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Pellicano, E., Jeffery, L., Burr, D., & Rhodes, G. (2007). Abnormal Adaptive Face-Coding Mechanisms in Children with Autism Spectrum Disorder. *Current Biology, 17*(17), 1508–1512. https://doi.org/10.1016/j.cub.2007.07.065

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience, 18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

SanMiguel, I., Saupe, K., & Schröger, E. (2013). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when". *Frontiers in Human Neuroscience, 7*, 407. https://doi.org/10.3389/fnhum.2013.00407

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schröger, E. (2013). Hearing silences: Human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience, 33*(20), 8633–8639. https://doi.org/10.1523/JNEUROSCI.5821-12.2013

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*(12), 1964–1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Stekelenburg, J. J., & Vroomen, J. (2015). Predictive coding of visual-auditory and motor-auditory events: An electrophysiological study. *Brain Research, 162*6, 88–96. https://doi.org/10.1016/j.brainres.2015.01.036

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders, 43*(7), 1517–1526. https://doi.org/10.1007/s10803-012-1695-5

Van de Cruys, S., De-Wit, L., Evers, K., Boets, B., & Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to Pellicano and Burr (TICS, 2012). *I-Perception, 4*(2), 95–97. https://doi.org/10.1068/i0580ic

van de Cruys, S., Evers, K., van der Hallen, R., van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review, 121*(4), 649–675. https://doi.org/10.1037/a0037665

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L. J. M., & Vroomen, J. (2019). Electrophysiological alterations in motor-auditory predictive coding in autism spectrum disorder. *Autism Research, 12*(4), 589–599. https://doi.org/10.1002/aur.2087

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2017). Temporal and identity prediction in visual-auditory events: Electrophysiological evidence from stimulus omissions. *Brain Research, 1661*, 79–87. https://doi.org/10.1016/j.brainres.2017.02.014

Chapter 7

Summary and general discussion

This chapter concludes the dissertation with a summary and general discussion of the main findings, directions for future research, potential implications for clinical practice, and conclusions.

7.1 MAIN FINDINGS

Temporal and identity predictability modulate electrophysiological correlates of visually-induced auditory predictions

Visual anticipatory motion may be incorporated in the internal predictive model and used to generate auditory predictions about when an auditory event is about to happen, and what the anticipated stimulus will sound like. Until now, it was unclear whether the neural correlates of visually-induced auditory predictions are primarily driven by the temporal characteristics, or by the identity features of the anticipated sound. Chapters 2 and 3 described two ERP studies that examined how temporal and identity predictability may affect the neural correlates of predictive coding of an ecologically valid audiovisual event (i.e. a handclap). The results of these studies showed that early electrophysiological markers of fulfilled prediction (i.e. the N1 suppression effect) and prediction error (i.e. the oN1 response) were most pronounced when visual anticipatory motion reliable predicted both the timing and identity of the anticipated sound. When the timing or identity of the sound could not be predicted by visual anticipatory motion, suppression of the auditory N1 was reduced (van Laarhoven, Stekelenburg, & Vroomen, 2020), and the auditory oN1 was abolished (van Laarhoven, Stekelenburg, & Vroomen, 2017). These findings show that early visual-auditory evoked brain responses are modulated by stimulus predictability, and provide support for the theoretical framework of predictive coding, which postulates that early brain responses are shaped by the precision of prior expectations (Friston, 2005). In addition, the study described in Chapter 3 showed that a natural match in identity between the visual and auditory stimulus is not required for inducing prediction error responses, which demonstrates that the internal predictive model is able to adapt to unnatural audiovisual stimulus pairings when utilizing visual motion to generate auditory predictions.

Taken together, the studies described in Chapter 2 and 3 demonstrate the flexibility of the perceptual system in audiovisual predictive coding, and indicate that both temporal and identity predictability modulate visually-induced auditory predictions.

The results of the studies described in Chapter 2 and 3 are in line with research in the motor-auditory domain (Baess, Jacobsen, & Schröger, 2008; Dercksen, Widmann, Schröger, & Wetzel, 2020; SanMiguel, Saupe, & Schröger, 2013). In a previous study by Baess et al. (2008), event-related potentials (ERPs) elicited by self-initiated tones where compared with ERPs elicited by externally-initiated tones. Frequency (pitch) and onset of the sound were either predictable or unpredictable in both conditions. Similar to the results of the study in the visual-auditory domain described in Chapter 2, suppression of the auditory N1 was largest when both sound onset and sound frequency were predictable, and reduced when the onset and frequency could not be predicted. Using an auditory omission paradigm similar to that of the study described in Chapter 3, SanMiguel et al. (2013) found that unexpected omissions of self-initiated tones only elicited prediction errors (i.e. the oN1 response) when the identity of the sound was predictable. More recently, Dercksen et al. (2020) reported reduced omission responses evoked by unexpected omissions of self-initiated tones with unpredictable identity, when compared with unexpected omissions of self-initiated tones with predictable identity (Dercksen et al., 2020). Unfortunately, temporal prediction was not manipulated in either of the studies that applied a stimulus-omission paradigm (Dercksen et al., 2020; SanMiguel et al., 2013), so the relative impact of temporal predictability on elicitation of the oN1 in the motor-auditory domain remains to be elucidated.

Both previous findings in the motor-auditory domain and the results of the current studies in the visual-auditory domain described in Chapter 2 and 3 suggest that predictions of timing and content are essential elements for predictive coding in audition. Future studies should examine if these findings apply to other intersensory domains as well. Other avenues for future research, including the potential role of neural oscillations underlying predictive coding mechanisms, are outlined in section 7.2.

Increased levels of sub-clinical autistic traits are associated with alterations in audiovisual speech processing

Individuals in the general population may display autistic-like behavior to varying degrees (Ruzich et al., 2015). Sub-clinical levels of autistic traits have previously been linked to alterations in multisensory integration of artificial audiovisual stimuli (Donohue, Darling, & Mitroff, 2012; Stevenson et al., 2017; Ujiie, Asai, & Wakabayashi, 2015). Until now, the impact of sub-clinical autistic traits on sensory processing of ecologically valid audiovisual stimuli was largely unclear. The study described in Chapter 4 examined the relationship between sub-clinical levels of autistic traits and audiovisual speech processing in a large non-clinical population. The results of this study showed that increased levels of sub-clinical autistic traits may be related to alterations in audiovisual speech processing (van Laarhoven, Stekelenburg, & Vroomen, 2019). Several associations between autistic traits and indices of audiovisual speech processing were found. Increased overall levels of autistic traits were associated with reduced visual enhancement of noise-masked speech, while more limited (autistic-like) imagination abilities were related to reduced audiovisual perceptual binding of incongruent audiovisual (i.e. McGurk) stimuli. In addition, increased difficulties with attention switching (resulting in more rigid and restricted behavior) were associated with a wider temporal binding window for audiovisual speech stimuli, while an increased tendency to focus on local aspects of sensory inputs (at the expense of global information) was associated with a more narrow temporal binding window. Alterations in audiovisual speech processing, including reduced audiovisual perceptual binding, reduced visual enhancement of noise-masked speech, and altered audiovisual temporal processing, are widely reported in autistic individuals (Feldman et al., 2018). The study described in Chapter 4 extends the existing literature by demonstrating that associations between autistic traits and alterations in audiovisual speech processing may also be observed in the general population.

Taken together, the results from previous work in clinical populations and the study described in Chapter 4 are in line with the notion of a heterogeneous spectrum of autistic symptoms that extends to the general population, and provide evidence for the Broader Autism Phenotype. Section 7.2 outlines several potential avenues for future research, including identification of potential subgroups of autistic individuals who share specific alterations in multisensory processing, and the development of training protocols that may ultimately reduce the impact of alterations in multisensory processing on daily life of autistic individuals.

Predictive coding of motor-auditory and visual-auditory events is altered in autism spectrum disorder

The studies described in Chapter 5 and 6 tested the predictive coding account of autistic symptomatology, which postulates that autistic individuals exhibit alterations in (multi)sensory processing and perception due to a decreased ability to anticipate upcoming sensory stimulation (Lawson, Rees, & Friston, 2014; Pellicano & Burr, 2012; Sinha et al., 2014; van de Cruys et al., 2014). While there is some behavioral evidence that suggests that predictive coding of sensory information may indeed be impaired in autistic individuals (Lawson, Aylward, White, & Rees, 2015; Turi, Karaminis, Pellicano, & Burr, 2016), it had yet to be examined whether these alterations are manifested at the neural level. The studies described in Chapter 5 and 6 addressed this hitherto underexamined issue by comparing the electrophysiological markers of fulfilled motor-auditory prediction (i.e. the N1 suppression effect) and visual-auditory prediction error (i.e. the oN1 response) between autistic individuals and age-matched individuals with typical development. The results of these studies showed clear group differences in the neural correlates of predictive coding in both the motor-auditory and visual-auditory domain. Unlike in age-matched participants with neurotypical development, self-initiation of tones through a button press did not attenuate the auditory N1 in autistic individuals (Chapter 5). In addition, unexpected omissions of a sound of which the timing and content could be predicted by preceding visual anticipatory motion elicited an increased oN1 prediction error response in autistic individuals when compared to their neurotypical peers (Chapter 6). Taken together, the results of these studies demonstrate that the neural correlates of predictive coding of motor-auditory and visual-auditory events are altered in ASD, and indicate that autistic individuals may have impairments in the ability to use predictive information provided by their own actions and those of others to anticipate upcoming auditory stimulation (van Laarhoven, Stekelenburg, Eussen, & Vroomen, 2019, 2020).

Since the emergence of the predictive coding account of autistic symptomatology (Pellicano & Burr, 2012), several elaborations of this hypothesis have been posited that each propose impairments in different levels of predictive processing in ASD. According to the theoretical framework of predictive coding, the brain attempts to predict the sensory environment by minimizing prediction errors resulting from discrepancies between top-down prior expectations and bottom-up sensory input (Friston, 2005). The influence of prior expectations on perception (relative to sensory input) is assumed to be controlled by the precision – or significance – ascribed to prediction errors at each level of the hierarchy (Feldman & Friston, 2010; Friston, 2010). Increased sensory precision is

assumed to bias perception towards sensory input and increases the influence of prediction errors, while decreased sensory precision will bias perception towards prior expectations. Some have argued that predictive processing of sensory information is impaired in autistic individuals due to a decreased ability to infer the probabilistic structure of sensory events (Pellicano & Burr, 2012). Failing to infer the statistics of the environment may result in reduced precision of prior expectations. Imprecise or (in Bayesian terms) 'hypo-priors' decrease the influence of prior knowledge on perception, which in turn may lead to an increased reliance on sensory input. The hypo-priors hypothesis might explain why autistic individuals are often overwhelmed by sensory stimulation, and seemingly experience every sensory event afresh (rather than mediated by prior knowledge). The absence of N1 attenuation for self-initiated tones in the ASD group of the study described in Chapter 5 suggests that perception in ASD may indeed be less affected by prior expectations, and hence more biased towards sensory input. This interpretation is in line with previous behavioral evidence showing little or no aftereffects of adaptation (i.e. recalibration) to audiovisual asynchronies in ASD (Turi et al., 2016).

However, the hypo-priors hypothesis is not in line with behavioral and neural evidence suggesting that autistic individuals are able to form prior expectations. For example, reduced behavioral and perceptual flexibility and insistence on sameness and routines are widely reported in ASD (Lord et al., 2020; Robertson & Baron-Cohen, 2017), which clearly shows that autistic individuals are able to have strong prior beliefs about what should happen in a particular situation. Furthermore, autistic individuals are able to learn and apply new contingencies (Dawson, Mottron, & Gernsbacher, 2008; Manning, Kilner, Neil, Karaminis, & Pellicano, 2017; Sinha et al., 2014), and they often excel in detail-focused tasks in which precise predictions are advantageous (Robertson & Baron-Cohen, 2017). In addition, several studies have reported enhanced auditory change detection abilities in autistic individuals, reflected in an increased neural response to (unexpected) deviant auditory stimulation (Ferri et al., 2003; Kujala et al., 2007), commonly referred to as the mismatch negativity (MMN) response - which suggests that precise prior expectations are formed in autistic individuals. It should be noted, however, that findings on the MMN response in ASD are highly inconsistent due to underpowered small sample sizes and differences in applied paradigms (Schwartz, Shinn-Cunningham, & Tager-Flusberg, 2018). Still, recent behavioral and neuroimaging evidence in the visual domain also suggests that prior knowledge for perceptual inference is preserved in ASD (Utzerath, Schmits, Kok, Buitelaar, & de Lange, 2019; van de Cruys, Vanmarcke, Van de Put, & Wagemans, 2018). Furthermore, the increased prediction error (i.e. the oN1 response) and typical deviancy detection (i.e. the oN2 response) in the ASD group of the study described

in Chapter 6 clearly demonstrate that autistic individuals are able to form (accurate) sensory predictions, and detect violations of those predictions. Taken together, both previous and current findings seem to argue against the hypo-priors account of ASD.

However, if formation of prior expectations is indeed intact in ASD, why did self-initiation of tones not attenuate the auditory N1 in the ASD group of the study described in Chapter 5? An alternative explanation for the results of the study described in Chapter 5 is that, rather than imprecise prior expectations (i.e. hypo-priors), the core deficit in predictive coding in ASD is an inability to attenuate the precision ascribed to sensory signals (Lawson et al., 2014). Uniformly high sensory precision may lead to an imbalance in precision ascribed to sensory input relative to prior expectations, resulting in perception that is less sensitive to contextual information acquired from prior knowledge, and more reliant on sensory input. Both decreased prior precision and increased sensory precision may lead to a perceptual bias towards sensory input. However, unlike the hypo-priors hypothesis, the increased sensory precision account is compatible with previous reports suggesting that autistic individuals are able to form accurate prior expectations (Ferri et al., 2003; Kujala et al., 2007; Lord et al., 2020; Robertson & Baron-Cohen, 2017). The results of the study described in Chapter 5 may thus be explained as a failure to attenuate predictable auditory stimulation. Similarly, the increased prediction error response observed in the ASD group of the study described in Chapter 6 suggests that there was little to no habituation to unexpected auditory omissions. The results of the studies described in Chapter 5 and 6 may thus be in line with the notion that autistic individuals show alterations in habituation to sensory stimulation because they systematically overweigh the significance of sensory input over prior expectations (Lawson et al., 2014).

If a failure to attenuate incoming sensory information is indeed the underlying cause of altered predictive coding in ASD, one would expect less habituation to repeated sensory stimulation in autistic individuals. A recent behavioral study has indeed shown that habituation to stimulus loudness is reduced in autistic individuals (Lawson et al., 2015), and there is some neural evidence for reduced habituation to repeated auditory stimulation in ASD (Kolesnik et al., 2019; Seery, Tager-Flusberg, & Nelson, 2014). Hence, one might expect that after many stimulus repetitions, autistic individuals have larger average ERP amplitudes than neurotypical individuals. However, the ERPs for both auditory and audiovisual stimulation were similar for the ASD and TD group in the study described in Chapter 6 – despite significant between group differences in prediction error response to unexpected auditory omissions. Although the paradigm applied in this study was not specifically designed to examine sensory habituation effects, this does suggest

that habituation to auditory and audiovisual stimulation is intact in ASD at the neural level – at least for the stimuli applied in this study (of note: visual inspection of the ERPs in Figure 6.2 shows a slightly increased N1 response, and decreased P2 response in the ASD group compared to the TD group for both auditory and audiovisual stimulation, but these differences were not statistically significant). Hence, the results of the study described in Chapter 6 might be more in agreement with a recently proposed predictive coding account of ASD, which suggests that, in addition to high sensory precision, autistic individuals may have alterations in the processing of prediction violations. Specifically, the precision – or significance – ascribed to prediction errors may be uniformly and inflexibly high in ASD (van de Cruys et al., 2014). An inability to flexible process prediction errors greatly affects the ability to learn the predictive value of environmental cues. Knowing which cues may be used to predict upcoming sensory events, and which cues are uninformative (and hence should be ignored) is crucial to be able to distinguish between newsworthy environmental changes and random variability in sensory input. Put differently, effective processing and contextualization of sensory information requires that prediction errors do not always need to be used to update and adjust prior expectations, but should sometimes be ignored entirely. Failing to ignore irrelevant prediction errors may generate predictions that are overfitted to specific contexts. This overfitting may significantly impair the ability to generate more global (as opposed to local) prior expectations that can be generalized to new sensory experiences. The consistently large prediction error response to unexpected auditory omissions observed in the ASD group in Chapter 6 could thus be indicative of an inability to adequately process or habituate to unexpected changes in sensory input. This interpretation is in line with a recent fMRI study in the visual domain, which reported typical repetition suppression effects in autistic individuals, but alterations in neural activity in response to unexpected sensory input (Utzerath, Schmits, Buitelaar, & de Lange, 2018). This notion is also in accord with a recent study that examined the influence of top-down contextual information on the neural response to deviant stimuli in autistic individuals (Goris et al., 2018). In individuals with typical development, the amplitude of the neural response to deviant stimuli (i.e. the MMN) is modulated by the probability of occurrence. Higher probability of occurrence of deviant tones typically attenuates the MMN (Wacongne et al., 2011). This modulatory effect is ascribed to context-dependent weighting of prediction errors and subsequent adjustment of prior expectations. The results of the study by Goris et al. (2018) showed that this context-dependent modulation of the MMN was less pronounced in ASD, suggesting that autistic individuals may indeed be less able to flexibly process prediction errors. It should be noted, however, that the signal-to-noise ratio of the data described in Chapter 6 did not allow for an analysis of prediction error response

over time, so whether the increased prediction error response in the ASD group indeed occurred due to an inability to flexibly process prediction errors and failure to habituate to unexpected changes in sensory input remains to be elucidated.

One noteworthy limitation of the studies described in Chapter 5 and 6 is that the included clinical samples were restricted to high-functioning (IQ ≥ 80) autistic individuals. The IQ restriction was necessary to ensure that participants fully comprehended the instructions and requirements of the administered experiments, and were able to attentively participate in the experiments over an extended period of time. Although intellectual functioning was not a confounding factor for the marked differences between the ASD and TD samples, this may limit the generalizability of the results of these studies to low-functioning autistic individuals with an intellectual disability. In addition, both clinical samples consisted of older adolescents and young adults, so it remains to be examined whether the neural correlates of motor-auditory and visual-auditory predictive coding are altered in autistic children. Differences in sensory processing and multisensory integration between neurotypical and autistic individuals are generally more pronounced in children and young adolescents than in older adults (Beker, Foxe, & Molholm, 2018; Feldman et al., 2018). Hence, it could be speculated that alterations in predictive processing in ASD may already occur in early childhood. A recent electrophysiological study has shown that visually-induced suppression of the auditory N1 in artificial audiovisual stimuli is reduced in autistic children (Brandwein et al., 2015), which suggests that this might indeed be the case. If future studies confirm this hypothesis, preferably across large clinical samples representing a wide range of intellectual functioning and age, neural correlates of predictive coding such as those described in Chapter 5 and 6 may potentially serve as electrophysiological biomarkers for ASD (see section 7.3).

Taken together, the studies described in Chapter 5 and 6 provide empirical evidence for alterations in motor-auditory and visual-auditory predictive coding in ASD. Future studies should further scrutinize whether these alterations occur due to imprecise priors, increased sensory precision, or an inability to flexibly process prediction errors – or a combination of these accounts. In addition, future research should investigate whether the neural correlates of predictive coding in ASD are also altered in other (multi)sensory domains, and if the extent to which these alterations occur can be related to clinical severity of ASD. These and other avenues for future research are further discussed in section 7.2. In addition, Section 7.3 outlines several potential implications for clinical practice of the notion that perception in autistic individuals may be overly reliant on sensory input due to alterations in predictive coding.

7.2 DIRECTIONS FOR FUTURE RESEARCH

Neural oscillatory mechanisms of predictive coding in multisensory integration and autism spectrum disorder

The studies described in Chapter 2 and 3 showed similar early electrophysiological brain responses for sounds with unpredictable timing, and sounds with unpredictable identity. The auditory N1 was equally suppressed when the timing or identity of the sound could not be predicted by visual anticipatory motion (Chapter 2), and disruptions in temporal or identity prediction both abolished the elicitation of the oN1 prediction error response (Chapter 3). These findings indicate that temporal and identity predictions are both crucial elements in visual-auditory predictive coding, and are in line with the notion that the brain generates predictions about timing (when) and identity (what) simultaneously (Arnal & Giraud, 2012).

An obvious approach to further investigate the relative contribution of temporal and identity predictions in motor-auditory and visual-auditory predictive coding is to examine the underlying neural oscillatory patterns in sensory attenuation and omission paradigms similar to those applied in Chapter 2 and 3. While temporal and identity prediction are assumed to be integral parts of a common predictive coding mechanism, the underlying neural mechanisms and rhythms may be distinct (Arnal & Giraud, 2012; Friston, 2005). Recently, it has been proposed that synchronization of neural oscillations is a crucial mechanism for effective predictive coding and integration of multisensory information (Bauer, Debener, & Nobre, 2020; Keil & Senkowski, 2018; Van Atteveldt, Murray, Thut, & Schroeder, 2014). Emerging evidence suggests that synchronization of neural oscillations in specific frequency bands may reflect different aspects of predictive coding in multisensory processing (for review, see Keil & Senkowski, 2018). Although the role of specific frequency bands of neural oscillations in temporal and identity prediction is not yet fully understood (Bauer et al., 2020), it has been hypothesized that temporal predictions may modulate evoked brain responses by phase-resetting or entrainment of delta-theta oscillations (1-8 Hz) prior to expected stimulus onset (Arnal & Giraud, 2012). From an ecological perspective this seems plausible, since the delta-theta band corresponds to the typical frequency range of ecologically valid rhythms, such as human body motion or audiovisual speech (Bauer et al., 2020). This notion is supported by a recent ERP study, which showed that viewing visual articulatory movements during perception of audiovisual speech increases delta-theta band coherence (Park, Kayser, Thut, & Gross, 2016). The role of neural oscillations in identity prediction is still relatively understudied, but is has been hypothesized that identity predictions are primarily

reflected in neural oscillations in the beta (10-30 Hz) and gamma (30-100 Hz) range (Arnal & Giraud, 2012; Schneider, Debener, Oostenveld, & Engel, 2008; Widmann, Gruber, Kujala, Tervaniemi, & Schröger, 2007). Future studies should investigate whether temporal and identity predictions of upcoming sensory events are indeed mediated at distinct frequency bands.

In addition, future research should examine whether the alterations in predictive coding in ASD observed in Chapter 5 and 6 can be attributed to alterations in underlying oscillatory activity. Several studies in neurotypical individuals have reported increased neural activity in the beta frequency range prior to onset of predictable sensory stimulation (Arnal, Wyart, & Giraud, 2011; Engel & Fries, 2010; Keil & Senkowski, 2018), which suggests that top-down predictions are primarily mediated by beta-band activity. Assuming that sensory processing in ASD is less affected by prior top-down expectations (as the results of the studies described in Chapter 5 and 6 suggest), prediction-related neural oscillatory activity in the beta-band could be reduced in ASD. Future work on neural oscillations in ASD might be able to reveal these alterations in neural activity prior to onset of predictable versus unpredictable sensory stimulation. Furthermore, future research on oscillatory patterns in predictive coding in ASD might be able to distinguish between neural activity related to sensory predictions and prediction-errors. Recent findings suggest that, in neurotypical individuals, fulfilled predictions are reflected in decreased evoked gamma-band activity, relative to pre-stimulus levels, while prediction errors induced by prediction violations (e.g. unexpected omissions of anticipated sounds) are reflected in increased evoked gamma-band activity (Arnal & Giraud, 2012; Gruber & Müller, 2005; Todorovic, van Ede, Maris, & de Lange, 2011). Both the absence of sensory attenuation for self-initiated tones (Chapter 5), and increased prediction error signaling evoked by unexpected auditory omissions (Chapter 6) suggest that gamma-band activity may be increased in autistic individuals compared to neurotypical individuals. While ASD has indeed been linked to alterations in neural oscillatory activity in the gamma-range (e.g. Gandal et al., 2010), findings are mixed whether neural activity in this frequency band is increased or decreased in ASD (David et al., 2016). Hence, further research is needed to clarify the role of gamma oscillations and other neural oscillatory patterns underlying predictive coding in ASD.

Sensory processing and multisensory integration in the broader autism phenotype

The results of the study described in Chapter 4 showed that different subdomains of autistic-like behavior may be linked to specific alterations in multisensory processing in the general population (van Laarhoven, Stekelenburg, & Vroomen, 2019). Recent evidence suggests that increased levels of sub-clinical autistic traits can be linked to alterations in other sensory domains as well, including tactile and visual perception (Lowe, Stevenson, Barense, Cant, & Ferber, 2018; Yaguchi & Hidaka, 2020). An obvious avenue of research to pursue would be to examine if similar associations between subdomains of autistic symptoms and alterations in multisensory processing can also be identified in clinical populations. Given the marked heterogeneity of autistic symptomatology (Lord et al., 2020; Robertson & Baron-Cohen, 2017), identifying potential subgroups of autistic individuals who share specific alterations in multisensory processing may have important implications for conceptualizations of ASD and potential interventions.

If specific alterations in multisensory processing can indeed be linked to distinct subdomains of autistic symptoms, the impact of these alterations might be reduced by explicit interventions aimed at improving multisensory integration abilities. In neurotypical individuals, several studies have demonstrated that audiovisual temporal processing and perception of noise-masked speech can be enhanced with training (Powers, Hillock, & Wallace, 2009; Song, Skoe, Banai, & Kraus, 2012; Stevenson, Wilson, Powers, & Wallace, 2013). While the impact of training on multisensory integration abilities in autistic individuals is still largely unclear, there is some evidence that audiovisual training may improve speech-in-noise perception in autistic children (Irwin, Preston, Brancazio, D'angelo, & Turcios, 2015). Further research is needed to corroborate these findings and should examine whether the perceptual improvements transfer beyond the context of the trained task and stimuli and – crucially – if the training effects are retained in the long-term. Still, the development of multisensory training protocols in ASD is a potentially promising avenue of research, that may ultimately reduce the impact of alterations in multisensory processing and perception on daily life of autistic individuals.

A crucial factor to be taken into account in the development of such interventions, is that differences in sensory processing and multisensory integration between autistic and neurotypical individuals are generally more pronounced in children and adolescents than in older adults (Beker et al., 2018; Feldman et al., 2018). This suggests that the alterations in sensory processing and multisensory integration associated with ASD may in part be

caused by delayed maturation of cognitive processes. Future studies incorporating a longitudinal or cross-sectional design should examine if the developmental trajectory of sensory processing and multisensory integration is indeed delayed in autistic individuals. Determining if and when the maturational delay of these cognitive abilities occurs during development may be crucial for effective implementation of multisensory training protocols, and may predict the extent to which autistic individuals are able catch up to their neurotypical peers.

Unraveling the neural basis of predictive coding in autism spectrum disorders

Based on the results of the studies described in Chapter 5 and 6, it could be hypothesized that alterations in predictive coding in ASD in the motor-auditory domain may occur due to a failure of sensory attenuation (Lawson et al., 2014), while alterations in visual-auditory predictive coding could be related to an inability to flexibly process prediction errors (van de Cruys et al., 2014). Future research should examine if the neural alterations in ASD discussed in Chapter 5 and 6 can indeed be linked to different mechanisms of predictive coding. An obvious approach to determine whether sensory attenuation in the visual-auditory domain is indeed intact in ASD (as suggested by the results of the study described in Chapter 6), is to examine the neural response to fulfilled prediction in an ecologically valid audiovisual event (as opposed to a motor-auditory event) in autistic individuals. A previous study on audiovisual speech processing in autistic individuals suggests that attenuation of early neural activity (i.e. the N1 and P2) is indeed intact in ASD (Magnée, De Gelder, Van Engeland, & Kemner, 2008). However, the relatively small sample size (N = 12) significantly reduces the statistical power and generalizability of the effects reported in this study, so further research is required to corroborate these findings. In addition, a future study based on an auditory omission paradigm including a motor-auditory event could determine whether prediction error responses are also atypically large in autistic individuals for unexpected omissions of self-initiated tones. Furthermore, future studies should examine whether the neural correlates of predictive coding in ASD are also altered in other intersensory domains (e.g. motor-visual), and in different ecologically valid stimuli (e.g. audiovisual speech).

Several behavioral studies have shown that increased symptom severity in autistic children is associated with decreased accuracy of audiovisual speech perception (Mongillo et al., 2008; Turi et al., 2016; Woynaroski et al., 2013). Associations between increased levels of autistic-like behavior and behavioral alterations in multisensory processing have also been reported in the general population for audiovisual speech (see Chapter 4), and artificial audiovisual stimuli (Donohue et al., 2012; Stevenson et al., 2017;

Ujiie et al., 2015). These findings suggest that the neural correlates of predictive coding in multisensory integration might also be linked to clinical severity of ASD. A notable limitation of the studies described in Chapter 5 and 6 is that additional diagnostic information regarding symptom severity (i.e. ADOS and SRS scores) was not available for all participants in the clinical samples of autistic individuals. Consequently, the data did not allow sufficiently powered correlational analyses to assess potential associations between measures of symptom severity and neural correlates of predictive coding. Future research should investigate whether the alterations in predictive processing described in Chapter 5 and 6 can be linked to clinical severity of ASD. A recent study in a clinical population of fifty-two autistic children (age range 6-17 years, 7 females) has shown that increased severity of ASD, derived from ADOS raw total scores (Gotham, Pickles, & Lord, 2009), is associated with reduced visually-induced suppression of the auditory N1 in artificial audiovisual stimuli (Brandwein et al., 2015). This suggests that increased autistic-like behavior is indeed linked to more pronounced alterations in predictive coding at the neural level. Further research is needed to replicate these findings in more natural stimuli with higher ecological validity (e.g. audiovisual speech), and to determine whether the neural correlates of predictive coding in other intersensory domains can also be linked to symptom severity in ASD.

In addition to examining electrophysiological correlates of predictive processing in other intersensory domains, future neuroimaging studies on structural and functional brain connectivity in autistic individuals may provide valuable insights into the underlying neural mechanisms of predictive coding in ASD. Alterations in brain connectivity are widely reported in ASD (Lord et al., 2020). Several longitudinal studies have shown alterations in structural brain connectivity in infants that later developed ASD, including increased cortical thickness and brain volume (Hazlett et al., 2017; Khundrakpam, Lewis, Kostopoulos, Carbonell, & Evans, 2017; Smith et al., 2016). While it is widely recognized that functional connectivity is altered in ASD (Lord et al., 2020), the underlying patterns are still a matter of debate (e.g. Picci, Gotts, & Scherf, 2016). A prominent hypothesis is that functional brain connectivity in ASD is characterized by long-range underconnectivity (i.e. between neural networks), combined with short-range overconnectivity (i.e. within neural networks). Long-range functional connectivity is assumed to reflect the integration of information between different neural networks, which is crucial for effective top-down processing and contextualization of sensory information in multisensory perception, social communication and interaction (Barber, Caffo, Pekar, & Mostofsky, 2013). If long-range functional connectivity is indeed reduced in ASD, this might explain why perception in ASD is less affected by prior knowledge.

Evidence from EEG and magnetoencephalography (MEG) studies indeed suggests that long-range functional connectivity is reduced in ASD, while short-range connectivity may be typical or increased (O'Reilly, Lewis, & Elsabbagh, 2017). However, recent functional magnetic resonance imaging (fMRI) studies have yielded evidence of both increased and decreased short-range and long-range connectivity in ASD (Hull et al., 2017; Oldehinkel et al., 2019; Picci et al., 2016; Rane et al., 2015), suggesting that alterations in functional brain connectivity in ASD might be network dependent.

Despite a lack of consensus about whether long- and short-range functional connectivity is increased or decreased in ASD, future studies examining the structural and functional brain connectivity in networks underlying (multi)sensory processing in autistic individuals may reveal valuable insights into the underlying neural mechanisms of predictive coding in ASD. One particular brain region of potential interest for future work on motor-auditory predictive coding in ASD is the cerebellum. Findings from two recent ERP studies examining N1 attenuation to self-initiated tones in patients with lesions in the cerebellum suggest that this particular brain region is involved in the generation of motor-auditory predictions (Knolle, Schröger, Baess, & Kotz, 2012; Knolle, Schröger, & Kotz, 2013). Using a paradigm similar to that of the study described in Chapter 5, it was found that the N1 evoked by self-initiated tones was attenuated in neurotypical controls but not in patients with cerebellar lesions, while P2 attenuation for self-initiated tones was similar in both groups. The study described in Chapter 5 showed a similar pattern in ERPs for self-initiated tones in a clinical ASD sample (i.e. absent N1 attenuation, but intact P2 suppression). Although the clinical phenomenology of cerebellar lesion patients and autistic individuals is fundamentally different, the similarities in neural activity between these clinical populations are noteworthy. While there is in fact an emerging literature on cerebellar alterations in ASD (for review, see Hampson & Blatt, 2015), future neuroimaging studies should examine if these similarities in neural correlates of motor-auditory predictive coding indeed stem from alterations in the same underlying neural networks.

Both the absence of sensory attenuation (Chapter 5) and increased prediction error signaling (Chapter 6) indicate that autistic individuals may experience difficulties in anticipating upcoming sensory events and seemingly process every new experience afresh, rather than mediated by prior expectations. A potential consequence of this failure to contextualize sensory information and attenuate sensory input and suppress prediction errors is a constant state of vigilance or sensory alertness – symptoms associated with sensory overload and hyperresponsiveness to sensory stimulation. Recent behavioral and pupillometric evidence indeed suggests that autistic individuals

show a tendency to overestimate the volatility of their sensory environment (Lawson, Mathys, & Rees, 2017). If autistic individuals are indeed hypervigilant to sensory stimulation, the neural networks involved in vigilance might be altered in ASD as well. An important brain region for exhibiting appropriate vigilant behavior to sensory stimulation is the amygdala (Adolphs, 2010). Exposure to salient stimuli typically results in increased amygdala activation. Stimuli can be salient because of their biological relevance (e.g. food, a fearful or happy face), or physical characteristics (e.g. loudness, brightness, predictability of occurrence), and the level of saliency of a particular stimulus may vary depending on the context (Sander, Grafman, & Zalla, 2003; Zalla & Sperduti, 2013). In neurotypical individuals, the neural response of the amygdala is typically increased for unpredictable stimuli, and rapidly decreases over time in response to predictable sensory stimulation (Herry et al., 2007). The amygdala is assumed to be controlled by input from the ventromedial prefrontal cortex (vmPFC). Stronger vmPFC-amygdala connectivity is associated with greater amygdala habituation (Hare et al., 2008). Two recent fMRI studies have shown that vmPFC-amygdala connectivity is reduced in ASD (Kleinhans et al., 2009; Swartz, Wiggins, Carrasco, Lord, & Monk, 2013). In addition, these studies demonstrated that autistic children, adolescents, and adults exhibit reduced habituation of amygdala activation in response to repeated exposure to faces. Furthermore, this decreased amygdala habituation was associated with increased clinical severity of ASD (measured by the SRS and ADOS). Although these findings suggest that autistic individuals may indeed be hypervigilant to sensory stimulation, future studies should examine if amygdala habituation and vmPFC-amygdala connectivity are also reduced in ASD for predictable (non-social) stimulation in other sensory domains (e.g. self-initiated tones).

7.3 POTENTIAL IMPLICATIONS FOR CLINICAL PRACTICE

Autism as a disorder of predictive coding

Despite conceptual differences (for an overview, see Brock, 2012; Friston, Lawson, & Frith, 2013; Lawson et al., 2014; van de Cruys, De-Wit, Evers, Boets, & Wagemans, 2013), the predictive coding accounts of ASD discussed in this dissertation all imply that perception in autistic individuals may be overly reliant on sensory input. The results of the studies described in Chapter 5 and 6 provide empirical evidence for this notion, which may have important implications for conceptualizations of ASD. Failing to contextualize and generalize sensory information in an optimal fashion may lead to atypical sensory perception, including hypo- and hyperresponsiveness to sensory stimulation - symptoms commonly seen in ASD (Lord et al., 2020; Robertson & Baron-Cohen, 2017). Alterations in predictive processing of sensory information will likely have cascading effects on higher-order processes related to social cognition and interaction. While basic rule learning may be intact or enhanced in autistic individuals (Dawson et al., 2008; Manning et al., 2017; Sinha et al., 2014), a perceptual bias towards sensory input might not be beneficial in learning from social contexts, which often require effective use of prior knowledge – as there is typically no clear one-to-one relationship between sensory inputs during social communication. Inferring sarcasm, for example, requires adequate integration of subtle changes in auditory (e.g. frequency and prosody) and visual information (e.g. facial expressions, posture, and gestures). Processing such ambiguous social cues is often challenging for autistic individuals, and they often experience great difficulties in attributing mental states to others (Lord et al., 2020; Robertson & Baron-Cohen, 2017). Understanding the neural basis of the alterations in predictive coding in ASD may therefore be a fundamental part of the explanation of why individuals with ASD often struggle with social communication and interaction with their environment.

In addition to increasing our understanding of the underlying neural mechanisms of autistic symptomatology, electrophysiological correlates of predictive coding in multisensory integration may potentially provide objective biomarkers for diagnostic purposes in clinical practice. EEG markers such as those described in Chapter 5 and 6 could potentially serve as a non-invasive and low-cost approach to objectively diagnose and subtype ASD. In addition, EEG markers can be used to objectively measure sensory processing abilities and perceptual improvements over time, and may thus provide an impetus for development of training protocols aimed at improving (multi)sensory processing in ASD (see section 7.2), which may ultimately improve social functioning and quality of life in individuals with ASD.

7.4 CONCLUSIONS

This dissertation showed that early visual-auditory evoked brain responses are modulated by temporal and identity predictability of the anticipated stimulus. These findings demonstrate that the brain generates auditory predictions about when an auditory event is about to happen and what the anticipated stimulus will sound like, and are in line with the notion that early brain responses are shaped by the precision of prior expectations (Friston, 2005).

In addition, this dissertation provided support for the notion of a heterogeneous spectrum of autistic symptoms that extends to the general population, by demonstrating that increased levels of sub-clinical autistic traits in specific subdomains may be linked to alterations in audiovisual speech processing. Future research should examine if similar associations between subdomains of autistic symptoms and alterations in multisensory processing can also be identified in clinical ASD populations.

Most importantly, this dissertation provided electrophysiological evidence for alterations in predictive coding of self-initiated auditory stimulation in ASD, thereby revealing a potential failure to attenuate predictable sensory input (Lawson et al., 2014). Furthermore, it was demonstrated that unexpected violations of visually-induced auditory predictions evoke an increased prediction error response in autistic individuals, which may be indicative of an inability to flexibly process prediction errors (van de Cruys et al., 2014). Taken together, these findings suggest that autistic individuals may experience difficulties in anticipating upcoming auditory stimulation initiated by their own actions and those of others. Future research should examine whether the neural correlates of predictive coding in ASD are also altered in other (multi)sensory domains, and should aim to identify the most plausible predictive coding account(s) of autistic symptomatology.

REFERENCES

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences*, *1191*, 42–61. https://doi.org/10.1111/j.1749-6632.2010.05445.x

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Arnal, L. H., Wyart, V., & Giraud, A. L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*(6), 797–801. https://doi.org/10.1038/nn.2810

Baess, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology*, *70*(2), 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Barber, A. D., Caffo, B. S., Pekar, J. J., & Mostofsky, S. H. (2013). Developmental changes in within- and between-network connectivity between late childhood and adulthood. *Neuropsychologia*, *51*(1), 156–167. https://doi.org/10.1016/j.neuropsychologia.2012.11.011

Bauer, A. K. R., Debener, S., & Nobre, A. C. (2020). Synchronisation of Neural Oscillations and Cross-modal Influences. *Trends in Cognitive Sciences*, *24*(6), 481–495. https://doi.org/10.1016/j.tics.2020.03.003

Beker, S., Foxe, J. J., & Molholm, S. (2018). Ripe for solution: Delayed development of multisensory processing in autism and its remediation. *Neuroscience and Biobehavioral Reviews*, *84*, 182–192. https://doi.org/10.1016/j.neubiorev.2017.11.008

Brandwein, A. B., Foxe, J. J., Butler, J. S., Frey, H. P., Bates, J. C., Shulman, L. H., & Molholm, S. (2015). Neurophysiological Indices of Atypical Auditory Processing and Multisensory Integration are Associated with Symptom Severity in Autism. *Journal of Autism and Developmental Disorders*, *45*(1), 230–244. https://doi.org/10.1007/s10803-014-2212-9

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: Comment on Pellicano and Burr. *Trends in Cognitive Sciences*, *16*(12), 573–574. https://doi.org/10.1016/j.tics.2012.10.005

David, N., Schneider, T. R., Peiker, I., Al-Jawahiri, R., Engel, A. K., & Milne, E. (2016). Variability of cortical oscillation patterns: A possible endophenotype in autism spectrum disorders? *Neuroscience and Biobehavioral Reviews*, *71*, 590–600. https://doi.org/10.1016/j.neubiorev.2016.09.031

Dawson, M., Mottron, L., & Gernsbacher, M. A. (2008). Learning in autism. In *Learning and Memory: A Comprehensive Reference* (pp. 759–772). https://doi.org/10.1016/B978-012370509-9.00152-2

Dercksen, T. T., Widmann, A., Schröger, E., & Wetzel, N. (2020). Omission related brain responses reflect specific and unspecific action-effect couplings. *NeuroImage*, *215*. https://doi.org/10.1016/j.neuroimage.2020.116840

Donohue, S. E., Darling, E. F., & Mitroff, S. R. (2012). Links between multisensory processing and autism. *Experimental Brain Research*, *222*(4), 377–387. https://doi.org/10.1007/s00221-012-3223-4

Engel, A. K., & Fries, P. (2010). Beta-band oscillations-signalling the status quo? *Current Opinion in Neurobiology*, *20*(2), 156–165. https://doi.org/10.1016/j.conb.2010.02.015

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *95*, 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Ferri, R., Elia, M., Agarwal, N., Lanuzza, B., Musumeci, S. A., & Pennisi, G. (2003). The mismatch negativity and the P3a components of the auditory event-related potentials in autistic low-functioning subjects. *Clinical Neurophysiology*, *114*(9), 1671–1680. https://doi.org/10.1016/S1388-2457(03)00153-6

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. , Lawson, R., & Frith, C. D. (2013). On hyperpriors and hypopriors: Comment on Pellicano and Burr. *Trends in Cognitive Sciences*, *17*(1), 1. https://doi.org/10.1016/j.tics.2012.11.003

Gandal, M. J., Edgar, J. C., Ehrlichman, R. S., Mehta, M., Roberts, T. P. L., & Siegel, S. J. (2010). Validating γ oscillations and delayed auditory responses as translational biomarkers of autism. *Biological Psychiatry*, *68*(12), 1100–1106. https://doi.org/10.1016/j.biopsych.2010.09.031

Goris, J., Braem, S., Nijhof, A. D., Rigoni, D., Deschrijver, E., Van de Cruys, S., … Brass, M. (2018). Sensory Prediction Errors Are Less Modulated by Global Context in Autism Spectrum Disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(8), 667–674. https://doi.org/10.1016/j.bpsc.2018.02.003

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(5), 693–705. https://doi.org/10.1007/s10803-008-0674-3

Gruber, T., & Müller, M. M. (2005). Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cerebral Cortex*, *15*(1), 109–116. https://doi.org/10.1093/cercor/bhh113

Hampson, D. R., & Blatt, G. J. (2015). Autism spectrum disorders and neuropathology of the cerebellum. *Frontiers in Neuroscience*, *9*, 420. https://doi.org/10.3389/fnins.2015.00420

Hare, T. A., Tottenham, N., Galvan, A., Voss, H. U., Glover, G. H., & Casey, B. J. (2008). Biological Substrates of Emotional Reactivity and Regulation in Adolescence During an Emotional Go-Nogo Task. *Biological Psychiatry*, *63*(10), 927–934. https://doi.org/10.1016/j.biopsych.2008.03.015

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., … Piven, J. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, *542*(7641), 348–351. https://doi.org/10.1038/nature21369

Herry, C., Bach, D. R., Esposito, F., Di Salle, F., Perrig, W. J., Scheffler, K., … Seifritz, E. (2007). Processing of temporal unpredictability in human and animal amygdala. *Journal of Neuroscience*, *27*(22), 5958–5966. https://doi.org/10.1523/JNEUROSCI.5218-06.2007

Hull, J. V., Jacokes, Z. J., Torgerson, C. M., Irimia, A., Van Horn, J. D., Aylward, E., … Webb, S. J. (2017). Resting-state functional connectivity in autism spectrum disorders: A review. *Frontiers in Psychiatry*, *7*, 205. https://doi.org/10.3389/fpsyt.2016.00205

Irwin, J., Preston, J., Brancazio, L., D'angelo, M., & Turcios, J. (2015). Development of an audiovisual speech perception app for children with autism spectrum disorders. *Clinical Linguistics and Phonetics*, *29*(1), 76–83. https://doi.org/10.3109/02699206.2014.966395

Keil, J., & Senkowski, D. (2018). Neural Oscillations Orchestrate Multisensory Processing. *Neuroscientist*, *24*(6), 609–626. https://doi.org/10.1177/1073858418755352

Khundrakpam, B. S., Lewis, J. D., Kostopoulos, P., Carbonell, F., & Evans, A. C. (2017). Cortical Thickness Abnormalities in Autism Spectrum Disorders Through Late Childhood, Adolescence, and Adulthood: A Large-Scale MRI Study. *Cerebral Cortex (New York, N.Y. : 1991)*, *27*(3), 1721–1731. https://doi.org/10.1093/cercor/bhx038

Kleinhans, N. M., Johnson, L. C., Richards, T., Mahurin, R., Greenson, J., Dawson, G., & Aylward, E. (2009). Reduced neural habituation in the amygdala and social impairments in autism spectrum disorders. *American Journal of Psychiatry*, *166*(4), 467–475. https://doi.org/10.1176/appi.ajp.2008.07101681

Knolle, F., Schröger, E., Baess, P., & Kotz, S. A. (2012). The Cerebellum Generates Motor-to-Auditory Predictions: ERP Lesion Evidence. *Journal of Cognitive Neuroscience*, *24*(3), 698–706. https://doi.org/10.1162/jocn_a_00167

Knolle, F., Schröger, E., & Kotz, S. A. (2013). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, *49*(9), 2449–2461. https://doi.org/10.1016/j.cortex.2012.12.012

Kolesnik, A., Begum Ali, J., Gliga, T., Guiraud, J., Charman, T., Johnson, M. H., & Jones, E. J. H. (2019). Increased cortical reactivity to repeated tones at 8 months in infants with later ASD. *Translational Psychiatry, 9*(46). https://doi.org/10.1038/s41398-019-0393-x

Kujala, T., Aho, E., Lepistö, T., Jansson-Verkasalo, E., Nieminen-von Wendt, T., von Wendt, L., & Näätänen, R. (2007). Atypical pattern of discriminating sound features in adults with Asperger syndrome as reflected by the mismatch negativity. *Biological Psychology*, *75*(1), 109–114. https://doi.org/10.1016/j.biopsycho.2006.12.007

Lawson, R. P., Aylward, J., White, S., & Rees, G. (2015). A striking reduction of simple loudness adaptation in autism. *Scientific Reports*, *5*, 16157. https://doi.org/10.1038/srep16157

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, *20*(9), 1293–1299. https://doi.org/10.1038/nn.4615

Lawson, R. P., Rees, G., & Friston, K. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, *8*, 302. https://doi.org/10.3389/fnhum.2014.00302

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., … Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, *6*(1). https://doi.org/10.1038/s41572-019-0138-4

Lowe, M. X., Stevenson, R. A., Barense, M. D., Cant, J. S., & Ferber, S. (2018). Relating the perception of visual ensemble statistics to individual levels of autistic traits. *Attention, Perception, and Psychophysics*, *80*(7), 1667–1674. https://doi.org/10.3758/s13414-018-1580-1

Magnée, M. J. C. M., De Gelder, B., Van Engeland, H., & Kemner, C. (2008). Audiovisual speech integration in pervasive developmental disorder: Evidence from event-related potentials. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(9), 995–1000. https://doi.org/10.1111/j.1469-7610.2008.01902.x

Manning, C., Kilner, J., Neil, L., Karaminis, T., & Pellicano, E. (2017). Children on the autism spectrum update their behaviour in response to a volatile environment. *Developmental Science*, *20*(5). https://doi.org/10.1111/desc.12435

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(7), 1349–1358. https://doi.org/10.1007/s10803-007-0521-y

O'Reilly, C., Lewis, J. D., & Elsabbagh, M. (2017). Is functional brain connectivity atypical in autism? A systematic review of EEG and MEG studies. *PLoS ONE*, *12*(5). https://doi.org/10.1371/journal.pone.0175870

Oldehinkel, M., Mennes, M., Marquand, A., Charman, T., Tillmann, J., Ecker, C., … Zwiers, M. P. (2019). Altered Connectivity Between Cerebellum, Visual, and Sensory-Motor Networks in Autism Spectrum Disorder: Results from the EU-AIMS Longitudinal European Autism Project. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(3), 260–270. https://doi.org/10.1016/j.bpsc.2018.11.010

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *ELife*, *5*, e14521. https://doi.org/10.7554/eLife.14521

Pellicano, E., & Burr, D. (2012). When the world becomes "too real": A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Picci, G., Gotts, S. J., & Scherf, K. S. (2016). A theoretical rut: revisiting and critically evaluating the generalized under/over-connectivity hypothesis of autism. *Developmental Science*, *19*(4), 524–549. https://doi.org/10.1111/desc.12467

Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual Training Narrows the Temporal Window of Multisensory Binding. *Journal of Neuroscience*, *29*(39), 12265–12274. https://doi.org/10.1523/jneurosci.3501-09.2009

Rane, P., Cochran, D., Hodge, S. M., Haselgrove, C., Kennedy, D. N., & Frazier, J. A. (2015). Connectivity in Autism: A Review of MRI Connectivity Studies. *Harvard Review of Psychiatry*, *23*(4), 223–244. https://doi.org/10.1097/HRP.0000000000000072

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, *18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism*, *6*(1), 1–12. https://doi.org/10.1186/2040-2392-6-2

Sander, D., Grafman, J., & Zalla, T. (2003). The Human Amygdala: An Evolved System for Relevance Detection. *Reviews in the Neurosciences*, *14*(4), 303–316. https://doi.org/10.1515/REVNEURO.2003.14.4.303

SanMiguel, I., Saupe, K., & Schröger, E. (2013). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when". *Frontiers in Human Neuroscience*, *7*, 407. https://doi.org/10.3389/fnhum.2013.00407

Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *NeuroImage*, *42*(3), 1244–1254. https://doi.org/10.1016/j.neuroimage.2008.05.033

Schwartz, S., Shinn-Cunningham, B., & Tager-Flusberg, H. (2018). Meta-analysis and systematic review of the literature characterizing auditory mismatch negativity in individuals with autism. *Neuroscience and Biobehavioral Reviews, 87,* 106–117. https://doi.org/10.1016/j.neubiorev.2018.01.008

Seery, A., Tager-Flusberg, H., & Nelson, C. A. (2014). Event-related potentials to repeated speech in 9-month-old infants at risk for autism spectrum disorder. *Journal of Neurodevelopmental Disorders, 6*(1), 43. https://doi.org/10.1186/1866-1955-6-43

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., … Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15220–15225. https://doi.org/10.1073/pnas.1416797111

Smith, E., Thurm, A., Greenstein, D., Farmer, C., Swedo, S., Giedd, J., & Raznahan, A. (2016). Cortical thickness change in autism during early childhood. *Human Brain Mapping*, *37*(7), 2616–2629. https://doi.org/10.1002/hbm.23195

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to improve hearing speech in noise: Biological mechanisms. *Cerebral Cortex*, *22*(5), 1180–1190. https://doi.org/10.1093/cercor/bhr196

Stevenson, R. A., Toulmin, J. K., Youm, A., Besney, R. M. A., Schulz, S. E., Barense, M. D., & Ferber, S. (2017). Increases in the autistic trait of attention to detail are associated with decreased multisensory temporal adaptation. *Scientific Reports*, *7*(1), 14354. https://doi.org/10.1038/s41598-017-14632-1

Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Experimental Brain Research*, *225*(4), 479–489. https://doi.org/10.1007/s00221-012-3387-y

Swartz, J. R., Wiggins, J. L., Carrasco, M., Lord, C., & Monk, C. S. (2013). Amygdala habituation and prefrontal functional connectivity in youth with autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*(1), 84–93. https://doi.org/10.1016/j.jaac.2012.10.012

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *Journal of Neuroscience*, *31*(25), 9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific Reports*, *6*, 21756. https://doi.org/10.1038/srep21756

Ujiie, Y., Asai, T., & Wakabayashi, A. (2015). The relationship between level of autistic traits and local bias in the context of the McGurk effect. *Frontiers in Psychology*, *6*, 891. https://doi.org/10.3389/fpsyg.2015.00891

Utzerath, C., Schmits, I. C., Buitelaar, J., & de Lange, F. P. (2018). Adolescents with autism show typical fMRI repetition suppression, but atypical surprise response. *Cortex*, *109*, 25–34. https://doi.org/10.1016/j.cortex.2018.08.019

Utzerath, C., Schmits, I. C., Kok, P., Buitelaar, J., & de Lange, F. P. (2019). No evidence for altered up- and downregulation of brain activity in visual cortex during illusory shape perception in autism. *Cortex*, *117*, 247–256. https://doi.org/10.1016/j.cortex.2019.03.011

Van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: Flexible use of general operations. *Neuron*, *81*(6), 1240–1253. https://doi.org/10.1016/j.neuron.2014.02.044

van de Cruys, S., De-Wit, L., Evers, K., Boets, B., & Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to Pellicano and Burr (TICS, 2012). *I-Perception*, *4*(2), 95–97. https://doi.org/10.1068/i0580ic

van de Cruys, S., Evers, K., van der Hallen, R., van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

van de Cruys, S., Vanmarcke, S., Van de Put, I., & Wagemans, J. (2018). The Use of Prior Knowledge for Perceptual Inference Is Preserved in ASD. *Clinical Psychological Science*, *6*(3), 382–393. https://doi.org/10.1177/2167702617740955

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L. J. M., & Vroomen, J. (2019). Electrophysiological alterations in motor-auditory predictive coding in autism spectrum disorder. *Autism Research*, *12*(4), 589–599. https://doi.org/10.1002/aur.2087

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L. J. M., & Vroomen, J. (2020). Atypical visual-auditory predictive coding in autism spectrum disorder: Electrophysiological evidence from stimulus omissions. *Autism, 24(7),* 1849–1859. https://doi.org/10.1177/1362361320926061

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2017). Temporal and identity prediction in visual-auditory events: Electrophysiological evidence from stimulus omissions. *Brain Research*, *1661*, 79–87. https://doi.org/10.1016/j.brainres.2017.02.014

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2019). Increased sub-clinical levels of autistic traits are associated with reduced multisensory integration of audiovisual speech. *Scientific Reports*, *9*(1), 9535. https://doi.org/10.1038/s41598-019-46084-0

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2020). Suppression of the auditory N1 by visual anticipatory motion is modulated by temporal and identity predictability. *Psychophysiology*, *58*(3), e13749. https://doi.org/10.1111/psyp.13749.

Wacongne, C., Labyt, E., Van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20754–20759. https://doi.org/10.1073/pnas.1117807108

Widmann, A., Gruber, T., Kujala, T., Tervaniemi, M., & Schröger, E. (2007). Binding symbols and sounds: Evidence from event-related oscillatory gamma-band activity. *Cerebral Cortex*, *17*(11), 2696–2702. https://doi.org/10.1093/cercor/bhl178

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(12), 2891–2902. https://doi.org/10.1007/s10803-013-1836-5

Yaguchi, A., & Hidaka, S. (2020). Unique Relationships Between Autistic Traits and Visual, Auditory, and Tactile Sensory Thresholds in Typically Developing Adults. *Perception*, *49*(4), 405–421. https://doi.org/10.1177/0301006620907827

Zalla, T., & Sperduti, M. (2013). The amygdala and the relevance detection theory of autism: An evolutionary perspective. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00894

Appendix

Nederlandse samenvatting
Dankwoord
List of publications
About the author

NEDERLANDSE SAMENVATTING

Achtergrond en doel

In het dagelijks leven worden onze hersenen voortdurend blootgesteld aan sensorische informatie. Ondanks de enorme verscheidenheid en complexiteit van deze signalen zijn onze hersenen in staat om al deze informatie adequaat te verwerken en in de juiste context te plaatsen.

De manier waarop we de wereld om ons heen waarnemen is niet alleen gebaseerd op informatie die we via onze zintuigen ontvangen, maar wordt ook gevormd door onze ervaringen uit het verleden. Een recent geïntroduceerde theorie over de verwerking en integratie van sensorische informatie en eerdere ervaringen, de zogeheten *predictive coding* theorie, gaat ervan uit dat ons brein continu een intern predictiemodel van de wereld om ons heen genereert op basis van informatie die we ontvangen via onze zintuigen en gebeurtenissen die we in het verleden hebben meegemaakt. Dit interne predictiemodel stelt ons in staat om de wereld om ons heen te 'begrijpen' en zorgt ervoor dat onze cognitieve vermogens vooral worden ingezet voor de verwerking van nieuwe of anderszins relevante informatie.

Ons brein maakt gebruik van verschillende signalen ('*cues*') om sensorische signalen te voorspellen. Het simpelweg herhalen van hetzelfde geluid in een vast ritme schept na verloop van tijd een sterke verwachting over wanneer we een geluid zullen horen, en hoe dat geluid zal klinken. Door informatie afkomstig van verschillende zintuigen te integreren is ons brein ook in staat om te anticiperen op meer complexe sensorische signalen. Zo levert het zelf initiëren van een geluid (bijvoorbeeld door op een knop te drukken), of het zien van een beweging die voorafgaat aan een geluid (bijvoorbeeld door te kijken naar iemand die in zijn/haar handen klapt) informatie op over de timing ('*wanneer*') en identiteit ('*wat*') van een aankomend geluid.

Eerder onderzoek heeft aangetoond dat mensen vaak sneller en accurater reageren op voorspelbare sensorische signalen. Op neuraal niveau zijn er ook aanwijzingen gevonden voor de aanwezigheid van een intern predictiemodel. Na het horen van een geluid genereert ons brein automatisch elektrische activiteit die met behulp van elektro-encefalografie (EEG) te meten is. Deze activiteit is over het algemeen *gereduceerd* als het geluid voorspelbaar is, en *verhoogd* als het geluid onvoorspelbaar is. Wanneer we luisteren naar een reeks voorspelbare geluiden waarin onverwacht een geluid wordt weggelaten genereert het brein ook een duidelijk elektrisch signaal, een zogeheten

predictie error. Zowel de relatieve afname in elektrische hersenactiviteit voor voorspelbare auditieve signalen, als het genereren van een predictie error na verstoring van een sterke auditieve verwachting worden gezien als duidelijke signalen voor het bestaan van een intern predictiemodel: zogeheten '*elektrofysiologische markers voor predictive coding'*. Het is vooralsnog niet duidelijk of deze markers voornamelijk optreden door voorspellingen met betrekking tot de *timing* of *identiteit* van het verwachte geluid. In **hoofdstuk 2** en **hoofdstuk 3** is dit nader onderzocht.

Het kunnen voorspellen wat we in bepaalde situaties zullen gaan zien, horen, voelen, ruiken en proeven, stelt ons in staat om te anticiperen op sensorische stimulatie. Het niet nauwkeurig kunnen voorspellen van sensorische prikkels resulteert in een verminderd vermogen om te anticiperen op sensorische stimulatie, met atypische gedragsreacties op sensorische stimulatie - waaronder *hypo*- en *hyperresponsiviteit* - tot mogelijk gevolg.

Hypo- en hyperresponsiviteit voor sensorische prikkels komt relatief vaak voor bij mensen met *Autisme Spectrum Stoornis* (ASS). ASS is een ontwikkelingsstoornis gekenmerkt door moeite met sociale interactie en communicatie, inflexibiliteit in denken en handelen, en veranderingen in sensorische informatieverwerking. Eerder onderzoek heeft aangetoond dat met name de verwerking en integratie van auditieve en visuele informatie ('beeld en geluid') anders verloopt bij mensen met ASS. Zo hebben mensen met ASS vaak moeite met het waarnemen van audiovisuele spraak. De mate waarin zij dit ervaren lijkt samen te hangen met de ernst van de ASS symptomen. Symptomen van ASS kunnen ook voorkomen in de algemene populatie bij mensen zonder een klinische diagnose. In **hoofdstuk 4** is onderzocht of het verband tussen de mate van aanwezigheid van symptomen van ASS en verwerking en integratie van audiovisuele spraak dat eerder is gevonden in klinische populaties ook voorkomt bij mensen in de algemene populatie.

Een recent geïntroduceerde hypothese stelt dat de veranderingen in verwerking en integratie van (multi)sensorische informatie in ASS mogelijk een gevolg zijn van een verminderd vermogen om sensorische prikkels te voorspellen. Hoewel er in gedragsonderzoek al enige bewijs is gevonden dat mensen met ASS inderdaad moeite lijken te hebben met het anticiperen op sensorische stimulatie, is het nog niet duidelijk of er ook op neuraal niveau aanwijzingen zijn dat mensen met ASS minder goed in staat zijn om sensorische prikkels te voorspellen. In **hoofdstuk 5** en **hoofdstuk 6** is dit onderzocht door de eerder genoemde elektrofysiologische markers voor predictive coding te vergelijken tussen mensen met en zonder ASS.

Belangrijkste bevindingen en conclusies

In **hoofdstuk 2** is in een groep jongvolwassenen uit de algemene populatie (*N* = 29) onderzocht of de relatieve afname van elektrische hersenactiviteit die optreedt na het horen van een voorspelbaar geluid afhangt van de mate waarin de timing en identiteit van dit geluid zijn te voorspellen door het zien van een voorafgaande beweging (een video van een handklap). Voor deze video is specifiek gekozen omdat het zien van handgebaren zoals een handklap vaak voorkomt in het dagelijks leven. Hierdoor laten de resultaten van deze experimentele studie zich beter vertalen naar de praktijk. De resultaten van deze studie lieten zien dat de reductie in elektrische hersenactiviteit het grootst was wanneer de voorafgaande visuele beweging zowel de timing ('*wanneer'*) als de identiteit ('*wat'*) van het geluid betrouwbaar voorspelde. Deze reductie was verminderd wanneer de timing of identiteit van het geluid niet kon worden voorspeld op basis van de video.

In **hoofdstuk 3** is in een groep jongvolwassenen uit de algemene populatie (*N* = 27) aan de hand van dezelfde video van een handklap bekeken in hoeverre predicties met betrekking tot de timing of identiteit van het verwachte geluid van invloed zijn op het error signaal dat het brein genereert in reactie op een onverwachte verstoring van de auditieve voorspelling. De resultaten van deze studie lieten zien dat het onverwachts weglaten van een geluid in een reeks geluiden alleen resulteerde in een duidelijk error signaal wanneer zowel de identiteit als de timing van het geluid betrouwbaar konden worden voorspeld aan de hand van de video.

De studies beschreven in hoofdstuk 2 en 3 tonen aan dat de elektrische activiteit in het brein in reactie op het horen van een geluid wordt beïnvloed door predicties met betrekking tot zowel de timing als de identiteit van het verwachte geluid.

In **hoofdstuk 4** is in een grote groep jongvolwassenen (*N* = 101) onderzocht of er een verband bestaat tussen de mate van aanwezigheid van symptomen van ASS en de verwerking en integratie van audiovisuele spraak bij mensen in de algemene populatie. In eerder onderzoek in klinische populaties is gevonden dat de verwerking en integratie van audiovisuele spraak anders verloopt bij mensen met ASS. De mate waarin mensen met ASS moeite ervaren met audiovisuele spraakwaarneming lijkt bovendien afhankelijk van de ernst van de ASS symptomen. De resultaten van de studie beschreven in hoofdstuk 4 lieten zien dat ook in de algemene populatie een toename in subklinische symptomen van ASS samenhangt met een verminderd vermogen om audiovisuele spraak te verstaan. Daarnaast wijzen de resultaten erop dat specifieke symptomen van ASS mogelijk samenhangen met de manier waarop audiovisuele spraak wordt geïntegreerd. Een

toename in aandacht voor detail lijkt samen te hangen met een lagere tolerantie voor asynchrone audiovisuele signalen, terwijl een verminderd vermogen om flexibel te denken en handelen mogelijk samenhangt met een verhoogde tolerantie voor asynchrone audiovisuele signalen. Daarnaast is gevonden dat een verminderd voorstellingsvermogen samenhangt met een verminderde gevoeligheid voor audiovisuele illusies.

De resultaten van de studie beschreven in hoofdstuk 4 laten zien dat er ook in de algemene populatie sprake lijkt te zijn van een verband tussen symptomen van ASS en de verwerking en integratie van audiovisuele spraak.

In **hoofdstuk 5** is de hypothese getest dat de veranderingen in verwerking en integratie van sensorische informatie gerelateerd aan ASS mogelijk een gevolg zijn van een verminderd vermogen om te anticiperen op sensorische stimulatie. Na het horen van een geluid genereert ons brein automatisch elektrische activiteit. Wanneer we datzelfde geluid zelf initiëren kunnen we normaliter beter anticiperen op het geluid, en zodoende is de elektrische hersenactiviteit die door ons brein wordt gegenereerd gereduceerd. De resultaten van de studie in hoofdstuk 5 laten zien dat het zelf initiëren van een geluid, door op een knop te drukken, bij de onderzochte groep jongeren en jongvolwassenen met ASS (n = 30) *niet* resulteerde in een afname in elektrische hersenactiviteit. Dit in tegenstelling tot een groep mensen uit dezelfde leeftijdscategorie zonder ASS (n = 30), waarbij het zelf initiëren van het geluid *wel* resulteerde in een reductie in hersenactiviteit.

In **hoofdstuk 6** is onderzocht in hoeverre de hersenactiviteit in reactie op een plotselinge verstoring van sensorische stimulatie verschillend is voor mensen met en zonder ASS. In deze studie werd steeds onverwachts het geluid weggelaten - een zogeheten *auditieve omissie* - in een reeks geluiden waarvan de timing en identiteit kon worden voorspelt op basis van een voorafgaande beweging (een video van een handklap). De resultaten van deze studie laten zien dat deze onverwachtse auditieve omissies bij de onderzochte groep jongeren en jongvolwassenen met ASS (n = 29) een *verhoogd predictie error* signaal induceerde in vergelijking met een groep mensen uit dezelfde leeftijdscategorie zonder ASS (n = 29).

De resultaten van de studies beschreven in hoofdstuk 5 en 6 suggereren dat mensen met ASS mogelijk minder goed in staat zijn om te anticiperen op auditieve prikkels. Daarnaast verloopt de neurale verwerking van onverwachtse verstoringen in sensorische stimulatie mogelijk anders bij mensen met ASS in vergelijking met mensen zonder ASS.

Implicaties voor de klinische praktijk en aanbevelingen voor vervolgonderzoek
De bevindingen van dit proefschrift die het meest relevant zijn voor de klinische praktijk zijn dat mensen met ASS minder goed in staat lijken te zijn om te anticiperen op auditieve prikkels en mogelijk meer moeite hebben met de verwerking van onverwachtse verstoringen in sensorische stimulatie.

Een verminderd vermogen om te kunnen anticiperen op sensorische prikkels kan niet alleen leiden tot atypische gedragsreacties, waaronder hypo- en hyper- responsiviteit voor sensorische stimulatie, maar heeft mogelijk ook gevolgen voor de sociale cognitieve vaardigheden. In sociale situaties is het kunnen anticiperen op hetgeen een ander zegt of doet van cruciaal belang. Het begrijpen van sarcasme vereist bijvoorbeeld de integratie van subtiele verschillen in auditieve (e.g. toonhoogte en prosodie) en visuele informatie (e.g. gezichtsuitdrukkingen, lichaamstaal). Het juist interpreteren van dergelijke ambigue sociale cues is vaak lastig voor mensen met ASS. De resultaten van hoofdstuk 5 en 6 van dit proefschrift suggereren dat de oorzaak hiervoor mogelijk ligt in veranderingen in interne neurale predictiemechanismen voor auditieve prikkels. Toekomstig onderzoek moet uitwijzen of mensen met ASS ook meer moeite hebben met het anticiperen op sensorische prikkels en verwerken van onverwachtse verstoringen in andere (multi)sensorische domeinen, en of de mate waarin zij dit ervaren samenhangt met de ernst van ASS symptomen.

Naast het vergroten van wetenschappelijke kennis over de onderliggende neurale mechanismen van de veranderingen in sensorische informatieverwerking in ASS kan verder onderzoek naar elektrofysiologische correlaten voor predictive coding mogelijk leiden tot een biomarker voor ASS die kan worden toegepast in de klinische praktijk. Elektrofysiologische correlaten zoals besproken in hoofdstuk 5 en 6 kunnen potentieel worden doorontwikkeld tot non-invasieve en relatief betaalbare aanvullingen op het huidige diagnostisch traject. Met name voor mensen waarbij de gedragskenmerken niet altijd goed te beoordelen zijn kunnen dergelijke elektrofysiologische correlaten mogelijk als objectief meetinstrument worden ingezet.

Tot slot kunnen dergelijke correlaten mogelijk worden toegepast om verbeteringen in sensorische informatieverwerking objectief in kaart te brengen, en kunnen zodoende een impuls geven aan de ontwikkeling van trainingsprotocollen gericht op het verbeteren van sensorische informatieverwerking van mensen met ASS.

DANKWOORD

Dit proefschrift was niet tot stand gekomen zonder de inzet, begeleiding, waardevolle adviezen en inzichten van mijn promotor, copromotor en collega's, de bijdragen van talloze deelnemers en student-assistenten, en de onvoorwaardelijke steun van mijn aanstaande vrouw, vrienden en familie. Graag wil ik iedereen hiervoor van harte bedanken. Een aantal mensen wil ik hier in het bijzonder noemen.

Op de eerste plaats gaat mijn dank uit naar mijn promotor **Prof. dr. Vroomen**. Jean, bedankt voor je begeleiding, het vertrouwen en de vrijheid die je mij hebt gegeven om mezelf als academicus te kunnen ontwikkelen. Ik bewonder je vermogen om altijd kritisch en to the point te blijven en waardeer je directe en snelle manier van denken en communiceren. Het is alweer acht jaar geleden dat ik mijn bachelor thesis onder jouw begeleiding schreef. Inmiddels zijn we een proefschrift verder, en wat mij betreft liggen er nog een hoop mooie publicaties in het verschiet. *"On we go!"*

Graag wil ik ook mijn copromotor **dr. Stekelenburg** in het bijzonder bedanken. Jeroen, bedankt voor al je adviezen, inspiratie en wijze EEG lessen. Ik bewonder je enthousiasme voor ons vakgebied, je humor en oprechtheid. De kritische vragen die je als 'advocaat van de duivel' menigmaal hebt gesteld hebben onmiskenbaar bijgedragen aan de inhoud van dit proefschrift. Ik waardeer het erg dat je deur ook altijd openstaat voor niet-werkgerelateerde serieuze en minder serieuze zaken. Ik hoop dat we nog lang samen zullen werken!

Daarnaast wil ik de leden van de promotiecommissie, **prof. dr. Geurts**, **prof. dr. Talsma, prof. dr. Wagemans** en **prof. dr. Swerts**, van harte bedanken voor de tijd en moeite die u genomen heeft voor het lezen en beoordelen van dit proefschrift en het opponeren bij de verdediging.

Dr. Eussen, Mart, bedankt voor je deskundige blik, inzet en hulp als co-auteur bij de klinische hoofdstukken van dit proefschrift. Het is jammer dat je het onderzoek gaat verlaten, maar wellicht dat onze paden elkaar nog kruisen in de toekomst zodat we de succesvolle samenwerking kunnen voortzetten.

Vanzelfsprekend gaat mijn dank ook uit naar alle deelnemers van de onderzoeken. In het bijzonder bedank ik **alle jongeren met autisme** voor hun inzet en deelname aan de verschillende studies. Jullie bijdrage is van onschatbare waarde voor dit proefschrift.

Erwin Ramadhin, dank voor je hulp bij alle logistieke zaken en het faciliteren van de onderzoeksruimtes bij Yulius. **Sandra Kint**, **Claudia Agenant**, **B.J. van Pelt**, en alle andere therapeuten en behandelaren van De Steiger, enorm bedankt voor jullie belangeloze inzet en alle tijd en moeite die jullie hebben gestoken in het benaderen, informeren en inplannen van alle jongeren.

Dank ook aan alle **collega's** van het Departement Cognitieve Neuropsychologie voor de fijne samenwerking, gezellige lunchpauzes, borrels en teamuitjes.

Mirjam, als stagebegeleider heb je mij op weg geholpen met mijn allereerste publicatie en heb je mij de fijne kneepjes van het thesis begeleidingsproces en later het coördinatorschap bijgebracht. Bedankt dat je altijd voor me klaar stond, niet alleen voor praktische zaken maar ook voor een goed gesprek. **Marion**, halverwege mijn PhD traject zijn we collega's geworden en inmiddels zitten we allebei op de tenure-track. Het is fijn om een collega uit hetzelfde bouwjaar te hebben. Je enthousiasme voor ons vakgebied en netwerk skills zijn inspirerend. Ik kijk uit naar het opzetten van ons ambitieuze onderzoeksproject! **Martijn**, halverwege mijn PhD traject werden we kamergenoten. Als fervente koffieliefhebbers wisten we de weg naar de automaat vaak samen te vinden. Dank voor de diepgaande gesprekken over ERPs, fonemen, heirlooms en over het (jaarlijkse terugkerende) voornemen om die sportkaart toch maar eens te gaan gebruiken. **Geert,** dank voor het beoordelen van mijn BKO portfolio en je advies bij het overnemen van de cursus Hersenen en Gedrag. **Wouter**, tot nu toe hebben we vooral samengewerkt op het gebied van onderwijs. Het lijkt me mooi om in de toekomst een onderzoeksproject op te zetten waarin jouw fMRI expertise en de bevindingen uit dit proefschrift samenkomen.

Johan en **Bram**, bedankt dat jullie achter mij staan tijdens de verdediging.

Johan, al 18 jaar zijn wij bevriend. We hebben in al die jaren een hoop mooie maar ook moeilijke momenten in onze levens met elkaar kunnen delen. Bedankt voor je oprechte vriendschap en onvoorwaardelijke steun.

Bram, we zijn samen opgegroeid en begrijpen elkaar daardoor als geen ander. Ik ben trots op ons doorzettingsvermogen en vastberadenheid om ondanks alles altijd de positieve kanten in te zien en te genieten van het leven – zoals Pa dat zou hebben gewild.

Mam, Marieke, bedankt voor jullie interesse en betrokkenheid en steun door de jaren heen. In Breda staat de deur altijd voor mij open en jullie staan altijd voor mij klaar, dat waardeer ik enorm.

Lieve **Irene**, al meer dan 10 jaar ben jij mijn maatje waar ik onvoorwaardelijk op kan rekenen en lief en leed mee deel. Zonder jouw liefde, steun, vertrouwen, geduld en relativeringsvermogen was dit proefschrift er niet geweest. Ik zou een heel proefschrift kunnen vullen met lieve woorden over jou. Een tweede proefschrift zal ik je niet aandoen. Die lieve woorden bewaar ik voor onze bruiloft en voor alle mooie momenten die nog voor ons liggen.

Tilburg, 17 maart 2021

LIST OF PUBLICATIONS

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2020). Suppression of the auditory N1 by visual anticipatory motion is modulated by temporal and identity predictability. *Psychophysiology*, *58*(3), e13749. https://doi.org/10.1111/psyp.13749

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L. J. M., & Vroomen, J. (2020). Atypical visual-auditory predictive coding in autism spectrum disorder: Electrophysiological evidence from stimulus omissions. *Autism*, *24*(7), 1849–1859. https://doi.org/10.1177/1362361320926061

Burgering, M. A., **van Laarhoven, T.**, Baart, M., & Vroomen, J. (2020). Fluidity in the perception of auditory speech: Cross-modal recalibration of voice gender and vowel identity by a talking face. *Quarterly Journal of Experimental Psychology*, *73*(6), 957–967. https://doi.org/10.1177/1747021819900884

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2019). Increased sub-clinical levels of autistic traits are associated with reduced multisensory integration of audiovisual speech. *Scientific Reports*, *9*(1), 9535. https://doi.org/10.1038/s41598-019-46084-0

van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L. J. M., & Vroomen, J. (2019). Electrophysiological alterations in motor-auditory predictive coding in autism spectrum disorder. *Autism Research*, *12*(4), 589–599. https://doi.org/10.1002/aur.2087

van Laarhoven, T., Keetels, M., Schakel, L., & Vroomen, J. (2018). Audio-visual speech in noise perception in dyslexia. *Developmental Science*, *21*(1), e12504. https://doi.org/10.1111/desc.12504

van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2017). Temporal and identity prediction in visual-auditory events: Electrophysiological evidence from stimulus omissions. *Brain Research*, *1661*, 79–87. https://doi.org/10.1016/j.brainres.2017.02.014

ABOUT THE AUTHOR

Thijs van Laarhoven was born on January 31, 1988 in Oudenbosch, The Netherlands. After completing secondary school at NewmanCollege in Breda in 2006, he moved to Delft to study at TUDelft. In 2010 he obtained his Bachelor degree in Industrial Design Engineering. It was only then that he started to take his education more seriously. Realizing he was on the wrong career path, he moved to Tilburg to study at Tilburg University and fulfill his ambition to pursue an academic career. He obtained his Bachelor degree in Psychology & Health in 2013 with distinction. After graduating cum laude for his Master's degree in Medical Psychology in 2015, he joined the department of Cognitive Neuropsychology at Tilburg University as a PhD student. During his PhD, he conducted several behavioral and electrophysiological studies on audiovisual speech perception and electrophysiological markers of predictive coding in multisensory integration. He successfully collaborated with Yulius Academy on several research projects on multisensory processing in autism spectrum disorder. He supervised many students working on their bachelor and master thesis, and was extensively involved in educational and teaching activities for several courses. In 2019, he obtained his University Teaching Qualification. After finishing his PhD, he continued to pursue his academic career and intellectual interests as an Assistant Professor (tenure track) at the department of Cognitive Neuropsychology at Tilburg University.