Neural entrainment to continuous speech and language processing in the early years of life.

Holly Louise Bradley

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Speech, Hearing and Phonetic Sciences

University College London

2021

Declaration

I, Holly Louise Bradley, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Holly Louise Bradley

Acknowledgements

I would firstly like to thank my fantastic supervisor Paul Iverson for providing me with this opportunity and believing that a small town Yorkshire girl could survive in London! I have learnt so much from Paul and I am so thankful for all the advice that he has given me over the past four years, he was also a great travel buddy to drink beers with at WILD conference in Berlin. It has truly been a wonderful experience having Paul as my mentor and friend, he has been a huge help in making me the researcher that I am today. Thank you to the Economic and Social Research Council who awarded me with the 1+3 scholarship that enabled me to conduct my research here at UCL and made all of this possible.

Dr. Kathleen McCarthy and Christine Hsu were invaluable in my first year and taught me everything I needed to know in order to run infant EEG, recruit babies at Rhyme Time and the importance of always having a supply of snacks when working with young children. I would also like to thank a countless amount of people in Chandler House for helping me along the way. Andrew Clark was always on hand to help me set up experiments and provide endless supplies of mascara brushes for EEG cap cleaning, Jyrki Tuomainen offered a lot of guidance and support through the PhD process, and Richard Jardine was always incredibly helpful with any questions and administrative worries that I ever had. Thank you to all the security and cleaning staff at Chandler House that help to keep the building running smoothly and letting me in every time that I forgot my student card (which was rather frequently).

Thanks to the other members of Paul's research group: Luke Martin was always there to help out with both technical and mental breakdowns and was easily coerced into helping me with one of my babies if we could go to the pub afterwards. Jieun Song has been a great help in answering my general queries and was always a friendly face to see. Begoña Pericas Herrero was a superstar helping with the babies in my second study and I very much enjoyed teaching her the ways of baby testing!

Thank you so much to all the lovely babies who took part in my research, and all the parents that brought them to the lab with a small suitcase of snacks in order to keep them happy, an especially big thanks to the mums that brought twins along because on those days I got a 2 for 1 deal on baby data. Without the willing babies and parents, this research would have not been possible and I truly enjoyed my time spent with every single one.

I would like to thank my mum and dad, Roger and Gillian Bradley, for always believing in me and being proud of me no matter what; all you ever wanted me to be was happy. The love and support from you both throughout my whole life has made me who I am today and I owe everything to you both. Perhaps most importantly, thank you to my other half Nathan Alormele for being with me every single step of this journey and always reminding me that I could do this even when everything felt like it was against me. Also, the last minute holidays and getting me Norman the cat helped a lot too, so thanks for that. Finally, thank you dear reader for spending your time to read this thesis.

For Grandma Joan

Abstract

This thesis aimed to explore the neural mechanisms of language processing in infants under 12 months of age by using EEG measures of speech processing. More specifically, I wanted to investigate if infants are able to engage in the auditory neural tracking of continuous speech and how this processing can be modulated by infant attention and different linguistic environments. Limited research has investigated this phenomenon of neural tracking in infants and the potential effects that this may have on later language development.

Experiment 1 set the groundwork for the thesis by establishing a reliable method to measure cortical entrainment by 36 infants to the amplitude envelope of continuous speech. The results demonstrated that infants have entrainment to speech much like has been found in adults. Additionally, infants show a reliable elicitation of the Acoustic Change Complex (ACC). Follow up language assessments were conducted with these infants approximately two years later; however, no significant predictors of coherence on later language outcomes were found.

The aim of Experiment 2 was to discover how neural entrainment can be modulated by infant attention. Twenty infants were measured on their ability to selectively attend to a target speaker while in the presence of a distractor of matching acoustic intensity. Coherence values were found for the target, the distractor and for the dual signal (both target and distractor together). Thus, it seems that infant attention may be fluctuating between the two speech signals leading to them entraining to both simultaneously. However, the results were not clear.

Thus, Experiment 3 expanded on from Experiment 2. However, now EEG was recorded from 30 infants who listened to speech with no acoustic interference and speech-in-noise with a signal-to-noise ratio of 10dB. Additionally, it was investigated whether bilingualism has any potential effects on this process. Similar coherence values were observed when infants listened to speech in both conditions (quiet and noise), suggesting that infants successfully inhibited the disruptive effects of the masker. No effects of bilingualism on neural entrainment were present.

For the fourth study we wanted to continue investigating infant auditory-neural entrainment when exposed to more varying levels of background noise. However, due to the COVID-19 pandemic all testing was moved online. Thus, for Experiment 4 we developed a piece of online software (the memory card game) that could be used remotely. Seventy three children ranging from 4 to 12 years old participated in the online experiment in order to explore how the demands of a speech recognition task interact with masker type and language and how this changes with age during childhood. Results showed that performance on the memory card game improved with age but was not affected by masker type or language background. This improvement with age is most likely a result of improved speech perception capabilities.

Overall, this thesis provides a reliable methodology for measuring neural entrainment in infants and a greater understanding of the mechanisms of speech processing in infancy and beyond.

Impact statement

During participant recruitment for this thesis I was required to approach parents and caregivers and speak at baby groups and classes in order to obtain participants. By engaging with infants' parents or caregivers and explaining the aims of the study, they were able to become more engaged with developmental research and understand more about the cognitive workings of their child. Parts of this thesis have been presented at conferences (Workshop of Infant Language Development, 2019) and will be continued to be used as a building block for future research into infant neural entrainment to speech and, more broadly, the general linguistic and cognitive capabilities of developing children.

This advanced research into infants' neural entrainment can result in a potentially wide range of social implications. An investigation into the social, cognitive and linguistic development of children allows us to improve social welfare and potentially enhance quality of life. This thesis helps us to acquire a greater understanding of how babies most effectively process and engage with speech, and the type of environment in which an infant can developmentally flourish. This new information can be used to inform parents and early years education practitioners with the aim of facilitating and enhancing child development. If we can improve knowledge of infant development, then future potential economic and social benefits may accrue arising from an improvement in learning strategies and UK, and worldwide, education policy for infants.

Table of Contents

List of figures

List of tables

1. General Introduction

From birth, infants possess the basic attentional and perceptive capabilities that allow them to engage with speech in order to ultimately acquire language. The first year of life is fundamentally important with regards to speech development. However, what we know about the mechanisms of this process within the early years remains extremely limited. In the first year of life, it seems that infant engagement, namely selective attention and the auditory-neural tracking of speech, is crucially important for learning language, but more work needs to be done in understanding the specific roles that these processes have on successful language development. Thus, this current thesis aims to investigate if infants are engaging with the auditory-neural tracking of continuous speech, if their selective attention is modulating this process, and what this means in terms of their language developmental journey.

Understanding how speech is processed in the auditory cortex has been a significant topic in cognitive neuroscience over the past several years. An increasing amount of research has focused on how information is extracted from the acoustic signal and subsequently construed in our mental lexicon. We know that we can use electroencephalography (EEG) in order to understand more about these neurophysiological processes and that when people listen to speech they are cortically tracking the amplitude envelope of the incoming information (Luo & Poeppel, 2007). We also know that selective attention plays an important role in facilitating this cortical tracking (Zion Golumbic et al., 2013); listeners typically have higher entrainment to attended talkers in multi-talker or noisy environments. However, research in this area has predominantly focused on adults and older children who already have proficient language systems and can follow explicit instruction. Studies exploring this phenomenon in pre-verbal infants remain limited thus making it a scientifically valid area of research. If infants under the age of 1 year are engaging in the cortical tracking of speech then this would mean that they are processing speech in the same way that adults are, despite lacking any clear knowledge of the linguistic properties of their native language. This phenomenon of entrainment could arguably work as a precursor for language acquisition and thus these findings would provide a neural insight into infant learning and development.

The overall aim of this thesis was to explore whether this cortical tracking of speech could be found in pre-verbal infants and, by modulating their attention, understand what factors can affect this tracking. In the first study (Chapter 2 *Infant neural entrainment to continuous speech: Initial methodological development*) I sought to measure neural entrainment by infants to the amplitude envelope of continuous speech by using EEG. In other words, I wanted to understand if infants were engaging in the cortical tracking of speech. While undertaking this study, we became aware of two other studies that had found this cortical tracking to speech in infants (Kalashnikova, Peter, Di Liberto, Lalor, & Burnham, 2018; Leong, Byrne, et al., 2017). Thus, we hypothesised that the infants in this study would be engaging with neural entrainment. In this study, we also investigated whether we could elicit the Acoustic Change Complex from these infants. The Acoustic Change Complex is a more general measure of auditory discrimination that has previously been recorded in infants and was used in this study as a reliable indicator of measuring coherence. The second study (*Neural entrainment and auditory selective attention in infants: A two talker study)* built upon the initial methodological development in order to understand how infants process speech in more complex auditory environments. We recorded infants' neural activity while they watched an engaging audio-visual film and listened to speech in a multi-talker environment. The infants listened to two-talkers simultaneously (a target and a distractor) and it was measured whether the infants were selectively attending and cortically tracking the amplitude envelope of the target speaker.

The third study (*Neural entrainment and auditory selective attention in infants: The effects of noise and bilingualism on speech perception*) presented infants with speech in noise in order to investigate whether they would be able to attend to a target talker while distractor of lesser acoustic intensity was present. The effects of bilingualism on cortical tracking were also investigated here.

For the fourth study we wanted to continue to use infant EEG in order to further investigate attention modulation in noise. However, due to the COVID-19 pandemic we were unable to continue with infant EEG testing. Therefore for the fourth study (*Listening development throughout childhood: An online study*) we created an online phonetic perception game that would measure children's (aged 4 and above) ability to identify vowel contrasts. We wanted to investigate whether performance on the memory card game improves with age and if it is affected by masker type (single-talker vs babble) or language (monolingual or bilingual participants). Additionally, during this period of no face-to-face testing the first study was re-visited. The infants that took part in the first study were contacted and asked to participate in a BPVS language assessment task administered by the researcher online or outside in a public space, it was then explored whether the infant's initial neural responses were correlated with their later language developmental outcomes.

1.1 The neural bases of speech processing

Speech processing can be investigated in a multitude of ways. Behavioural measures, such as speech recognition tests, can be used to tell us about an individual's auditory perceptive and discriminative abilities; however, these measures cannot give us any information about the brain bases and underlying neural mechanisms at work when auditory processing is occurring. Some techniques which allow us to do this, such as fMRI, can be expensive, time-consuming and are quite difficult to conduct with infants who can't follow explicit instruction or remain still on demand (Ellis & Turk-Browne, 2018). One non-invasive method of tracking speech processing is electroencephalography (EEG), this is the method that is consistently used throughout this thesis and is arguably the most efficient and scientifically popular method to investigate neural mechanisms in young children. EEG records brain activity at the scalp level that is generated by electrical impulses produced by neurons in the brain. Although EEG is considered to have poor spatial resolution, it has excellent temporal resolution; we are able to see the brain's response milliseconds after a stimulus onset which is particularly useful when investigating neural speech processing.

EEG studies can vary largely in experimental paradigm and electrode placement which allows for a relatively inexpensive and comprehensive understanding of neural responses to a potentially wide range of research stimuli. Up until recently, most infant EEG studies have relied on measuring Event Related Potentials (ERPs) in order to understand the workings of the infant brain. ERP studies involve the repetition of stimuli in order to elicit a certain response that can be compared either between individuals or conditions. A frequently researched ERP response to auditory stimuli is the P300 which is typically elicited through an oddball paradigm (Donchin & Coles, 1988; Näätänen, Gaillard, & Mäntysalo, 1978; Polich, 2007) in which a listener will be presented with a repeated sequence of tones that is interrupted by a sudden change of loudness or pitch, the P300 response will be elicited shortly following the 'oddball' tone. These studies are often unnatural, as they cannot tell us about naturalistic continuous speech, and often result in a quick drop in participant attention levels due to the repetitive nature of the auditory stimuli. In more recent years, new approaches for studying adult EEG data have become popular; one of these approaches is auditory-neural tracking which enables experimenters to measure a relationship between the EEG signal and an ongoing auditory signal.

EEG studies with adults have found that neural oscillations have a major role in contributing to human cognition, they facilitate a number of cognitive tasks ranging from memory (Gruber, Tsivilis, Giabbiconi, & Müller, 2008) to motor planning (Donoghue, Sanes, Hatsopoulos, & Gaál, 1998) and they also play a significant role in the processing of continuous speech (Ahissar et al., 2001; Luo & Poeppel, 2007). It is thought that speech processing can be divided into two distinct stages: analysis and synthesis (Howard & Poeppel, 2010). The analysis stage concerns the dissociation of speech sounds into their primary auditory structures, these are then combined to enable speech perception in the synthesis stage. Neural processes in the analysis stage are thought to be reflective of the brain's responses to speech onsets and offsets resulting in the neural tracking of time-fluctuating acoustic features (Shamma et al., 2011). In adult listeners, theta oscillations (4-8Hz) align to speech segments in-between vowels in the synthesis stage and thus it is hypothesised that these neural processes reflects the packaging of speech into syllabic level representations that are thought to enable speech processing. This neural process is referred to in this thesis as neural entrainment and is thus defined as a measurable correlation between neural oscillations and an ongoing auditory stimulus.

1.1.1 Neural entrainment

Brain operations are principally rhythmic and neural oscillations are believed to inhibit or allow the detection of external stimulus. The timing and frequency of neural oscillations are optimised by external rhythmic stimuli so that neural excitability is high or low when stimuli is expected. This explains why predictable and rhythmic stimuli are perceived more easily than unexpected and non-rhythmic stimuli (Calderone, Lakatos, Butler, & Castellanos, 2014). Despite the growing research on neural entrainment, its precise underlying mechanisms are still unclear. Neural entrainment has been identified in monkeys in addition to humans (Kikuchi et al., 2017), as well as to non-speech sounds (Millman, Prendergast, Hymers, & Green, 2013) and reversed speech (Howard & Poeppel, 2010), potentially illustrating that it is an automatic response elicited by other types of auditory stimulus and not just intelligible continuous speech. However, this thesis will only refer to neural entrainment as the degree of synchronisation of low frequency neural oscillations with ongoing continuous speech and will be measured through a machine learning correlation technique (mTRF toolbox; Crosse, Di Liberto, Bednar, & Lalor, 2016).

Neural entrainment is thought to concern the temporal alignment of a neural process with regularities in an external stimuli such as speech. The underlying neural process is meant to arise from a rhythmic neural generator which resets with the occurrence of specific event. However, if this neural process is oscillatory in a physical manner is unknown. In the relevant literature neural entrainment is often used synonymously with speech tracking but these terms can also be used to refer to distinct aspects of the neural process (Obleser & Kayser, 2019). In this thesis, neural entrainment and the cortical tracking of speech will be used interchangeably.

Previous studies with EEG have aimed to investigate the neural mechanisms underlying the processing of continuous speech. Low frequency (1-8Hz) neural oscillations in the auditory cortex have a significant role in the processing of continuous speech and these neural oscillations track, or phase-lock to, slow temporal fluctuations in the speech signal (Ahissar et al., 2001; Ghitza, 2011; Giraud & Poeppel, 2012; Gross et al., 2013; Luo & Poeppel, 2007). These temporal fluctuations make up the amplitude envelope of speech. The amplitude envelope is crucial for conveying linguistic information, both segmental (e.g. voicing and manner of articulation) and suprasegmental (e.g. stress; Rosen, 1992). Previous studies have found that speech remains intelligible to listeners if spectral information is limited as long as the low-frequency envelope cues remain intact (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) which demonstrates the importance of the acoustic information within the amplitude envelope for speech processing. Different time scales are contained in the amplitude envelope that relay different linguistic information: faster oscillations of 30-50Hz constitute the low gamma band of neural oscillations which carries phonological information such as voice onset times and formant transitions, the theta band of 4-8Hz carries syllabic rate information, and <2Hz are responsible for lexical rate information in the delta band such as intonational patterns (Doelling, Arnal, Ghitza, & Poeppel, 2014). Theta temporal modulations (4-8Hz) are typically associated with syllabic level speech representation and are considered to be the “master oscillators” with regards to speech perception (Giraud & Poeppel, 2012). Oscillations of the auditory cortex track low-frequency amplitude modulations (1- 8Hz) in the speech by aligning in such a way that the acoustic input arrives at a time of high neural excitability (perhaps via phase-resetting by stimulus onsets; Peelle & Davis, 2012) with a latency of approximately 100-200ms (Baltzell et al., 2016). This neural entrainment to theta-band syllabic-level amplitude fluctuations in the speech signal is thought to increase the overall efficiency of speech processing.

The phenomenon of neural entrainment has been used in adult studies in order to explore speech processing in more complex auditory scenes. Ding & Simon (2013) used magnetoencephalography (MEG) to illustrate that speech processing can be robust to competing background noise; neural entrainment remains stable to attended speech until the background noise is greater than twice as strong as the speech itself. The auditory cortex is able to selectively entrain to a target speech signal in order to achieve effective speech perception (Ding & Simon, 2012). Additionally, it has been shown that entrainment is dependent upon intelligibility. When stimuli are varied in intelligibility by using noise-vocoding techniques, it has been found that entrainment relies on both acoustic characteristics and linguistic information in order to enable speech perception (Peelle, Gross, & Davis, 2013). Studies measuring neural entrainment in different acoustic environments allow us to gain a greater understanding of the effects of attention on the processing of continuous speech.

1.1.2 The Acoustic Change Complex

Electrophysiological measures of cortical activity in response to discrete sounds are known as cortical auditory evoked potentials. A class of these known as exogeneous responses include the P1-N1-P2 complex, the acoustic change complex (ACC) and the mismatch negativity complex (MMN, Cheek & Cone, 2020). The ACC is an auditory evoked potential that is elicited to acoustic changes in an ongoing stimulus. It reflects the ability to engage in sound discrimination at the auditory cortex which can thus provide insight into the brain's ability to process speech (Martin, Boothroyd, Ali, & Leach-Berth, 2010; Ostroff, Martin, & Boothroyd, 1998). Analysis of the ACC is undertaken by measuring multiple P1-N1-P2 peaks of the evoked potential in response to an acoustic transition (e.g /i/ to /a/). The later peaks are thought to be associated with attention and learning as opposed to solely auditory sensitivity (Tremblay, Ross, Inoue, McClannahan, & Collet, 2014).

The ACC is typically recorded from mid-central electrode sites which represents neural contributions from the primary auditory cortex (Ponton, Eggermont, Khosla, Kwong, & Don, 2002) and the elicitation of the ACC is resultative of changes in basic stimulus properties, namely phase, frequency and intensity (Dimitrijevic, Michalewski, Zeng, Pratt, & Starr, 2008) in addition to more naturalistic changes such as intensity changes in speech sounds (Ostroff et al., 1998). In summary, the ACC is able to demonstrate cortical encoding to acoustic features that are necessary for speech perception (Martin, Tremblay, & Stapells, 2007).

The ACC has been well-studied in adults and has been recorded in response to speech stimuli such as consonant-vowel syllables in which the frequency, periodicity and amplitude cues are similar to those in typical conversational speech (Ostroff et al., 1998). A constantly changing auditory stimulus, such as continuous speech, can produce several overlapping ACC responses. The ACC can also be used to investigate auditory discriminative abilities in cochlear implant users (Liang, Houston, Samy, Abedelrehim, & Zhang, 2018) and is thought to be a useful tool in assessing speech processing capabilities in this population of people.

Additionally, it has been claimed that the ACC can be used as a mechanism for assessing speech perception skills in infants and young children (Martinez, Eisenberg, & Boothroyd, 2013). The ACC does not require overt listener attention or response. It is a passive listening mechanism which makes it a useful tool for measuring infant speech processing. Although the ACC has not been extensively studied with infants, as is the case with most electrophysiological responses, Chen & Small (2015) found evidence for the elicitation of the ACC to long duration speech stimuli in 24 4 month old infants presented with consonant contrasts. Additionally, Uhler et al. (2018) found that sleeping infants elicited the ACC response when they heard vowel onsets but not when the vowel sound changed in their auditory stimulus. A more recent study investigated infants' neural responses for spectral changes between all pairs of a set of English vowels (McCarthy, Skoruppa, & Iverson, 2019). The neural response to each vowel change was measured using the

ACC via a measurement of long concatenated vowel sequences with several vowel category changes per second. This improves time-efficiency of the experiment which is crucial when recording EEG in infants. Eighty three infants aged between 4 and 11 months old were tested and the results showed the ACC is a robust measure of infant neural sensitivity to acoustic changes in vowels.

We know from the aforementioned previous research that it is possible to elicit the ACC response in infants in response to auditory changes and so we included this in the first study as our robust measure of entrainment. The infants in our study heard a short concatenated vowel sequence in-between each repetition of the continuous story that aimed to elicit neural entrainment from the infants. However, as previously mentioned, existing research into infant neural entrainment is limited and it was unclear whether we would be able to measure it using our proposed methodology. Therefore, ACC elicitation would be an indicator that the infants are engaging with the auditory stimuli and our EEG measures were working. If we are consistently able to measure the ACC in our infants, but are unable to find evidence for neural entrainment to continuous speech in the same participants, then we will conclude that the infants are perhaps not engaging with auditory-neural tracking.

1.2 Speech processing in the first year of life

Most infants are seemingly able to acquire language effortlessly and rapidly and it is thought that infants begin to process speech from as early as in the womb (Moon, Lagercrantz, & Kuhl, 2013); speech perception begins from the earliest of developmental stages. Infants are able to pay attention to the sounds that they hear through the amniotic fluid (DeCasper & Fifer, 1980) and before they perceive streams of sound as language they are intuitively attentive to the patterns of sounds produced when people talk. Neonates are able to respond differently to native/non-native vowels from birth; the ambient language that a foetus is exposed to is already shaping their perception of speech. New-born infants are able to show a preference for the language that they were exposed to prenatally (Mehler et al., 1988) and show a sensitivity to statistical probabilities in strings of speech sounds (Saffran, Aslin and Newport, 1996). Once an infant is born, the first year of life is absolutely crucial in their ability to acquire language and many major developmental milestones, both linguistic and cognitive, are reached within this period. Infants are immediately able to discriminate between a universal set of phonetic contrasts (Werker, Gilbert, Humphrey, & Tees, 1981). However, this ability declines at 10-12 months when babies become more ‘perceptually tuned’ to their native language (Kuhl et al., 2008). Similarly, at 9 months of age infants are able to distinguish between words that comply with, or violate, the phonetic and phonotactic rules of their native language (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). Thus, it can be assumed that by the end of the first year of life, infant non-native speech perception seems to have declined while their native language speech perception skills are showing improvement. This is a fundamental step in order for them to become masters of their native language.

For a long period of time it was thought that infants could only reliably be tested by using behavioural methods. Common behavioural methods to investigate speech processing in infants included the high-amplitude sucking technique, in which infants would suck harder when they heard a new sound, or the head-turn preference procedure in which infants' looking duration was a marker for their interest in the sound. This could be used to show, for example, that infants prefer to listen to sentences with words that they had received prior familiarisation to (Jusczyk & Aslin, 1995). Behavioural methods can tell us if babies are engaging and discriminating between different sounds. However, they are unable to tell us anything about the neural mechanisms underlying these processes and lack adequate temporal precision. Neural imaging techniques, such as fMRI, would be able to give us a deeper insight, but are not very infant-friendly; fMRI machines are noisy and this method requires subjects to be extremely still. Over the past decade, non-invasive techniques such as EEG have become much more prevalent in infant studies and have been able to give us more detailed information about the neural processes underlying speech processing in infants.

1.2.1 EEG measures of infant speech processing

Event-related Potentials (ERPs) have been commonly used in investigating infant speech processing. An ERP shows a neural response that is time-locked to specific events or stimuli. An example of a typical ERP study is the oddball paradigm in which subjects are presented with a sequence of repetitive stimuli (e.g. repetition of the phoneme /p/) that is infrequently interrupted by a deviant stimulus (e.g. the phoneme /b/), the neural reaction to this 'oddball' is what is recorded and is known as the Mismatch Negativity ERP (MMN; Squires, Squires, & Hillyard, 1975). ERP studies have been popular with regards to investigating infant speech processing, specifically at the phoneme level. It has been shown that a component closely resembling the MMN could be elicited in sleeping new-borns to vowel contrasts

(Cheour-Luhtanen et al., 1995), although this had a different scalp distribution to that recorded in adults. ERPs have also been used to show infants' ability to discriminate vowel contrasts in different languages (Cheour et al., 1998). Finnish infant ERPs showed a discriminatory response between Finnish and Estonian vowel contrasts at 6 months old. However, this response was attenuated in infants aged 12 months. In addition, some studies have used ERPs in order to investigate how early phonetic perception is linked to later language development (Molfese & Molfese, 1985). Shortly after birth, ERPs were recorded to syllables and it appeared that these neural measures predicted language ability at 3, 5 and 8 years of age. They were even able to link to reading disabilities at the age of 8 (Molfese, 2000). Although ERP studies have been useful in informing us about infant speech perception, they can only be used to investigate infant neural responses to discrete stimuli and so are not able to tell us about the mechanisms involved in the processing of natural continuous stimuli.

It has been suggested that young infants aged 3 months old are able to entrain to sounds with high temporal precision. Werner et al., (2001) measured responses to intermittent gaps in noise and found no significant differences in the auditory brainstem response in 3 month old babies versus adults. Previous studies using EEG have shown that the infant brain is sensitive to the temporal modulations of sound (S. Telkemeyer et al., 2009; S. Telkemeyer et al., 2011). EEG were recorded in new-borns to temporally modulated non-speech and showed evidence for neural entrainment. However, this does not reflect what happens in natural continuous speech environments. EEG has also been used to measure neural entrainment to auditory rhythms in 7 and 15 month old infants (Cirelli, Spinelli, Nozaradan, &

Trainor, 2016). The 7 month old infants were exposed to an ambiguous rhythmic pattern that could be perceived in two different meters while the 15 month old infants listened to a rhythmic pattern with an unambiguous meter. Both groups of infants demonstrated clear neural entrainment to the rhythms. Interestingly it was found that the 7 month old infants who regularly participated in music classes had enhanced entrainment to the stimuli, perhaps illustrating that environmental factors can affect neural oscillatory activity. Additionally, 6 month old infants experience enhanced gamma oscillations while listening to native phoneme contrasts but not when listening to non-native phoneme contrasts (Ortiz-Mantilla, Hamalainen, Musacchia, & Benasich, 2013).

Not all aspects of auditory processing, such as the processing of complex speech sounds, develop in the early years. Johnson et al., (2008) measured brainstem responses to click tones and speech syllables in 3-12 year old children. All of the children displayed identical neural responses to the clicks. However, the younger children showed delayed and less synchronous neural activity to the speech syllables in comparison to the older children. In addition to this, although event-related potentials (e.g. P1, N1, P2) can be elicited in premature infants from 24 weeks gestational age, the waveforms do not resemble those of adults' until late childhood (Wunderlich & Conewesson, 2006). Thus, although many auditory perceptual abilities are in place from birth, some mechanisms behind more complex auditory processing skills may continue to develop throughout infancy and early childhood.

It is claimed that language acquisition is underpinned by rhythmic processing (Goswami, 2019) which highlights the potential relationship between neural entrainment, selective attention and language processing in the first year of life. Although theta-band entrainment is associated with successful reading skills in children (Power, Mead, Barnes, & Goswami, 2012) and considered by some to be the master oscillator with regards to speech processing, information encoded in other frequency bands may also be important for speech processing. Children with developmental dyslexia have an impaired encoding of speech envelopes in the delta band (0-2 Hz) (Power, Colling, Mead, Barnes, & Goswami, 2016) and dyslexic children show atypical phase entrainment to rhythmic speech in the delta band. Arguably, it is possible that developmental phonological disorders may be a result of atypical oscillatory mechanisms in low frequency bands other than theta. This illustrates the importance of multiple low frequency oscillatory bands (namely theta and delta) in the early language acquisition process.

Previous studies have been useful in providing evidence for infants' ability to cortically track features of auditory stimuli and inform us about basic infant neural development. However, few studies have aimed to investigate how infants process natural continuous speech i.e. if infants, like adults, are also engaging in the auditory-neural tracking of spoken speech. This is what we aim to investigate in this current thesis.

While completing the research for this thesis, a limited number of other studies have also aimed to investigate how infants are cortically tracking continuous speech. Leong et al., (2017) investigated how the infant brain processes the temporal structure of popular nursery rhymes in comparison to their mothers. Infants and their mothers watched videos of sung nursery rhymes while their neural activity was recorded by EEG. They calculated phase-locking values (PLVs) across the

EEG-speech frequency spectrum in order to measure the degree of synchrony between the amplitude envelope of speech and the neural signal. PLV results showed that infants have an enhanced cortical entrainment to theta-rate frequency bands of speech in comparison to their mothers. More recently, Kalashnikova et al., (2018) assessed the cortical tracking of 12 seven month old infants to both adult-directed speech (ADS) and infant-directed speech (IDS) lasting for approximately eight minutes. They found evidence for low-frequency (theta-band) cortical tracking to both speech signals, but this was stronger for IDS than ADS. This suggests that infants have a neural preference for the processing of IDS.

In addition, Jessen et al., (2019) recorded the EEG signal of 52 7 month old infants while they watched a 5 minute episode of an age-appropriate cartoon. Forward encoding models (Crosse et al., 2016) were employed to predict the EEG signal from the modulation spectrum of the audio-visual stimulus. They found neural entrainment to the amplitude envelope of the speech signal as well as defined response patterns to the cartoon’s motion content, concluding infants are engaging in auditory-neural tracking and that mTRFs, as used in this thesis, are a suitable tool for analysing continuous EEG responses in infants.

The multivariate temporal response function (mTRF) (Crosse et al., 2016) toolbox is used in this thesis in order to assess the degree of synchronisation of neural oscillations with an auditory stimulus. The mTRF output resembles a traditional ERP and essentially extracts a neural component that is time-locked to the amplitude envelope of the acoustic signal. It is measured by a correlation technique based on machine learning and can be applied in two directions of analysis: forwards and backwards. The forward model predicts the neural response to the stimulus and calculates the correlation between the predicted and actual response. On the other hand, the backwards model reconstructs the auditory stimulus based on the neural response and then calculates the degree of correlation between the predicted and actual stimulus. The backwards model, as used in this thesis, has potential benefits over the forwards model. For example, the backwards model does not require a pre-selection of neural response channels and instead automatically assigns low weightings to the channels that don’t contribute as much to the signal reconstruction. Additionally, the backwards model is able to account for inter-channel correlation in the data because the mapping is done from all channels simultaneously rather than each channel individually. The toolbox also uses the technique of ridge regression in order to account for multicollinearity in the data. This is often the case with EEG data in which single electrodes are frequently highly linearly related to one another.

1.3 The role of attention in speech processing

Attention can broadly be defined as the ability to shift between and then maintain focus on events or tasks. The role of attention in speech processing is important because naturalistic listening scenarios are made up of multiple competing sound sources which generally, in every day listening situations, include an abundance of irrelevant background sounds. Listeners must extract the necessary acoustic features from a mixture of sounds in order to focus on a particular target.

The ability of the human auditory system to organise sounds into perceptually meaningful elements has been referred to as auditory scene analysis

(ASA) (Bregman, 1994); sounds are perceptually organised with the objective of identifying individual sound sources held within the signal. The mechanisms of auditory scene analysis can be illustrated by the cocktail party effect (Cherry, 1953). The cocktail party effect refers to the brain's real-life ability to focus its auditory attention on one particular speech stream while filtering out a range of other competing stimuli, much like listening to a single speaker when stood in a noisy cocktail bar. Interestingly, when someone calls your name from the other side of this environment we automatically notice and respond to this immediately. This illustrates our communicative ability to selectively attend to important information even when we are in less than ideal auditory environments. In other noisy scenarios, such as those including multiple talkers, listeners are able to attentively entrain to a target speaker while filtering out other competing auditory information in order to facilitate a robust neural representation of the attended speech signal (Rimmele, Zion Golumbic, Schröger, & Poeppel, 2015). In addition, Ding and Simon (2012) used magnetoencephalography (MEG) to record brain activity in adult listeners who were engaged in a multi-talker task. It was found that the target and unattended speech could be decoded separately; the listeners produced individual neural representations for both auditory streams that they heard. This suggests that in multi-talker scenarios listeners have the neural ability to encode auditory objects as individual speech streams despite them being heard concurrently, and this is dependent upon the focus of our selective attention.

Selective attention refers to the enhanced representation of some stimuli over others. It is defined in this thesis as the process of focusing on an external stimuli, namely auditory, for a sustained period of time. Neural entrainment seems to be a core mechanism of, and underlie, selective attention; deficiencies in entrainment are apparent in people with attention-deficit/hyperactivity disorder (ADHD; Calderone et al., 2014). Schizophrenia and dyslexia have also been linked to deficits and abnormalities with entrainment: Schizophrenia patients exposed to deviant auditory stimuli among rhythmic streams were unable to entrain to the rhythmic stimuli showing that they were unable to predict the incoming information which may show a loss of selective attention (Lakatos, Schroeder, Leitman, & Javitt, 2013). Recent evidence suggests that attention deficits in ADHD patients affect their early stage sensory processing (Ortega, López, Carrasco, Anllo-Vento, & Aboitiz, 2013). For example, adults with ADHD struggled to detect visual targets amongst distractors (Stevens et al., 2012). This potential attentional deficit has been shown also through abnormalities in EEG data. Typically developing children demonstrated gamma-level synchronisation to remembered vs novel visual stimuli whereas children with ADHD did not (Lenz et al., 2010) which shows early stage attentional deficits. This evidence illustrates that neural entrainment is a fundamental capacity for selective attention.

Auditory selective attention also plays an important role in other areas of speech processing such as second language learning. The role of attention here is to only select necessary information among several items of input in order to increase the effectiveness of learning and speech comprehension. Selective attention as a brain function has been investigated in second language learning. Neuroimaging techniques, such as EEG and positron emission tomography (PET) have shown that attentional networks can be related to the linguistic areas of the brain (Posner & Petersen, 1990). Oishi (2007) claimed that increased blood flow in language-related brain areas, such as the Wernicke's area, was associated with L2 comprehension tasks in comparison to L1 tasks. It was suggested therefore that easier tasks require less attention and thus less blood flow to the relevant neural areas in comparison to a more difficult L2 processing task.

Wickens (1980) claimed that attentional demands become exacerbated when simultaneous tasks draw on the same attentional mechanisms, for example holding two conversations at the same time. In these cases, it is necessary for attention to shift from one task to the other and this is referred to as serial processing. In other cognitive tasks in which attentional demands differ, parallel processing may occur. However this may inhibit task processing effectiveness. Individual attentional processing can vary depending on memory and cognitive load and these individual variables ought to be accounted for in attentional studies. The aforementioned evidence stresses the importance of auditory attention when processing speech; we are constantly attending to what we want to hear even when this is done subconsciously. This seemingly automatic ability to attend to relevant or important auditory stimuli is a fundamentally important cognitive mechanism when considering speech processing effectiveness.

1.3.1 Attention and speech processing: Infants

Infants are born with attentional systems that direct them towards salient stimuli in their environment: Pre-birth infants pay attention to voices through amniotic fluid (Decasper & Fifer, 1980) and are attentive to regularities of patterns in sounds when people speak; infants are attending to speech even when in the womb. New-born infants can distinguish between different spoken languages and show preference for those heard prenatally (Mehler et al., 1988) and are sensitive to statistical probabilities in strings of speech sounds.

Sustained attention involves parts of the brain including the brainstem, thalamus and cardio-inhibitory centers in the frontal cortex (Reynolds & Romano, 2016). The general arousal/attention system is functional in infancy but develops considerably throughout later infancy and early childhood. This explains why as children get older they are able to engage in longer periods of sustained attention. In order to gain fluency in one's native language an infant must participate in a number of different cognitive and perceptual tasks including auditory discrimination, categorisation and sound structure representation. All of these abilities rely on attention; an infant must first selectively attend to the phonetic representation of a word before they can discriminate and categorise it (de Diego-Balaguer, Martinez-Alvarez, & Pons, 2016).

Jusczyk et al., (1990) investigated how infants' focus of attention can affect their perception of speech. Infants aged 4 days and 2 months old were familiarised with a set of stimuli that was comprised of items judged to be close in perceptual space ([pa], [ta], and [ka]) and then measured if they were able to detect new syllable addition ([ma]) that was perceptually dissimilar to the set. Both sets of infants were able to do this. They then exposed the infants to perceptually dissimilar vowel sets ([bi], [ba], and [bu]) and measured if they could detect the addition of the vowel [bʌ] which is perceptually similar to [ba]. The 4-day old new-borns were unable to perceive this distinction, whereas the 2-month old infants could. This shows that in the course of acquiring one's native language an infant's perceptual distinctions become more apparent, and also how attention must be fine-tuned to the parts of speech that carry meaningful distinctions. The presence of aspiration at syllable onset is a primary indicator for a voiceless consonant in English, but this is not the case in French. In this case infants must learn to attend to the presence of aspiration vs non-aspiration in order to perceive parts of the acoustic signal that carry meaning. As an infant's sensitivity increases to distinctions in their native language that mark meaning, they become less sensitive to non-native contrasts (Werker & Lalonde, 1988) and this is arguably resultative of an infant's selective attention to the parts of language that carry maximal meaning distinctions.

We can attempt to modulate and manipulate infant auditory attention by making language more acoustically appealing. It is widely known that infant-directed speech (IDS) is something that babies respond extremely positively to, and this may also be a factor that aids their overall development. IDS is characterised by several distinct features that make it different from typical Adult-Directed speech (ADS). These include slower speech rate, raised pitch and a more variable prosody – this is thought to be a cross-culturally universal phenomenon (Narayan & McDermott, 2016). Additionally, the typical locations of the extreme point vowel sounds (/a/, /i/ and /u/) are maximally distanced in the acoustic space in order to enhance perceptual development. Other aspects of IDS that are perceptually salient for young infants' attention are higher volume and extreme intonation contours (De Boer, 2005). Infants attend to sounds that they like to hear, and this in turn assists in their perceptual development. It has also been shown that IDS, in comparison to ADS, enhances attention to speech in typically developing infants, and that this is also the case for deaf infants with cochlear implants (Wang, Bergeson, & Houston,

2017) which shows the general reliability of IDS for assisting speech perception through enhancing infant attention.

We decided to use IDS in all of our studies for measuring neural entrainment to continuous speech as we wanted our infants to hear maximally engaging speech that would automatically focus their attention and it has been shown that IDS can do this in neural entrainment studies (Kalashnikova et al., 2018; Leong, Kalashnikova, Burnham, & Goswami, 2014). Another form of acoustic modulation that has been shown to hold infant attention is engagement with a female voice; this is typically specific to mother preference (Lee & Kisilevsky, 2014) due to familiarity and high pitch. We had to record our set stimuli in advance and so were unable to use each infant's own mother's voice per study, but future studies with infants could try to do this to optimise attention and engagement.

Infants seem to attend to auditory stimuli that is phonetically engaging, this attention seems to work as a potential precursor for language acquisition and thus is fundamentally important in infant development. Infants are active participants of the language learning process. They do not learn successfully from passive listening and selective attention can support this by focusing on the parts of language that carry maximum meaning distinctions and help them to learn (de Diego-Balaguer et al., 2016).

1.3.2 Attention and speech processing: Bilingualism

Evidence exists to suggest that bilingual and monolingual attention differs. In the third study in this thesis we look at the effect of bilingualism on infant neural entrainment to continuous speech. It is difficult to define an infant as a bilingual or not, as they cannot yet speak. However in this thesis we define a bilingual as an infant or child that receives 25% or more of their daily exposure to two (or more) languages. It has been found that children who receive less than 25% of daily language exposure to a second language do not become proficient in the second language (Pearson, Fernández, Lewedeg, & Oller, 1997), and this is why we have chosen this criterion. More recent research has identified the difficulty regarding measuring bilingualism in young children and has provided suggestions for measuring and reporting a child's language exposure and proficiency (Byers-Heinlein et al., 2019). Researchers should conduct detailed interviews with a child's caregiver in order to understand sufficient information about the child's language exposure and capabilities e.g. MAPLE: A Multilingual Approach to Parent Language Estimates (Byers-Heinlein et al., 2020). This will allow participants to be accurately classified into monolingual and bilingual groups. In this thesis, detailed questionnaires regarding infant language exposure were administered to parents and caregivers in order to obtain a fully detailed and coherent picture of the infant's language experience.

The Bilingual Interaction Activation model (BIA) (Dijkstra & van Heuven, 2002) posits that bilingualism improves attentional control by enhancing the ability to inhibit unnecessary information, bilinguals must engage in this frequently in order to avoid cross-language interference. In the bilingual brain it is thought that both languages are simultaneously active and that the bilingual must choose when to supress one language and use the other. This constant state of inhibiting the unwanted language(s) is thought to heighten attentional control.

It is argued that bilingual toddlers show a greater attention to speech than their monolingual peers (Kuipers & Thierry, 2015). Kuipers & Thierry (2015) measured ERPs to spoken words preceded by pictures in monolingual and bilingual toddlers. Words that matched the images elicited early frontal positivity in the bilingual children showing that bilingualism may positively affect overall attention during speech perception.

It has been found that bilingual infants have an enhanced cognitive control system and that bilingual pre-verbal infants have displayed improved cognitive control abilities compared with monolinguals (Kovács & Mehler, 2009). In 3 different eye tracking studies it was found that 7 month old monolinguals and bilinguals were both able to respond to an auditory or visual cue in anticipation of an on-screen reward. However, only infants who had experienced two different languages from birth were able to redirect their anticipatory looks when the cue began signalling a reward on the opposite side of the screen. This illustrates that an exposure to two languages from birth has an improvement on domain-general cognitive control in young infants. Additionally, bilingual infants are able to switch their attention more frequently than monolinguals and this presents them with advantaged executive functioning skills (D'Souza, Brady, Haensel, & D'Souza, 2020). In this study, both bilingual and monolingual infants aged between 7 – 9 months of age participated in four gaze-contingent eye-tracking studies. They failed to replicate the exact results of the previous study (Kovács & Mehler, 2009). However, they did find that the bilingual infants were able to disengage attention from one stimulus more quickly than monolinguals in order to focus on another stimulus. Consequently, bilingualism seems to be modulating pre-verbal infant attentional processes in the first year of life. On the other hand, a more recent study (Kalashnikova, Pejovic, & Carreiras, 2021) failed to find a specific advantage in bilingual attentional control in seven month old infants. However, it was found that bilingual infants showed uniquely different patterns in their allocation of attention in compared to monolinguals. This illustrates that the encoding of two or more languages from birth in the first year of life has a significant effect with regards to their attentional processing.

It is thought that early bilingualism affects the neural mechanisms of auditory selective attention in adults. Spanish-English and Dutch-English adults participated in a cocktail-party paradigm in which they were instructed to attend to one of two competing speech streams while their neural activity was recorded using EEG. Their results were compared to that of monolingual English speakers (Olguin, Cekic, Bekinschtein, Katsos, & Bozic, 2019). All participants elicited stronger neural encoding to the attended speech stream. However the bilingual participants experienced no significant differences in attentional encoding between any of the attended streams across different interference conditions, whereas monolinguals did. This may show that bilingual modulation of attention occurs at the neural level.

1.3.3 Attention and speech processing: Early childhood and beyond

We cannot explicitly measure selective attention in infancy. We can use behavioural measures such as gaze fixation and neural responses such as EEG, which is much more reliable, but attention is something that becomes easier to explicitly modulate and measure as children become older and can follow explicit instructions.

In our fourth study, due to the COVID-19 pandemic, we were forced to move our research online. As infants are unable to participate in online tasks, it was decided that we would investigate slightly older children in order to understand how attention develops throughout childhood from the age of 4 years old, more specifically how children perform in cognitive tasks with and without a language element, and how differing background noise (intelligible speech and unintelligible noise) affects this. Thus, it is important to be aware of how attention typically develops beyond infancy and into early childhood.

Attention is thought to be a fundamental foundation for most cognitive functions to develop and so is crucially important in infancy and childhood. Several studies have indicated that selective attention develops throughout childhood and can be evident from the age of 3 years old (Kannass, Oakes, & Shaddy, 2006) as long as the task is engaging or rewarding for the child. Temporally selective attention was measured using ERPs in 3-5 year olds (Astheimer & Sanders, 2012) and it was found that they were able to direct their attention to times relative to word-onsets during speech perception, and the distribution and latency of this effect resembled that which is found in adults.

Selective attention in children is an important factor to consider in terms of their overall development; children must focus on the voice of the teacher in the classroom while surrounded by noisy peers and playground noise and this can affect learning from pre-school up until high school. The fact that children seem to be able to do this shows that they are selectively attending to an auditory stream while inhibiting a distractor. However, this can be difficult for children with learning or auditory problems and can become problematic for all children after sustained attentional periods. It has been shown that both behavioural and neural responses, including the P300 response, can be indicators of attentional fatigue in children (Key, Gustafson, Rentmeester, Hornsby, & Bess, 2017); after a sustained period of listening tasks lapses of attention, both behaviourally and neural, were measured showing that speech processing fatigue exists and may have a detrimental impact after a certain point of children listening to classroom instruction.

It is argued that both younger and older children are able to process irrelevant distractor stimuli but that the older children have an advantage in separating the channels of stimuli in the memory and thus selectively reporting on the target stimuli. Older and younger children are able to remember the same amount of irrelevant stimuli in a surprise recognition task (Doyle, 1973). However when the performance of this task is represented as a percentage of relevant/target stimuli recognised then the older children are better; the younger children recognised a considerably higher amount of irrelevant information. This supports the theory that older children are more proficient at blocking out irrelevant information and thus have a more advanced attentional system.

1.4 This current thesis

In this thesis, we wanted to explore the neural mechanisms of language processing in infants under 12 months of age by using EEG measures of auditory-neural processing. More specifically, we wanted to investigate if infants were engaging in neural entrainment and how this processing can be modulated by infant attention and different linguistic environments e.g. bilingualism and multi-talker scenarios.

Chapter 2 (*Infant neural entrainment to continuous speech: Initial methodological development*) aimed to investigate whether we were able to find robust evidence for infant cortical tracking to the amplitude envelope of speech and the Acoustic Change Complex (ACC) in 36 babies. It was hypothesised that we would be able to elicit infant cortical tracking to continuous speech as other recent papers had also found evidence for this (Kalashnikova et al., 2018; Leong, Byrne, et al., 2017). The infants listened to an infant-appropriate story read in child directed speech while their neural activity was recorded with EEG. Results showed evidence for neural entrainment, much like has been found in adults. In addition to this, the infants showed a reliable elicitation of the ACC response. Later language assessments (BPVS) were conducted on these infants approximately two years later; however, we found no significant predictors, including initial infant EEG coherence levels, on later language developmental outcomes.

Chapter 3, *Neural entrainment and auditory selective attention: A two-talker study*, built upon the first study by using our successfully developed toolkit of measuring neural entrainment in 20 infants in order to investigate infant speech processing in more complex auditory scenarios. Infants were presented with an audio-visual film containing a target and a distractor speaker of equal intensity. Due to infants typically listening to language in noisy environments, it was hypothesised that infants would be able to selectively attend to a target while inhibiting a distractor. However, it seemed as though the infants were entraining to both speech signals simultaneously and thus not selectively attending to the target. It is possible that the listening conditions in this study were too complex for the infants.

Thus, in Chapter 4 (*Neural entrainment and auditory selective attention in infants: The effects of noise and bilingualism on speech perception)* 30 infants were exposed to an audio-visual film comprised of two listening conditions: no-interference and 10dB signal-to-noise babble. It was hypothesised that the infants would be able to entrain to the speech while in the noisy condition as this is an easier listening environment than in Study 2. In addition, the effect of bilingualism was also explored here and it was hypothesised that a bilingual advantage would exist due to their ability to reliably inhibit one language while their other is active (Bialystok, Craik, Green, & Gollan, 2009a). Our results showed that infants were able to attend to a target talker in both noisy and quiet conditions, but this may perhaps be a consequence of the noise levels being too low to make a significant difference in the infants' auditory processing. No significant predictors, including bilingualism, were found on coherence levels.

For our fourth study, we wanted to continue to investigate infant auditory-neural entrainment. However, due to the COVID-19 pandemic in-person testing had to be stopped and research had to be moved online. Chapter 5 (*Listening development throughout childhood: An online study*) was an online investigation in order to explore how a speech recognition task interacts with masker type and how this changes with age during childhood. Children aged between 4-12 years old were required to participate in the online memory card game. Results showed that performance on the memory card game improved with age but was not affected by masker type or language background. The age effect was the most prominent for the condition in which the cards had to be matched on sound alone. This general improvement is thought to be a consequence of improved speech perception capabilities rather than an overall memory effect.

Overall, this thesis provides a reliable methodology for measuring neural entrainment in infants and a greater understanding of the mechanisms of speech processing and attention modulation in infancy and beyond. From the information learnt in this thesis, we can understand more about the general language developmental process and consequently how to optimise auditory environments in order for infants and children to acquire language successfully.

2. Infant neural entrainment to continuous speech: Initial methodological development

2.1 Introduction

The primary aim of this study was to see if we can reliably measure coherence to the amplitude envelope of continuous speech in pre-verbal infants. While conducting this research, we became aware of other work that had also measured this cortical tracking of speech in infants. However, it is valuable for this thesis to demonstrate that we can also reliably measure this phenomenon in infants. As this was our first attempt of measuring speech entrainment, we also wanted to include a robust measure of infant coherence that has been previously measured with infants in our lab: The Acoustic Change Complex (ACC). Besides from demonstrating that we are getting a reliable neural response from babies, this also gives us an opportunity to compare the infant speech entrainment with more of an auditory response. Finally, approximately two years later this data was revisited by collecting later language scores from these children. This was an exploratory analysis to see whether coherence in the first year of life is related to later language abilities.

Leong et al., (2017) measured coherence to continuous speech in 58 infants and claimed that infants actually have higher entrainment to speech than adults do. Their results showed that infants were able to exhibit better phase-locking to the temporal structure of IDS in nursery rhymes than their mothers at the theta and alpha rates. Infant entrainment levels matched the adults in terms of delta waves, or prosodic stress patterns. However, infants were not as accurate as adults in terms of tracking slow phrasal patterns of <0.5Hz. The analytical measures used in this study departed from regular adult coherence measures. Their methods involved creating phase locking values (PLVs) between 0 and 1 in order to assess the degree of entrainment between the neural EEG signal and the amplitude envelope of speech. The PLV index may be affected by the latency of the neural response because adult neural latencies are generally shorter than those of infants (Wunderlich & Cone-Wesson, 2006). Consequently, if the time window of the PLV analysis is not large enough (e.g. 60ms) then the infant neural response to the stimulus may be missed. This would produce incorrectly low entrainment values for infants thus making it difficult to compare to adult entrainment.

The other two studies measuring infant coherence used machine learning techniques (multivariate temporal response functions; mTRFs, Crosse et al., 2016) that have been previously used in adult studies and will be also used in this thesis. Firstly, Jessen et al., (2019) recorded EEG from 52 infants while they watched an audio-visual film. Their results showed that they were able to measure a clearly defined response for both the audio and the visual stimulus. This demonstrates that mTRFs are a useful tool in analysing infant coherence to a continuous stimuli and that they can reliably measure this neural tracking of continuous speech. Kalashnikova et al., (2018) also used mTRFs in order to assess the cortical tracking of naturalistic infant directed speech (IDS) in comparison to adult directed speech (ADS) by 12 infants. Their results showed reliable infant coherence to the continuous speech. More specifically, higher cortical tracking for IDS in comparison to ADS was found suggesting that perhaps IDS has an advantaged role in facilitating infant speech perception. This may illustrate that infant attention is modulating their coherence and that the infants are really engaging with the speech that they are hearing. However, IDS and ADS have different acoustic properties in terms of the modulation spectrum. It is difficult to know if the infant IDS preference is because of its attention-grabbing properties or if this is driven more by the acoustic properties of the speech signal. This creates a confound because if the two speech signals were equated in terms of acoustic properties then there wouldn't be any difference between the two. It is thus unclear whether infant coherence is driven by attentional or acoustic properties.

This present study includes a comparison of the speech entrainment with the ACC. This can help us to understand whether the speech entrainment is just an auditory response, or if this is tracking something slightly different. The auditory-evoked P1-N1-P2 complex involves a positive peak at 50ms after stimulus onset, a negative peak at 100ms, and another positive peak at 180ms (Martin et al., 2007). When this is measured in response to a change in an on-going sound then it is referred to as the Acoustic Change Complex (ACC). The P1 is thought to reflect the detection of sound by the auditory cortex (Anderson, Chandrasekaran, Yi, & Kraus, 2010), N1 is thought to reflect time-varying aspects of the sound such as amplitude (Hoonhorst et al., 2009), and the P2 is thought to have a role in the categorisation of auditory stimuli (Novak, Ritter, & Vaughan, 1992). Previous work with babies has reliably shown elicitation of the ACC. McCarthy et al., (2019) investigated 83 infants' neural responses to spectral changes between a set of English vowels by using the ACC. The ACC can be measured by using multiple vowel changes per second which makes it a useful tool for assessing speech perception in young infants. It was found that when babies are presented with these repetitive sequences they exhibit a response that is more like the P2 and seems to be modulated by the spectral differences in the sounds. It was also found that the neural responses of the younger infants corresponded more strongly to the first formant frequency and were more sensitive to nearby vowel pairs. Furthermore, the ACC responses became less driven by raw acoustical differences as the infants got older (McCarthy et al., 2019). This change, which occurs at around 6 months of age, may be due to infants' emerging development of their native phonology and lexicon. Not only was the ACC included in this present study as a robust indicator of coherence, as it is something that we know we can reliably measure, but we also wanted to measure this and compare it to speech entrainment in the same babies in order to see if the neural response was the same. If the neural response for speech entrainment is the same for the ACC then it raises the question of whether the infants really are listening to speech or just engaging in the basic auditory processing of the stimuli.

The ACC is typically analysed as an ERP (P1, N1, P2), and is traditionally seen as a response to an acoustic change. In this thesis, to elicit the ACC we used stimuli with well-defined acoustic changes much like a typical ACC experiment. However, in order for us to compare the ACC to the story coherence we analysed them in the same way by using backwards mTRFs. Coherence, like the ACC, is also an evoked response. However, whereas the story coherence in this thesis is evoked to acoustic changes in continuous speech, the ACC response is evoked to more controlled stimuli e.g. vowel changes in a concatenated sequence. As we can't use a traditional ERP analytical technique in order to analyse the response to the continuous acoustic changes in speech, we decided to use mTRFs in order to analyse both conditions the same way. In addition to this, the backward mTRF analysis is more powerful than a traditional ERP analysis due to the denoising techniques and how it accounts for inter-channel correlation in the data. Using the mTRF method for both conditions is just a different way of analysing the coherence, and it is sensible that these conditions are analysed the same way in order to accurately compare the results (ACC and story coherence). Both of the responses will be regarded specifically in terms of theta-band coherence. Although theta-band coherence is considered by some to be a master oscillator for speech perception in terms of an oscillation framework, it is also considered as more of an attention-modulated evoked auditory response (Keller, Payne, & Sekuler, 2017). Thus, in this thesis theta-band coherence will be regarded as the most important oscillatory band for analysis in both conditions.

This infant EEG data was revisited approximately two years later and language assessments were collected from these infants. This exploratory analysis was undertaken in order to understand if a link exists between early infant coherence and later language abilities. Recent neurophysiological research has shown that early oscillatory measures can be predictive of later expressive language abilities (Cantiani et al., 2019). Rapid Auditory Processing (RAP), the ability to rapidly discriminate between successive auditory stimuli, is a crucial process in the first year of life and Cantiani et al., (2019) argued that this is a robust predictor of later language outcomes. Thus it seems plausible that performance on these categorisation type measures of phonetic processing seem to link to better language learning. Additionally, behavioural research has attempted to investigate the links between early infant speech processing and later language. Tsao et al., (2004) tested 6 month old infants on their ability to discriminate between a simple vowel contrast. They found that early auditory discriminative abilities were significantly correlated with word production and understanding at 13, 16 and 24 months of age (Tsao, Liu, & Kuhl, 2004). Moreover, it has been illustrated that proficient native-language discrimination at 7.5 months old predicts enhanced later language skills whereas better non-native language discrimination at 7.5 months old predicts reduced later language abilities (Kuhl, 2004). In support of this, Newman et al., (2006) found that performance on a speech segmentation task before 12 months of age was positively related to expressive vocabulary development at 24 months of age. This implies that an infant's ability regarding early phonetic discrimination has an effect on their ability to acquire language.

In this study, EEG was recorded from 36 infants while they listened to blocks of a story read in IDS followed by a concatenated vowel sequence designed to elicit the ACC. The primary aim of this study is to see if we are able to measure reliable coherence to the story. If we are able to reliably measure coherence in infants then we can use these methods throughout this thesis in order to understand more about the specific roles that this phenomenon may have on successful language development. ACC was included in this study because this is a response we know we are able to measure. Inclusion of the ACC also allows for us to compare the speech entrainment with the acoustic response in terms of processing differences. Later language scores were collected from the infants approximately two years after their EEG in order to see if their later linguistic abilities were related to their ACC or story coherence.

2.2 Methodology

2.2.1 Participants

Thirty six infants (16 = male, 20 = female) with a mean age of 7.8 months (SD = 2.01), ranging from 5.1 to 11.6 months old, participated in this study. All infants were exposed to English as a first language and were living in London at the time of testing. All infants were born full-term, were healthy with no identified developmental problems and had normal hearing, as stated by their caregivers in a language questionnaire given to them prior to the testing. Along with the language questionnaire, the caregiver of each infant also completed an informed consent form in which they were made aware that the experiment could be discontinued immediately if they wished so.

2.2.2 Stimuli

The stimuli used was the infant friendly story *Oh, Baby, The Places You'll Go!* (Dr Seuss., 2017) which was read in infant-directed speech by a native British female speaker. Each block of the story lasted for approximately 334 seconds (5.6 minutes). After each story block the infants heard a concatenated vowel sequence of /a/ and /i/ that was designed to elicit the ACC and lasted for 20 seconds. During this 20 second period, the infant would hear 40 repetitions of each vowel. The story was repeated to the infant until they became restless and testing ended. The average number of blocks that the infants listened to was 3 which means that on average an infant listened to 1,002 seconds (16.7 minutes) of the continuous speech in the story and 60 seconds (1 minute) of the ACC stimuli. The stimuli was set at 60dB.

2.2.3 Apparatus

We conducted an EEG experiment using a Biosemi EEG system with 32 (Ag/AgCl) electrodes) at a sampling rate of 2048Hz. All analysis of the EEG data was performed in Matlab (version R2016b; MathWorks, Natick, MA) after being converted into MATLAB format with the function pop_biosig from EEGLab. For the cortical analyses, the recordings were high-pass filtered at 0.1Hz, and down-sampled to 64Hz in order to improve computation speed. The filters were zero-phase causal Butterworth filters as applied in the ERPlab toolbox in EEGlab. The EEG recordings were re-referenced to the two mastoid electrodes. To handle any artefacts, the data was hard-limited manually and anything +/- 200uV was removed. A manual independent component analysis was also conducted in order to remove components associated with eye blinks. In addition to this, data recorded from the back 10 electrodes (Pz, P3, P4, P7, P8, PO3, PO4, OZ, O1 and OZ) were removed as these channels are often noisy due to infants hitting their heads on their caregivers while seated on their lap. The pre-processing procedures of the data, except for filtering, were all performed in Matlab using the Fieldtrip toolbox.

2.2.4 Procedure

Prior to testing, infants and their caregiver(s) used our infant-friendly waiting area while they became comfortable with their surroundings. Here, the infant was played with, fed and, if necessary, changed. While the infant was playing, their head was measured and the correctly sized EEG cap was chosen and tried on. The infant continued playing while the EEG cap was gelled and the electrodes were attached. Once the cap and all of the EEG equipment had been prepared, the infant was seated on their caregiver’s lap in a comfortable air-conditioned and sound-attenuated room. The EEG cap and reference electrodes were then fitted once the infant was comfortable. The infant’s caregiver was instructed not to speak to their child and to keep them as still as possible once the recording began. The researcher was present for the whole duration of the testing, sat out of the infant’s field of vision and made notes on the infant’s behaviour for the duration of the testing. During the experiment, the infant was entertained with bubbles, puppets and soft toys. A silent cartoon film was also continuously running in the room. Short breaks were taken in-between the story blocks if necessary in order to feed the infant, communicate with them and allow them to move around in order to avoid restlessness. The breaks were decided on by the experimenter when the infant was inattentive. The experiment ended when the infant was too tired or restless to sit on their caregiver’s lap or participate in the experiment any longer.

2.2.5 Coherence Analysis

In this thesis, coherence is operationalised as the degree of synchronisation between neural oscillations (as measured by the EEG) and the amplitude envelope of speech. Backwards multivariate temporal response functions (mTRFs; Crosse et al., 2016) were calculated to map the EEG neural response back to the amplitude envelope of speech. The mTRF resembles a traditional ERP and essentially extracts a neural component that is time-locked to the amplitude envelope of the acoustic signal. The mTRFs were fitted for each subject and then these fitted functions were used to calculate the predicted acoustic waveforms. The extracted neural component was then compared to random data in terms of coherence (i.e., the phase synchrony of the two signals across frequency bands). This was done by segmenting the data in to 1-sec Hann windows with 50% overlap, and calculating coherence from the cross-spectral density of the FFT of the two signals, divided by the power spectrum of each signal. The random data refers to randomly selected sections of the amplitude envelopes of the stimulus the babies heard. In order to measure coherence, mTRFs to random amplitude envelopes were created 1000 times and if the mTRFs modelled on the actual data are better at predicting the signal than 950 of the random envelopes, then we can be sure that we are measuring reliable coherence.

2.3 Results

A large range in amount of data was collected from the infant participants ranging from 5.5 minutes to 22.2 minutes (M=14.97 minutes, SD = 4.7) as illustrated in the following boxplot (Figure 2-1). It is not surprising that this large range in data amounts exists. Babies often don’t sit still for very long and sometimes fussiness results in the testing session being cut short. As a result of this, periods of noise will exist in every infant EEG recording and not all of the data will be usable. In order to try and account for this, we decided to look at coherence on an individual basis in order to see which babies had significant neural tracking. The analysis used in this study is slightly unconventional and involves using permutation statistics in order to find the best and cleanest 30 second segment of data from each infant. This will be explained shortly. However, prior to running the new analysis a more conventional approach was taken in order to ensure that we could measure reliable coherence in the baby data and that it was not dependent upon our analytical choice of choosing the best 30 seconds of data. For this analysis, mTRFs were fitted for each subject and then these fitted functions were used to calculate the predicted acoustic waveforms (windows of -150 to +700ms). The extracted neural component was then compared to the speech amplitude envelope in terms of coherence. The recordings were split into 10 sections and leave-one out cross-validation was used in order to assess how well the neural responses could be predicted. If the EEG signal could be predicted with significant accuracy (greater than zero) then we can be confident that it is encoding the speech amplitude envelope. A prediction was derived for each tenth of a signal based on modelling the other 90% of the signal. The predicted envelope and the actual envelope were compared in terms of coherence which assessed the degree of phase locking of the EEG signal to the original amplitude envelope of the stimulus. We calculated the coherence of each individual baby and compared this to random data created by calculating permutation statistics. We ran 100 random combinations of amplitude envelopes and EEG and then picked out the 95th highest coherence result at each frequency as the significance criterion. If the actual infant coherence is higher than the permutation statistics then this shows that the infants are eliciting a significant level of coherence to the stimulus. A multiple linear regression model was ran in order to predict the relationship between theta coherence and amount of data, age and gender. A significant regression equation was found (F(3,22) = 8.934, p<.001) with this model predicting 49% of the variance in theta coherence. Amount of data significantly and positively predicts theta coherence (p<.001). Gender is also a significant predictor (p=.049) with male infants being more likely to elicit significant theta coherence. Once amount of data and gender are included in the model, age does not significantly predict coherence (p=.826). Amount of data can affect the magnitude of coherence. As a consequence of this, we decided that it may be more suitable to look at shorter, more comparable chunks of infant data. This new, more unconventional analysis, is described next. To begin with, we again calculated individual coherence and compared this to permutation statistics. Permutation statistics were created by running 1000 random combinations of amplitude envelopes and EEG and we then calculated where the obtained actual coherence ranked in the random data. For example, if the real data is better than all of the random permutations then a value of 100% would be obtained.

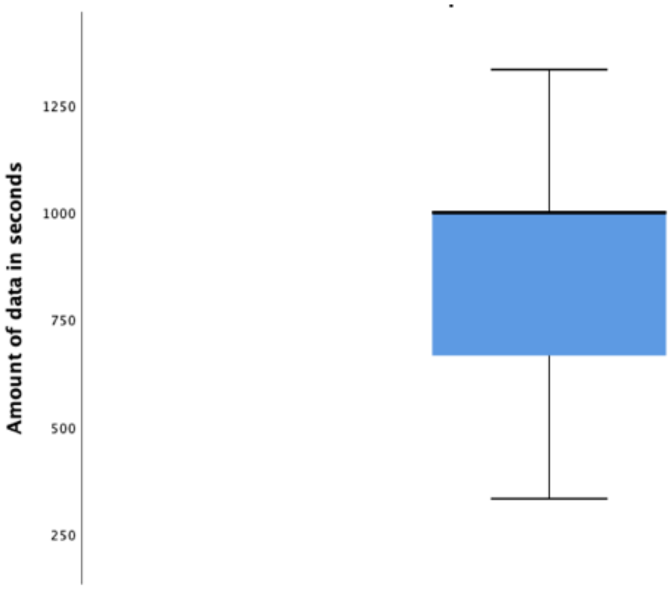


Figure 2-1: A boxplot to show the range in amount of data (s) recorded from the infant participants.

The following boxplots (Figure 2-2) show the range in overall coherence for both the story and the ACC and the corresponding permutation statistics. Coherence is illustrated here as the sum of coherence in the 2-8Hz range as this is the range in which we would expect to find neural tracking. This theta level range of neural processing has been identified in previous infant studies as the most important range with regards to oscillatory entrainment (Kalashnikova, Peter, Di Liberto, Lalor, & Burnham, 2018; Leong et al., 2017) and is regarded as the master oscillator with regards to speech processing. Overall, coherence can be measured in both conditions but seems slightly more reliable for the ACC condition. With regards to the permutation statistics, both averages are around the 95th percentile which illustrates that the majority of infants had individual significant neural tracking. ACC was included in this study as we thought that it might be a more reliable response. Our results indicate that the babies have a strong overall ACC coherence but that not every baby had good overall coherence for the story.

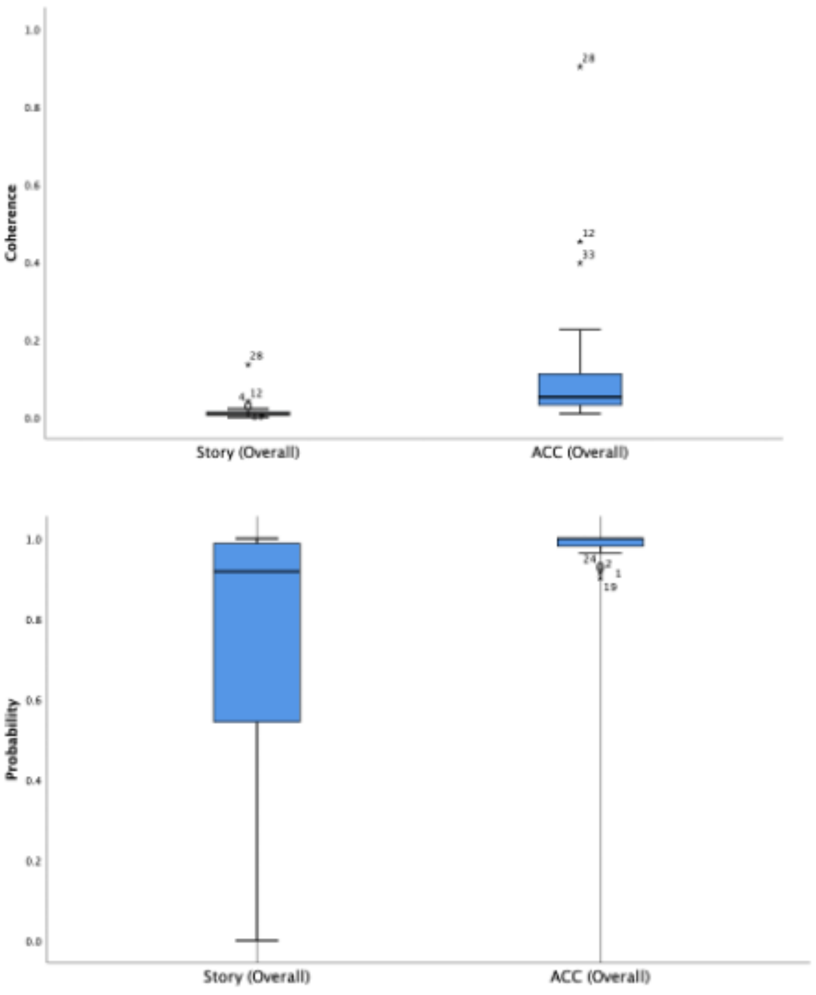


Figure 2-2: Boxplots to show overall story and ACC coherence and their corresponding permutation statistics

Although we were able to observe coherence for both the story and the ACC when looking at overall data, as previously mentioned an issue with the above calculation is that amount of data can affect the magnitude of coherence. In our data, we have large ranges in length of recording time and comparing coherence over these big ranges may not be appropriate. We know that not every minute of recorded data from an infant will be analysable due to their often extensive periods of fussiness. Additionally, we are not interested in how long an infant can sit still for and we want to measure coherence independent of these factors. Consequently, we decided that instead of looking at overall coherence we would analyse a shorter, more cleaner, segment of each infant's EEG data recording. Baby data is noisy and we wanted to have equal amounts of data for the story coherence and ACC in order to create an objective criterion, see how good the data is for each individual baby and compare the ACC and story coherence fairly. The data from each infant was divided into 30 second segments and then permutation statistics were further ran on these segments in order to obtain the best 30 seconds of data from each infant. In order to ensure a fair analysis, the best 30 seconds were also picked from 1000 separate random combinations of amplitude envelopes and EEG and the 95% threshold was used once again as our criterion. Comparing to random data by using permutation statistics helps us to see which data differs the most from random data i.e. at the 95% threshold. Once this selection has been completed, the permutation statistics are no longer used in the analysis as this would classify as 'double dipping' (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). In order to avoid this, only the infant coherence values are used in the coherence analysis. Although this is an unconventional analysis, it may not be appropriate to analyse all of the data from each infant. Although analysing all of the data is arguably the most conservative way of looking at the data, and one in which we were still able to observe coherence to both the story and the ACC, it doesn't seem appropriate to look at overall data when considering the long periods of noise and inconsistencies in length of data between babies. The first analysis of looking at the overall data was a check to ensure an indication of finding coherence, then we decided to look at the best 30 seconds of each baby's data in order to create a fair comparison between babies and conditions. As we were able to observe coherence in the overall condition before selecting the best data, this ensures that we were not manipulating the data in order to find any significant coherence.

Figure 2-3 illustrates coherence to the story and the ACC when calculated over the best 30 seconds of data, and the corresponding permutation statistics. Higher permutation statistics can be observed, indicating that more reliable coherence is being recorded here. One infant (HN28) had extremely high ACC and story coherence which skewed the data. In addition to this, three infants (HN17, HN13 and HN35) were only tested for a short period due to fussiness. These babies will be removed from hereon in the analysis. HN13 and HN35). Correlation analyses were ran on coherence values recorded per participant in the entire session and in their best 30 second segment of data for both the story and the ACC. With regards to the story coherence across the overall data and the best 30 seconds, the two are correlated (r=.719, p=.001), as are ACC coherence across the two conditions (r=.908, p<0.001). This confirms that infant coherence is able to be calculated across both the overall recording and the best 30 seconds. Therefore, this unconventional analysis is not using the best 30 seconds in order to manipulate the data. Infant coherence can be observed when looking over all of the data, but appears to be more reliable when calculated across standardised segments e.g. 30 seconds.

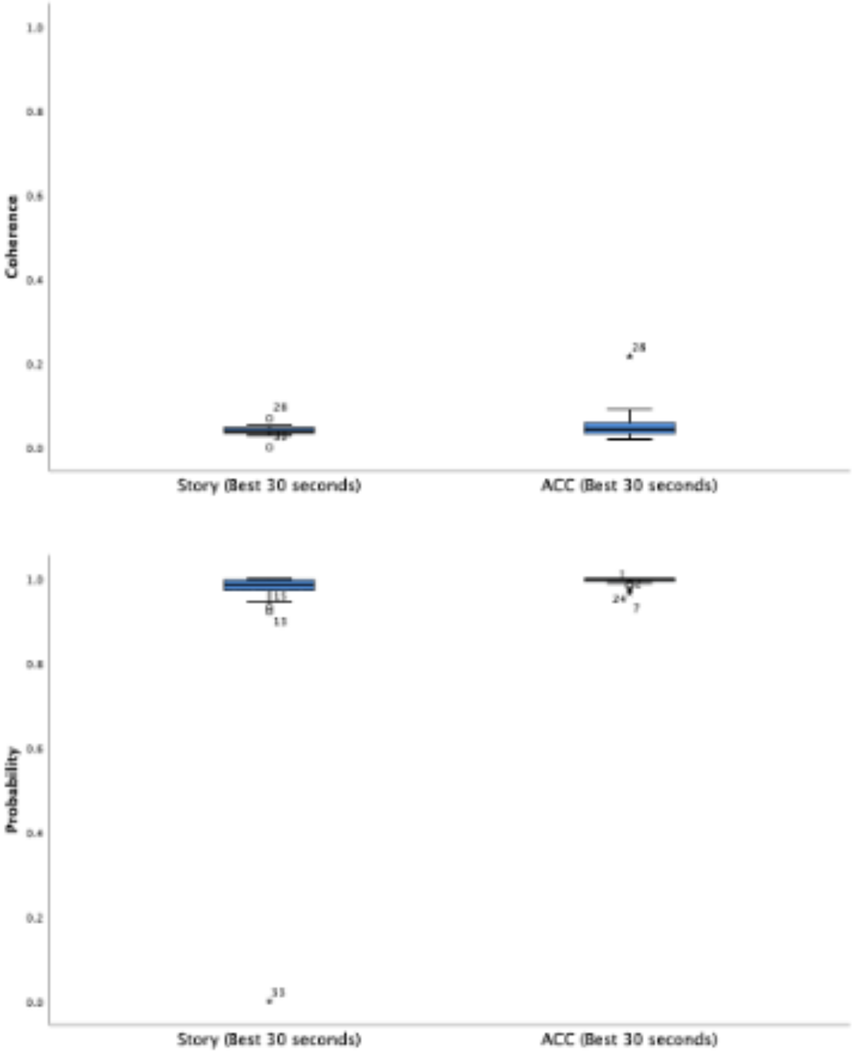


Figure 2-3: Boxplots to show story and ACC coherence calculated across the best 30 seconds of data and their corresponding permutation statistics.

The following coherence plot (Figure 2-4) shows average ACC and story coherence when calculated over the best 30 seconds of data for all infants. Low frequency peaks in the theta range (4-8Hz) can be observed which is consistent with what we would find in adult participants and is what has been observed in previous infant studies (Kalashnikova, Peter, Di Liberto, Lalor, & Burnham, 2018; Leong et al., 2017). The average ACC peak is much higher indicating more reliable coherence to the ACC than the story. The ACC peak is perhaps due to a higher due to a small number of outliers with very high coherence. However, when these are removed paired sample t-tests show a significant difference between coherence between the two conditions, t(32)=-2.338, p=.026. Additionally, ACC and story coherence are correlated with one another, r(32)=.429, p=.014. This illustrated that even though the two responses are different, they are correlated within individual infants.

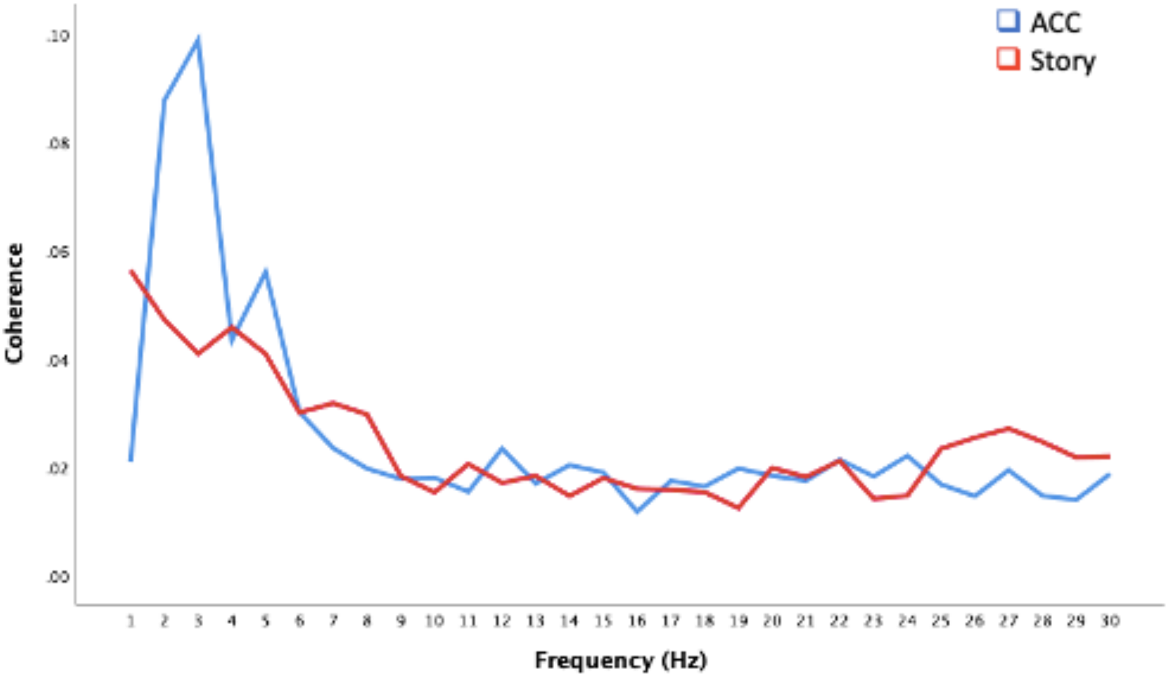


Figure 2-4: Average coherence plots for the ACC and story coherence. Blue represents the ACC condition and red represents the story condition.

Figure 2-5 shows the mTRF model weights and sensor space plots for the story and ACC. The mTRF resembles a traditional ERP and essentially extracts a neural component that is time-locked to the amplitude envelope of the acoustic signal. The mTRFs were fitted for each subject and then these fitted functions were used to calculate the predicted acoustic waveforms. The TRF latency window used in the backward modelling analysis is 0 – 40ms.

With regards to the infant ACC response, we would expect to see a peak at around 200ms after stimulus onset and a frontal scalp distribution (McCarthy et al., 2019) and this is what is illustrated in our results. However, our sensor space plots show some areas of negativity. This is most likely due to the denoising calculation technique used in the backwards mTRF model. The denoising negativity effects can also be seen in the sensor space plots for the story coherence. The frontal distribution for the story coherence is similar to previous findings in infant neural entrainment studies to IDS (Kalashnikova et al., 2018). However, it can be seen that story coherence has a shorter latency at approximately 100ms. The two measures are highly related but the latency differences may be illustrating that the ACC and the story may elicit slightly different neural responses from the infants. Although previous research has shown entrainment to continuous speech in babies, it is unclear whether this is simply an auditory response or if they are truly engaging with the speech. Kalashnikova et al., (2019) measured stronger entrainment for IDS than ADS which provides evidence that infants really are engaging with the speech that they are listening to. ACC was included in this study in order to validate the speech coherence measure with a non-meaningful speech spectral contrast. The high correlation between the story and the ACC signifies that our measures are working. However, the difference in latencies signify that even though the measures are related, they may not be tracking the exact same thing. If this is the case, the correlation between the two may be due to shared variants between the conditions such as data quality and general infant hearing abilities.

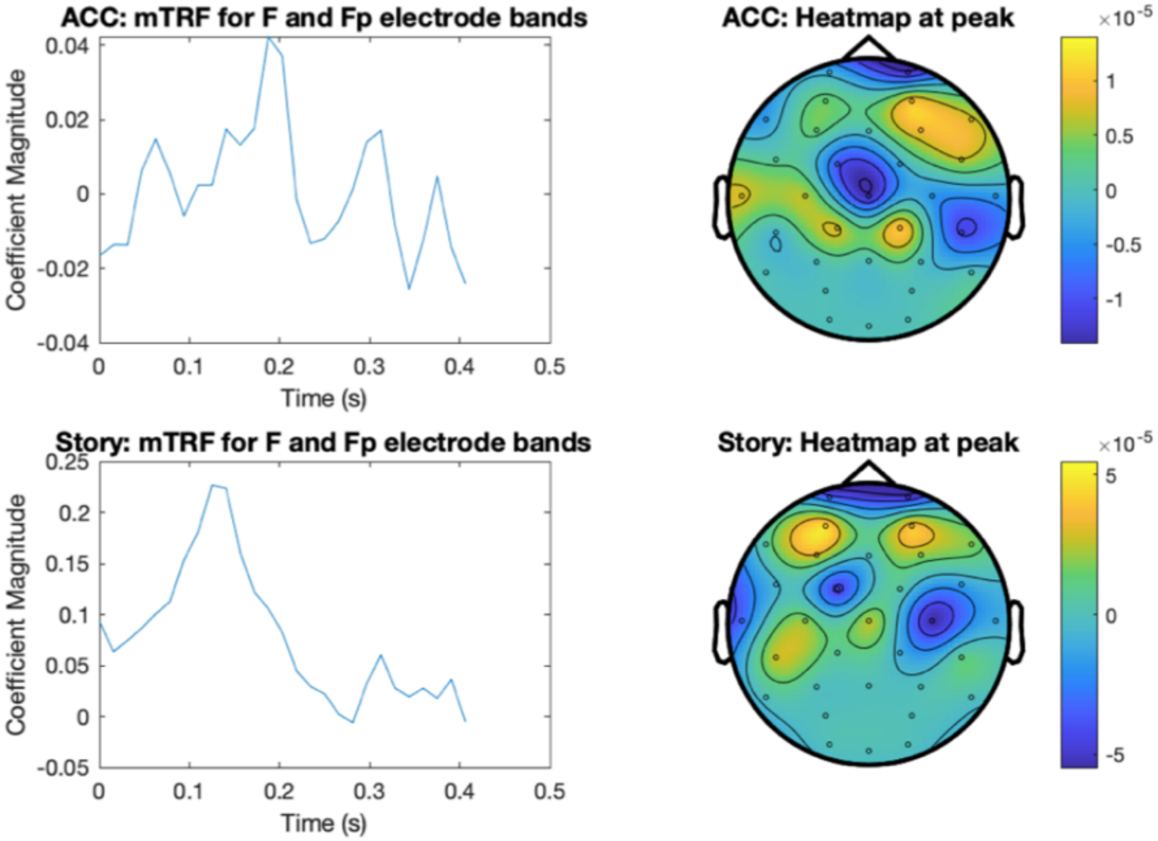


Figure 2-5: Model mTRF weights and sensor space plots for both the story and the ACC.

2.3.1 Later language assessments

Infants that participated in the EEG study were re-contacted and asked if they would like to take part in a follow-up study. A total of 18 toddlers (Mean age = 2.8, SD = 0.295) took part in the language assessments. In order to assess the children’s receptive language skills the standardised British Picture Vocabulary Scale (Dunn, Dunn, Whetton, & Burley, 1997) was administered to them. The BPVS is a widely used assessment tool of children’s language skills used between the ages of 3;0 to 15;08 years. From the data collected the researcher is able to calculate norms and age equivalents of scores in order to assess their receptive linguistic ability. Some of our toddlers fell slightly below this age threshold; our youngest child was 2.4 years old at the point of test administration. As a consequence of this we added a correction factor of (3.5-age)*10 to each child's score. Research exploring the performance of typically-developing school-age children on the BPVS showed that children are roughly increasing their scores by about 10 points per year of age (Mahon & Crutchley, 2006). The BPVS assessments with the children either took place online due to restrictions of the COVID-19 pandemic. The test was administered in a single testing session, the time of this session would vary on the child's concentration level but normally lasted for approximately 25 minutes. The test was scored according to the BPVS instruction manual. The BPVS scores reported in the analysis were calculated by working out each child's raw score (the sum of all correct answers) and then transferred into a standardised score for a particular age level as stated in the BPVS scoring conversion tables.

We wanted to investigate whether later language abilities were related to infant coherence. A correlation analysis was ran in order to see if BPVS result was positively correlated with the following variables: Age at BPVS, Age at EEG, ACC coherence and story coherence. No positive correlations were found between BPVS and any variables, as illustrated in the correlation table below (Table 2-1). This shows no strong relationship between any of the variables. A potential power issue may be occurring here illustrated by the fact that some variables that account for approximately 12% of the variance in the model were not significant. With this smaller number of babies it seems that we can only detect correlations over a certain magnitude. A few correlations seem to be approaching significance e.g. BPVS score and age at BPVS, and ACC and BPVS. However, no solidly significant relationships can be established from the correlation model. Although story and ACC coherence were correlated with one another in the main analysis, not all infants from the original study participated in the later language assessments.

A multiple linear regression was also ran in order to explore if there were any potential predictors of later language outcomes scores obtained by administering the BPVS. This regression model was non-significant, F(4,17)=1.006, p=.439, R^2 =.236 with no variables contributing to the model: Age at BPVS (B=16.162, p=.184), (Age at EEG B=.846, p=.618), Story coherence (B=-354.870, p=.461), ACC coherence (B=50.835, p=.631). The relationships between all variables are displayed in Figure 2-6. From this evidence, it can be concluded that in this study infant coherence was not predictive of an infant's later linguistic abilities.

Variable	1. BPVS	2. Age at BPVS	3. Age at EEG	4. Story coherence	5. ACC coherence
1. BPVS	-	.398	.297	.054	.350
2. Age at BPVS	.398	-	.470	.190	.050
3. Age at EEG	.297	.470	-	.122	-.074
4. Story coherence	.054	.190	.122	-	.116
5. ACC coherence	.350	.050	-.074	.116	-

*Table 2-1: A correlation table to show the relationships between BPVS, Age at BPVS, Age at EEG, Story coherence and ACC coherence. The numbers represent the Pearson correlation coefficient and significant numbers are represented by *.*

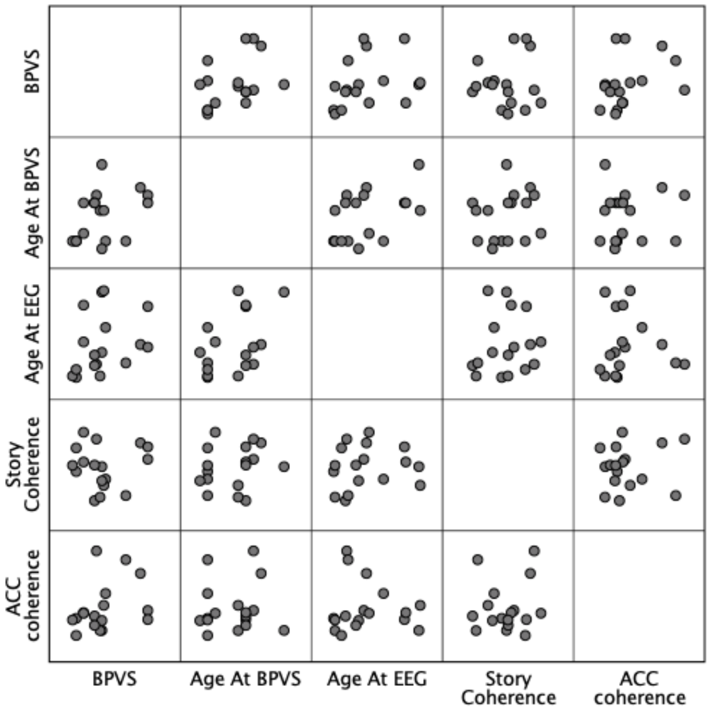


Figure 2-6: A scatterplot matrix to show the relationships between the following variables: BPVS, Age at BPVS, Age at EEG, Story coherence and ACC coherence.

2.4 Discussion

This present study demonstrated that we are able to reliably measure coherence to continuous speech and the ACC in pre-verbal infants. This study used backwards mTRFs in an unconventional analysis and focused on the best 30 seconds of infant data in order to measure infant coherence. In order to ensure that we were not manipulating the results, we first looked at coherence across each infant's overall data. It was found that we could reliably measure coherence in infants when the analysis is ran over their whole testing duration. However, due to long periods of noise in infant recordings coupled with the issue of large ranges between infant testing times potentially affecting the magnitude of coherence, we decided to divide the data up into smaller 30 second segments for the analysis. When calculated across the whole recording it is evident that average infant coherence looks reliable and it seems that most babies are cortically tracking the amplitude envelope of speech. We can still see this reliable coherence when looking at the best 30 seconds. However, the results seem slightly more robust. Average coherence plots illustrate low- frequency peaks in the theta-range (4-8Hz). The theta-band is thought to be the master oscillator in speech-brain entrainment and corresponds to the rhythmic properties of syllabic information (Greenberg, Carvey, Hitchcock, & Chang, 2003). It is thought to be a dominant oscillatory band in adult speech processing and has also been found in similar research studies when infants process speech (Kalashnikova, Peter, Di Liberto, Lalor, & Burnham, 2018). In addition to this, the theta-band is considered by some to be more of an attention-modulated evoked auditory response (Keller et al., 2017) which makes it an appropriate oscillatory band to compare the two evoked responses of story coherence and the ACC. Overall the story coherence presented in this study matches what previous infant literature has found in terms of theta-band phase locking and a frontal scalp distribution (Kalashnikova et al., 2018a).

The ACC was initially included in this study as a validity measure of coherence as it is something that we know we are able to measure in infants. Our analysis showed that we were able to reliably measure coherence to both the story and the ACC in this study. The two coherence measures were independent of each other which supports the conclusion that our overall story coherence measures are reliable and robust. A complex relationship seems to exist between story coherence and the ACC. Statistical analyses showed the mean difference between the two conditions to be significantly different yet they are correlated with one another. The same babies that elicited the ACC coherence also did so for the stories. The ACC response that we recorded looked extremely similar to that which has been recorded in previous infant studies (McCarthy et al., 2019) in terms of the scalp distribution and the P2-like peak in our mTRF model weights. The neural generators looked similar for the two when looking at the scalp topographies as both responses show a frontal distribution. Additionally, considering that the ACC blocks were a lot shorter than the story blocks (20 seconds vs 5.6 minutes), it seemed to be a more reliable response. However, the latencies of the story and ACC response are completely different. The ACC is a response to a meaningless acoustic signal and the aim of this comparison with the story was to understand whether infants are actually engaging with the speech that they are hearing or if this is simply just an auditory response. Previous infant EEG literature has argued that infants prefer listening to IDS than ADS (Kalashnikova et al., 2018) which would suggest that infants' attentional preferences play a role in their speech entrainment and that they really are processing the speech that they are hearing. However, as previously mentioned we cannot be sure if this is due to infant attention or the different acoustic properties of the speech signals. From our results, the ACC and story coherence seem to be highly related but are perhaps tracking different things. This may show that infants really are processing speech differently to the ACC, and that their speech processing is more than just an auditory response.

The results from the language assessments showed that there was no link between ACC or story coherence and later language abilities. This does not mean that cortically tracking speech in the first year of life does not help the language acquisition process. All of the infants tested, even those with lower individual coherence, still managed to acquire language to a successful degree. Data collection issues may have arisen with regards to the BPVs results due to conducting the language assessments online or in a socially distanced manner. However, the small individual differences in the language assessment scores may be due to environmental factors such as the children's social environments from birth. It has been argued that as children of talkative parents grow older they typically vocalise more than that of taciturn parents (Panagos, 1998) and that IDS in a 1:1 context is positively correlated with later word production (Ramírez-Esparza, García-Sierra, & Kuhl, 2014). However, other more recent research has argued that the quantity of speech heard by a child in the first year of life does not affect children's later language outcomes (d'Apice & von Stumm, 2020). Future work could explore this avenue further by investigating if any potential links exist between infant neural entrainment and more broader socio-linguistic developmental processes. This study showed that we are able to measure coherence to the amplitude envelope of continuous speech in pre-verbal infants. Previous research has shown that this speech engagement could be driven by attention (Kalashnikova et al., 2018). In support of this, we found evidence of differences in processing when comparing this to an auditory response (the ACC), illustrating that the infants really are engaging with the speech that they are listening to. Going forwards in this thesis, we will continue to use our reliable measures of speech coherence in order to fully understand the effects of attention on speech modulation. How infants process language in more complex auditory scenes, such as multi- talker or noisy environments, can now be reliably investigated. This will allow us in this thesis to learn more about infant selective attention and speech processing in more naturalistic environments and consequently understand more about language development in the first year of life.

3. Neural entrainment and auditory selective attention in infants: A two-talker study

3.1 Introduction

Our previous study, *Infant neural entrainment to continuous speech: Initial methodological development*, illustrated that we were able to successfully measure coherence in pre-verbal infants to both the ACC and to the amplitude envelope of continuous speech. This current experiment aims to build upon this foundation in order to investigate how infants process speech in more complex auditory scenes and how infant attention modulates this. In this study, EEG was used to record infants' neural activity while they watched an audio-visual film and listened to speech in a multi-talker environment. The infants listened to two- talkers simultaneously: The target and a distractor. From our previous study we know that we are able to measure neural entrainment to the amplitude envelope of speech in infants, and that this had also been achieved in a small number of other infant studies (Jessen, Fiedler, Münte, & Obleser, 2019; Kalashnikova et al., 2018; Leong, Byrne, et al., 2017). One of these studies, Jessen et al., (2019), also recorded infants' EEG response to dynamic audio-visual stimuli. Infants watched a 5 minute cartoon video while EEG was recorded. The infant visual ERP to complex stimuli such as faces is made up of 3 main components (Pb, Nc and Slow Wave) (Webb, Long, & Nelson, 2005). In addition, a frontocentral negativity between 400-800ms after stimulus onset is thought to be linked to the allocation of visual attention (Reynolds & Guy, 2012). The results of Jessen (2019) showed neural entrainment to the amplitude envelope of speech as well as a clear response pattern to the cartoon's motion content. They concluded that the infant brain is able to simultaneously process neural responses to different sensory modalities: auditory and visual. This present study also used an audio-visual stimulus but EEG analyses were not run on the visual processing of our infants.

The audio-visual film that was presented to the infants in this present study was created in order to ensure that the babies were attending to the target speaker. This stimulus was comprised of 5 different Southern British English speaking mothers using child-directed speech and maintaining direct eye contact with the camera. This also enabled us to keep the infants engaged. In order to create the multi-talker stimulus, an audio-visual clip was overlaid with the matched audio of another clip, so that while the original audio-video clip was playing a matched distractor audio was playing simultaneously. The infant would be able to see and hear one of the mothers speaking to them and would need to attend to this while ignoring the simultaneously played distractor audio. We wanted to explore if infants were able to exhibit neural entrainment to the target audio of continuous speech while also exposed to an equally engaging distractor. In other words, we wanted to investigate whether infants are able to attend to a stimulus when a distractor is present, similar to that illustrated in the Cocktail Party Effect (E. Colin Cherry, 1953).

The Cocktail Party Effect refers to the phenomenon that when we are in a noisy environment, such as in a bar in which many people are speaking simultaneously, we are able to follow a target conversation or be able to hear our name above the background noise. This phenomenon is more broadly related to our ability to selectively attend to acoustic stimulus. This is something that has been extensively investigated in the adult literature (Bregman, Abramson, Doehring, & Darwin, 1985; Brokx & Nooteboom, 1982; Treisman, 1960) and it is widely accepted that adults are able to engage in this process. However, research in this area concerning infants is limited. This is an important area of consideration as infants do not learn language in quiet and isolated environments but instead in the presence of competing acoustic signals and noise. Infants must attend to the speech of their parent or caregiver while ignoring irrelevant and distracting audio in order to efficiently process the acoustic properties of the target speech and consequently learn language. Even at home it is estimated that two thirds of parent-infant interactions occur when other members of the household are speaking simultaneously (Barker & Newman, 2004).

In order to be able to separate one stream of speech (the target) from the other (a distractor), the infant must be able to engage in a process known as stream segregation. Stream segregation requires an efficiently sensitive auditory system and the ability to engage in selective attention and both of these processes are still immature in infancy. With regards to auditory system sensitivity, it is thought that infants have a poorer auditory threshold to both pure tones (Nozza & Wilson, 1984) and speech (Trehub, Bull, & Schneider, 1981) than adults. In other words, infants tend to need both tones and speech to be louder in comparison to adults before they demonstrate evidence of detecting them. This may mean that infants are particularly disadvantaged at attending to a target in a multi-talker or noisy environment. Previously, 7.5 month old infants were tested on their ability to attend to a female speaker while a male talker was speaking simultaneously (Newman & Jusczyk, 1996). The target speaker repeated isolated words while a distractor spoke fluently throughout in one of three different intensity levels and then the infants heard passages of text containing the familiar or novel words. The infants listened longer to the familiar words when the target voice was more intense than the distractor by 5dB but not when they were equally intense. This suggests that infants may need a more intense target in order to be able to discriminate it from distracting background noise. This paper (Newman & Jusczyk, 1996) used the behavioural testing technique of a head-turn paradigm in order to measure infant attention. Although behavioural methods can give us an insight into infant preference, they cannot explicitly tell us about the underlying neural mechanisms representing infant selective attention and auditory neural tracking. To our knowledge, no other study has measured infant selective attention to a target and distractor acoustic stimuli by using EEG.

Unlike when testing older children and adults, we cannot explicitly instruct the infants to focus on a target speaker and therefore our target speech had to be engaging as possible in order to maximise infant attention to it. This was done by using mothers for the audio-visual film who spoke in Infant Directed Speech and maintained eye contact with the camera (and so it would seem as if they were looking at the infant watching the film). Direct eye contact has shown to be an important factor in infant language learning and development and it has been proven that infants like to look at faces with direct gaze (Farroni, Menon, & Johnson, 2006). Eye gaze is a fundamental part of human communication and in the first year of life infants begin to understand that eye gaze provides significant information in terms of emotional and social learning (Striano, Kopp, Grossmann, & Reid, 2006). From birth, infants are inherently social creatures and this is reflected in their tendency to attend to socially engaging features such as IDS and direct eye contact. The primary aim of our audio-visual stimulus was to try to ensure that the infants were attending to the target speaker.

There are several reasons why understanding infants’ ability to perceive speech in noisy environments is important to the more general understanding of how they acquire language. It is evident that the majority of infants’ linguistic input occurs in non-ideal acoustic environments in the presence of background noise or other speakers (Barker & Newman, 2004). The ability to separate the target voice of the parent or caregiver from competing acoustic streams would benefit an infant’s ability to learn language. Additionally, perceiving speech in a multi-talker environment combines auditory stream segregation with selective attention capabilities. Not only must infants be able to separate competing acoustic streams, they must then be able to choose to attend to the target stream as opposed to one, or more, distractor streams. Thus, in order for an infant to attend to a target in a multi-talker environment they must be proficient in a multitude of cognitive capabilities.

In this present study, 20 infants participated in an EEG experiment in which they were exposed to an audio-visual stimulus. They were required to visually and acoustically attend to the mother on screen while ignoring a distractor audio of equal intensity. It was hypothesised that the infants would be able to separate the target from the distractor. In other words, that neural entrainment would be greater to the target amplitude envelope of speech than the distractor. Although previous studies have illustrated that infants struggle separating an auditory target from a distractor when they are of equal intensity, these studies have used behavioural techniques in order to measure this phenomenon and have therefore failed to fully explore the neural mechanisms underlying infant stream segregation and selective attention. By using EEG in order to measure infant neural entrainment to the amplitude envelopes of the target and the distractor, we will be able understand if infants are successfully able to inhibit a distractor while attending to a target. If they are able to do this, then it will show that infants in everyday, naturalistic environments are still successfully able to attend to the linguistic input of their caregiver while surrounded by competing audio streams both in the playground and at home.

3.2 Methodology

3.2.1 Participants

Twenty infants (9 = male, 11 = female) with a mean age of 8.2 months (SD = 1.91), ranging from 5.3 to 11 months old, participated in this study. All of the infant participants were being raised in monolingual English homes and were solely exposed to English on a daily basis. Infants were recruited at in-person baby groups e.g. Rhyme Time and Baby Yoga and through online platforms such as Facebook. All infants were born full-term, were healthy with no identified developmental problems and had normal hearing, as stated by their caregivers in a language questionnaire given to them prior to the testing. Along with the language questionnaire, the caregiver of each infant also completed an informed consent form in which they were made aware that the experiment could be discontinued immediately if they wished so.

3.2.2 Stimuli

Audio-visual recordings were obtained from 5 southern British English speaking mothers who were known to the experimenter as their infants had participated in our previous study. They were asked to return to help with creating this experiment's materials as they had previously expressed an interest to help out with future studies. While an experimenter was playing with the infant, the mother was welcomed into the sound- attenuated booth in our Infant and Child Language Research Centre. A DJI Osmo Pocket camera, equipped with an internal microphone, was attached to a stuffed animal's body using its flexible arms. The mothers were instructed to interact with the stuffed animal as if it were their infant by using infant-directed speech and looking directly into the Osmo as it was recording them in order to replicate infant and mother direct eye contact. The mothers were left alone in the booth for approximately 30 minutes. During this time they had a box of books and toys that they could use to structure their dialogue and simulate mother-infant interaction. Once all of the 30 minute recordings from each mother had been obtained, they were split by hand into coherent 20 second clips using iMovie. Each 20 second clip was required to include continuous infant directed speech and direct eye contact with the camera; sections in which the mothers paused or changed toys were removed. A total of 73 twenty-second clips were obtained. Each clip was overlaid with the audio of another random clip, so that while the original audio-video clip was playing a distractor audio was playing simultaneously. The infant would be able to see and hear one of the mothers speaking to them and would need to attend to this while ignoring the simultaneously played distractor audio (recorded from another mother in the same set-up). All of the new clips overlaid with a distractor audio were then combined back together using iMovie in order to create an audio-visual film lasting for a total duration of 24 minutes and 20 seconds. The order that the mothers appeared on screen was kept constant: mother 1 would be followed by mother 2,3,4 and 5 before appearing back on screen again. The main design of this experimental stimuli was to keep the infants attentive for as long as possible by taking into account distinct characteristics of infant attentional systems e.g. infant directed speech and eye contact and focus on the target.

3.2.3 Apparatus

The audio-visual stimuli presented using the multimedia player software iMovie. The visuals were presented via a 35inch television screen placed in the booth approximately 60 centimetres from the ground while the audio signal was played via a speaker in the booth to the left of the television screen. The sound of the audio was set to 60 decibels. The experiment took place in an infant-friendly sound attenuated booth. The television was placed approximately 1 metre away from the infant. We conducted an EEG experiment using a Biosemi EEG system with 32 (Ag/AgCl) electrodes) at a sampling rate of 2048Hz. All analysis of the EEG data was performed in Matlab (version R2016b; MathWorks, Natick, MA) after being converted into MATLAB format with the function pop_biosig from EEGLab. For the cortical analyses, the recordings were high-pass filtered at 0.1Hz, and down-sampled to 64Hz in order to improve computation speed. The filters were zero-phase causal Butterworth filters of the 3rd order (response slope of - 18dB per octave) as applied in the ERPlab toolbox in EEGlab. The EEG recordings were re-

referenced to the two mastoid electrodes and the noisy channels. To handle any artefacts, the data was hard-limited manually and anything +/- 200uV was removed. A manual independent component analysis was also conducted in order to remove components associated with eye blinks. In addition to this, data recorded from the back 10 electrodes (Pz, P3, P4, P7, P8, PO3, PO4, OZ, O1 and OZ) were removed as these channels are often noisy due to infants hitting the backs of their heads on their caregivers while seated on their lap. The pre-processing procedures of the data were all performed in Matlab using the Fieldtrip toolbox.

3.2.4 Procedure

Before the EEG began, the infants used our infant-friendly waiting area while they became comfortable with their surroundings. This is a space where the infants can be fed, played with and changed. It is important to ensure that the infants are happy and comfortable before testing begins so that they can be tested for the maximum amount of time. While the infant was playing, their head was measured and the correctly sized EEG cap was chosen and tried on. The infant then continued playing while the EEG cap was gelled and the 32 electrodes were attached. Once the cap and all of the EEG equipment was fully prepared, the infant was seated on their caregiver’s lap facing the television screen in a comfortable air-conditioned and sound-attenuated room. The EEG cap and reference mastoid electrodes were then fitted once the infant was comfortable. The infant’s caregiver was instructed not to speak to their child and to keep them as still as possible once the recording began. The distance between the screen and the positioning of the infant was approximately 90 centimetres. A researcher was present for the whole duration of the testing; they were sat out of the infant's field of vision and made notes on the infant's behaviour for the duration of the testing e.g. when the child was vocalising or looking away. It was necessary to ensure that the infant was looking at the screen at all times in order for them to be attending to the target. If the infant looked away from the screen then the recoding was paused. During the experiment, it was important to ensure that the infant was looking at the screen and paying attention to the experimental stimuli without making any abrupt movements. Once the baby, caregiver and researcher were comfortable in the booth, and a clear EEG signal was being received from all electrodes, then the door to the booth was closed and the experiment began. Breaks were taken in-between the story blocks if necessary in order to feed the infant, communicate with them and allow them to briefly move around so that they didn't get restless. During these breaks, the experimental film was paused but the EEG cap remained on. The experiment would only end when the infant was too tired or restless to sit on their caregiver's lap or participate in the experiment any longer. The EEG cap was then removed and the infant was taken back to the child friendly room to play and receive their gift and certificate.

3.2.5 Coherence Analysis

As in the previous study, coherence is operationalised as the degree of synchronisation between neural oscillations (as measured by the EEG) and the amplitude envelope of speech. Backwards multivariate temporal response functions (mTRFs; Crosse et al., 2016) were calculated to map the EEG neural response back to the amplitude envelope of speech. The mTRF resembles a traditional ERP and essentially extracts a neural component that is time-locked to the amplitude envelope of the acoustic signal. The mTRFs were fitted for each subject and then these fitted functions were used to calculate the predicted acoustic waveforms. The extracted neural component was then compared to random data in terms of coherence (i.e., the phase synchrony of the two signals across frequency bands). This was done by segmenting the data in to 1-sec Hann windows with 50% overlap, and calculating coherence from the cross-spectral density of the FFT of the two signals, divided by the power spectrum of each signal. The random data refers to randomly selected sections of the amplitude envelopes of the stimulus the babies heard. In order to measure coherence, mTRFs to random amplitude envelopes were created 1000 times and if the mTRFs modelled on the actual data are better at predicting the signal than 950 of the random envelopes, then we can be sure that we are measuring reliable coherence. In this study, coherence is analysed to three conditions: target, distractor and both. The both condition refers to the sum of the two other speech signals added together.

3.3 Results

Due to issues with the trigger box, 7 babies were dropped from the analysis: IN08, IN09, IN15, IN16, IN17, IN19, IN20. For the remaining 13 infants coherence values were calculated for three different conditions: to the target, to the distractor and to both speech streams simultaneously. The both condition envelope is the sum of the two other speech signals (target and distractor) added together. The amount of data recorded from the infant participants varied from 101 seconds (1.6 minutes) to 1405 seconds (23.4 minutes), mean = 1054.2, SD = 520.7. Due to the large range with regards to testing times, we decided to look at coherence on an individual basis in order to see which infants had significant neural tracking of the attended (target) speech signal.

If the infants are able to successfully attend to the target while ignoring the distractor acoustic signal then we would expect to see significant neural tracking to the target in comparison to the distractor. To investigate this, we calculated individual coherence and compared this to permutation statistics. Permutation statistics were created by running 1000 random combinations of amplitude envelopes and EEG and we then calculated where the obtained actual coherence ranked in the random data. For example, if the real data is better than all of the random permutations then a value of 100% would be obtained. The following boxplots (Figure 3-1) show the range in overall coherence for the target, the distractor and both speech signals simultaneously, and their corresponding permutation statistics. Once again, we are looking at coherence here as the sum of coherence in the 2-8Hz range as this is the range in which we would expect to find neural tracking. In these boxplots we can see that overall coherence is higher for the 'both' condition illustrating that the infants were perhaps processing the target and the distractor simultaneously. Additionally, for the permutation statistics the averages are around the 95th percentile for the 'both' condition, in comparison to the independent target and distractor, which also shows that the infants were most likely tracking a combination of the target and distractor speech signals.

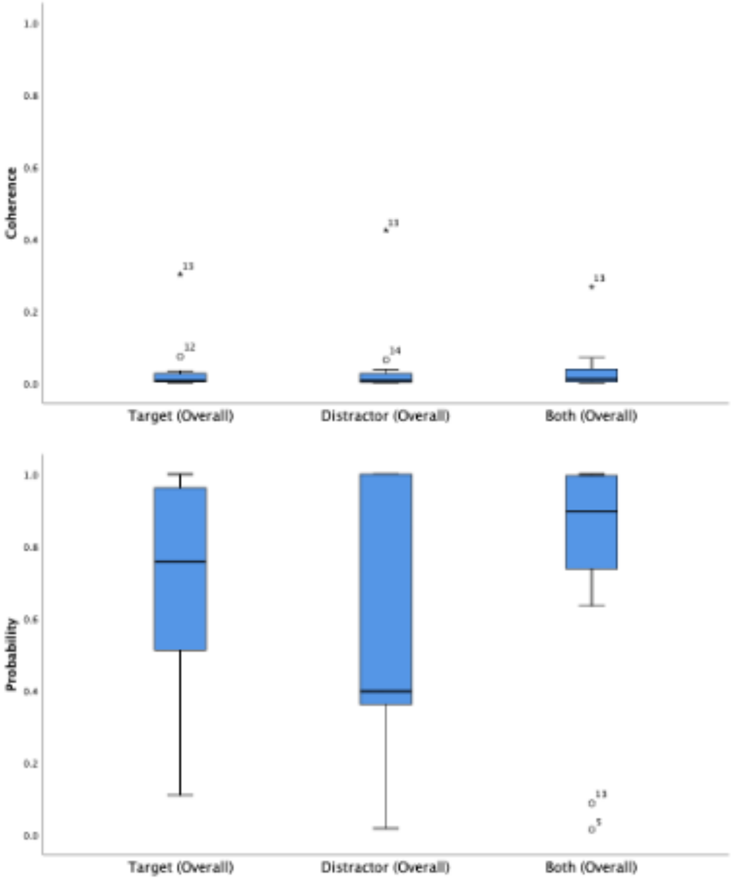


Figure 3-1: Boxplots to show overall coherence to the target, distractor and both speech streams simultaneously and their corresponding permutation statistics.

As per the statistical analysis in the previous study, it was decided that we would analyse infant coherence over a shorter segment of their data. Amount of data can affect the magnitude of coherence and thus comparing coherence over these big ranges may be inappropriate. Each infant's data was split into 30 second segments and then permutation statistics were ran on these segments in order to obtain the best 30 seconds of data from each infant. The following boxplot (Figure 3-2) illustrates coherence to the target, distractor and both speech signals when calculated over the best 30 seconds of data, and the corresponding permutation statistics. Greater coherence on average can now be seen for all conditions when regarding the permutation statistics. Paired sample t-tests were ran in order to see if any significant difference exists between the coherence in the conditions. However, no differences existed between coherence to the target in comparison to the distractor t(13)=-.951, p=.361. The following figure (3-3) shows the average coherence plots for the target, the distractor and both speech signals simultaneously when calculated over the best 30 seconds of data. The blue line represents the target speaker, the red line represents the distractor and the green line represents the 'both' condition. All three conditions have a low-frequency peak which represents the theta-range of neural processing and is associated with speech-brain entrainment and attention. All of the average entrainment plots look similar and no clear difference between the conditions can be observed. The coherence for the target, the distractor and both signals are not significantly different from one another. It seems as though on average infants are not successfully able to inhibit a distractor of equal intensity. This may be a result of infant attention fluctuating between the two signals leading to a neural tracking of both signals simultaneously. Conversely, different neural generators may exist for each of these things resulting in parts of the brain responding to the joint signal, specifically the target or specifically the distractor.

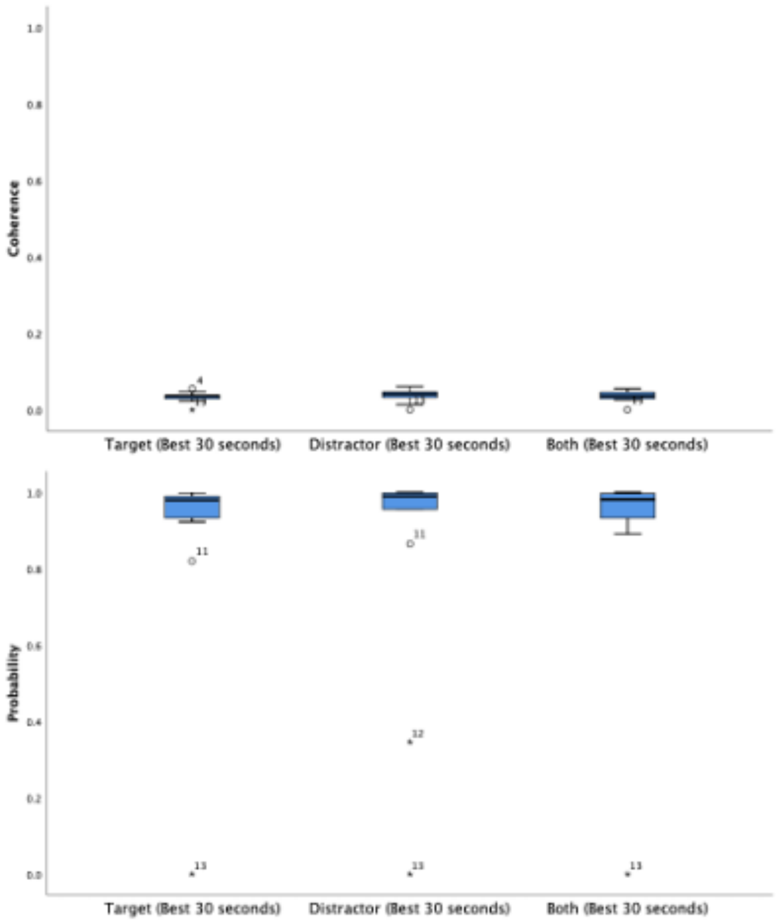


Figure 3-2: Boxplots to show coherence to the target, distractor and both signals simultaneously, and their corresponding permutation statistics.

It could be argued that we ought to get very different coherence for the ‘both’ envelope, as opposed to the target and distractor conditions, because it has a different modulation spectrum. The ‘both’ condition is created from the sum of the single talker (target condition) envelope plus a single talker with a distractor (distractor condition) envelope and so in total is made up of 3 speakers. This, in theory, would mean different modulation envelopes leading to potential difference in coherence values in the three conditions. However, it is unclear to what extent this is true. As the entrainment is not particularly strong for any of the conditions, it is not giving us a clear representation of the modulation in the original envelopes (target and distractor). It is clear that we are picking up on some coherence in the analysis, but are unable to pick up on a specific response. In typical adult continuous speech, we would expect that the an added second distractor talker would fill in the gaps in the original recording thus overall reducing the envelope modulation. The potential difference in this study is the use of IDS as opposed to ADS. The speech in our recordings are quite sparse; people don't speak to babies in regular continuous sentences but instead in short phrases or words with elongated pauses in-between. This may mean that, when a second speaker is added, the change in modulation is unlikely to be the same as if the recordings were made up of adult continuous speech. This may account for why there are no significant differences in coherence between the three conditions.

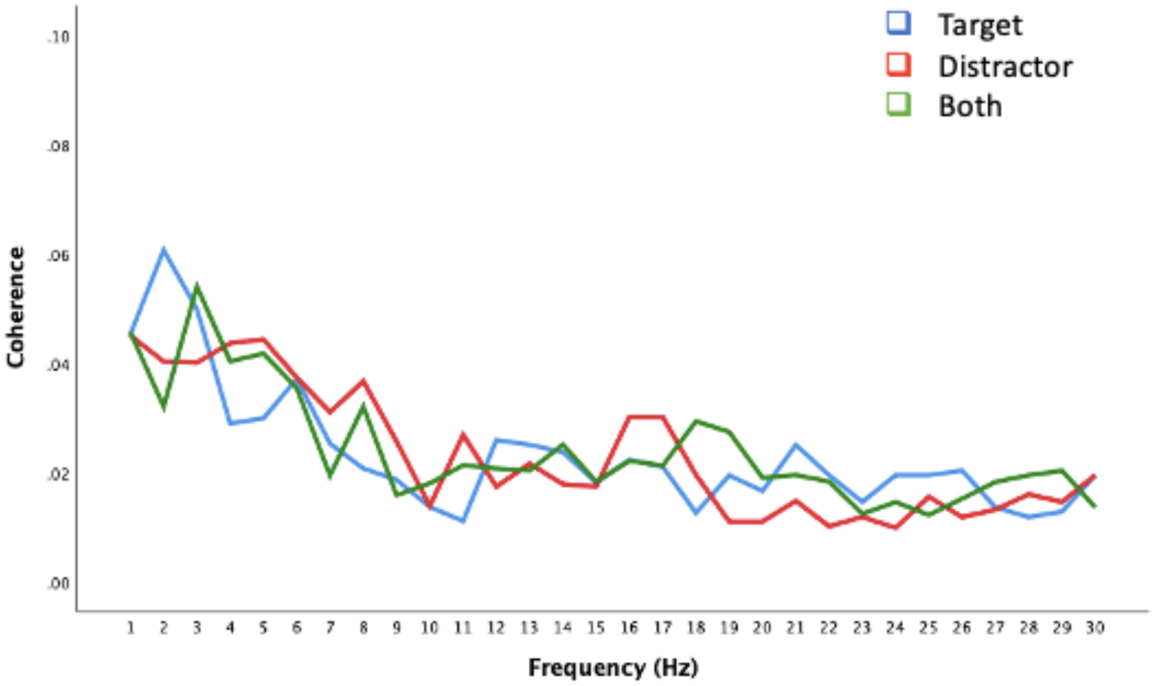


Figure 3-3: Average coherence plots for the target (blue), distractor (red) and dual both signal (green).

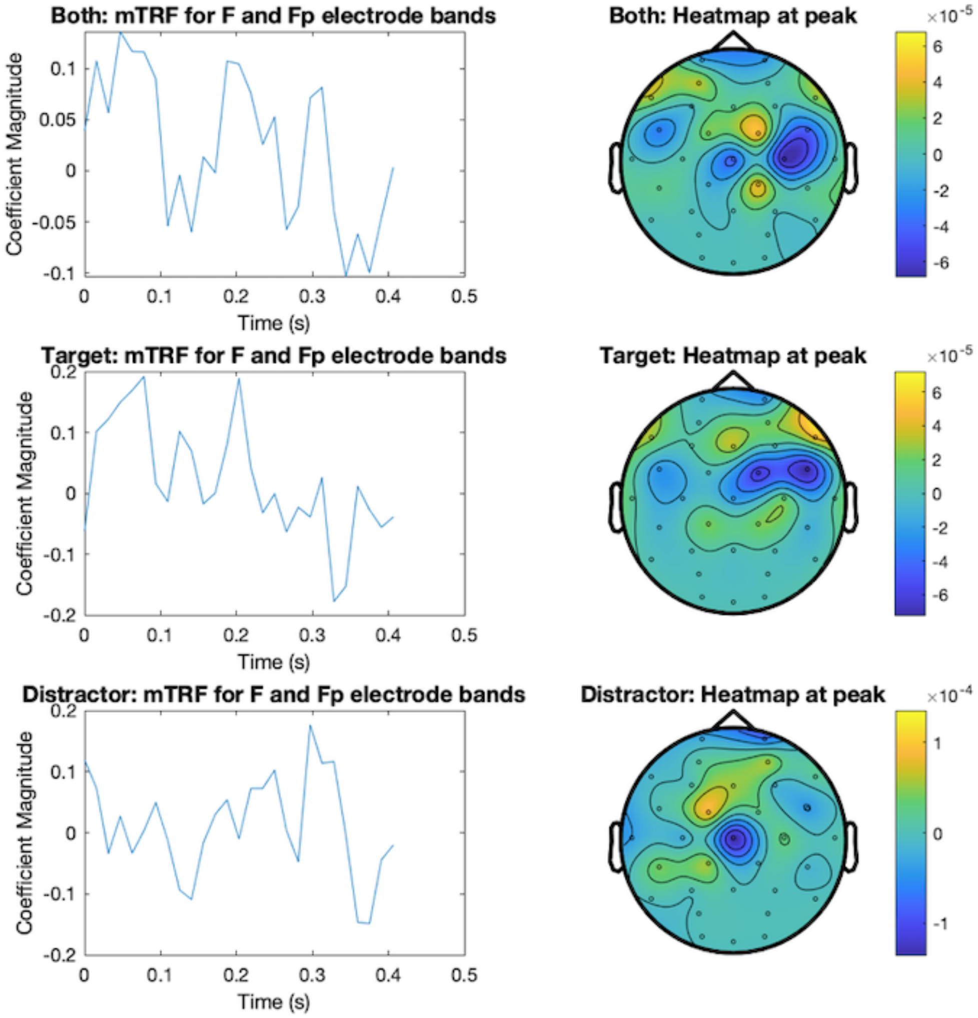


Figure 3-4 Model mTRF weights and sensor space plots for the target, distractor and dual both signal.

Figure 3-4 shows the sensor space plots and mTRF model weights. The TRF latency window used in the background model here is 0 – 40ms. All three of the mTRF plots are unclear with regards to coherence peaks. Additionally, the sensor space plots are also unclear in terms of a clear scalp distribution of the response. In comparison to the study 1 sensor space plots, these are much less regular and orderly. This supports the idea that the overall coherence was poorer in this study potentially because of infant attention fluctuation or differences in the neural generators of the target, distractor and both condition.

In order to understand whether the babies have three separate neural generators for the target, distractor and both speech streams simultaneously, we extracted the neural component associated with the target from the dual ‘both’ signal (bXt) and measured how well this compares to the separate target speech entrainment we obtained from the infants when coherence is calculated across the best 30 seconds of data. If three separate generators exist then we would expect to find that the ‘both’ neural signal shouldn't be very well correlated with the target or distractor stimuli. That is, the ‘both’ neural signal should correlate best with the ‘both’ acoustic signal, the target with target signal, and distractor with distractor signal. If our neural generators are not very specific like this, then all three types of extracted neural signals could be related. No significant difference in coherence values exists between the extracted neural component from the dual ‘both’ signal and the target entrainment t(13)=1.357, p=.198 as illustrated in the following boxplot (Figure 3-5). The coherence for the target was approximately the same regardless of whether the neural signal was extracted based on the ‘both’ or target signal. This illustrates that the infants were not attending more to the target than the distractor. This may be a result of infant attention fluctuation, or that the experiment was simply too challenging for the infants.

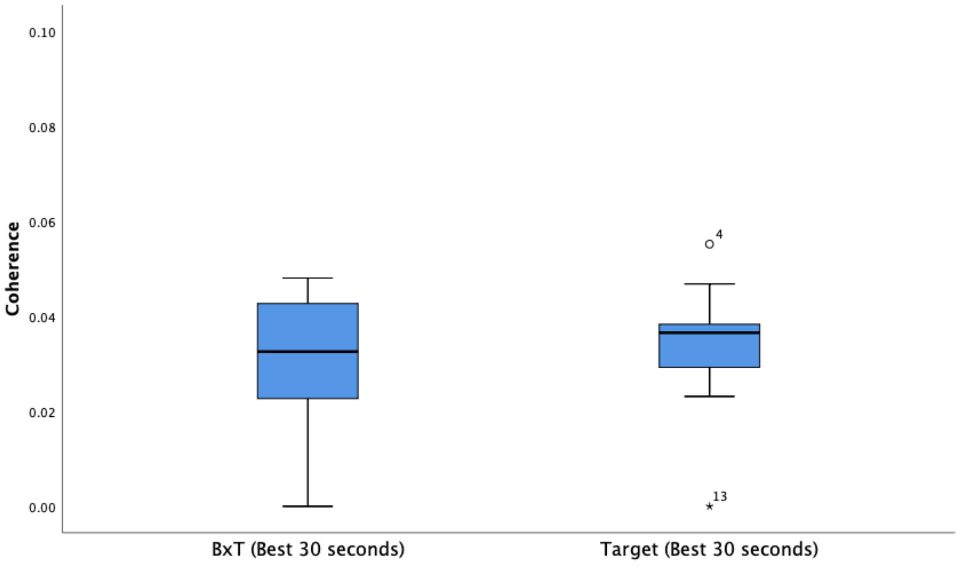


Figure 3-5 Boxplots to show average coherence of the neural component associated with the dual signal (bXt) in comparison to the target coherence.

A multiple regression was run in order to see if there were any variables influencing coherence calculated across the best thirty seconds of data to the target speaker. Variables included age (in months), amount of data (in seconds) and gender. The multiple linear regression model was statistically significant f(3,10)=8.563, p=.004 with an R^2= .720 (age B=-.003, p=.038, amount of data B=1.963E-5, p=.001, gender B=.005, p=.272). This model illustrates that age and amount of data are significantly contribute to coherence. Correlation analyses were ran in order to see if target coherence across the best 30 seconds was correlated with amount of data, age and gender. Target coherence was only positively correlated with amount of data (r=.745, p=.002) and not age (r=.-.104, p=.725) or gender (r=.065, p=.826). The following scatterplot shows the relationships between and amount of data and coherence to the target speaker (Figure 3-6). The babies that were able to sit still for longer were more likely to have better coherence along with the older children. Although the results showed no evidence of selective processing to the target speaker, it seems as though these two factors are able to contribute overall to measuring significant coherence in infants.

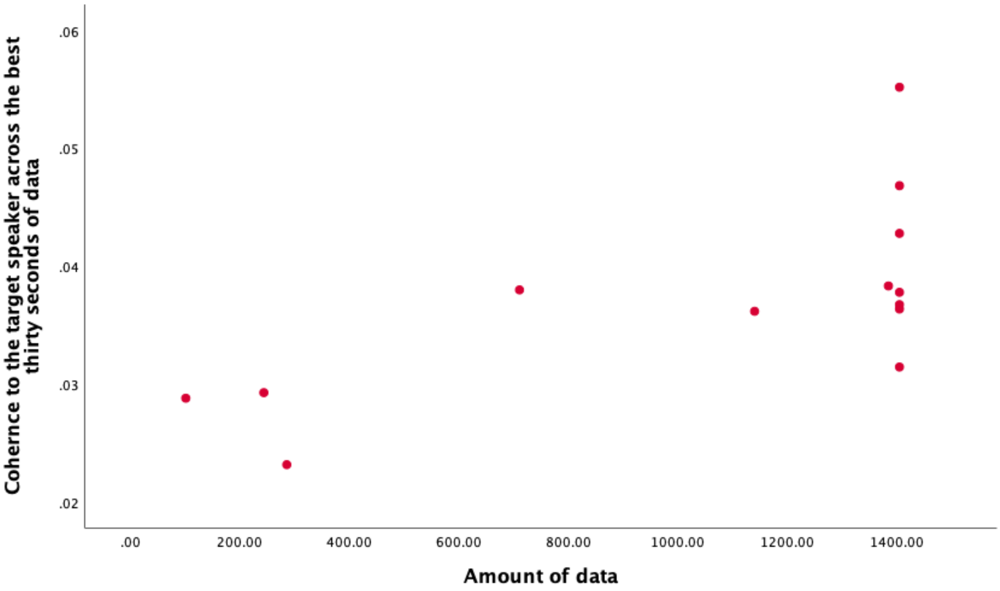


Figure 3-6 A scatterplot to show the relationship between target coherence calculated across the best 30 seconds of data and amount of data in seconds.

3.4 Discussion

This study aimed to investigate whether infants are able to attend to a target speaker when a distractor of the same acoustic intensity is present. This is a valid area of research with regards to infant language acquisition in which background noise may hinder a developing infant’s ability to become proficient in their linguistic capabilities. The results of this study were unclear and it seemed as though the infants were entraining to both of the acoustic signals at once and that they are unable to separate them. We found coherence for the target, the distractor and for the dual signal in which the target and the distractor were combined. Additionally, the coherence for the target was approximately the same regardless of whether the neural signal was extracted based on the ‘both’ or the target signal illustrating no differences in neural generators for the conditions. It seems as though attention is perhaps drifting between the two signals consequently resulting in mixed simultaneous entrainment to both the target and the distractor. From these findings, we can assume that these present listening conditions are perhaps too difficult for the infant to successfully inhibit a distractor.

Perhaps both the target and the distractor in this study were too engaging. Both voices were female and speaking in IDS. Perhaps infants would find it easier to ignore a distractor if that voice were male. We know that infants have a preference for a female over a male's voice (Decasper & Prescott, 1984) and so perhaps a distractor male voice would have been easier for the babies to inhibit. Additionally, the preference for a woman's voice in infants is typically mother specific (Lee & Kisilevsky, 2014) due to aspects concerning familiarity. We had to record our set stimuli in advance and so were unable to use each infant's own mother's voice per study, but future studies with infants could try to do this to optimise attention and engagement with regards to a target speaker. Another way of creating a maximal difference in terms of infant selective attention between our target and distractor would be if one spoke IDS and the other ADS. We know from previous research that infants prefer the temporal modulation structure of IDS and that the theta-band entrainment for this is greater than for ADS (Leong, Kalashnikova, Burnham, & Goswami, 2017). IDS may most likely act as a facilitator for language acquisition through it's highly engaging properties such as rhythm, repetition and stressed syllables. Thus, if our distractor of equal intensity was speaking ADS instead of IDS, the infants may have been more likely to selectively attend to the target.

With regards to variables that may influence target speaker coherence, the multiple regression showed that age and amount of data positively contributed to an infant's coherence. Although it remains unclear whether any infants were really selectively attending to the target or just entraining to everything they were hearing, it could be possible that the older infants have slightly advanced selective attention skills. This would make sense as we know that a sensitive auditory system and ability to engage in selective attention is maturing throughout the first year of life. Going forwards, older infants could be investigated with this experimental paradigm in order to explore at what point the ability to attend to a target while inhibiting a distractor occurs.

Overall this study seems to show that the current listening conditions of a target and a distractor of matching acoustic intensity are too difficult in order for an infant to reliably entrain to the target. Instead, our results show that they are most likely processing both acoustic signals simultaneously perhaps due to fluctuations in attention. Previous behavioural research (Newman & Jusczyk, 1996) had illustrated that infants are only able to selectively attend to a target speaker when it is at a higher intensity than the distractor, but not when they are of an equal intensity. Our results would support this argument. Consequently, the next study will use a distractor of a lesser intensity in order to see if infants are able to attend to a target that is more acoustically salient. A wider variety of noise levels and speaker types needs to be used going forwards in order to understand more what drives infant selective attention and how this can modulate their coherence levels to continuous speech in non-ideal listening environments.

4. Neural entrainment and selective attention in infants: The effects of noise and bilingualism on speech perception.

4.1 Introduction

In our previous study, *Neural entrainment and selective attention in infants: A two talker study,* we wanted to explore if infants were able to attend to a target speaker when a distractor speaker was present. The results showed that when infants are presented with two talkers of matching acoustic intensity the infants struggle to separate the competing speech streams. Infant attention seemed to fluctuate between the two signals resulting in mixed simultaneous entrainment to both the target and the distractor. From these findings, it is reasonable to assume that the listening conditions in the previous experiment were too challenging for the infants. In this present study, we still wanted to explore if infants are able to inhibit a distractor and entrain to a target speaker. However, as a result of our previous findings it was decided to make the listening conditions easier. In one condition, the target speech was played with no acoustic interference. In the other condition, the target speech was played along with a 10-talker babble with a signal-to-noise ratio of 10dB. The target speech the infants needed to attend to was that of the speaker in the audio-visual film. We wanted to explore if infants were able to exhibit neural entrainment to the target audio of continuous speech while also exposed to the 10-talker babble. However, in contrast to our previous study the distractor audio is now being presented at a lower acoustic intensity. Additionally, in this study it was investigated whether bilingualism has any effects on this process.

Infants in everyday naturalistic environments must process speech in less than ideal listening conditions and they are faced with this problem while using an auditory system that is not yet fully developed. Although much of the previous literature has focused on how adults are able to segregate target from distractor speech in a multitude of environments (Cherry, 1953; Pollack & Pickett, 1958; Poulton, 1953), few studies have investigated how infants process speech in noisy scenarios. Because of this, uncertainties existed as to what level of noise should be employed in this study. Previous research using behavioural measures found that infants from 5 months old can attend to their own name among other names when the signal-to-noise ratio between the target and the distractor is set at 10dB, but not when this is at 5dB (Newman, 2005; Newman & Jusczyk, 1996). More recent research (Bernier & Soderstrom, 2018) using behavioural measures have also found that multi-talker speech at a 10dB SNR is accessible for children aged 5 to 12 months old to perceive their own name. However, it isn't until 9 months old that they show sensitivity to the phonetic details that discriminate their own name from others. Taken together, these findings reasonably suggest that infants should be able to attend to a target speaker when the SNR is 10dB. This is thus the noise level that will be used in this study.

In addition to exploring whether infants are able to entrain to a target speaker when a less acoustically intense distractor is present, one factor that we wanted to consider in our investigation was the effects that bilingualism may or may not have on the degree of neural entrainment to the target talker. Bilingualism is difficult to define in infants who cannot yet speak and so here we define a bilingual infant as one who is exposed to another language for at least 25% of their typical daily language exposure (Pearson et al., 1997). More than two thirds of children worldwide are raised in bilingual environments and in London, where our research project was set, an exceptionally high number of multilingual inhabitants are raising children. This makes this investigation highly relevant to real-world applications of infant language learning. Bilingual language acquisition seems to be as seamless as monolingual acquisition; infants pass the same developmental milestones at the same ages regardless of how many languages they are exposed to. Acquiring two (or more) separate language systems from birth undoubtedly has an impact on the way that the infant brain acquires and processes information. For example, all infants are born with the capability to discriminate between a universal set of phonetic contrasts. Monolinguals begin to lose this skill at 6 months old (Janet F. Werker & Tees, 1984), whereas bilinguals maintain the ability to develop phonetic distinctions in all of the languages that they are exposed to. A bilingual's ability to perceive speech and their cognitive capabilities must be linked as one language must be activated while the other remains supressed. The Bilingual Interaction Activation (BIA) model poses the idea that bilinguals have increased selective attentional control and are better at inhibiting unwanted information (Bialystok, Craik, Green, & Gollan, 2009b). Bilingual and monolingual children were tested on this process in a dimensional change card sort task (DCCS). In this task cards vary on two dimensions (normally shape and colour) and children are first asked to sort the cards by one dimension and then the other. Bilingual children aged between 4 and 5 are markedly better than monolinguals at switching to the new rule and researchers claimed that this was due to the constant need of the bilinguals to inhibit one language which generalises to their effective inhibition of nonverbal information (Bialystok & Martin, 2004).

In addition, more recent research regarding the effects of bilingualism on attentional control in pre-verbal infants has found that a bilingual advantage exists (D'Souza et al., 2020; Kovács & Mehler, 2009). While other research has found that while bilinguals aren't necessarily advantaged in comparison to monolinguals, substantial differences in attentional processing are evident (Kalashnikova et al., 2021). Consequently, the attentional processing differences that arise from experiencing two or more languages in the first year of life may be evident in this study. The bilingual infants may have an advantage at supressing the distractor in this present experimental condition and the differences in attentional control between the monolinguals and bilinguals may be clear. If this was the case, then we would expect to see greater coherence to the target speech in the 'noise' condition in which the infants must inhibit a 10-talker background babble in order to attend to the target speech.

It has been reported that the same neural network underpins the processing of both languages in the bilingual brain and evidence indicates that the L2 seems to be acquired by the same neural structures as L1 (Abutalebi, 2008). The constant need to inhibit one language while keeping the other active within the same neural network may be the reason that bilingual children seem to show a consistent advantage when it comes to selective attention in comparison to monolingual children. However, there are issues regarding the reliability of this bilingual advantage of enhanced cognitive control. Some studies have concluded that monolinguals and bilinguals actually perform equally on tasks that aimed to measure their ability to inhibit irrelevant information (Duñabeitia et al., (2014), Paap, Johnson, & Sawi, (2015)) and that it is actually the specific contexts of language learning and language use that may lead to the differences in attentional control.

Regarding auditory processing, bilinguals show stronger subcortical entrainment of fundamental frequency (F0) (Krizman et al., 2012) and an earlier frontal positivity for primed spoken words (Kuippers & Thiery, 2015). Olguin, Cekic, Bekinschtein, Katsos & Bozic (2019) investigated how bilingualism influences the neural mechanisms of auditory selective attention and additionally how this is affected if the two languages are typologically similar (e.g. Dutch-English) or not (e.g. Spanish-English). The participants (monolingual and early bilingual adults) were instructed to listen to a story in their native language while ignoring four different distractor conditions in their other ear and their neural activity was recorded using EEG. The distractor was either: a narrative in the subject's language (most distracting), a narrative in an unknown language, musical rain or no interference (least distracting). Results showed that all participant groups experienced stronger encoding of the target stream in all conditions. However differences existed between monolinguals and bilinguals regarding the degree to which each interference type affected entrainment. For monolinguals, the less intelligible the distractor (e.g. musical rain, or a narrative in an unknown language), the stronger the neural entrainment, with the strongest entrainment to the target being elicited when no distractor was present. This shows that in monolinguals the presence of a distracter somewhat impacts their ability to attend to a target and that the strength of this impact overall depends on the intelligibility of the distractor. In the bilingual group, however, there were no significant differences in entrainment across the distractor conditions and so it is argued that bilingualism modulates selective attention at the neural level. As part of this present study, we wanted to see if these identifiable differences in selective attention and neural entrainment between monolinguals and bilinguals infants are also observable.

In this present study, 30 infants were tested on their ability to attend to a target speaker in the presence of distracting background noise. The potential effects of bilingualism were also investigated as both monolinguals and bilinguals were included in this experiment. In our previous study we found that infants were unable to inhibit a distractor. However, the listening conditions in the previous study were much more acoustically challenging. This study used a SNR that previous behavioural research has shown to be easily inhibited by infants. By doing this, it was hypothesised that the infants would be able to attend to a target while the noisy distractor is present; infants often attend to a target speaker in noisy environments and are rarely ever exposed to speech in isolation. Noisy nurseries and playgrounds often make up the background of the auditory environment in which infants learn speech. The secondary hypothesis, relating to bilingualism, is that we would perceive a slight bilingual advantage with regards to attending to a target and suppressing the distractor. Bilinguals are constantly in a state of supressing one (or more) language(s) while they are processing/producing speech and so it is assumed that they might have a slight neural advantage in ignoring the distracting audio in this study.

4.2 Methodology

4.2.1 Participants

Thirty infants (12 = male, 18 = female) with a mean age of 8.7 months (SD = 1.75) ranging from 5 to 11 months participated in this study. Out of the 30 infants tested, 4 were removed from the analysis due to technical issues with the trigger box. Out of the remaining 26 participants, 13 were being raised in a monolingual English home and were solely exposed to English on a daily basis. The remaining 13 subjects received daily exposure to two or more languages and will be referred to here as bilingual infants. The parents or caregivers of the bilingual children provided a detailed overview of the infant's language exposure in terms of what percentage of the day they would hear each language. In order to classify as a bilingual infant, the infant would have to be exposed to at least 25% daily of one other language in addition to English. The additional languages that the infants were exposed to were: Polish (N=3), Spanish (N=3), Bengali (N=1), Czech (N=1), French (N=1), German (N=1), Hungarian (N=1), Indonesian (N=1), Italian (N=1), Japanese (N=1), Lithuanian (N=1), Norwegian (N=1), Swedish (N=1) and Turkish (N=1). Infants were recruited at in-person baby groups such as Rhyme Time and Baby Yoga and through online platforms such as Facebook. All infants were born full-term, were healthy with no identified developmental problems and had normal hearing, as stated by their caregivers in a language questionnaire given to them prior to the testing. Along with the language questionnaire, the caregiver of each infant also completed an informed consent form in which they were made aware that the experiment could be discontinued immediately if they wished so.

4.2.2 Stimuli

The same audio-visual recordings from the previous study were used. Approximately half of these recording clips (37) were combined with a 10-talker babble speech signal over the original audio using Praat; the signal-to-noise ratio was set at 10dB for all clips. The other half of the recordings (36) belonged to the no-interference speech condition and therefore the audio was not manipulated. All of the clips were then combined back together using iMovie in order to create an audio-visual film lasting for a total duration of 24 minutes and 20 seconds. The order in which the clips were presented was randomised so that no-sound interference would be followed by a clip with babble and so on. The order that the mothers appeared on screen was kept constant: mother 1 would be followed by mother 2,3,4 and 5 before appearing back on screen again. The main design of this experimental stimuli was to keep the infants attentive for as long as possible by taking into account distinct characteristics of infant attentional systems e.g. infant directed speech and eye contact.

4.2.3 Apparatus

The audio-visual stimuli was presented using the multimedia player software iMovie. The visuals were presented via a 35inch television screen placed in the booth approximately 60 centimetres from the ground while the audio signal was played via a speaker in the booth to the left of the television screen. The infants were seated approximately 1 metre from the television and the volume of the audio was set at 60 decibels. We conducted an EEG experiment using a Biosemi EEG system with 32 (Ag/AgCl) electrodes) at a sampling rate of 2048Hz. All analysis of the EEG data was performed in Matlab (version R2016b; MathWorks, Natick, MA) after being converted into MATLAB format with the function pop_biosig from EEGLab. For the cortical analyses, the recordings were high-pass filtered at 0.1Hz, and down-sampled to 64Hz in order to improve computation speed. The filters were zero-phase causal Butterworth filters of the 3rd order (response slope of - 18dB per octave) as applied in the ERPlab toolbox in EEGlab. The EEG recordings were re-referenced to the two mastoid electrodes and the noisy channels. To handle any artefacts, the data was hard-limited manually and anything +/- 200uV was removed. A manual independent component analysis was also conducted in order to remove components associated with eye blinks. In addition to this, data recorded from the back 10 electrodes (Pz, P3, P4, P7, P8, PO3, PO4, OZ, O1 and OZ) were removed as these channels are often noisy due to infants hitting their heads on their caregivers while seated on their lap. The pre-processing procedures of the data were all performed in Matlab using the Fieldtrip toolbox.

4.2.4 Procedure

Experimental procedure followed the same routine as the previous EEG experiment (Study 2).

4.2.5 Coherence Analysis

The process for analysing coherence follows the same use of the mTRF toolbox as in the previous two studies. In this study, coherence is analysed with regards to the two conditions: quiet and noise.

4.3 Results

Due to issues with the trigger box, 4 infants (H304, H321, H324, H330) had to be dropped from the analysis. With regards to the remaining 26 infants, testing duration ranged from 246 seconds (4.1 minutes) to 758 seconds (12.6 minutes), with a mean testing time of 643 seconds (10.7 minutes), SD = 175.9. On average we were able to collect approximately 10 minutes of data from each baby and 16 babies completed the maximum testing time of 12.6 minutes. This may illustrate that the infants were happy and engaged with the presented audio-visual stimulus and this captured their attention sufficiently.

In order to see which infants had significant neural tracking we looked at infants on an individual level. We calculated individual coherence to two conditions: the condition to speech with no noise interference (the quiet condition) and the condition to speech with a 10dB SNR (the noise condition). These individual coherence results were compared to permutation statistics in order to see if reliable coherence could be measured. Permutation statistics were created by running 1000 random combinations of amplitude envelopes and EEG and we then calculated where the obtained actual coherence ranked in the random data. For example, if the real data is better than all of the random permutations then a value of 100% would be obtained. For our analysis, we decided to look at the 95th percentile in order to observe significant neural tracking. The following boxplots (Figure 4-1) show overall coherence in the quiet and noise condition in comparison to their corresponding permutation statistics. The coherence boxplots show overall low coherence to the speech in both of the two conditions, but when looking at the probability boxplots it seems as though there is a greater chance of significant neural tracking occurring in the noise condition, meaning that the infants are reliably tracking speech in the presence of background noise.

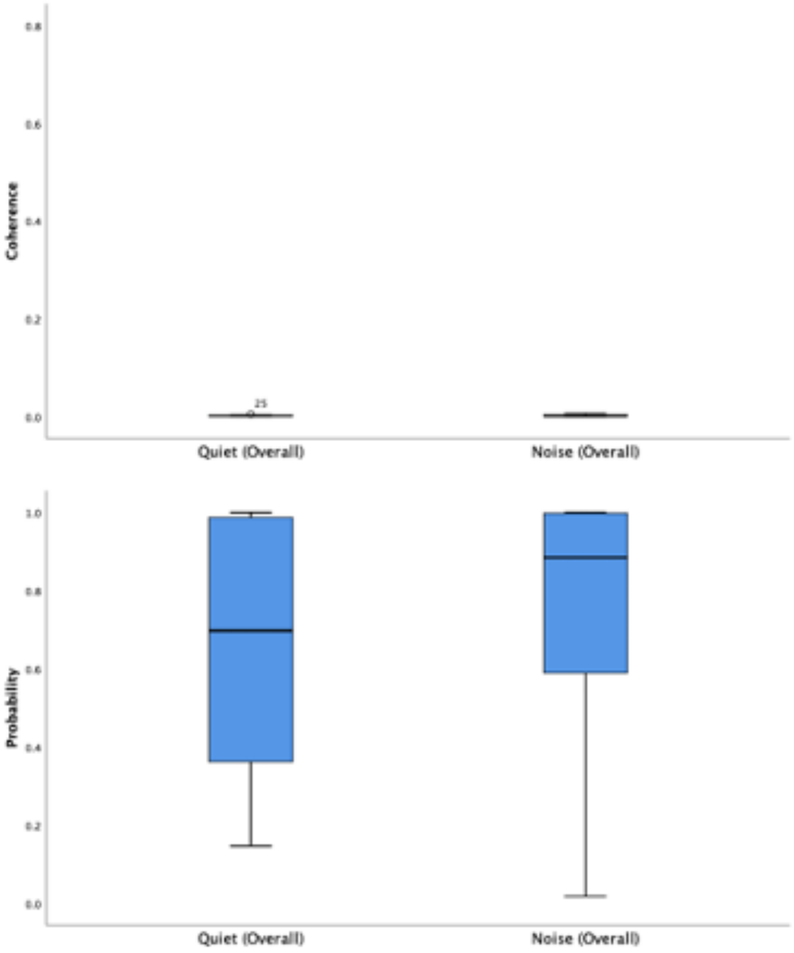


Figure 4-1: Boxplots to show overall quiet and noise coherence and their corresponding permutation statistics.

As per the other successful analyses in this thesis, it was decided to pick the best 30 seconds of data per infant. We know that amount of data can affect the magnitude of coherence and comparing over big ranges may be inappropriate. Additionally, infant data is noise and so looking at their best 30 seconds of data allows us to observe their best potential speech tracking. The following boxplots (Figure 4-2) show coherence to quiet and noise across the best 30 seconds of data, and the corresponding permutation statistics.

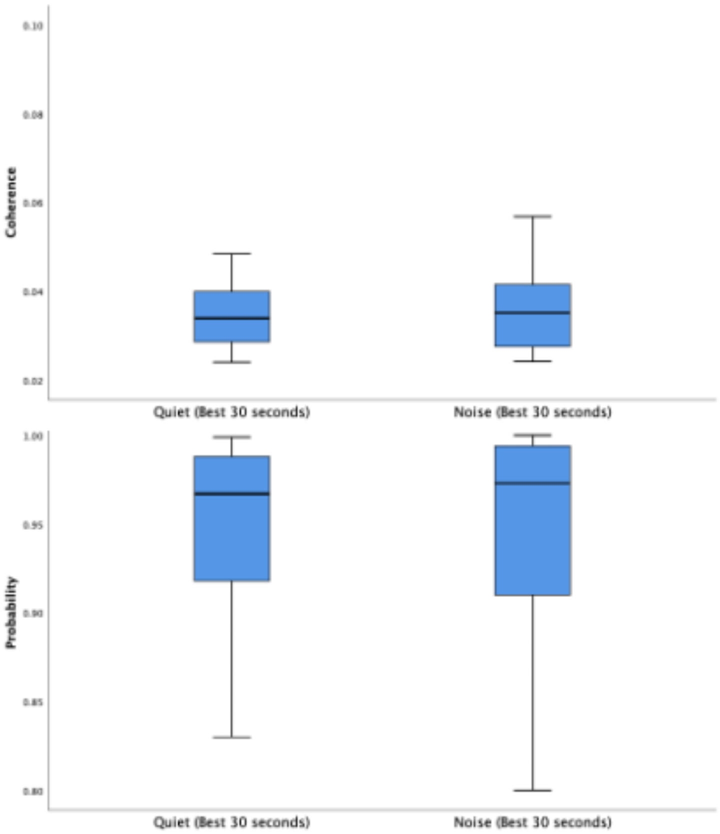


Figure 4-2: Boxplots to show noise and quiet coherence calculated across the best 30 seconds of data and their corresponding permutation statistics.

Both the coherence plots and permutation statistics have improved. Permutation statistics show that the real data is better than 95% of the random permutations for both conditions illustrating that on average individual significant coherence is occurring for both the quiet and noise conditions. This can be illustrated by the following average coherence plots (Figure 4-3). Here, the noise condition is represented by the blue line and the quiet condition by the red line. Although not as clearly defined peaks as in Study 1, peaks in the low-frequency range (2-8Hz) can be observed. Theta-range (4-8Hz) entrainment is typically associated with speech-brain entrainment and we can see evidence for this in both the quiet and the noise condition. From a visual inspection there seems to be a slightly larger delta-range low-frequency peak (<2Hz) for the noise condition.

However, the two conditions are not significantly different t(26)=-.631, p=.534, and also are not correlated with one another r(26)=.012, p=.954.

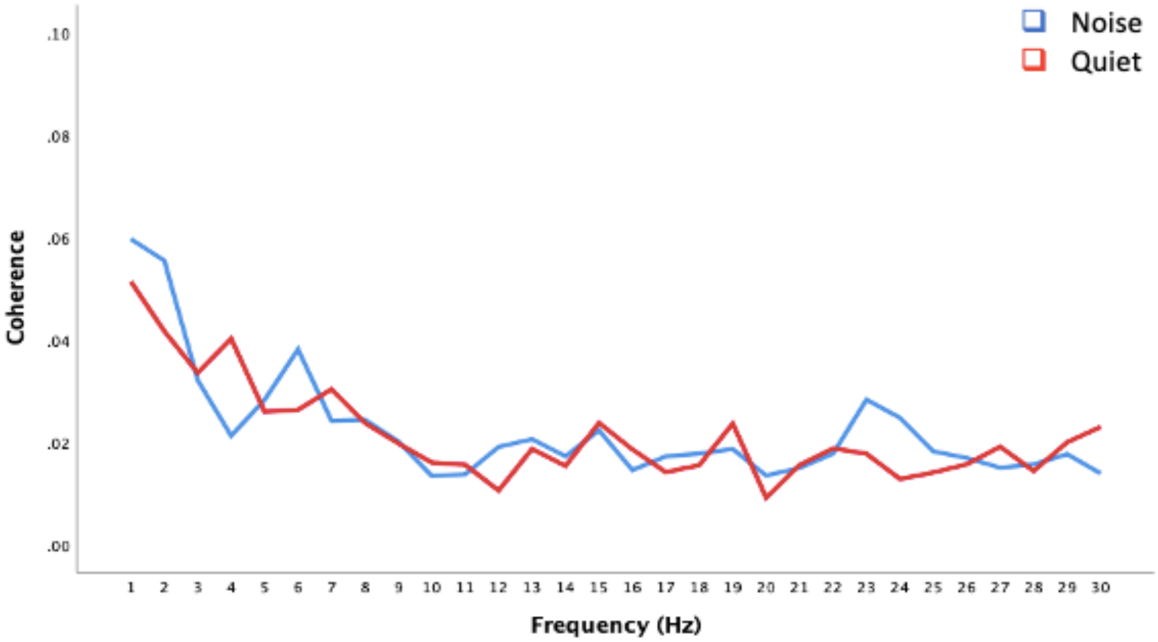
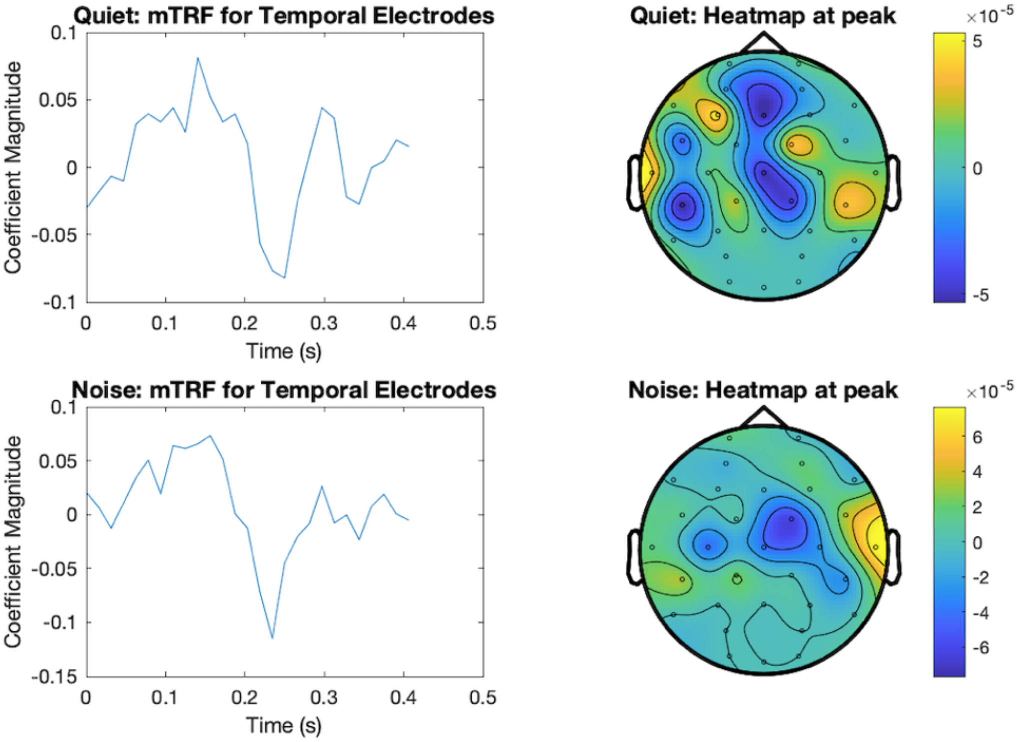


Figure 4-3: Average coherence plots for both quiet (red) and noise (blue) conditions.

The following figure (4-4) shows the mTRF model weights and sensor space plots for both the noise and quiet condition. Once again, the latencies of the TRF used was 0 – 40ms. Both the quiet and noise mTRF model weights show the same peak as we observed in Study 1 at around 100ms which may be showing the speech-brain entrainment. However, in these mTRF model weights we can see a much larger negativity following the peak and a different scalp distribution in the sensor space plots. With regards to the scalp topographies, the greatest positivity can be seen at the temporal electrodes. The differences between these two mTRF model weights and the sensor space plots in comparison to study 1 may be due to the audio-visual aspect of the study. Jessen et al., (2019), who used an audio-visual stimulus, observed clearly defined response functions for both the audio and motion regressors. Part of the motion response included a larger negativity after the peak, as seen in our model mTRF weights. Perhaps our mTRF is suppressing a video-related response resulting in this negativity.


Quiet: mTRF for Temporal Electrodes

Quiet: Heatmap at peak

Noise: mTRF for Temporal Electrodes

Noise: Heatmap at peak

Figure 4-4 Model mTRF weights and sensor space plots for both the quiet and noise conditions

In this present study we also wanted to investigate the effects of bilingualism on the cortical tracking of speech. A multivariate linear regression was ran in order to see of language (monolingual or bilingual) had an effect on coherence calculated across the best 30 seconds in both the quiet and the noise condition. It was found that language did not have a significant effect on coherence in the quiet condition (f(1)=.004, p=.951) or the noise condition (f(1)=.976, p=333).In addition, a correlation analysis was ran in order to see if both noise and quiet coherence was positively correlated with age at test. No positive correlations were found, as illustrated in the correlation table below (Table 4-1).

Variable	1. Quiet Coherence	2. Noise Coherence	3. Age
1. Quiet Coherence	-	.012	-.041
2. Noise Coherence	.012	-	-.009
3. Age	-.041	-.009	-

*Table 4-1: A correlation table to show the relationships between quiet coherence, noise coherence and age. The numbers represent the Pearson correlation coefficient and significant numbers are represented by *.*

The following boxplots (Figure 4-5) show the differences in average coherence for both noise and quiet coherence calculated over the best 30 seconds of data for both bilinguals and monolinguals. The blue boxplots represent quiet coherence and the red boxplots represent noise coherence. From a visual inspection, the monolinguals seem to be slightly better at entraining to speech in noise whereas the bilinguals are better at entraining to speech in quiet, but not significantly so (t(25)=-.631, p=.534.

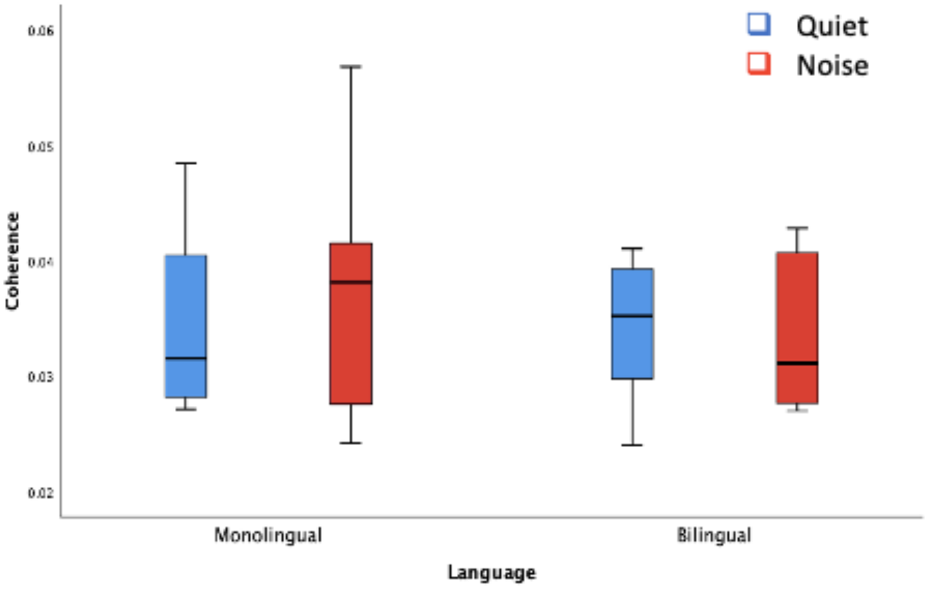


Figure 4-5: Boxplots to show the range in noise (red) and quiet (blue) coherence for bilinguals and monolinguals.

4.4 Discussion

Our results, similar to Study 1, showed reliable evidence for neural entrainment to the amplitude envelope of speech in pre-verbal infants. Coherence was observed in infants when they were listening to the target speech in both quiet and noise conditions, this suggests that infants were able to successfully inhibit the distractive effects of the 10-talker babble that was used in the study. In other words, infants were successfully able to attend to our target speaker when a distractor audio was present in this scenario.

Coherence was not significantly higher to the target in the quiet condition than in the noisy condition and it seemed as though the babies were tracking the speech similarly in both conditions. It was hypothesised that the infants would have stronger coherence to the target speech in the quiet condition than in the +10dB noise condition as typically the degree of neural entrainment declines as the speech intelligibility decreases (Luo & Poeppel, 2007). However, there were no significant differences between levels of coherence in the two speech conditions. The level of noise used in the study (+10dB) may have not been high enough to cause any major acoustic interference. Previous behavioural research has shown that 5 and 9 month old infants are able to attend longer to their own names than other names when the signal to noise ratio is 10dB but not 5dB, whereas 13 month old infants listened longer to their own name at both a 5dB and 10dB signal to noise ratio (Newman, 2005). This present study was slightly more difficult due to the infants' need to attend to continuous speech and not just repeated isolated words. Perhaps the 10dB masker in our study was not disruptive enough to measure any potential neurophysiological effects on infant coherence levels. However, recent EEG research (Song, Martin, & Iverson, 2020) has demonstrated that even though a distractor with a high SNR may not affect perceptive performance for native speakers, large differences in the EEG measures can still be observed. In other words, there are attentional effects occurring that can be measured by EEG even when the distractor audio is not intense enough to affect intelligibility. Thus, even with the high SNR used in this study, it was possible that we would have seen effects in the recorded EEG measures.

There was no effect of bilingualism on coherence levels. Our hypothesis that bilingual infants would elicit stronger neural entrainment to the target speech than monolingual infants was not met. Infants in both language groups (monolingual and bilingual) exhibited similar coherence levels in all conditions. We assumed that bilinguals would have an unconscious advantage in inhibiting a distractor due to the fact that they are constantly suppressing one language while the other is active (Bialystok et al., 2009). We know that bilingual pre-verbal infants display differences in attentional control in comparison to monolinguals (D'Souza et al., 2020; Kalashnikova et al., 2021; Kovács & Mehler, 2009) and so were expecting to see some differences with regards to the processing of the target and the distractor. Bilingual adults have been shown to have more advanced auditory selective attention when listening to both continuous speech-in-noise (Olguin et al., 2019) and pure vs deviant tones (Rämä et al., 2018). However, it seems as though this bilingual advantage with regards to auditory-neural speech coherence is not yet present in infancy. We classified infants as bilingual if they were exposed to another language for at least 25% of their typical daily language exposure. This may not be enough exposure to classify an infant as bilingual. Going forwards, perhaps our definition of bilingualism in infancy needs to be changed to ensure that they receive 50% or more of their daily language exposure in another language. We asked the parents or caregivers of the infants to estimate the amount of English that the infant was exposed to and this was expressed by the main caregiver as a percentage of total daily language exposure as an approximate figure. Perhaps more accurate measures of the infant's daily language exposure could be recorded and analysed in future studies in order to obtain a more coherent picture of infant language learning and the consequences of this on neural development.

Although this study did not find a bilingual advantage with regards to the neural processing of continuous language, it may be because of the young age of the infants. The suppression of one language while the other is activated is typically common in communication as speakers converse with people around them (De Bruin, Roelofs, Dijkstra, & FitzPatrick, 2014). Perhaps infants at the age of 8 months old, as in this present study, are not yet participating in language-inhibitory control and so have not yet acquired bilingual auditory selective attention. Bilingual infants could regularly be tested throughout the later stages of infancy and into early childhood in order to understand at what point bilingualism may have an effect on the neural processing of language, and if this has any consequences on their wider cognitive or social developmental abilities.

The age range in our study was 4 to 11 months old with a mean age of 8.7 months, no significant differences existed between the levels of coherence between the older and younger infants - there was no age effect on coherence in either condition. Newman's (2005) study measured the ability of infants from the age of 5 months old to filter out 10dB background noise, perhaps an infant's ability to selectively attend to speech with this level of background noise occurs in the first months of life. Younger infants should be tested in order to explore at what point this phenomenon develops.

To summarise, the infants in this study were able to demonstrate neural entrainment to the amplitude envelope of speech in both conditions. They were successfully able to suppress a 10dB signal-to-noise ratio distractor in order to attend to the target speech, as well as entraining to the speech in the quiet condition. There were no effects of bilingualism, age or amount of data coherence levels. This study built upon the foundations set in Studies 1 and 2, namely that we could measure neural entrainment in pre-verbal infants, and further investigated the topics of distractor noise and bilingualism. We have discovered that infants are able to inhibit a distractor while attending to a target, but this largely depends on the type and intensity of the distractor.

Our findings bring us one step closer in knowing more about infant language processing and the neural mechanisms that underlie this. This can give us a deeper insight into how infants acquire language in typical every day environments and how attention modulation can affect this. It is important to attempt to understand infant language processing in as naturalistic environments as possible because infants don’t typically learn language in perfect acoustic scenarios without background distractor noise. It seems that we can be quite confident in infants’ ability to attend to a target talker while a distractor is present but this all comes down to their selective attention abilities. Infants are active participants of their language acquisition process, they are not just passive listeners. More research must be conducted on how all of these attentional processes affect their developing brains, and the consequences that this has on speech processing and overall learning development. Future research must investigate more varying levels of background noise in order to understand the full extent of infant cortical tracking and selective attentional abilities.

5. Listening development throughout childhood: An online study

5.1 Introduction

Due to the COVID-19 pandemic, face-to-face testing had to be stopped and research had to be moved online. As a consequence of this, we were no longer able to investigate infant auditory neural tracking to speech in different noise levels with EEG. As part of our lab's approach to the pandemic, we developed a piece of online software (the memory card game) that could be used remotely and in a number of different experiments including second language training. In this present study, the memory card game was used in order to assess speech in noise abilities in young children.

The memory card game required the children to match pars of cards while either a single or babble masker was playing in the background. The children needed to find 7 matching pairs of cards per game trial. Each child would be required to complete 6 trials of a memory card game that varied in terms of masker (single vs multi-talker babble) and cognitive demand: 2 games would have symbols on the cards when they were flipped over, 2 games would have cards that said a word when flipped over, and 2 games would have a combination of symbols and speech when flipped over. This type of game can have a number of ways in which participants can make errors; there may be difficulty hearing the differences in sounds and participants might get 'lucky' depending on their initial card selection.

In order to continue with the developmental theme of my research, I wanted to test children as young as possible. Because of this, it was decided that the phonetic discrimination task involved would be that of single words with a vowel contrast. Typical research with children using a masker use sentence materials as the target stimulus (Baker, Buss, Jacks, Taylor, & Leibolda, 2014; Buss, Leibold, Porter, & Grose, 2017; Song et al., 2020). However, due to wanting to test younger children it was decided that a simpler vowel contrast would be used. This means that the children didn't necessarily have to understand the presented acoustic materials and it was not necessary for the children to track the meaning of the speech in order to process the memory card game. From the pilot trials using this game with four year olds, they seemed to find the memory card game both doable and fun. However, as this is a new measure it is unclear whether if a single vowel contrast would work in the same way as sentence materials do and whether the attentional demands of the game would make it less sensitive to basic speech measures.

Previous research has shown that children from as young as five are able to successfully complete picture-matching and object-matching tasks (Giezen, Escudero, & Baker, 2016) and based on our existing knowledge of children's abilities, we decided that the lower age limit for this task would be 4 years old. This is also the approximate age that children start learning phonics and so are becoming more aware of the fine-grained phonetic distinctions between words. However, this type of experiment with young children is typically conducted in person. This means that children are able to benefit from both explicit instructions and social interaction. One issue of our experiment being online, and thus remote, is that we have no experimental control over the game and it is easier for children to give up or get distracted. However, the main benefit of online testing is that we can gather a much larger scale of online participants. This is particularly useful in a study like this in which we are unsure whether this new game is suitable and sensitive enough to measure differences in phonetic perception. Remote testing allows us to assess a large range of children with different language backgrounds and while using different masker types.

The main goals of the memory card game in this study were to assess speech in noise abilities in children and how this changes with age, if this is sensitive to masker type (single vs babble), and if this is affected by bilingualism. With regards to age, the ability to perceive speech in non-ideal auditory scenarios, such as speech-in-noise, seems to be a difficult process for children to engage with. This phenomenon doesn’t seem to become developmentally mature until approximately ten years of age, perhaps due to improvements in auditory and attentional abilities (Thompson, Woodruff Carr, White-Schwoch, Otto-Meyer, & Kraus, 2017). However, this process is thought to be a gradual rather than a sudden process that develops throughout childhood (Nishi, Lewis, Hoover, Choi, & Stelmachowicz, 2010). In a recent study by Ghinst et al., (2019), children aged between 6-9 were instructed to attend to four different stories with differing levels of background signal-to-noise ratios (noiseless, +5, 0, and -5 dB). Their coherence to the attended speech was significantly lower in comparison to adult coherence in all of the conditions (Ghinst et al., 2019). From this, it was argued that children’s difficulty to understand speech relates to their immature selective attention of the target speech stream. This is concerning when considering that the majority of a child’s learning occurs in noisy environments; school-age children must acoustically attend to their teacher when dealing with a multitude of background noise in a busy classroom. In other words, it seems as though children are posed with typical ‘cocktail party’ (E C Cherry, 1953) challenges in everyday life and this is potentially a barrier for educational and social development. Distracting background noise resulting in attentional shifts and the failure to encode incoming information can cause challenges for academic success in areas such as maths and history in which the content is primarily orally delivered by the teacher (Erickson & Newman, 2017).

In this study we also wanted to investigate whether performance on the memory game is affected by masker type. With regards to masker type in this study, it was decided that a babble vs a single talker masker would be used. It has been shown that children show worse performance in comparison to adults on a wide range of masked speech perception tasks, but this effect is particularly significant when the masker itself is also composed of speech (Corbin, Bonino, Buss, & Leibold, 2016). It has been argued that the ability to perceive speech effectively when in the presence of steady-state background noise develops at approximately the age of 10, (McCreery & Stelmachowicz, 2011) whereas the perception of speech amongst background speech can be delayed up until the age of 16 (Wightman & Kistler, 2005). In addition, research has shown that differences exist between one-talker maskers and two or multi-talker maskers in terms of how effective they are in masking target speech. Single-talker maskers have shown to be less effective in masking speech than a two-talker masker (Buss, Leibold, Porter, & Grose, 2017) and while some studies have found evidence that a two-talker masker is maximally disruptive when attending to a target speaker (Freyman, Balakrishnan, & Helfer, 2004).

Single-talker and babble maskers place different demands on auditory processing. A single talker masker has a natural amplitude fluctuation but their linguistic content can cause additional interference. They often cause more difficulties in auditory organisation due to resembling the target more similarly than babble (Brungart, 2001). Recent research (Song et al., 2019) illustrated that speech is processed differently when there is a single talker masker present in comparison to unintelligible babble at a neurophysiological level. Results showed that a single-talker masker makes it more difficult to understand the target speech at a lexical-semantic level and that this affects the ability of a listener to track the speech acoustics. This may mean that the single-talker distractor in the present study may be more distractive in comparison to the babble and consequently negatively affect performance on the game.

We also wanted to investigate whether performance on the memory card game was affected by language. We included both monolinguals and bilinguals in this present study in order to investigate if any potential differences between the two exist. Previous research investigating the effects of bilingualism on speech intelligibility (Reetzke, Lam, Xie, Sheng, & Chandrasekaran, 2016) found no differences between monolinguals or bilinguals on recognising speech across a number of masker types (pink noise, steady-state and two-talker babble). This may indicate that we will not observe any language effects on task performance.

In this study, 73 children aged between 4 to 13 years old remotely participated in the online memory card game. We wanted to explore if the memory card game was a useful tool for assessing speech in noise abilities in young children, how performance on the game changes with age and if this is affected by masker type or bilingualism. Alongside this new memory card game, the children also completed a more traditional phonetic perception game (three-interval oddity task) as a control measure. This is a more standardised task that has been used in research previously (Halliday, Tuomainen, & Rosen, 2017; Moore, Ferguson, Edmondson-Jones, Ratib, & Riley, 2010). Our pilot study established that the young children found the three-interval oddity task tedious in comparison to the memory card game. This perhaps indicates that the latter is a more stimulating phonetic perception game for children. However, the memory card game requires a higher memory load and has a greater amount of potential errors that may occur during the testing process.

5.2 Methodology

5.2.1 Participants

A total of 73 children participated in all of the tasks involved in the experiment. The mean age of the children was 7.1 years old (SD = 2.641), ranging from 4 years to 13 years old. Out of these participants 29 were bilingual and 44 were monolingual English speakers. It was required that the bilingual speakers were exposed to English in the household or school that they attend, they must also experience at least 25% of their daily linguistic exposure to 1 other language. Additional languages included: Italian (n=5), Spanish (n=4), Polish (n=4), French (n=3), Punjabi (n=3), Russian (n=3), Mandarin (n=3), Albanian (n=2) and Portuguese (n=2). The children were mainly recruited through Facebook and by word-of-mouth. If a parent was interested in the study then they would email me and I would provide them with more information and then send them the relevant materials. Children could not participate if they had any hearing or visual problems, any developmental or learning problems or any neurological disorders; parents were required to confirm that their child met the eligibility criteria for the study before participation.

5.2.2 Stimulus

As part of our lab’s approach to the pandemic, we developed a piece of software platform in Javascript for the memory card game that the lab has used in a number of studies including second language training. Here, we are using this same software in order to assess speech in noise abilities in young children, understand if this measure is sensitive to masker type, and if this is affected by bilingualism.

The consent forms, parental questionnaire and cognitive assessment task (Coloured Progressive Ravens Matrices) were all created and issued on Gorilla. The control study (three-interval oddity task) consisted of 36 trials with breaks every 12, half of the trials had a single-talker background distractor and the other half had a multi-talker babble background distractor. The child must ignore the background noise and listen to the three frogs each say a word and then click the frog that said a different word to the other two. All of the fourteen words heard in the study were in the controlled phonological environment of /bVt/: "bet", "bart", "bat", "but", "beat", "bit", "bite", "bait", "boot", "bout", "bert", "bot", "boat", "bought". For example, the child would hear “bet” from the first frog, “but” from the middle frog and “bet” from the last frog, and thus the child would need to click the middle frog in order to be correct and have picked out the ‘odd one out’ of the three words. Each frog’s word would also be produced by a different speaker (out of a possibility of 10) and so the child needed to focus on the difference in the vowel sound irrespective of gender and accent: the two same stimuli that the child hears are thus not identical.

The memory card game consisted of 6 games, the order of which was randomised for each child. The background distractor noise differed between the games; in games 1-3 the background noise was made up a multi-talker babble whereas in games 4-6 the background noise was that of a single speaker. The aim of each game was to match 14 cards into 7 correct pairs. The games varied in difficulty with regards to the card matching requirements: In games 1 and 4 (condition 1) the cards had to be matched on symbols, in games 2 and 5 (condition 2) the cards had to be matched on both the symbol and the sound heard when turning over the card, and in games 3 and 6 (condition 3) there were no symbols and the cards could only be matched on the sound heard when turning over the card. This would be the most difficult as the children were only relying on the auditory cues here.

5.2.3 Procedure

Parents were emailed with a set of instructions and unique individualised computer links in order for their child to participate in each task. Each task had its own separate link which meant that all tasks did not have to be completed in one sitting. This was beneficial for the younger children who would get tired or distracted easily. The first link took the parent to a series of online forms to complete on Gorilla. All parents were required to read and fill in a consent form and confirm that their child did not fall under the exclusion criteria before their child participated in the tasks. The parents were then required to complete an information sheet about their child that included information such as date of birth and details of language exposure. Immediately following this, the child was required to engage in the Coloured Progressive Ravens Matrices task which was made up of 1 practice trial set and 3 actual trial sets. This is a widely used cognitive assessment task and was used in order to ensure that the children in our study were able to make sense and patterns out of confusing, non-verbal data. The next link took the child to a three-interval oddity task. The children were given the instructions that they needed to ignore the background noise while listening to the three frogs, each frog would say a word and the child would have to pick which one said a different word to the others. This task was made up of 36 trials with breaks every 12. The final task that the children had to participate in was a memory card game. The children were instructed to find the matching pairs of cards and that some pairs would be determined by sounds and others by pictures, they were again instructed to ignore the background noise. There were 6 games for each child to complete, each game had 14 cards and thus 7 pairs to make, and the games varied in difficulty. The games were scored with regards to the number of clicks taken to complete them. The lower the number of clicks is, the better the score is. For example, a child who completed the memory card game in 20 clicks would have performed better on the task than a child who completed the memory card game in 50 clicks. A high score number may also be indicative of the child randomly pressing the screen in order to find the correct answer, and thus not fully understanding the rules of the game.

5.3 Results

The following boxplots (Figures 5-1 and 5-2) show the relationship between memory card condition type and the numbers of moves taken and the time elapsed to complete the game. Condition type is as follows: 1 is matched on symbols, 2 is matched on both symbols and speech sounds and 3 is matched on only speech sounds. This allows us to investigate performance on the task with and without a language element. With regards to number of moves taken (number of clicks on the screen), it can be clearly seen here that condition 3 (the condition in which children must match the pairs on the basis of sound only) causes the highest amount of difficulty with one child taking more than three times as many moves than the average moves of conditions for one and two to complete. The average moves for each condition were calculated across both trial types (multi-talker and single distractor babble) in order to first observe specific differences relating to condition type. As the tests were taken by the children remotely we cannot be sure that the children were not just randomly clicking cards in order to find pairs, but on average we can see that the memory card game with solely a language element caused the children the most confusion. When looking at the time elapsed to complete the game, we can also see that condition 3 takes the children on average a longer time (approximately 20 seconds) to complete. Again, potential discrepancies in time taken may exist due to children taking breaks or being distracted in the home.

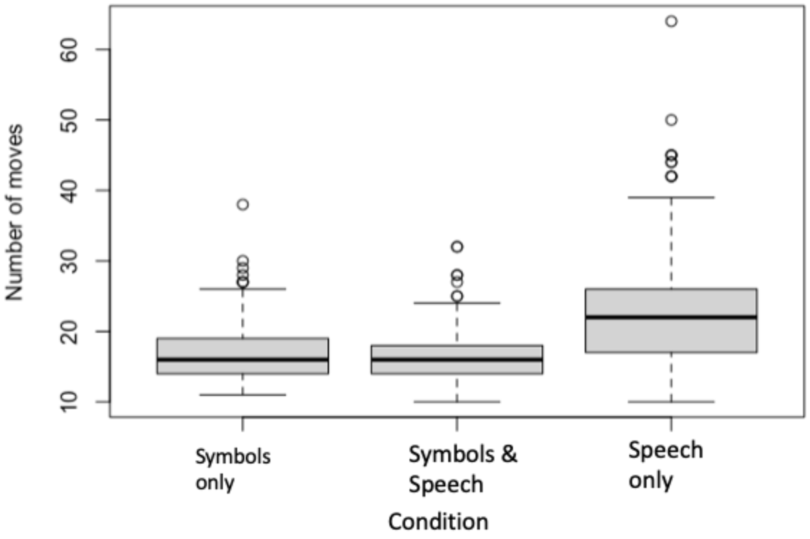


Figure 5-1 Boxplots to show the relationship between conditions and number of moves taken to complete the memory card game.

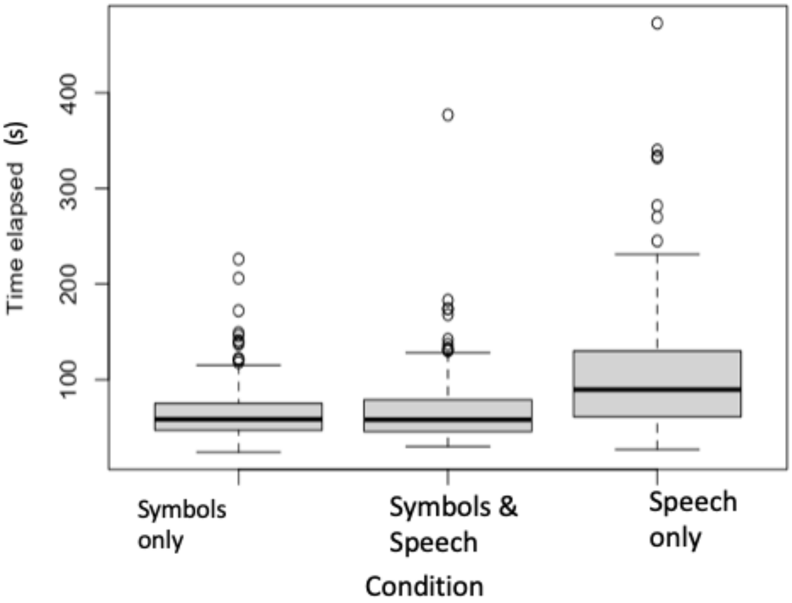


Figure 5-2 Boxplots to show the relationship between condition and time elapsed (in seconds) to complete the memory card game.

Figure 5-3 shows the relationship between the average number of moves made per game and the age in years. Here the black filled in points represent a single talker distractor while the unfilled points represent the multi-talker babble.

From a visual inspection of the data we can see that the older children on average take less moves to complete the game which reflects an overall better performance on the game. The relationship between age, number of moves and condition type is illustrated in Figure 5-4. Here, we can see a gradual decrease in number of moves taken to complete the game associated with an increase in age. The decrease in number of moves is greatest for condition 3 in which the cards had to be matched on sounds alone.

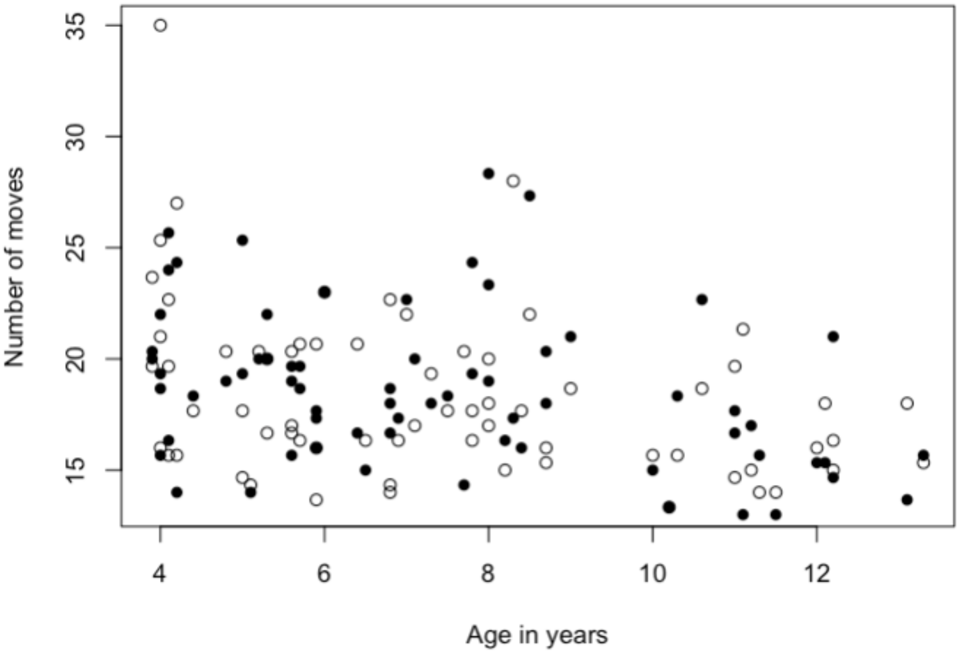


Figure 5-3 A scatterplot to show the relationship between age in years, number of moves taken to complete the memory card game and noise type (black filled in points represent a single-talker whereas the unfilled points represent the multi-talker babble

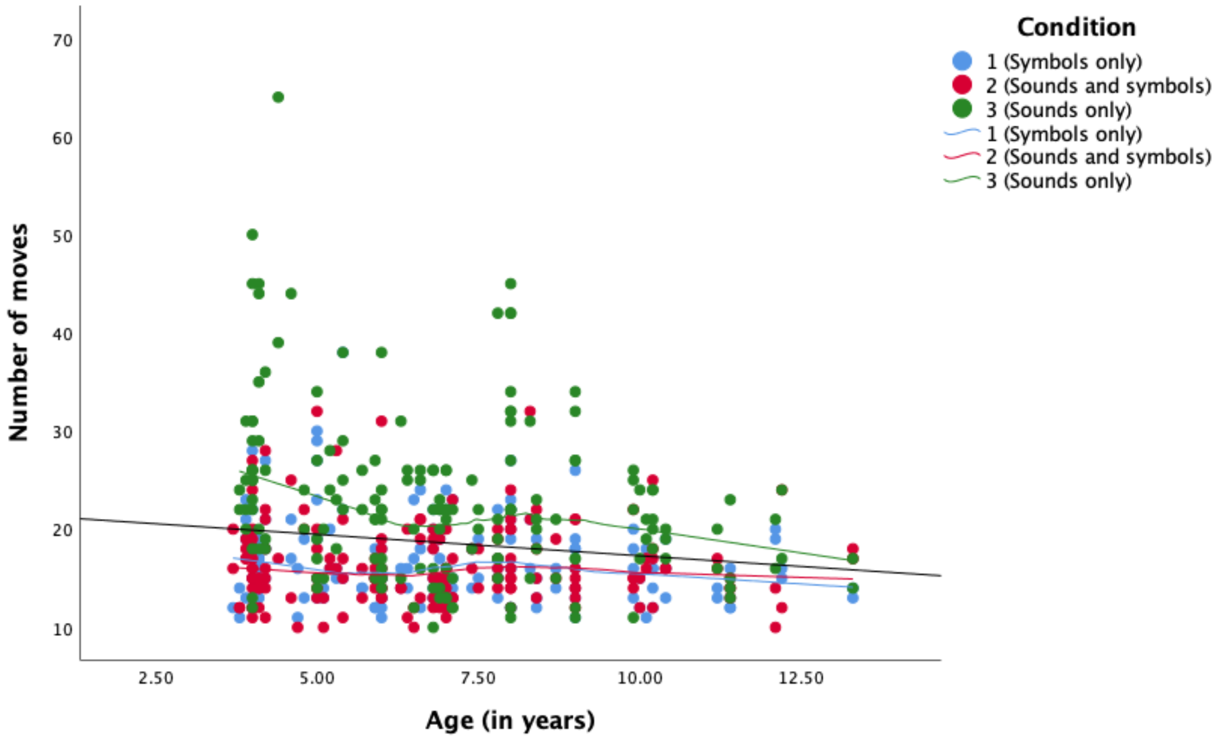


Figure 5-4 A scatterplot to show the relationship between number of moves, age in years and condition type.

Linear mixed model regressions were run in order to investigate whether there were any significant effects of language, task condition, noise type and age on number of moves taken and time elapsed to complete the task, and if any interactions existed between these variables. With regards to moves taken, significant effects were found for condition, χ^2 (2) = 153.7401, $p < 0.001$, and age, χ^2 (1) = 13.7794, $p < 0.001$, and these two variables interacted, χ^2 (2) = 15.8431, $p < 0.001$. This means that the older children on average had a greater performance on one of the conditions in terms of number of moves taken to complete the game. However, no effects were found for masker type, χ^2 (1) = 0.0938, $p = 0.422$, or language, χ^2 (1) = 0.0028, $p = 0.95$. Similarly, for time elapsed significant effects were found for condition, χ^2 (2) = 150.5747, $p < 0.001$ and age, χ^2 (1) = 47.9849, $p <$

0.001, and these two variables interacted, χ^2 (1) = 27.7949, $p < 0.001$. Again, this means that the older children on average took less time to complete the trial under a certain condition than the younger children. However, there were no effects here of masker type, χ^2 (1) = 0.0938, p = 0.759, or language, χ^2 (1) = 0.0037, p = 0.95. With regards to the improvement with age, this could be a speech effect or just to do with general improvements in attention and cognition. However, the interaction between condition and age here specifically concerns condition 3 (the game in which the pairs had to be matched solely on the speech component) which was on average the most difficult condition in the game. The game in which the cards had to be matched on symbols did not change with age. However, condition 3 in which the cards had to be matched only on speech sounds did improve with age which may illustrate a developmental improvement in speech perception abilities. No significant effect of noise type was found, indicating no differences in interference between a single and multi-talker babble, and so this will not be investigated any further.

Variable	**ChiSq**	**Df**	**Pr(>Chisq)**
Language	0.002	1	0.9575571
Condition	153.7401	2	<2.2e-16
Noise	0	1	0.4225171
Age	13.70.4225171794	1	0.0002056
Language:Condition	4.2706	2	0.1182073
Language:Noise	1.8327	1	0.1758101
Condition:Noise	2.7841	2	0.2485704
Language:Age	0.0177	1	0.8941852
Condition:Age	15.8431	2	0.0003628

Noise:Age	0.4250	1	0.5144409
Language:Condition:Noise	0.8735	2	0.6627407
Language:Condition:Age	0.1216	2	0.9409886
Language:Noise:Age	0.0188	1	0.8909816
Condition:Noise:Age	0.4088	2	0.8151565
Language:Condition:Noise:Age	0.8227	2	0.6627407

Table 5-1: Linear mixed effects model output for the number of moves taken to complete the game.

Variable	**ChiSq**	**Df**	**Pr(>Chisq)**
Language	0.0037	1	0.9515
Condition	150.5747	2	<2.2e-16
Noise	0.0938	1	0.7594
Age	47.9849	1	4.295e-12
Language:Condition	3.6900	2	0.1580
Language:Noise	0.7846	1	0.3757
Condition:Noise	2.3933	2	0.3022
Language:Age	0.0378	1	0.8458
Condition:Age	27.7949	2	9.213e-07
Noise:Age	0.0781	1	0.7798
Language:Condition:Noise	1.0227	2	0.5997
Language:Condition:Age	1.3012	2	0.5217
Language:Noise:Age	0.1980	1	0.6564
Condition:Noise:Age	0.5472	2	0.7606
Language:Condition:Noise:Age	0.4961	2	0.7803

Table 5-2: Linear mixed effects model output for the time taken to complete the game

Taken together, these findings illustrate that age has a positive effect on game performance with the older children performing better on all of the conditions, specifically condition 3. There were no effects of masker type. In other words, performance did not vary significantly depending on if the masker was a single-talker or babble. Additionally, there were no significant effects of language illustrating that performance did not vary greatly depending on if the children were monolingual or bilingual. However, the number of monolinguals and bilinguals in the study were unequal with 15 more monolingual speakers which could have affected this result.

In addition to this, we looked at the results from the three-interval oddity control study. Figure 5-4 illustrates the positive relationship between age and the percentage of frogs correctly identified. The children overall seem to become better at identifying the correct odd frog out as they became older. However, even children from the age of 4 were able to correctly identify this in over 50% of trials. In this figure we can see the relationship between age, language and percentage of frogs correctly identified. Again, no large discrepancies exist between monolingual and bilingual performance on the task. A multiple linear regression was run in order to investigate whether there were any significant effects of language or age on the average number of trials correct in the frog game, and if any interactions existed between these variables. No significant effects were found for language ($p= 0.59$). However, a significant effect was found for age ($p<0.001$). This main effect of age is illustrated in Figure 5-4. It seems that performance on the memory card game improved with age. More specifically, the memory card game in which the cards were matched purely on sounds improved the most with age illustrating that this is potentially a result of enhanced speech perception abilities.

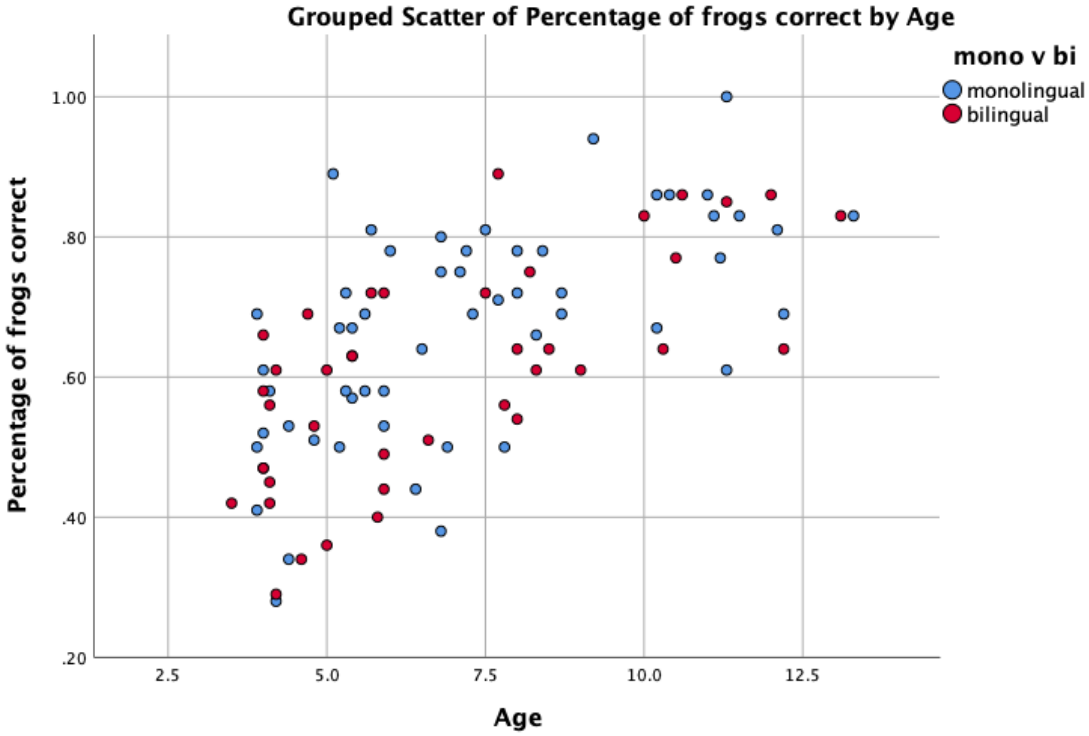


Figure 5-5 A scatterplot to show the relationship between percentage of frogs correctly identified in the three-interval oddity task and age. Monolingual children are represented by blue and bilingual children by red.

5.4 Discussion

This study aimed to use a piece of online software, that had been recently developed in our lab, in order to assess the phonetic perception of vowel contrasts in young children. We wanted to explore if the newly developed memory card game was both suitable and sensitive enough to measure phonetic perception in young children remotely. In addition to this, we were interested in exploring whether performance on the memory card game improves with age and if it is affected by masker type (single-talker vs babble) or language (monolingual or bilingual participants). Results showed that performance on the memory card game improved with age but was not affected by masker type or language background. The main effect of age illustrated that the older children were better across all conditions, specifically condition 3. Young children's performance on the task seemed to be affected more when the task was language related and required a higher memory load. It could be argued that this improvement with age may be a result of children's increased general attention and memory abilities. However, the increased ability of the older children to perform better on the sounds only condition seems to show a more general improvement in speech perception rather than just an overall memory or maturational effect.

As previously mentioned, perceiving speech in non-ideal auditory scenes is a gradual developmental process that doesn't seem to become fully mature until approximately ten years of age (Thompson et al., 2017). Our experiment included participants ranging from 4 to 13 years old and, as the previous literature has suggested, saw a gradual improvement (Nishi et al., 2010) in game performance with age. We did not see any sensitivity regarding masker type in the game. Previous research has suggested that speech is processed differently in the brain when there is a single talker masker present in comparison to unintelligible babble (Song et al., 2020). If we were to run EEG while the children were engaged with the task then perhaps we would have seen some neural processing differences. Additionally, we did not find any effects of language. Bilinguals and monolinguals performed similarly on the memory card game. This supports previous research that did not find any effect of language when recognising speech in varying masker types (Reetzke et al., 2016). However, 15 more monolinguals participated in the study than bilinguals which may have affected our results.

The memory card game is overall a useful tool for measuring phonetic perception in young children. Very young children, from the age of four, were able to complete the game and found it enjoyable. This is a slightly younger age than typically used in child phonetic perception tasks. However, standard tasks normally take place in person whereas our experiment had to take place remotely online. This posed a number of potential problems. As we were unable to monitor the experiment it is unclear if the caregivers offered the child any assistance in completing the task. In addition, number of moves may have been influenced by children randomly hitting the screen and the time taken to complete the game may have been influenced by children getting bored and leaving the game to go play elsewhere. Going forwards, the memory card game is a useful tool that could be included in a battery of phonetic perception tasks on young children. However, the most accurate results would accrue if the task could be monitored in-person by an experimenter in order to account for any discrepancies in time or number of moves taken to complete the memory card game.

6. General Discussion

The main aim of this thesis was to investigate whether pre-verbal infants in the first year of life are engaging with the auditory-neural tracking of continuous speech, how this can be modulated by factors such as language and attention and understand the wider role that this plays in their language developmental process.

For the first study *Infant neural entrainment to continuous speech: Initial methodological development* we wanted to see if we could measure infant neural entrainment to IDS continuous speech in 36 infants. This was something that at the time had only been reliably been found in adults and a limited number of infant research (Kalashnikova, Peter, Di Liberto, Lalor, & Burnham, 2018; Leong et al., 2017). We also played these infants a concatenated vowel sequence with the aim of eliciting the ACC response, as this is something that has been more reliably measured in infants in previous research and is measures a more general auditory response (Chen & Small, 2015; McCarthy et al., 2019; Uhler, Hunter, Tierney, & Gilley, 2018). We found reliable coherence by the infants to both continuous speech and the ACC. The two coherence measures were independent of each other yet correlated which this supports the conclusion that our overall story coherence measures are reliable and robust. However, differences were apparent with regards to the latencies of the two responses. This could potentially be taken as evidence to show that the infants really are processing speech differently to the ACC, and that their speech processing is more than just an auditory response.

The backwards multivariate temporal response functions (mTRFs; Crosse, Di Liberto, Bednar, & Lalor, 2016) were an efficient mechanism for analysing infant neural responses to auditory continuous speech by reconstructing the auditory stimulus based on the multi-channel, time varying neural response. This signal reconstruction technique has successfully been used before in infant entrainment studies (Jessen et al., 2019; Kalashnikova et al., 2018). This model was used again in studies two and three and could be used in other infant entrainment research going forwards as a successful toolbox for measuring the cortical tracking of continuous speech.

Tracking the acoustic envelope of continuous speech by neural oscillations has been shown in adult research to be involved with the successful encoding and thus the processing of natural continuous speech (Ghitza, 2011; Gross et al., 2013; Peelle, Gross, & Davis, 2013). From the evidence presented in this thesis it can be argued that pre-verbal infants, from as young as four months old, are also engaging in this process. The ability of infants to track the amplitude envelope of speech despite not yet being able to engage with the linguistic meaning of speech may be working as a potential bootstrapping mechanism for the language acquisition process. Infants are active participants of their language acquisition process and learn most efficiently through social interaction. As a consequence of this, their attention modulation can have a significant impact on their processing of speech and thus the developmental process of acquiring their native language.

Language assessments were undertaken on the children from the first study approximately two years after their initial EEG. However, no significantly predictable variables were found on language developmental outcomes. Despite this finding, we can still assume that infant engagement with auditory-neural coherence could perhaps be acting as a potential bootstrapping mechanism for later language development. The language assessments we administered were not perfect, we may not have been able to gauge the child's full potential as the tests had to be administered on Zoom or in a socially distanced manner due to the COVID-19 pandemic. Similarly, we used a modified BPVS scoring system due to some of the children falling beneath the standardised age range for the test. Future research should look into how later language, and more general cognitive abilities, can be related to coherence levels in the first year of life. This will allow us to cast important new light on the developmental role that neuronal oscillations may have in the first year of life with regards to acquiring both language and more general cognitive capabilities.

Study two *Neural entrainment and auditory selective attention in infants: A two-talker study* used the same measures as study one in order to build upon these findings. For this study, we wanted to explore how infants are able to process speech in more complex auditory environments. We played 20 infants an audio-visual film lasting approximately 20 minutes - the infants were presented with a series of engaging video clips that were overlaid with matching audio from a separate clip. We wanted to explore whether the infants were able to engage in neural entrainment to the amplitude envelope of the target audio stream while inhibiting the distractor audio. To our knowledge, no other EEG research has attempted to investigate how, and if, infants are able to inhibit a distractor when attending to a target speech. The results from this study showed that perhaps these listening conditions were simply too difficult for the infants to follow. It seemed as though infant attention was fluctuating between the two speech signals and consequently not successfully selectively attending to a target. However, previous behavioural research had shown that infants can only inhibit a distractor when the target is more acoustically intense (Newman & Jusczyk, 1996) and thus for the next study we decided we would make the listening conditions slightly easier.

For study 3, *Neural entrainment and auditory selective attention in infants: The effects of noise and bilingualism on speech perception* we wanted to explore how neural entrainment to continuous speech could be modulated by infant selective attention when the speech was in a less acoustically intense background noise and if language (monolingual vs bilingual) had any effects on this. We played 30 infants the same audio-visual film as in study 2. However, the audio-visual film was now comprised of two listening conditions: no-interference and 10dB signal-to-noise babble in order to see if they would still be able to attend to the target engaging speech while a less intense background distractor was present. A signal-to-noise ratio of 10dB was chosen due to evidence from previous behavioural research that suggests that infants are able to attend to a target when the distractor is set at 10dB but not 5dB (Bernier & Soderstrom, 2018; Newman, 2005; Newman & Jusczyk, 1996). Similar coherence values to the amplitude envelope of the target speech were observed across all infants when they listened to speech with no interference and speech masked with noise. This illustrates that infants are able to inhibit a distractor that is 10dB lower in intensity than the target noise. There were no effects of bilingualism on the ability to inhibit a distractor, and no variables were significant predictors of entrainment. Arguably the biggest limitation of this study was that perhaps the level of noise was not enough in order to perceive a significant difference in entrainment levels between the two conditions. Infants need to be presented with more varying distractor noise levels in order to fully understand the extent of their selective attention and acoustic inhibitive abilities. This was the initial plan for study four, however due to the COVID-19 pandemic we were no longer able to test infants in-person and research needed to be moved online.

Thus, in study four *Listening development throughout child: An online study* we investigated how a phonetic perception task interacts with distractor type and how this changes with age during childhood. Children aged between 4-13 years old were required to participate in an online memory card game. Results showed an interaction between condition and age; the younger children only performed significantly worse than the older children when the cognitive task had a language element showing that the improvement with age is a result of improved speech perception capabilities. Although this study was not part of the initial thesis plan, it was still able to provide interesting results with regards to how performance on a cognitive task is affected by background noise and how this develops throughout childhood. Obvious limitations existed due to the remote nature of the experiments; the children couldn’t be monitored while undertaking the experiment and so aspects of the study, like the time taken for a child to complete the experiment, may be unreliable.

With regards to the whole thesis, limitations will always exist when conducting infant EEG; the data is typically noisy and infants cannot follow explicit instructions which makes it hard to measure their engagement and selective attention. Future infant EEG work should always aim to keep the babies as happy and as comfortable as possible while providing an engaging stimulus in order to maintain infant attention. Due to the COVID-19 pandemic our infant EEG work was cut short, our fourth study would have used different noise levels in order to gain a better understanding of how infants inhibit a distractor when attending to a target speaker. This should be done in future research and could be done again by using audio-visual stimulus or in-person entrainment to an experimenter or caregiver. Measuring different levels of speech-in-noise with different signal-to-noise ratios is crucial in understanding infants' threshold of acoustic attention in noisy environments. This will allow us to apply our findings to everyday learning environments, such as nurseries and playgroups, in order to explore how varying levels of background noise may impact a child's speech processing abilities. Again, language background should be explored here. We failed to find any significant difference between monolingual and bilingual's coherence levels in this thesis, but this may be a processing difference that surfaces in certain noise environments or at a certain developmental stage. Additionally, from our parental questionnaires across all of the studies, we discovered that the vast majority of parents and caregivers whose infants participated in the research were all educated to at least undergraduate degree level. Future research must aim to be socially inclusive of infants growing up in a range of backgrounds in order to be representative of a wider population of typically developing infants.

To summarise, this thesis has demonstrated that infants are engaging with the cortical tracking of continuous speech, and presents some evidence that they are able to selectively attend to target stimuli in certain listening environments. The techniques presented in this thesis have shown to be useful for measuring infant neural entrainment to continuous speech. However, we are only just beginning to understand the effects of infant attention modulation on this phenomenon. Going forwards, the techniques employed in this study could be used in future research in order to continue the line of research presented in this thesis in terms of the infant auditory-neural tracking of speech and how this can be modulated by infant attention. This thesis has provided new and exciting research into the mechanisms of this process in the early years. However, this is only just the beginning. Our findings illustrate the important yet complex relationship between auditory-neural entrainment, selective attention and language processing abilities in the first year of life. It is crucial that future research investigates this further in order to understand even more about the neural mechanisms underlying language acquisition in infancy. The more we know about what is happening in the infant brain from birth, the more we can optimise learning environments that provide infants with the best start to life and start them on a successful journey for language acquisition.

7. References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.201400998

Anderson, S., Chandrasekaran, B., Yi, H. G., & Kraus, N. (2010). Cortical-evoked potentials reflect speech-in-noise perception in children. *European Journal of Neuroscience*, *32*(8). https://doi.org/10.1111/j.1460-9568.2010.07409.x

Astheimer, L. B., & Sanders, L. D. (2012). Temporally selective attention supports speech processing in 3- to 5-year-old children. *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/j.dcn.2011.03.002

Baker, M., Buss, E., Jacks, A., Taylor, C., & Leibolda, L. J. (2014). Children's perception of speech produced in a two-talker background. *Journal of Speech, Language, and Hearing Research*, *57*(1). https://doi.org/10.1044/1092-4388(2013/12-0287)

Baltzell, L. S., Horton, C., Shen, Y., Richards, V. M., D'Zmura, M., & Srinivasan, R. (2016). Attention selectively modulates cortical entrainment in different regions of the speech spectrum. *Brain Research*. https://doi.org/10.1016/j.brainres.2016.05.029

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2). https://doi.org/10.1016/j.cognition.2004.06.001

Bernier, D. E., & Soderstrom, M. (2018). Was that my name? Infants' listening in conversational multi-talker backgrounds. *Journal of Child Language*. https://doi.org/10.1017/S0305000918000247

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009a). Bilingual minds. *Psychological Science in the Public Interest, Supplement*. https://doi.org/10.1177/1529100610387084

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009b). Bilingual minds. *Psychological Science in the Public Interest, Supplement*. https://doi.org/10.1177/1529100610387084

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*. https://doi.org/10.1111/j.1467-7687.2004.00351.x

Bregman, A. S., Abramson, J., Doehring, P., & Darwin, C. J. (1985). Spectral integration based on common amplitude modulation. *Perception & Psychophysics*, *37*(5). https://doi.org/10.3758/BF03202881

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, *10*(1). https://doi.org/10.1016/s0095-4470(19)30909-x

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3). https://doi.org/10.1121/1.1345696

Buss, E., Leibold, L. J., Porter, H. L., & Grose, J. H. (2017). Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4979936

Byers-Heinlein, K., Esposito, A. G., Winsler, A., Marian, V., Castro, D. C., & Luk, G. (2019). The Case for Measuring and Reporting Bilingualism in Developmental Research. *Collabra: Psychology*, *5*(1). https://doi.org/10.1525/collabra.233

Byers-Heinlein, K., Schott, E., Gonzalez-Barrero, A. M., Brouillard, M., Dubé, D., Jardak, A., … Tamayo, M. P. (2020). MAPLE: A Multilingual Approach to Parent Language Estimates. *Bilingualism*, *23*(5). https://doi.org/10.1017/S1366728919000282

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

Cantiani, C., Ortiz-Mantilla, S., Riva, V., Piazza, C., Bettoni, R., Musacchia, G., … Benasich, A. A. (2019). Reduced left-lateralized pattern of event-related EEG oscillations in infants at familial risk for language and learning impairment. *NeuroImage: Clinical*. https://doi.org/10.1016/j.nicl.2019.101778

Cheek, D., & Cone, B. (2020). Evidence of Vowel Discrimination Provided by the Acoustic Change Complex. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000809

Chen, K. H., & Small, S. A. (2015). Elicitation of the Acoustic Change Complex to Long-Duration Speech Stimuli in Four-Month-Old Infants. *International Journal of Otolaryngology*. https://doi.org/10.1155/2015/562030

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., …

Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*. https://doi.org/10.1016/0378-5955(94)00164-L

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*. https://doi.org/10.1038/1561

Cherry, E. Colin. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1907229

Cherry, E C. (1953). Some experiments on the recognition of speech, with one and with 2 ears. *JASA*. https://doi.org/10.1121/1.1907229

Cirelli, L. K., Spinelli, C., Nozaradan, S., & Trainor, L. J. (2016). Measuring neural entrainment to beat and meter in infants: Effects of music background. *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2016.00229

Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000000201

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2016.00604

d'Apice, K., & von Stumm, S. (2020). The Role of Spoken Language and Literacy Exposure for Cognitive and Language Outcomes in Children. *Scientific Studies of Reading*, *24*(2). https://doi.org/10.1080/10888438.2019.1641505

D'Souza, D., Brady, D., Haensel, J. X., & D'Souza, H. (2020). Is mere exposure enough? The effects of bilingual environments on infant cognitive development. *Royal Society Open Science*, *7*(2). https://doi.org/10.1098/rsos.180191

De Boer, B. (2005). Infant-directed speech and the evolution of language. *Language Origins Perspectives on Evolution*.

de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a Scaffold for language development. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.00044

Decasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*. https://doi.org/10.1126/science.7375928

Decasper, A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology*, *17*(5). https://doi.org/10.1002/dev.420170506

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/s1366728902003012

Dimitrijevic, A., Michalewski, H. J., Zeng, F. G., Pratt, H., & Starr, A. (2008). Frequency changes in a continuous tone: Auditory cortical potentials. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2008.06.002

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1205381109

Ding, N., & Simon, J. Z. (2013). Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.5297-12.2013

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X00058027

Donoghue, J. P., Sanes, J. N., Hatsopoulos, N. G., & Gaál, G. (1998). Neural Discharge and Local Field Potential Oscillations in Primate Motor Cortex During Voluntary Movements. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1998.79.1.159

Doyle, A. B. (1973). Listening to distraction: A developmental study of selective attention. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/0022-0965(73)90134-3

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale (BPVS-II) 2nd ed. NFER-Nelson Publishing Company*.

Ellis, C. T., & Turk-Browne, N. B. (2018). Infant fMRI: A Model System for Cognitive Neuroscience. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2018.01.005

Erickson, L. C., & Newman, R. S. (2017). Influences of Background Noise on Infants and Children. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721417709087

Farroni, T., Menon, E., & Johnson, M. H. (2006). Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child*

Psychology. https://doi.org/10.1016/j.jecp.2006.08.001

Ghinst, M. Vander, Bourguignon, M., Niesen, M., Wens, V., Hassid, S., Choufani, G., … De Tiège, X. (2019). Cortical tracking of speech-in-noise develops from childhood to adulthood. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1732-18.2019

Giezen, M. R., Escudero, P., & Baker, A. E. (2016). Rapid learning of minimally different words in five- to six-year-old children: Effects of acoustic salience and hearing impairment. *Journal of Child Language*, *43*(2). https://doi.org/10.1017/S0305000915000197

Goswami, U. (2019). Speech rhythm and language acquisition: an amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.14137

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech - A syllable-centric perspective. *Journal of Phonetics*. https://doi.org/10.1016/j.wocn.2003.09.005

Gruber, T., Tsivilis, D., Giabbiconi, C. M., & Müller, M. M. (2008). Induced electroencephalogram oscillations during source memory: Familiarity is reflected in the gamma band, recollection in the theta band. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.2008.20068

Halliday, L. F., Tuomainen, O., & Rosen, S. (2017). Auditory processing deficits are sometimes necessary and sometimes sufficient for language difficulties in children: Evidence from mild to moderate sensorineural hearing loss. *Cognition*. https://doi.org/10.1016/j.cognition.2017.04.014

Hoonhorst, I., Serniclaes, W., Collet, G., Colin, C., Markessis, E., Radeau, M., & Deltenre, P. (2009). N1b and Na subcomponents of the N100 long latency auditory evoked-potential: Neurophysiological correlates of voicing in French-speaking subjects. *Clinical Neurophysiology*, *120*(5). https://doi.org/10.1016/j.clinph.2009.02.174

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00251.2010

Jessen, S., Fiedler, L., Münte, T. F., & Obleser, J. (2019). Quantifying the individual auditory and visual brain response in 7-month-old infants watching a brief cartoon movie. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2019.116060

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology*. https://doi.org/10.1006/cogp.1995.1010

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language*. https://doi.org/10.1093/jmp/20.3.233

Kalashnikova, M., Pejovic, J., & Carreiras, M. (2021). The effects of bilingualism on attentional processes in the first year of life. *Developmental Science*, *24*(2). https://doi.org/10.1111/desc.13011

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018).

Infant-directed speech facilitates seven-month-old infants’ cortical tracking of speech. *Scientific Reports*. https://doi.org/10.1038/s41598-018-32150-6

Kannass, K. N., Oakes, L. M., & Shaddy, D. J. (2006). A longitudinal investigation of the development of attention and distractibility. *Journal of Cognition and Development*. https://doi.org/10.1207/s15327647jcd0703_8

Keller, A. S., Payne, L., & Sekuler, R. (2017). Characterizing the roles of alpha and theta oscillations in multisensory attention. *Neuropsychologia*, *99*. https://doi.org/10.1016/j.neuropsychologia.2017.02.021

Key, A. P., Gustafson, S. J., Rentmeester, L., Hornsby, B. W. Y., & Bess, F. H. (2017). Speech-processing fatigue in children: Auditory event-related potential and behavioral measures. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/2016_JSLHR-H-16-0052

Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A. E., Nourski, K. V., Gander, P. E., … Petkov, C. I. (2017). Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.2000219

Kovács, Á. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16). https://doi.org/10.1073/pnas.0811323106

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5). https://doi.org/10.1038/nn.2303

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature*

Reviews Neuroscience. https://doi.org/10.1038/nrn1533

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2007.2154

Kuipers, J. R., & Thierry, G. (2015). Bilingualism and increased attention to speech: Evidence from event-related potentials. *Brain and Language*. https://doi.org/10.1016/j.bandl.2015.07.004

Lakatos, P., Schroeder, C. E., Leitman, D. I., & Javitt, D. C. (2013). Predictive suppression of cortical excitability and its deficit in schizophrenia. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0010-13.2013

Lee, G. Y., & Kisilevsky, B. S. (2014). Fetuses respond to father's voice but prefer mother's voice after birth. *Developmental Psychobiology*. https://doi.org/10.1002/dev.21084

Lenz, D., Krauel, K., Flechtner, H. H., Schadow, J., Hinrichs, H., & Herrmann, C. S. (2010). Altered evoked gamma-band responses reveal impaired early visual processing in ADHD children. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2010.03.019

Leong, V., Byrne, E., Clackson, K., Harte, N., Lam, S., Barbaro, K. de, & Wass, S. (2017). Infants' neural oscillatory processing of theta-rate speech patterns exceeds adults'. *BioRxiv*. https://doi.org/10.1101/108852

Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2014). Infant-directed speech enhances temporal rhythmic structure in the envelope. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*.

Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind*. https://doi.org/10.1162/opmi_a_00008

Liang, C., Houston, L. M., Samy, R. N., Abedelrehim, L. M. I., & Zhang, F. (2018). Cortical Processing of Frequency Changes Reflected by the Acoustic Change Complex in Adult Cochlear Implant Users. *Audiology and Neurotology*. https://doi.org/10.1159/000492170

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*. https://doi.org/10.1016/j.neuron.2007.06.004

Mahon, M., & Crutchley, A. (2006). Performance of typically-developing school-age children with English as an additional language on the British Picture Vocabulary Scales II. *Child Language Teaching and Therapy*. https://doi.org/10.1191/0265659006ct311xx

Martin, B. A., Boothroyd, A., Ali, D., & Leach-Berth, T. (2010). Stimulus presentation strategies for eliciting the acoustic change complex: Increasing efficiency. *Ear and Hearing*. https://doi.org/10.1097/AUD.0b013e3181ce6355

Martin, B. A., Tremblay, K. L., & Stapells, D. R. (2007). Principles and applications of cortical auditory evoked potentials. In *Auditory evoked potentials: Basic principles and clinical application*.

Martinez, A., Eisenberg, L., & Boothroyd, A. (2013). The acoustic change complex in young children with hearing loss: A preliminary study. *Seminars in Hearing*. https://doi.org/10.1055/s-0033-1356640

McCarthy, K. M., Skoruppa, K., & Iverson, P. (2019). Development of neural perceptual vowel spaces during the first year of life. *Scientific Reports*. https://doi.org/10.1038/s41598-019-55085-y

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*. https://doi.org/10.1016/0010-0277(88)90035-2

Millman, R. E., Prendergast, G., Hymers, M., & Green, G. G. R. (2013). Representations of the temporal envelope of sounds in human auditory cortex: Can the results from invasive intracortical “depth” electrode recordings be replicated using non-invasive MEG “virtual electrodes”? *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2012.09.017

Molfese, D. L. (2000). Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain and Language*. https://doi.org/10.1006/brln.2000.2287

Molfese, D. L., & Molfese, V. J. (1985). Electrophysiological indices of auditory discrimination in newborn infants: The bases for predicting later language development? *Infant Behavior and Development*. https://doi.org/10.1016/S0163-6383(85)80006-0

Moon, C. M., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica, International Journal of Paediatrics*. https://doi.org/10.1111/apa.12098

Moore, D. R., Ferguson, M. A., Edmondson-Jones, A. M., Ratib, S., & Riley, A. (2010). Nature of auditory processing disorder in children. *Pediatrics*. https://doi.org/10.1542/peds.2009-2826

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

Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4944634

Newman, R. S. (2005). The cocktail party effect in infants revisited: Listening to one’s name in noise. *Developmental Psychology*. https://doi.org/10.1037/0012-1649.41.2.352

Newman, R. S., & Jusczyk, P. W. (1996). The cocktail party effect in infants. *Perception and Psychophysics*. https://doi.org/10.3758/BF03207548

Nishi, K., Lewis, D. E., Hoover, B. M., Choi, S., & Stelmachowicz, P. G. (2010). Children’s recognition of American English consonants in noise. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.3377080

Novak, G., Ritter, W., & Vaughan, H. G. (1992). Mismatch Detection and the Latency of Temporal Judgments. *Psychophysiology*, *29*(4). https://doi.org/10.1111/j.1469-8986.1992.tb01713.x

Nozza, R. J., & Wilson, W. R. (1984). Masked and unmasked pure-tone thresholds of infants and adults: Development of auditory frequency selectivity and sensitivity. *Journal of Speech and Hearing Research*, *27*(4). https://doi.org/10.1044/jshr.2704.613

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

Olguin, A., Cekic, M., Bekinschtein, T. A., Katsos, N., & Bozic, M. (2019). Bilingualism and language similarity modify the neural mechanisms of selective attention. *Scientific Reports*. https://doi.org/10.1038/s41598-019-44782-3

Ortega, R., López, V., Carrasco, X., Anllo-Vento, L., & Aboitiz, F. (2013). Exogenous orienting of visual-spatial attention in ADHD children. *Brain Research*. https://doi.org/10.1016/j.brainres.2012.11.036

Ortiz-Mantilla, S., Hamalainen, J. A., Musacchia, G., & Benasich, A. A. (2013). Enhancement of Gamma Oscillations Indicates Preferential Processing of Native over Foreign Phonemic Contrasts in Infants. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3260-13.2013

Ostroff, J. M., Martin, B. A., & Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. In *Ear and Hearing*. https://doi.org/10.1097/00003446-199808000-00004

Panagos, R. J. (1998). Meaningful Differences in the Everyday Experience of Young American Children. *Remedial and Special Education*, *19*(6). https://doi.org/10.1177/074193259801900609

Pearson, B. Z., Fernández, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*.

https://doi.org/10.1017/s0142716400009863

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhs118

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

Pollack, I., & Pickett, J. M. (1958). Stereophonic Listening and Speech Intelligibility against Voice Babble. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1909505

Ponton, C., Eggermont, J. J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of human central auditory system activity: Separating auditory evoked potentials by dipole source modeling. *Clinical Neurophysiology*. https://doi.org/10.1016/S1388-2457(01)00733-7

Poulton, E. G. (1953). Two-channel listening. *Journal of Experimental Psychology*. https://doi.org/10.1037/h0057873

Power, Alan J., Colling, L. J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, *160*. https://doi.org/10.1016/j.bandl.2016.06.006

Power, Alan James, Mead, N., Barnes, L., & Goswami, U. (2012). Neural entrainment to rhythmically presented auditory, visual, and audio-visual speech in children. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2012.00213

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who’s talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6). https://doi.org/10.1111/desc.12172

Reetzke, R., Lam, B. P. W., Xie, Z., Sheng, L., & Chandrasekaran, B. (2016). Effect of simultaneous bilingualism on speech intelligibility across different masker types, modalities, and signal-to-noise ratios in school-age children. *PLoS ONE*, *11*(12). https://doi.org/10.1371/journal.pone.0168048

Reynolds, G. D., & Guy, M. W. (2012). Brain-behavior relations in infancy: Integrative approaches to examining infant looking behavior and event-related potentials. *Developmental Neuropsychology*. https://doi.org/10.1080/87565641.2011.629703

Reynolds, G. D., & Romano, A. C. (2016). The development of attention systems and working memory in infancy. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/fnsys.2016.00015

Rimmele, J. M., Zion Golumbic, E., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex*. https://doi.org/10.1016/j.cortex.2014.12.014

Song, J., Martin, L., & Iverson, P. (2020). Auditory neural tracking and lexical processing of speech in noise: Masker type, spatial location, and language experience. *The Journal of the Acoustical Society of America*, *148*(1). https://doi.org/10.1121/10.0001477

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*.

https://doi.org/10.1016/0013-4694(75)90263-1

Stevens, A. A., Maron, L., Nigg, J. T., Cheung, D., Ester, E. F., & Awh, E. (2012). Increased sensitivity to perceptual interference in adults with attention deficit hyperactivity disorder. *Journal of the International Neuropsychological Society*. https://doi.org/10.1017/S1355617712000033

Striano, T., Kopp, F., Grossmann, T., & Reid, V. M. (2006). Eye contact influences neural processing of emotional expressions in 4-month-old infants. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsl008

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., … Wartenburger, I. (2009). Sensitivity of Newborn Auditory Cortex to the Temporal Structure of Sounds. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1246-09.2009

Telkemeyer, Silke, Rossi, S., Nierhaus, T., Steinbrink, J., Obrig, H., & Wartenburger, I. (2011). Acoustic processing of temporally modulated sounds in infants: Evidence from a combined near-infrared spectroscopy and EEG study. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2011.00062

Thompson, E. C., Woodruff Carr, K., White-Schwoch, T., Otto-Meyer, S., & Kraus, N. (2017). Individual differences in speech-in-noise perception parallel neural speech processing and attention in preschoolers. *Hearing Research*. https://doi.org/10.1016/j.heares.2016.11.007

Trehub, S. E., Bull, D., & Schneider, B. (1981). Infants’ detection of speech in noise. *Journal of Speech and Hearing Research*, *24*(2). https://doi.org/10.1044/jshr.2402.202

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4). https://doi.org/10.1080/17470216008416732

Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/fnsys.2014.00028

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*. https://doi.org/10.1111/j.1467-8624.2004.00726.x

Uhler, K. M., Hunter, S. K., Tierney, E., & Gilley, P. M. (2018). The relationship between mismatch response and the acoustic change complex in normal hearing infants. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2018.02.132

Wang, Y., Bergeson, T. R., & Houston, D. M. (2017). Infant-directed speech enhances attention to speech in deaf infants with cochlear implants. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/2017_JSLHR-H-17-0149

Webb, S. J., Long, J. D., & Nelson, C. A. (2005). A longitudinal investigation of visual event-related potentials in the first year of life. *Developmental Science*. https://doi.org/10.1111/j.1467-7687.2005.00452.x

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*. https://doi.org/10.1111/j.1467-8624.1981.tb03051.x

Werker, Janet F., & Lalonde, C. E. (1988). Cross-Language Speech Perception: Initial

Capabilities and Developmental Change. *Developmental Psychology*. https://doi.org/10.1037/0012-1649.24.5.672

Werker, Janet F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*. https://doi.org/10.1016/S0163-6383(84)80022-3

Wunderlich, J. L., & Cone-Wesson, B. K. (2006). Maturation of CAEP in infants and children: A review. *Hearing Research*. https://doi.org/10.1016/j.heares.2005.11.008

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., … Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party.” *Neuron*. https://doi.org/10.1016/j.neuron.2012.12.037