Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Investigating Melodic Perception of Sung Nursery Rhymes in Adults and Infants with EEG

Sean Brieffies, B.A. (Mod)

A Dissertation

Presented to the University of Dublin, Trinity College

in partial fulfilment of the requirements for the degree of

Master of Science in Computer Science (Data Science)

Supervisor: Dr. Giovanni Di Liberto

August 2022

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

Sean Brieffies

August 18, 2022

Permission to Lend and/or Copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

Sean Brieffies

August 18, 2022

Abstract

The ability to effectively study neural responses to continuous naturalistic stimuli is a relatively new development. It is an important development as these stimuli are most akin to the stimuli we encounter in everyday life, making the study of them of great interest and importance to those seeking to better understand the processes underlying our responses to them. These methods involve using linear models which treat neural responses as a linear combination of stimulus features. This methodology has recently shown promising results in the study of neural responses to speech, melodic and visual stimuli, as well as the process of attention. It has been shown in the past that neural responses are shaped in part by our expectations of stimuli. In a musical context, these expectations are in turn shaped by our ability to learn the statistics of the musical stimuli at hand, as well as from our preexisting notions of musical structure that we have learned from past exposures. Our wider ability to learn the statistics and structure of the world around us is thought to be a generalisable and fundamental element of the brain. Hence, new information about this process from a specific context could better inform us of this broader phenomenon. Recent work has showed encoding of melodic expectations in subjects' EEG responses for the first time using continuous naturalistic stimuli. The current work sought to expand upon these developments and ideas to study how EEG responses from the brains of adult and infant subjects reflected their melodic expectations of sung nursery rhymes. Nursery rhymes had never previously been used for these purposes, and this was also the first time that melodic perception had been studied in infants using these new methodologies. Infant data was recorded across a longitudinal period at 3, 7, and 11 months of age. To perform this investigation, melodic expectation features were derived using a statistical model of melodic surprise, which has been shown to mimic the statistical learning properties of the brain in a physiologically plausible way, and then regressed with subjects' EEG responses using linear models in the form of temporal response functions. Significant encoding of these features was not found within the adult subjects' responses. The infants showed a non-significant improvement in their encoding of melodic expectation features as they aged. This effect was still interesting however as previous works on this same dataset suggest that it could not be explained purely by the attention of the subjects.

Acknowledgements

I wish to thank a number of people who have helped me this year.

Firstly, I would like to thank my supervisor Dr. Giovanni Di Liberto for taking me on and introducing me to this area. I sincerely appreciate your time throughout the past few months in our one-on-one and lab meetings, and for constantly steering me in the right direction and answering my many questions.

I would like to thank my friends for their support during the year. This academic year has been very busy and having people to talk, laugh, and disconnect with was a very important respite.

I would like to thank Sherin Cherian for helping me transcribe the melodies of the stimuli and for having such a positive impact on my life over the last ten-and-a-half months. P.S. your self-proclaimed musical talents really came through in a big way.

Last but not least, I would like to thank my family for their support throughout my education and life in general, without you none of this would be possible and I am forever grateful for the start in life that you have provided me with.

Sean Brieffies

University of Dublin, Trinity College
August 2022

Contents

List of Figures

Chapter 1

Introduction

1.1 Project Description & Motivation

The overall goal of this dissertation is to better understand how the brain works, more specifically, how it learns and processes melodies. Music has great cultural, societal, and perceptual importance. This work seeks to determine to what extent adults and infants in the first year of life perceive musical melodies in nursery rhymes. This will involve analysing electroencephalogram (EEG) data collected from the brains of adults as well as longitudinal data collected from infants while they listened to sung nursery rhymes at 4, 7, and 11 months of age respectively. Existing methodologies such as Temporal Response Functions (TRF) (Crosse et al., 2016) will be used to map the stimuli to the resulting EEG signal. These methods will be first applied to the adult dataset and then extended to the infant dataset.

Studying how people learn and process melodies is part of a larger question of how we learn the statistical properties of our environment. It is thought that this *Statistical Learning* plays an automatic, generalisable, and fundamental role in how we learn and hence predict the world around us (Perruchet and Pacton, 2006). Evidence for this has been found in the learning of language (Erickson and Thiessen, 2015), music (Pearce, 2018), and visual information (Jost et al., 2015). Investigating melodic perception with nursery rhymes will explore a new way in which this important phenomenon can be studied.

This work will make a number of novel contributions to the field. While EEG analysis using nursery rhyme stimuli has been done before (Soufineyestani, 2020; Soufineyestani et al., 2021), including with infants (Schabus et al., 2016; Attaheri et al., 2022), to the best of my knowledge there is no existing work on using nursery rhymes to investigate melodic perception in adults or infants. Monophonic or polyphonic music is typically used for these analyses (Di Liberto et al., 2020; Reynolds, 2022). There also appears to be no literature on linear modelling of melodic features for continuous naturalistic stimuli in infants.

Nursery rhymes exist at an interesting intersection between song and speech from which to study melodic perception. These are two of the main processes studied with neural signal analysis and so the relationship between the two is of great interest in the field. Speech and music have been shown to produce particularly strong EEG responses compared to other natural sounds (Zuk et al., 2020), and existing research has also previously shown that the brains response to speech can be modulated by the melodic properties of the stimuli (Gordon et al., 2010).

This work will also help to establish a framework for investigating melodic perception in infants. This will involve investigating new ways to account for the differences in infants' neurological make-up relative to adults. This in turn will empower others to explore these questions further.

With a view to the future, measuring a person's perception of music and the rhythms and melodies contained therein, could also potentially tell us a lot about their perceptual faculties and ability to learn the statistics of the world around them. For example, Dzafic et al. (2020) found a relationship between inhibited statistical learning and prediction, and psychotic-like episodes. Previous studies by Jentschke et al. (2008) and Sallat and Jentschke (2015) found that language-impaired children also had trouble processing musical features. They suggested that music could be used in the diagnosis and treatment of language-impaired individuals. Flagging these irregularities would be especially useful for infants where, at an earlier stage of life, more can be done to address potential problems. In this way, melodic perception could potentially act as a biomarker for measuring perceptional abilities. After conducting a review of electrophysiological biomarkers, Jeste et al. (2015) concluded that "insights gained from EEG studies will contribute significantly to a more mechanistic understanding of these [neurodevelopmental] disorders and to the development of biomarkers that can assist with diagnosis, prognosis, and intervention".

1.2 Research Objectives

This research has two primary objectives:

1. Determine whether melody perception can be studied with nursery rhymes and EEG.

2. Determine to what extent infants in the first year of life perceive melodies and how this relationship changes across the different months of early life.

1.3 Dissertation Layout

1.3.1 Background

This chapter provides information necessary to understand key concepts mentioned later in the report. This will include descriptions of the methodologies, science, software libraries, and other tools needed to perform the analysis along with references to relevant and/or complementary works in the wider field.

1.3.2 Methods

The process to go from stimuli and EEG data to the eventual analysis will be laid out and described in detail.

1.3.3 Results

This chapter will put forth the results of the analysis undertaken as per the description in the Methods section.

1.3.4 Discussion

A more in depth discussion of the "why" of the results will take place. This will include relating the results to the original hypothesis of the dissertation and discussing if there is evidence to accept/reject them. This will be followed by a wider discussion of key takeaways, potential improvements, thoughts towards future works in this area, and a personal reflection on the project as a whole.

Chapter 2

Background

2.1 Statistical Learning

Section 1 describes statistical learning as a general feature of human learning whereby our brains continually update an internal model of the patterns, properties, and statistics of the surrounding world and uses this model to make predictions about the future. These predictions are consciously and unconsciously experienced as our *expectations* about future events. It may be useful to think of this statistical learning process like a machine learning model learning from training samples, although of course the brain learns generalisable knowledge from it's "training data" far better than current machine learning methods, which typically excel at tasks of a far more narrow scope.

One specification of statistical learning is *Predictive Coding*. This extends beyond the domain of learning and rather posits the generation of internal predictive models as a foundational process of how the entire brain works. It proposes that the brain is continuously predicting and then refining these predictions based on the error between these predictions and the eventual input stimuli. The theory also says that there are distinct neurons in the brain responsible for models of encoding and decoding respectively. Predictive coding is the most widely acknowledged framework of statistical learning and has even been called a "grand unified theory of the brain" (Friston, 2010).

2.2 Music and the Brain

Functional Magnetic Resonance Imaging (fMRI) studies have frequently shown that different parts of the auditory cortex respond particularly to speech or music (Belin et al., 2000; Angulo-Perkins et al., 2014; Norman-Haignere et al., 2015; Overath et al., 2015). These findings were recently replicated in a study by Norman-Haignere et al. (2022), which also found evidence of populations of neurons in the auditory cortex which respond to singing but not instrumentals or speech, which would suggest that these segregations of duty go even further, and that different subgroups within the auditory cortex are responsible for responding to different types of music. In the previously cited work by Zuk et al. (2020), in which the authors demonstrated stronger neural responses to speech and music, acoustically identical but incomprehensible stimuli were used as a control and these same stronger responses were not found. This indicates that *something* about the higher-level features of speech and music brings about these stronger responses, further highlighting their "special status" in the brain.

While the studies referenced above are from adult subjects, which are more common in part because they are easier to perform, it is well established that infants do also perceive melody (Trehub, 1987; Dowling, 1999; Ilari, 2002; Stalinski and Schellenberg, 2012). We know from fetal studies that this is true even before they are born (Ilari, 2002). Two month old infants have demonstrated the ability to remember a melody and tell it apart from similar ones (Plantinga and Trainor, 2009). Like most adults, infants at six months of age have been shown to remember the relative pitch of a piece (Plantinga and Trainor, 2005). Rather than remembering purely the pitch that they heard, they can recognise the same piece played in different keys by recalling the differences between the notes. Ilari (2002)'s review paper: "Music Perception and Cognition in the First Year of Life" is an excellent resource for a more comprehensive overview of the literature on early-life music perception. In this review, the need for longitudinal data and "real music", which requires multiple musical features to be processed at once as opposed to the segmented artificial pieces used to study these features in isolation, was highlighted. While this review is older and came from a psychological position, I believe this still encouraging for the current study as fortunately I have access to longitudinal data which was collected while infants listened to "real music".

The aforementioned concepts of statistical learning and predictive coding are thought to play an important role in human auditory processing. In a review of the literature by Heilbron and Chait (2018) to investigate the presence of predictive coding in the auditory cortex, the authors found repeated evidence that neural responses were shaped by expectations, which is a foundational concept for the current study. The idea of violating expectations has long been used to study auditory cortex processes. Researchers have often studied the auditory cortex by inserting purposefully incorrect sounds amongst otherwise correct ones. These incorrect sounds produce much larger neural responses than the correct ones. This difference can be quantified as a subject's mismatch negativity (MMN). Exploiting the phenomenon known as Neural Adaptation is another means by which to study MMN. Neural adaptation says that a repeated stimulus will gradually result in smaller and smaller responses as the brain 'adapts' to this stimulus. Playing repeated stimuli and inserting purposefully incorrect or surprising sounds, will produce much larger responses relative to the now-adapted stimulus.

Outside of the more heavy-handed nature of MMN experiments, the concept of neural responses being shaped by expectations can be useful in a more naturalistic experimental set-up. As listeners, our expectations of music can vary depending on our prior musical experiences (long-term expectations) as well as the melodic features of the song at hand that we 'pick up' (short-term expectations) (Pearce, 2018). Previous studies have also shown that the degree of melodic prediction error, or surprise, is reflected in a persons neural activity (Vuust et al., 2012; Moldwin et al., 2017; Quiroga-Martinez et al., 2019a). If neural responses are indeed being shaped by long and short-term melodic expectations, being able to model these expectations would open an important avenue for the study of the neural responses associated with them. Models which attempt to do this will be discussed in Section 2.3. Using one of these models, Di Liberto et al. (2020) recently showed encoding of melodic expectations of pitch and note-onsets in subjects' neural responses to naturalistic stimuli for the first time. This current work will follow a similar approach and extend it to infants for the first time.

2.3 Modelling Melodic Expectations

The need was outlined for models which can model a listener's melodic expectations in Section 2.1. There are several methods which attempt to do this. The most established of these is the Information Dynamics of Music (IDyOM) model from Marcus Pearce (Pearce, 2005). This model is based on variable-order Markov model which seeks to mimic the statistical learning paradigm of the brain and learn the statistical properties of musical pieces. After learning these properties, it can then be used to compute the likelihood of certain events occurring in a musical piece e.g. a note with a certain pitch appearing within a piece or a note occurring at a given time in the piece. IDyOM can compute the information content and entropy of a multitude of musical features. A musical note with high information content can be thought of as being very surprising, while high entropy denotes high unpredictability, and vice versa. In accordance with the literature referenced in Section 2.2, highly surprising or unpredictable notes elicit stronger neural responses.

IDyOM's models come in three forms: long-term(LTM), short-term(STM) and a mixture of these, known as "Both". Long-term models learn the statistics of a corpus of music and use these pre-trained statistics when calculating how surprising or unpredictable notes from a new piece of music are. This is intended to represent the learned structure an individual has acquired through their exposure to music over time. This exposure could come in many forms. Listeners' long-term musical experiences do not have to be of a formal nature, as the acquisition of music has been shown to be an automatic process (Rohrmeier and Rebuschat, 2012). Short-term models do not pre-train and rather learn the properties of a new piece in real-time. It has been shown that more complex musical structures are harder to remember and hence predict (Cohen et al., 1977; Bartlett and Dowling, 1980; Cuddy and Lyons, 1981; Bartlett et al., 1995). IDyOM has been shown to reflect this and to model listener's melodic expectations in a physiologically plausible way (Omigie et al., 2013; Egermann et al., 2013; Halpern et al., 2017; Agres et al., 2018; Quiroga-Martinez et al., 2019b, 2020).

While IDyOM might be the most tried-and-true method of modelling a listeners melodic expectations, there are other newer methods of approaching this same problem. Music is a sequential stream of information. In recent years, Recurrent Neural Networks (RNNs) and more recently Transformer models have achieved excellent results operating on sequential data such as speech translation and text-to speech applications (Cho et al., 2014; Wang et al., 2019; Karita et al., 2019). Huang et al. (2018) proposed a Music Transformer model to model melodic surprise features from music. There is also the Probabilistic Model of Melody Perception based on gestalt-like rules Temperley (2008). Kern et al. (2022) compared the effectiveness of the IDyOM, Probabilistic Model of Melody Perception, and Music Transformer models at reconstructing musical pieces based on their likelihood predictions. The computational IDyOM and Music Transformer models were found to outperform the rule-based Probabilistic Model of Melody Perception model. Despite being simpler models, IDyOM's STM and Both models performed comparably with the Music Transformer model. This study also replicated the results from Di Liberto et al. (2020) by showing neural encoding of melodic expectations, this time using surprise features generated by the Music Transformer model.

2.4 Analysing the Brain's Response to External Stimuli

Analysing brain responses to external stimuli like speech or music was typically done by looking at Event-Related Potentials (ERPs) until recently. An 'event' being a stimulus, and the related potentials being the brain's response to it. In experiments, these events were usually highly segregated as there was no way to decompose a response into its constituent parts. The responses resulting from specific neurological processes were often identified using pre-existing knowledge about when these processes occurred in time. Figure 2.1 shows ERPs in response to speech. Higher level processes such as lexical-semantic analysis are thought to occur after acoustic/phonetic feature processing (Salmelin, 2007), so the different responses present across the EEG signal could then be attributed to the appropriate process. This presents a limitation however, as it has been postulated that overlaps exist between the timings of both high and low level processes (Di Liberto, 2017). This makes ERPs less effective for studying naturalistic speech or other continuous stimuli, as these overlaps will continually occur. This deficiency necessitated another way of studying brain responses to continuous stimuli. Continuous stimuli are of particular interest because our brains are continuously being stimulated by the world around us and hence continuously responding to it. In addition, naturalistic stimuli like speech or music are usually continuous and not broken up into neat chunks which we can easily segment and analyse.

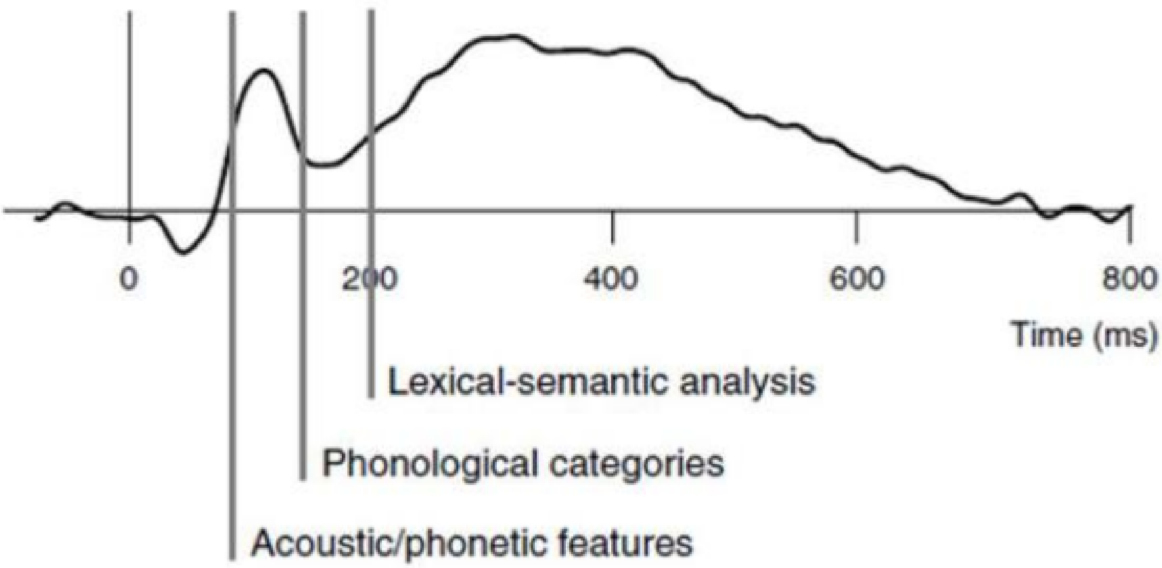

Figure 2.1: Higher level processing of speech occurring at subsequently further time points (Di Liberto, 2017)

A more recent approach is to assume the EEG signal is a linear combination of continuous stimulus features. This allows us to model the signal with techniques such as linear regression. Linear models have the advantage of being computationally inexpensive and far more interpretable than a deep learning model for example. Feature weights from a linear model can be interpreted as a reflection of what the brain is doing, and even decomposed to the individual electrodes to get a spatial understanding of these processes. Linear models also require far less data, which is often a feature of these experiments. There are two kinds of "directions" when modelling the relationship between stimuli and EEG data; forward and backward. A forward, or encoding model, predicts the EEG response that will result from an input stimulus, e.g. predicting the EEG response of my brain when I listen to a piece of music. A backward, or decoding model, takes an EEG response as input and attempts to reconstruct the preceding stimulus, e.g. reconstructing the speech I listened to based off of my EEG response. A popular linear regression-based method to model these relationships is the Temporal Response Function (TRF) (Crosse et al., 2016), which implement a form of linear ridge regression.

This research involves modelling and analysing the response of subjects while they listened to nursery rhyme stimuli and hence encoding/forward TRF models can be used for this purpose. This model is of the form:

$$r(t, n) = \sum_{\tau} w(\tau, n)s(t - \tau) + \epsilon(t, n)$$

such that t is a time point, n is a channel e.g. an EEG electrode, r(t,n) is the predicted neural response, s(t - τ) is the stimulus, w(τ , n) is the TRF, and ϵ(t,n) is the behaviour not captured by the model.

The TRF, w, seeks to minimise the mean-square error between the actual and predicted responses. Given that this minimisation typically involves little data, this can be solved with the closed formula:

$$w = (s^T s)^{-1} s^T r$$

It was mentioned previously that different neural processes occur at different times after the onset of a stimulus. To account for this, the TRF includes a time-lagged window when mapping stimuli to responses. A stimuli s is replaced with a time-lagged convolution matrix of that stimuli. This lag window is in the range [t_{min}, t_{max}]. In addition to choosing values for t_{min} and t_{max}, we can also pass a range of lambda regularisation parameters to be used when training a TRF model.

The encoding model attempts to reconstruct the original EEG responses and ultimately model what the brain is doing. The quality of this encoding can be evaluated in several ways. We can compare the reconstructed response with the original response, and compute average correlation coefficients across individual subjects or individual electrodes. Looking at this at a per-electrode level provides another means of evaluation. It allows us to see what areas of the scalp are most strongly responding to the stimuli. Comparing this with existing neurological theory can tell us whether results are physiologically plausible.

Figure 2.2 shows the output of a TRF predicting the response to the envelope of a stimulus. The x-axis shows the time-lag window between -100 and 400ms, with 0ms being when the subject first hears a stimulus. We can see that the EEG response is strongest at the peak around 150ms.

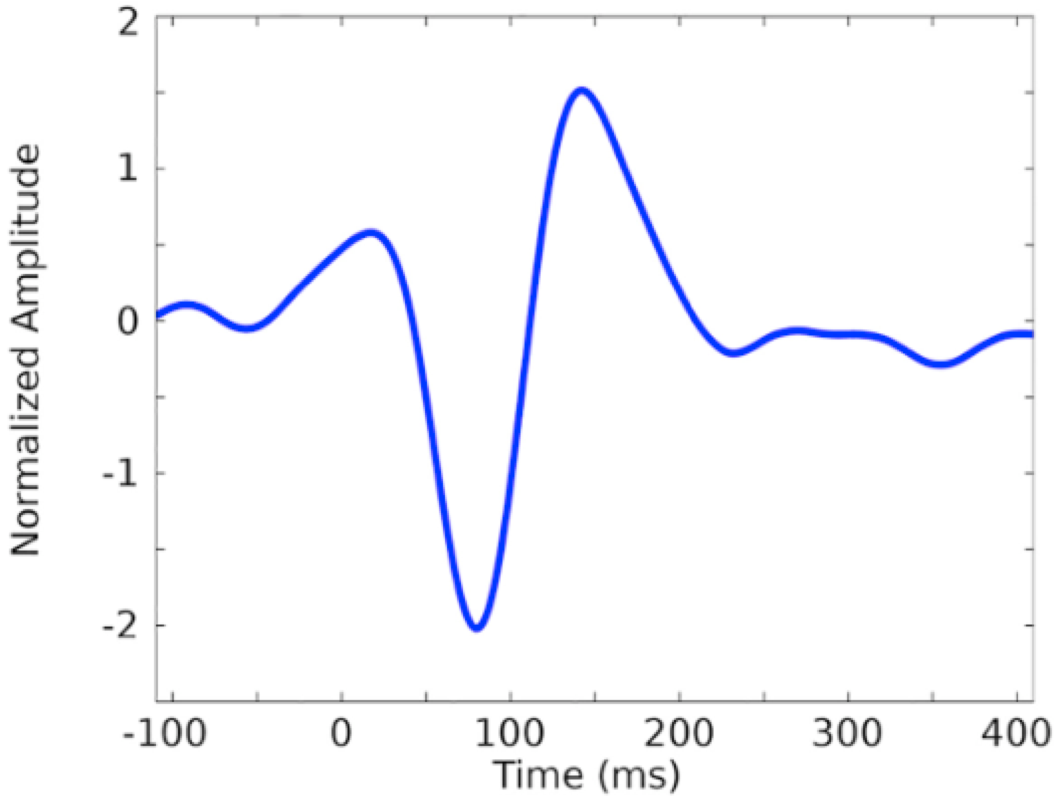

Figure 2.2: Example plot of a TRF (Di Liberto et al., 2015)

TRF encoding models have shown promising results in a variety of use cases including studying acoustic envelope tracking (Attaheri et al., 2022), responses to melodic surprise (Di Liberto et al., 2020), the effects of attention to visual (Jia et al., 2017) and speech stimuli (Vanthornhout et al., 2019; Kurthen et al., 2021; Reetzke et al., 2021), and speech processing during sleep (Harimohan, 2021).

TRFs can be easily implemented with the use of Matlab's mTRF Toolbox (Crosse et al., 2016) or Python's Eelbrain libraries (Brodbeck et al., 2021).

Chapter 3

Methods

3.1 Overview of Approach

Figure 3.1 shows a high-level overview of the approach taken to investigate the hypotheses. It closely follows the methodologies outlined by Crosse et al. (2021) and those used by Di Liberto et al. (2020) to investigate melodic expectations. First, the EEG is preprocessed in various ways to make the signal more usable. A statistical model is then used to get representations of the melodic surprise features of the stimuli. The preprocessed EEG can then be regressed with these melodic features and/or acoustic features using a TRF. Weights and correlation values can be derived from this, and analysis can be performed on these. Each of these steps will be described in more detail in the proceeding sections. The mTRF Toolbox (Crosse et al., 2016) in Matlab was used for EEG preprocessing and analysis. IDyOM (Pearce, 2005) was used to model the melodic surprises of the stimuli. Python, Microsoft Excel, and Tableau were used for additional data analysis and visualisation.

3.2 Dataset Description

The dataset used in this dissertation was originally collected as part of a previous study by Attaheri et al. (2022). It consists of EEG data collected from 60 electrodes placed on the scalps of 17 adults and 47 infants at 4, 7, and 11 months of age respectively. These EEG recordings were made while each subject listened to recordings of 18 different nursery rhymes. Ten of the nursery rhyme recordings were sung, while the remaining 8 were chanted. For this investigation, only the ten sung pieces were used as these contain the most melodic information. After removing empty trials and responses to non-sung stimuli, two of the infant subjects at the 4 month interval had no remaining EEG data,

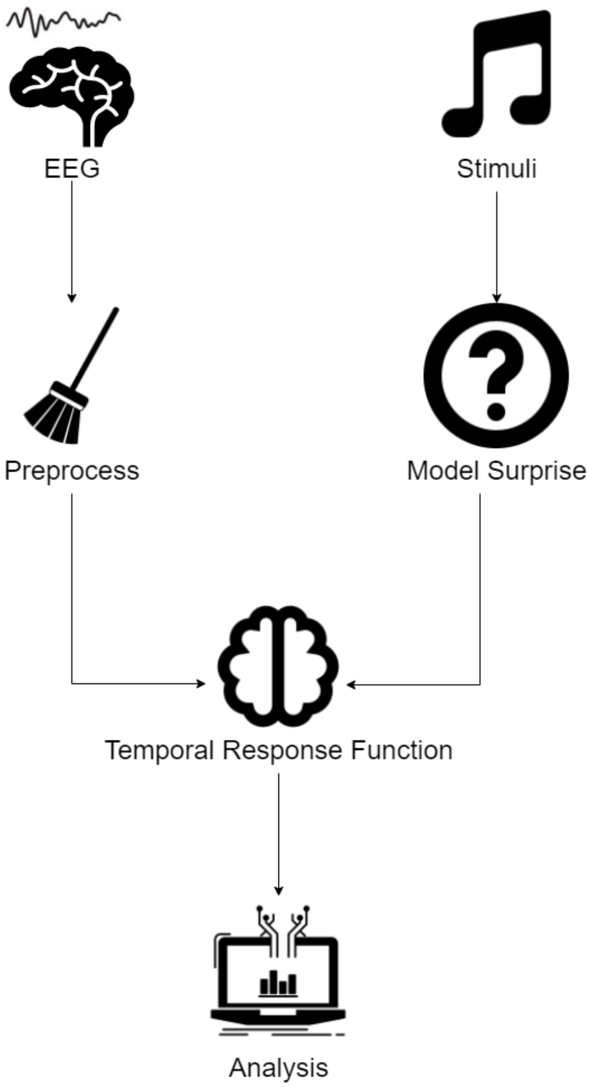

Figure 3.1: Schematic of Approach

so these subjects were omitted entirely from the analysis to allow for a full longitudinal comparison between subjects.

3.3 EEG Data Preprocessing

EEG affords us a relatively low-cost, non-invasive method for collecting signals from a subjects' brain with high temporal resolution. It also allows for mobility of the subject during the experiment, which could especially be a problem in experiments with young infants. A draw-back of a non-invasive method like EEG compared to invasive methods like ECoG/intracranial EEG, is the additional noise captured due to the physical barrier of the scalp between the electrodes and the surface of the cortex. Before any meaningful analysis of the data can be performed, it is necessary to preprocess the EEG recordings. This is particularly important due to the extra noise inherent in EEG signals, as the signal of interest is small relative to the electrical signal as a whole. There are a variety of established techniques to remedy this issue.

Firstly, it must be decided which frequencies from the signal will be analysed. Different neural processes are best analysed in specific frequency bands. This is typically one or more of the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–100 Hz) bands. In their previous work outlining methodological considerations discussing methodological considerations for linear modelling of neurophysiological responses, Crosse et al. (2021) noted that speech responses are mostly reflected in the delta and theta bands, citing the works of Poeppel (2003), Giraud and Poeppel (2012), and Ding and Simon (2014). Previous works have also shown significant responses to melodic surprise between 1 Hz - 8 Hz (Di Liberto et al., 2020). Despite the high level of noise at lower frequencies, responses at frequencies as low as 0.1 Hz have been shown to be important when analysing melodic expectations with EEG (Di Liberto et al., 2021). With these considerations, a High-Pass Filter (HPF) and Low-Pass Filter (LPF) were applied at 0.1 Hz and 8 Hz respectively.

Secondly, Artifact Subspace Reconstruction (ASR) was applied to the EEG signals. EEG is used here to study neural activity. However, a portion of an EEG signal results from the multitude of different physiological processes occurring in the body at that time. It is important to isolate as much of the signals of interest as possible. Artifact correction methods seek to remove signals that come from non-neural processes such as eye or muscle movements. ASR has been shown to remove large amounts of these unwanted signals while preserving a desirable amount of the neural response (Chang et al., 2019). It has also been shown to correct artifacts significantly better than other artifact correction methods such as independent component analysis (ICA) or principal component analysis (PCA) (Plechawska-Wojcik et al., 2019).

Bad EEG channels were then removed. If the standard deviation of a channel is more than three times the average standard deviation across all channels, it is classified as a bad channel. These bad channels are replaced with a spline interpolation of the remaining channels.

Finally, all channels are rereferenced to the mastoid channels. Of the 60 electrodes placed on the subjects' scalps, two of these are mastoid channels. These are electrodes which sit on the mastoid bones behind each ear, which are areas with little neural activity. By rereferencing to these channels, we use their electrical activity as a baseline and only consider the amplitudes of the other electrodes which are above this baseline.

3.4 Modelling the Melodic Surprise of the Stimuli

The IDyOM model Pearce (2005) was used to model the melodic surprise features of the stimuli. There were several reasons for this. As mentioned in Section 2, IDyOM has been shown many times to be a physiologically plausible way of modelling surprise to music. It has also been used to produce features for a TRF models which have in turn showed significant effects on the ability of the TRF models to predict the subjects' EEG signals (Di Liberto et al., 2020; Marion et al., 2021). The original LISP implementation of IDyOM was used [1]. There is also a more accessible Python version currently in development [2]. At first, the "Both" model was used, where the LTM component was trained on a corpus of 185 Bach choral pieces [3]. This dataset was one of the datasets used to evaluate IDyOM in its original publication (Pearce, 2005). This was later replaced by a purely STM as this was shown to better reflect the melodic expectations of the subjects (See Section 4). The features generated by the model were Pitch Information Content, Pitch Entropy, Onset Information Content, and Onset Entropy. These will be collectively referred to as the "Melodic Surprise Features".

3.4.1 Generating Musical Scores from the Audio Stimulus Files

In order to use the IDyOM model, MIDI files of the stimuli were required. This was particularly challenging as the stimuli were sung by humans. Stimuli with pre-defined sheet music such as piano music would be more convenient. This necessitated a method of going from the WAV audio file of a stimulus to a MIDI file.

Initially, an automated approach was attempted using Melodia (Salamon and Gomez, 2012). This algorithm extracts the pitch from audio files, and these pitch frequencies could then be converted to musical notes to make a MIDI file by taking the frequency values at the note onset times, which I was provided with. Much time was spent on this and various different methods were tried to improve results, including averaging the frequency over a window around the onsets, but satisfactory results could not be achieved. Later, Melodyne 5 Studio [4] was also tested for this purpose. It provides a feature to go directly from an audio file to a MIDI file. Again, this did not achieve satisfactory results. It would have been possible to use these results as a base and manually tweak the files to get the desirable results, however, I have no musical training and hence doing this would have taken a long time and may not have been as accurate as required. This led to me

[1] https://github.com/mtpearce/idyom
[2] https://guimarion.github.io/IDyOMpy
[3] https://github.com/GuiMarion/IDyOMpy/tree/master/dataset/bach_Pearce
[4] https://www.celemony.com/en/melodyne

consulting a classmate who did have musical training. She regularly sings in choirs and has completed various grades of piano exams, which included listening tests. She manually transcribed the notes while listening to the stimuli. This was particularly doable as the stimuli were very simple, being nursery rhymes. Qualitatively, this resulted in far more faithful recreations of the stimuli than any other method attempted.

3.5 Stimuli Feature Preprocessing

The envelope and melodic surprise features were standardised before being used to train a TRF model. The derivative envelope was also half-way rectified. A new vector was created for each of the four surprise features. The note onset vectors had been provided, which tell us at what point during the listening experiment a note occurred. Using the onset vectors, four new vectors were created for each stimuli i.e. one vector for each surprise feature. These vectors were the note onset vectors, with the onset markers replaced by the appropriate surprise value.

3.6 Encoding the Stimuli to the EEG Data

After preprocessing the EEG data and stimuli features, and modelling the melodic surprise of the stimuli using IDyOM, it was time to fit a regression model to the EEG data using the stimuli features. As mentioned in Chapter 2, there are two kinds of models for relating stimuli and EEG data. These being forward/encoding models, and backward/decoding models. The focus in this dissertation was mapping the melodic properties of nursery rhyme stimuli to EEG responses, so a forward/encoding model was used. Multiple features were used so a multivariate TRF (mTRF) was used. This was implemented using the mTRF Toolbox in Matlab (Crosse et al., 2016). Sample code from the Cognition and Natural Sensory Processing Workshop (2021) was used as a starting point. The time-lag window was set between 0 and 600ms. Leave-one-out cross-validation across a range of lambda regularisation values was used to train a ridge regression model. These models are relatively simple and very explainable, which is important in this use-case. Models were trained using acoustic features e.g. the note onset times. Another model was then trained with the acoustic feature(s) and the four melodic surprise features. This allowed the difference brought about by the introduction of melodic surprise features to be assessed. To explore how encoding of melodic surprise features changed overtime, reconstructions were generated from TRF models trained at each of the 3, 7, and 11 month intervals respectively using just the melodic surprise features.

3.7 Evaluating the Encoding

After having obtained these models, an attempt could be made to reconstruct the EEG responses using one of the models. The Pearson's correlation coefficient (r) between the reconstructed and true signals was computed. This allowed for a comparison between the various models. T-tests were performed to check for statistically significant correlation improvements when introducing melodic surprise features to TRF models trained on the adult subjects' data, in accordance with one of the hypothesis. These improvements were sought over using solely acoustic features. For example, a desirable outcome would be a statistically significant increase in correlation values when melodic surprise features were introduced to a model which previously used acoustic features, such as the acoustic envelope and envelope derivative of the stimuli. Di Liberto et al. (2020) used the acoustic envelope and envelope derivative as the acoustic features for their comparisons. These statistical tests were ran on the average correlation across subjects and the average correlation across individual electrodes. Investigating the effect across individual electrodes also allowed the cortical topography of the subjects to be plotted, and for comparisons to be made with previous literature as well as preexisting beliefs about where certain processes should occur in the brain. This domain knowledge informs the physiological plausibility of the results. The significance between individual electrodes was False Discovery Rate (FDR) corrected. The analysis was also repeated with randomly shuffled surprise vectors to ensure any differences were not the result of merely introducing an additional higher-dimensional feature into the acoustic models. A one-way ANOVA was performed on the correlation values from each of the three data collection intervals to check for a statistically significant increase in melodic encoding for the infant subjects across 3 to 7 and then 11 months. Topography plots were not useful here as the neural structure of infants changes considerably during the early months of life.

Chapter 4

Results

4.1 Modelling the Melodic Surprise of the Stimuli

As mentioned in Chapter 3, two different IDyOM models were used to model the melodic surprise of the stimuli. First, a Both model, with the LTM portion of the model trained on a large corpus of Bach music was used. A TRF trained with these surprise features and the onset vectors showed no significant improvement on the average r across subjects($p = 0.079$, permutation test). Similarly, using these features in tandem with the envelope and its derivative showed no significant improvement for the same metric ($p = 0.478$, permutation test). There were also no significant improvements on a per-electrode level. This was before controls such as controlling for the added dimensionality were even introduced. Models trained with Bach music were clearly not modelling the melodic surprise of the stimuli effectively.

It was then decided to explore the use of a STM. The results here were far superior as seen in Figure 4.1, where it can be seen that the average correlation across adult subjects for a TRF trained with just the STM surprise features was almost twice as high compared to the Bach-trained Both Model. Kern et al. (2022) had found in their comparison that the STM to slightly outperform a Both model which was trained on a large corpus of monophonic pieces. Possible reasons for this in relation to the current study are discussed in Section 5.

4.2 Encoding of Surprise in EEG

As mentioned before, nursery rhymes have never been used to study melodic perception before. The primary aim of this work is to investigate this possibility. The melodic surprise features generated from the STM were regressed with acoustic features to ascertain if these

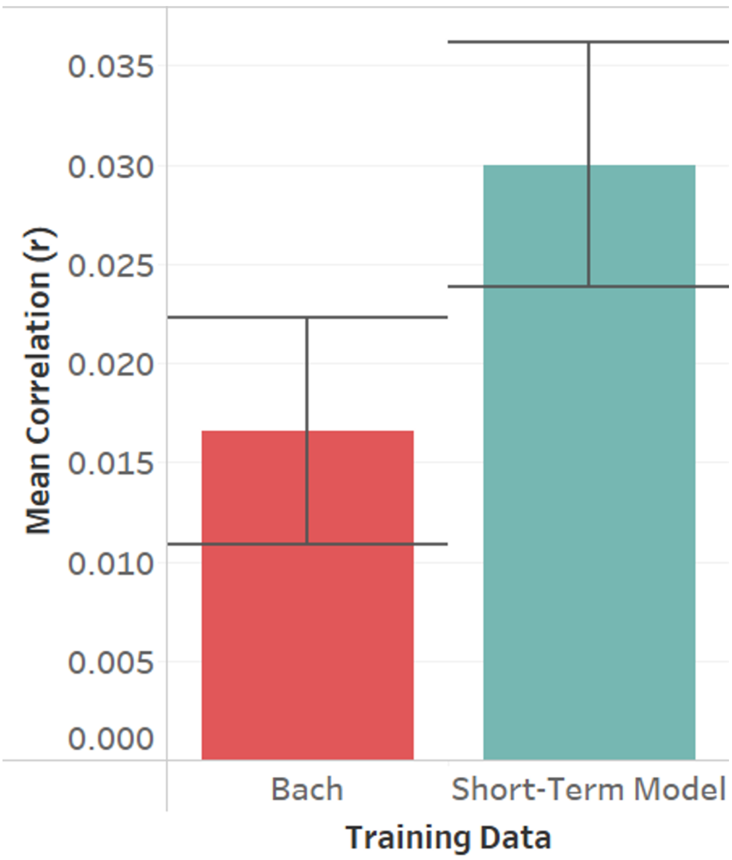

Figure 4.1: Average correlation (r) of the reconstructed signal across all electrodes for two TRFs fitted to the adult EEG data. One fitted with melodic surprise features from an IDyOM Both model where the LTM portion was trained on a corpus of Bach music, and the other using an IDyOM STM

features resulted in a significant improvement in the ability of a TRF to reconstruct the original EEG responses from the 17 adult subjects who took part in the experiment. These surprise features were then regressed against EEG responses from infants at 4, 7, and 11 months of age to explore how the relationship between these features and the infants' neural responses changed overtime.

4.2.1 Adult Subjects

Unlike with melodic surprise features from the Bach-trained Both model, adding surprise features from the STM to acoustic models did result in significant increases in reconstruction correlation. The increase in this metric when introducing the surprise features to a TRF previously trained on the Acoustic Envelope and Envelope Derivative was significant at $p = 0.033$ (permutation test). However, when adding new multidimensional features to a TRF model, it is important that any improvements in performance are not merely due to adding additional complexity to the model. To control for this, comparisons were made between a TRF model trained on acoustic features and a randomly rearranged version of the surprise vectors and a TRF model trained on an acoustic features and the original unshuffled version of the vectors. Figure 4.2 shows this comparison for models regressed on the Acoustic Envelope(E), Envelope Derivative(D), and then Shuffled and Unshuffled Melodic Surprise Features (S_shuf and S respectively). We can see in Figure 4.2A that on the average across all electrodes across all subjects, the model with shuffled surprise features performed slightly worse. This difference was not statistically significant however ($p = 0.15$, permutation test).

Looking at differences at the per-electrode level in Figure 4.2B, the distributions of the correlation values are largely the same. Three electrodes showed a statistically significant difference from t-test (circled in white in Difference topography plot), however after correcting for multiple comparisons, these were shown to not be significant (FDR-adjusted p-values for all three electrodes = 0.547).

These comparison were also repeated with the Note Onset Timings as a singular acoustic feature. No significant results were found here either. See Appendix Figure 2.

4.2.2 Infant Subjects

To assess whether the infants' encoding of the melodic features improved overtime, a TRF model was created for each of the three ages. These models regressed the melodic surprise features against the EEG signals at each of the data collection intervals. Improvements in the average reconstruction correlations across all subjects and electrodes at each consecutive age can be seen in Figure 4.3. One-way ANOVA found these improvements to be non-significant [$F(2, 132) = 1.28$; $p = 0.28$]. Individual electrodes could not be compared like in previous Section 4.2.1, as the infants' neural structure would have changed across the longitudinal timeline of the experiments as their brains continued to develop with age.

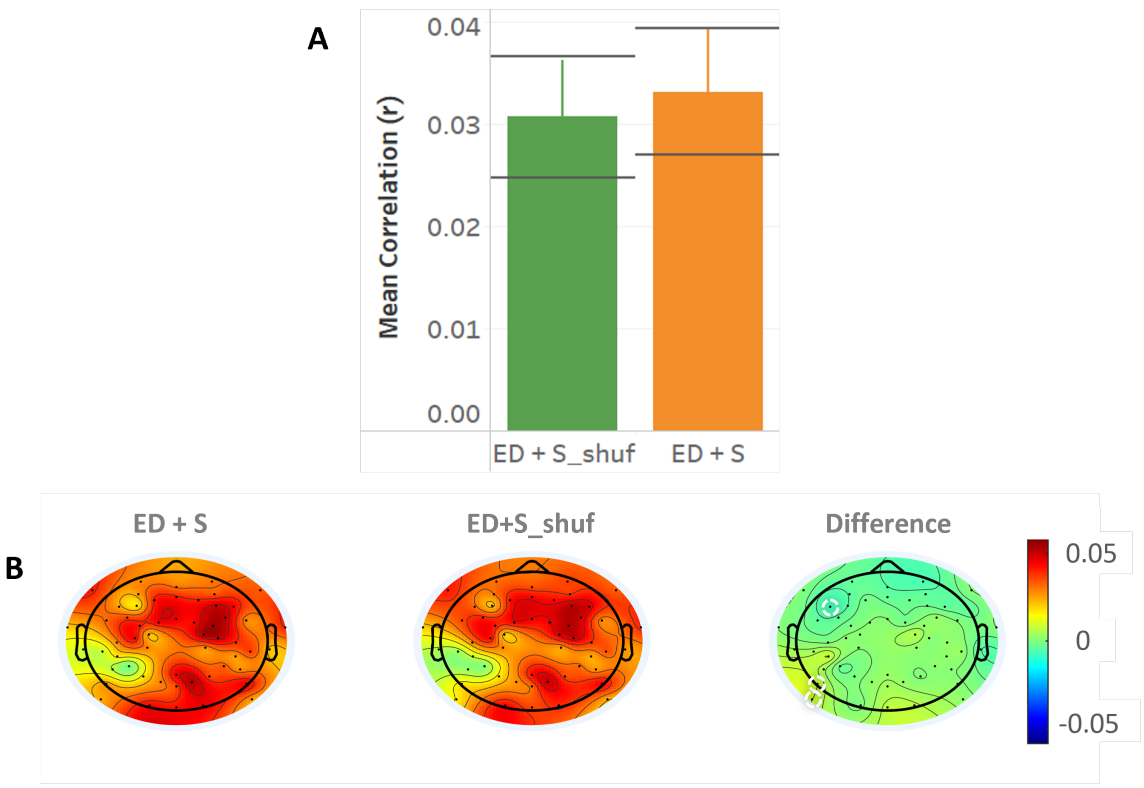

Figure 4.2: (**A**) Average correlation (r) of reconstructed EEG signal across all electrodes for TRF models trained with Envelope + Derivative + Randomly Shuffled Surprise (ED + S_shuf) and Envelope + Derivative + Surprise (ED + S). (**B**) Average correlation (r) of reconstructed EEG signal for each individual electrode.

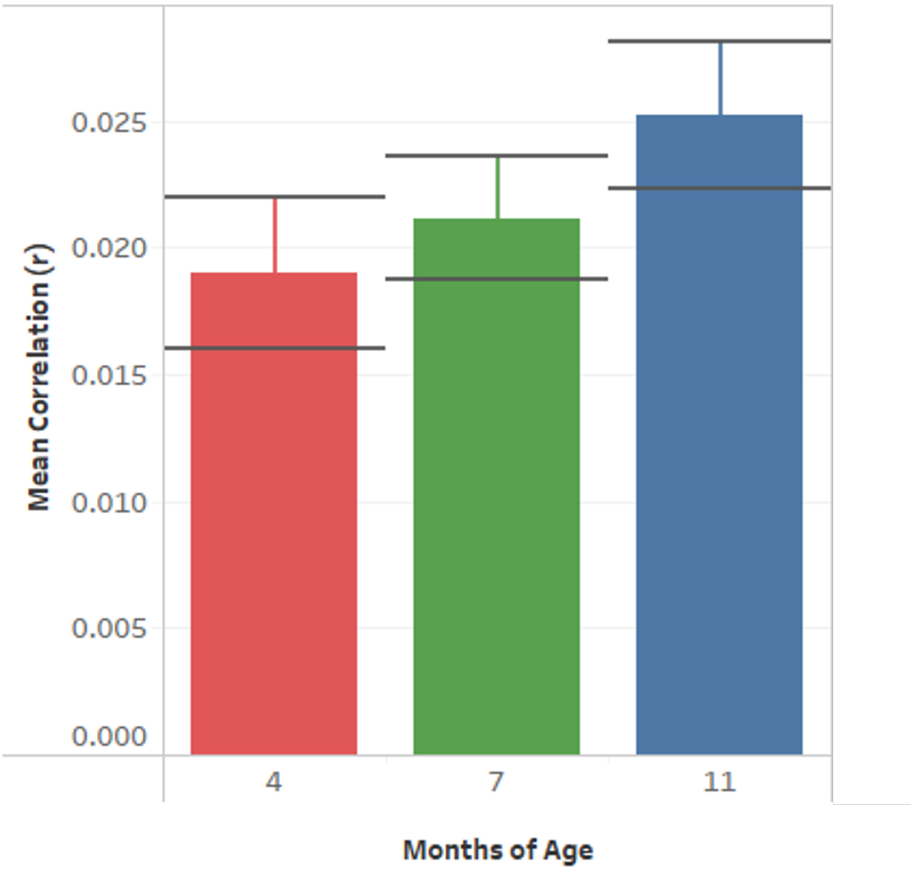

Figure 4.3: Average correlation (r) of reconstructed signal across all electrodes for infants at 3, 7, and 11 months of age respectively.

Chapter 5

Discussion

This work sought to use nursery rhymes for the first time ever to study melodic perception in humans. This was done by following novel methodologies from Di Liberto et al. (2020), the framework of which is further outlined by Crosse et al. (2021). Musical scores were manually transcribed from audio files of sung nursery rhymes. IDyOM, a statistical model of melodic surprise, used these musical scores to model how surprising the pitch and onset timings of these stimuli were to the listening subjects. Surprise values obtained from this model were regressed with acoustic features against the subjects' EEG responses using a TRF to learn if introducing these features led to a significant improvement in the ability of a TRF to reconstruct the original EEG responses. While there did appear to be *some* effect, this difference was not statistically significant. This work was also the first time melodic perception in infants has been studied using these methods. Having access to a longitudinal dataset of EEG responses collected from infants at 4,7, and 11 months old meant that it was possible to see how the encoding of the melodic features in the infants' EEG responses changed overtime. Again, there did appear to be some effect, which will be further discussed below, but there was no statistically significant improvement over the timeline of the data.

As outlined in Chapter 4, training on Bach music did not effectively capture an effect of melodic surprise being present in subjects' EEG responses. While there is no exact science to choosing a training set, it was certainly surprising that the simpler STM model performed so much better than a model that combined a STM with a LTM trained on Bach Music. In hindsight however, there may be a simple reason for this: the stimuli are *extremely* simple. The LTM component of IDyOM attempts to model the statistical learning of a listener with a particular musical experience/enculturation by learning the statistical properties of a training set. One could consider nursery rhymes as being purposefully constructed to appeal to those with little/no prior musical listening experience e.g. infants/children. These pieces use a deliberately small range of notes (little pitch variation) at a predictable rhythm (consistent note onsets). Using Bach music as a training set may have been too high a standard for these purposes. A simpler training corpus would likely have been better suited. We can see in Figure 5.1A, that while there were some variations in the average pitch surprise and pitch entropy across the stimuli, there is almost no variation between the average onset surprise and onset entropy. This is very much in contrast to Figure 5.1B from Di Liberto et al. (2020), showing onset surprise values estimated on a corpus of Bach music after training an IDyOM Both model on a corpus of western music. We can also see this "lack" of surprise looking at the onset surprise values for every note in stimulus 1 in Figure 5.1C, there is an initial jump in surprise between the first and second notes, from which point the surprise stays constant as the model has already learned the simple rhythm of the stimulus. The other stimuli also shared a similarly narrow onset surprise value distribution, which was not very surprising given the similarities present across the stimuli in Figure 5.1A.

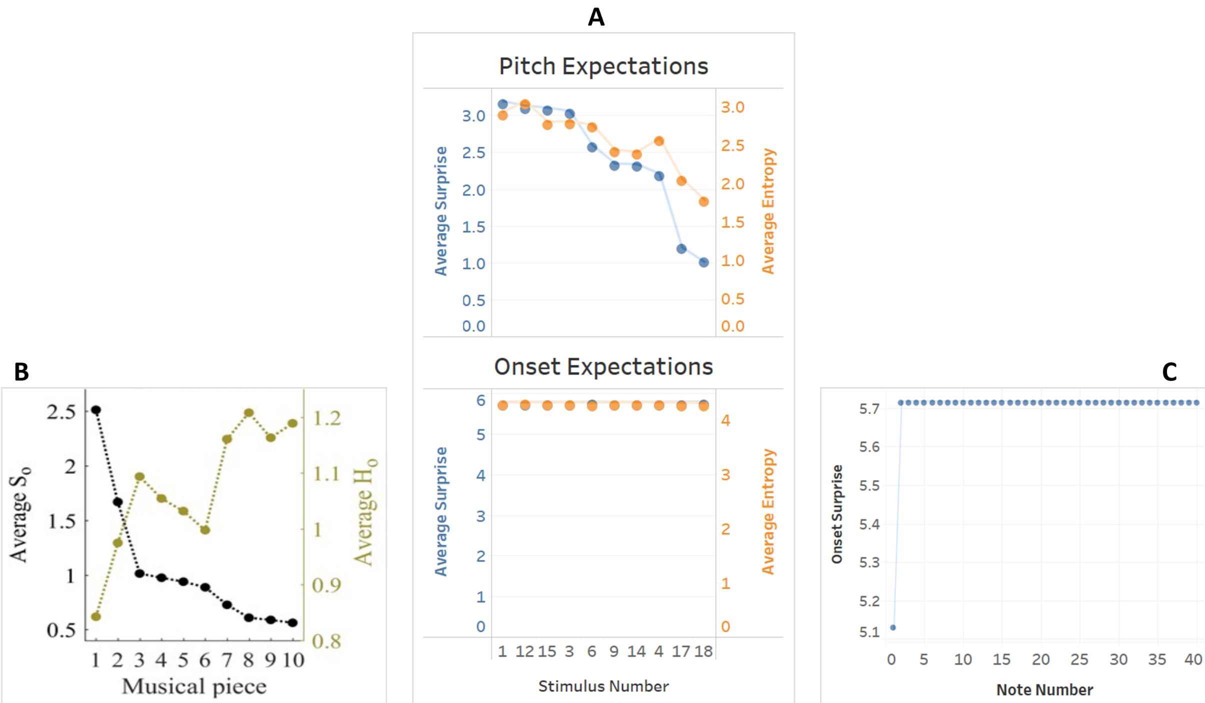

Figure 5.1: **(A)** Average pitch and onset surprise and entropy from IDyOM STM for every stimulus. **(B)** Onset surprise and entropy values from Di Liberto et al. (2020). This IDyOM model was trained on a large carpus of western music and estimated surprise values for a corpus of Bach stimuli. Onset surprise (S_O) is shown on the left axis and onset entropy (H_O) on the right axis. **(C)** Onset surprise for every note in stimulus 1 from IDyOM STM.

This should be considered in future works when using very melodically simple stimuli such as nursery rhymes. It was mentioned in Section 2 that more surprising notes are thought to elicit larger brain responses, and simpler ERP studies would often insert deliberately wrong and hence highly surprising notes. In keeping with this idea, predictable-by-design stimuli such as nursery rhymes will not produce as noticeable a response. This introduces some additional experimental and analytical considerations. For example, in future, it may be worthwhile considering more carefully what melodic features to use in a regression model. In this case, the rhythmic predictability of the stimuli renders the onset surprise less useful than it would be for stimuli with less predictable rhythms. Stimuli with greater and more complex rhythmic properties could also be included deliberately as a point of contrast. There is also the option of deliberately making the pitch of the stimuli more surprising by using more complex renditions of the stimuli with a wider range of notes.

This lack of surprise also presented another problem. When adding higher dimensional features to an existing model, adding multiple melodic surprise features to acoustic models in this case, there is always the risk that performance gains may merely be due to the added dimensionality. To account for this, the melodic features can be randomly shuffled. If the performance gains are still present, one could say it is due to the added dimensionality and not the information contained within the new features. This shuffling can be done by shifting the indexes of the values in the vector by a random increment (Di Liberto et al., 2020), which still preserves the order of the values but misaligns them with the neural response. In this case that did not work, which may have been because the stimuli were so repetitive that shifting indexes could result in surprise values effectively appearing in roughly the right place. This is particularly true for the onset surprise and onset entropy values, as these were very repetitive and narrow in range. To work around this, the values instead needed to be completely reordered at random to be used as a control.

There is also the speech component of nursery rhymes which means that the listener must also process this important non-melodic information. Maidhof and Koelsch (2010) found decreased ERP responses to irregular chords when speech was presented to subjects simultaneously in addition to music. An interesting point of comparison in future work would be to regress the EEG with speech features, such as phonemes and word onsets, in addition to melodic features to explore the individual and combinatory effects of these elements.

This combination of speech and music could also potentially alter the attention of the listener, which could be reflected in their responses. Maidhof and Koelsch (2010) found evidence of this. Attention can be thought of as paying specific attention to part or all of a stimulus. Attention has been shown to enhance neural responses to speech compared to unattended speech (Obleser and Kayser, 2019) e.g. when focussing on listening to a specific person in a room full of talking people, the brain responds most strongly to the speech of that person. It is possible that different responses could emerge depending on a listeners attention to the speech or melodic components of the stimuli. This would be another interesting area to investigate using continuous naturalistic stimuli.

The concept of attention raises an interesting point about the results from the infants' data. While the effect was not significant, there is an increase in the encoding of the melodic features as the subjects aged. It could be argued that this was because they were older and hence paid more attention to the stimuli while they were playing. Previous work on this same dataset by Attaheri et al. (2022) investigated how well the acoustic envelope was encoded in the subjects' EEG responses. In this study, they actually found a decrease in the reconstruction correlation between 4 and 7 month old subjects between 0.5 - 8 Hz (See Appendix Figure 3). So, if this increase in the encoding of the melodic surprise features was purely caused by attention, this increase over time should also have been evident in this previous work, which it was not.

There is an open question as to how best to model the melodic surprise of stimuli for infants. Previous works on estimating melodic surprise have focused on adult perception. IDyOM studies have also focused on adults and not infants. As such, there is no current literature on how best to approach this. While a STM model was used for this purpose here because it's outputs were shown to work best when regressed with the adults' EEG data, one could make the intuitive argument that a STM is indeed suitable for young infants, as they will have had very little prior musical experience and hence less of a preconceived notion of musical structure and must "learn as they go" like a short-term model.

IDyOM does have parameters intended to simulate the amount of short-term memory available to a listener. We know that short-term memory improves hugely as children develop (Gathercole, 1999), so one would expect that simulating a listener with smaller amounts of short-term memory would more accurately represent a young child. This can even occur over relatively short intervals. Oakes et al. (2013) found a significant difference between the visual short-term memory of 6 and 8 month old children in an eye-tracking experiment. Hence, this simulated amount of short-term memory could be increased in line with the children's age at the different longitudes of the experiment as they age and get closer to adult perception. Using simpler and smaller training sets for long-term models would be another possibility here. The size of these training sets could also increase with age to model the increasing amount of musical exposure experienced by the ageing infants. Investigating how effectively other models of melodic perception such as the Music Transformer Model performed at this task could also be of interest. Due to the time constraints of this project, comparing many different training sets, memory configurations or models was not able to be explored, but these are factors which I believe should be considered for future works in this area.

A dataset with responses to a greater number of stimuli would also be useful in future works. In Di Liberto et al. (2020)'s study of melodic perception of monophonic music, 30 stimuli were played to 20 subjects. For this project, by focussing on the sung subset of the stimuli, there were EEG responses from 10 stimuli to 17 subjects. This always made it likely that the size of any effect found, if any, would be relatively smaller than this previous study.

It should be noted with the data used in this study that the same stimuli were played to the infants at each age interval. Ideally, this would not have been the case as it runs the risk of the subjects remembering parts of the stimuli and being less surprised by them. The simplicity of the stimuli increases the likelihood of this as mentioned in Section 2. It may have been possible to somewhat account for this by including some of the stimuli, or pieces with similar melodic properties, in the training set of a LTM, perhaps in accordance with age.

The note onset times used here to create the surprise feature vectors were also manually created by the original data collectors, so there is a chance these were slightly misaligned with the EEG responses. While the TRF does look inside a time lag window and so misalignments likely would not cause values to be totally "missed", it would still be desirable for these to be a close as possible to the true note onset times. This highlights one difficulty of using naturalistic stimuli over artificially created stimuli, where the note onset timings would be known precisely.

Different types of linear regression besides the ridge regression-based TRF model used here would be another possibility to explore in future. Banded ridge regression is one of these options. Rather than applying a singular regularisation parameter to every feature, banded ridge regression can apply varying amounts of regularisation to different feature sets to reduce correlation between features. This is also useful when features are of different scales whereby the optimal amount of regularisation required could be different for each feature. It has been shown in the past to produce superior prediction accuracy over ridge regression on fMRI data (Nunez-Elizalde et al., 2019). An implementation of banded ridge regression is available in the mTRF Toolbox in Matlab.

Lastly, Section 3 highlighted the usefulness of a tool to convert sound files to musical sheets. IDyOM and the Music Transformer model require their inputs to be in this form. Two current solutions which attempt to do this were discussed in Section 2, although the output of these would have needed manual adjustments which require a degree of skill, time, and perhaps subjectivity as well. Manually correcting or creating MIDI files will become increasingly difficult as moves are made away from artificial monophonic stimuli, for which a midi file can be more easily attained (if it was not used to create the stimulus in the first place), to more complex naturalistic forms of music such as live polyphonic music with multiple instruments and/or voices.

Looking back on this work on a personal level, I have learned a lot over the course of this dissertation. Stepping largely outside of computer science and into a completely new subject area presented a lot of challenges but was something which drew me to this project in the first place. Initially, even reading literature was a huge struggle as I was unfamiliar with even basic jargon. I had also not used Matlab in four years. This was all quite intimidating at first. Looking back, there is a great sense of accomplishment seeing what has been achieved. On a technical level, I gained experience in Matlab, Python, and the various libraries used therein, as well as Excel and Tableau for data processing and visualisation. I believe the project has been very beneficial and rewarding. There have been very few times in my formal education to date where there has been an opportunity to work on a large-scale, longer-term project of this nature. It requires a greater degree of self-motivation and discipline to stay on target, experiences which are valuable in all walks of life.

Bibliography

Agarwal, A., Dowsley, R., McKinney, N. D., Wu, D., Lin, C.-T., De Cock, M., and Nascimento, A. C. A. (2019). Protecting privacy of users in brain-computer interface applications.

Agres, K., Abdallah, S., and Pearce, M. (2018). Information-theoretic properties of auditory sequences dynamically influence expectation and memory. *Cognitive Science*, 42(1):43–76.

Angulo-Perkins, A., Aubé, W., Peretz, I., Barrios, F. A., Armony, J. L., and Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. *Cortex*, 59:126–137.

Attaheri, A., Áine Ní Choisdealbha, Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., and Goswami, U. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, 247:118698.

Bartlett, J. and Dowling, W. J. (1980). Recognition of transposed melodies: a key-distance effect in developmental perspective. *Journal of experimental psychology. Human perception and performance*, 6 3:501–15.

Bartlett, J. C., Halpern, A. R., and Dowling, W. J. (1995). Recognition of familiar and unfamiliar melodies in normal ageing and alzheimer's disease. *Memory & Cognition*, 23(5):531–546.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–312.

Brodbeck, C., Das, P., Kulasingham, J. P., Bhattasali, S., Gaston, P., Resnik, P., and Simon, J. Z. (2021). Eelbrain: A python toolkit for time-continuous analysis with temporal response functions. *BioRxiv*.

Centre for Internet Security (2019). Eternal blue, https://www.cisecurity.org/wp-content/uploads/2019/01/security-primer-eternalblue.pdf.

Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., and Jung, T.-P. (2019). Evaluation of artifact subspace reconstruction for automatic artifact components removal in Multi-Channel EEG recordings. *IEEE Trans Biomed Eng*, 67(4):1114–1121.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Cognition and Natural Sensory Processing Workshop (2021). Resources. https://cnspworkshop.net/resourcesCNSP2021.html.

Cohen, A. J., Cuddy, L. L., and Mewhort, D. J. K. (1977). Recognition of transposed tone sequences. *The Journal of the Acoustical Society of America*, 61(S1):S87–S88.

Crosse, M. J., Di Liberto, G. M., Bednar, A., and 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*, 10:604.

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., and Lalor, E. C. (2021). Linear modeling of neurophysiological responses to speech and other continuous stimuli: Methodological considerations for applied research. *Frontiers in Neuroscience*, 15.

Cuddy, L. and Lyons, H. (1981). Musical pattern recognition: A comparison of listening to and studying tonal structures and tonal ambiguities. *Psychomusicology: A Journal of Research in Music Cognition*, 1.

Di Liberto, G. M. (2017). *Isolating neural indices of continuous speech processing from multivariate neural data*. PhD thesis.

Di Liberto, G. M., Marion, G., and Shamma, S. A. (2021). The music of silence: Part ii: Music listening induces imagery responses. *Journal of Neuroscience*, 41(35):7449–7460.

Di Liberto, G. M., Pelofi, C., Bianco, R., Patel, P., Mehta, A. D., Herrero, J. L., de Cheveigné, A., Shamma, S., and Mesgarani, N. (2020). Cortical encoding of melodic expectations in human temporal cortex. *eLife*, 9:e51784.

Ding, N. and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in Human Neuroscience*, 8.

Di Liberto, G., O'Sullivan, J., and Lalor, E. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25(19):2457–2465.

Dowling, W. J. (1999). 15 - the development of music perception and cognition. In Deutsch, D., editor, *The Psychology of Music (Second Edition)*, Cognition and Perception, pages 603–625. Academic Press, San Diego, second edition edition.

Dzafic, I., Randeniya, R., Harris, C. D., Bammel, M., and Garrido, M. I. (2020). Statistical learning and inference is impaired in the nonclinical continuum of psychosis. *Journal of Neuroscience*, 40(35):6759–6769.

Egermann, H., Pearce, M. T., Wiggins, G. A., and McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience*, 13(3):533–553.

Erickson, L. C. and Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, 37:66–108.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3(11):410–419.

Giraud, A.-L. and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, 15(4):511—517.

Gordon, R. L., Schön, D., Magne, C., Astésano, C., and Besson, M. (2010). Words and melody are intertwined in perception of sung words: Eeg and behavioral evidence. *PLOS ONE*, 5(3):1–12.

Guardicore Labs (2021). Threats making wavs - incident response to a cryptomining attack, https://www.akamai.com/blog/security/threats-making-wavs-incident-reponse-cryptomining-attack.

Halpern, A. R., Zioga, I., Shankleman, M., Lindsen, J., Pearce, M. T., and Bhattacharya, J. (2017). That note sounds wrong! age-related effects in processing of musical expectation. *Brain and Cognition*, 113:1–9.

Harimohan, A. (2021). Understanding neural signals related to speech processing in humans during sleep.

Heilbron, M. and Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*, 389:54–73. Sensory Sequence Processing in the Brain.

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A. M., Hoffman, M. D., Dinculescu, M., and Eck, D. (2018). Music transformer.

Ilari, B. S. (2002). Music perception and cognition in the first year of life. *Early Child Development and Care*, 172(3):311–322.

Jentschke, S., Koelsch, S., Sallat, S., and Friederici, A. D. (2008). Children with Specific Language Impairment Also Show Impairment of Music-syntactic Processing. *Journal of Cognitive Neuroscience*, 20(11):1940–1951.

Jeste, S. S., Frohlich, J., and Loo, S. K. (2015). Electrophysiological biomarkers of diagnosis and outcome in neurodevelopmental disorders. *Current opinion in neurology*, 28 2:110–6.

Jia, J., Liu, L., Fang, F., and Luo, H. (2017). Sequential sampling of visual objects during sustained attention. *PLOS Biology*, 15(6):1–19.

Jost, E., Conway, C. M., Purdy, J. D., Walk, A. M., and Hendricks, M. A. (2015). Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain Research*, 1597:95–107.

Karita, S., Chen, N., Hayashi, T., Hori, T., Inaguma, H., Jiang, Z., Someki, M., Soplin, N. E. Y., Yamamoto, R., Wang, X., et al. (2019). A comparative study on transformer vs rnn in speech applications. In *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 449–456. IEEE.

Kern, P., Heilbron, M., de Lange, F. P., and Spaak, E. (2022). Cortical activity during naturalistic music listening reflects short-range predictions based on long-term experience. *bioRxiv*.

Kurthen, I., Galbier, J., Jagoda, L., Neuschwander, P., Giroud, N., and Meyer, M. (2021). Selective attention modulates neural envelope tracking of informationally masked speech in healthy older adults. *Human Brain Mapping*, 42(10):3042–3057.

Maidhof, C. and Koelsch, S. (2010). Effects of selective attention on syntax processing in music and language. *J Cogn Neurosci*, 23(9):2252–2267.

Marion, G., Di Liberto, G. M., and Shamma, S. A. (2021). The music of silence: Part i: Responses to musical imagery encode melodic expectations and acoustics. *Journal of Neuroscience*, 41(35):7435–7448.

Moldwin, T., Schwartz, O., and Sussman, E. (2017). Statistical learning of melodic patterns influences the brain's response to wrong notes. 2114–2122.

Norman-Haignere, S., Kanwisher, N. G., and McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, 88(6):1281–1296.

Norman-Haignere, S. V., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J. H., Schalk, G., and Kanwisher, N. (2022). A neural population selective for song in human auditory cortex. *Current Biology*, 32(7):1470–1484.e12.

Nunez-Elizalde, A. O., Huth, A. G., and Gallant, J. L. (2019). Voxelwise encoding models with non-spherical multivariate normal priors. *NeuroImage*, 197:482–492.

Oakes, L., Baumgartner, H., Barrett, F., Messenger, I., and Luck, S. (2013). Developmental changes in visual short-term memory in infancy: evidence from eye-tracking. *Frontiers in Psychology*, 4.

Obleser, J. and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends Cognitive Science*, 23(11):913–926.

Omigie, D., Pearce, M. T., Williamson, V. J., and Stewart, L. (2013). Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia*, 51(9):1749–1762.

Overath, T., McDermott, J. H., Zarate, J. M., and Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience*, 18(6):903–911.

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. PhD thesis, City University London.

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, 1423(1):378–395.

Perruchet, P. and Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in cognitive sciences*, 10(5):233–238.

Plantinga, J. and Trainor, L. J. (2005). Memory for melody: infants use a relative pitch code. *Cognition*, 98(1):1–11.

Plantinga, J. and Trainor, L. J. (2009). Melody recognition by two-month-old infants. *The Journal of the Acoustical Society of America*, 125(2):EL58–EL62.

Plechawska-Wojcik, M., Kaczorowska, M., and Zapała, D. (2019). *The Artifact Subspace Reconstruction (ASR) for EEG Signal Correction. A Comparative Study*, pages 125–135.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, 41(1):245–255. The Nature of Speech Perception.

Quiroga-Martinez, D., Hansen, N., Højlund, A., Pearce, M., Brattico, E., and Vuust, P. (2020). Decomposing neural responses to melodic surprise in musicians and non-musicians: Evidence for a hierarchy of predictions in the auditory system. *NeuroImage*, 215:116816.

Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M. T., Brattico, E., and Vuust, P. (2019a). Reduced prediction error responses in high-as compared to low-uncertainty musical contexts. *Cortex*, 120:181–200.

Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M. T., Brattico, E., and Vuust, P. (2019b). Reduced prediction error responses in high-as compared to low-uncertainty musical contexts. *Cortex*, 120:181–200.

Reetzke, R., Gnanateja, G. N., and Chandrasekaran, B. (2021). Neural tracking of the speech envelope is differentially modulated by attention and language experience. *Brain and Language*, 213:104891.

Reynolds, K. (2022). Investigating melodic expectation with eeg and polyphonic music. Master's thesis, Trinity College Dublin.

Rohrmeier, M. and Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science*, 4(4):525–553.

Salamon, J. and Gomez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6):1759–1770.

Sallat, S. and Jentschke, S. (2015). Music perception influences language acquisition: Melodic and rhythmic-melodic perception in children with specific language impairment. *Behavioural Neurology*, 2015:606470.

Salmelin, R. (2007). Clinical neurophysiology of language: The meg approach. *Clinical Neurophysiology*, 118(2):237–254.

Schabus, M., Del Giudice, R., Wislowska, M., Lang, A., Lithari, C., Weisz, N., and Calvet, C. (2016). Processing of nursery rhymes in the newborn brain: A hdeeg study. 20th International Conference on Biomagnetism, BIOMAG2016 ; Conference date: 01-10-2016 Through 06-10-2016.

Soni, A., Barth, J., and Marks, B. (2019). Blackberry blogs: Malicious payloads - hiding beneath the wav, https://blogs.blackberry.com/en/2019/10/malicious-payloads-hiding-beneath-the-wav.

Soufineyestani, M. (2020). *Physiological and psychological effects of listening to nursery rhymes*. PhD thesis.

Soufineyestani, M., Khan, A., and Imtiaz, R. (2021). Physiological and psychological reactions to a musician, robot, or boombox music player: Comparison between eda, hrv, and eeg. *Neurol Res Surg*, 4(1):1–8.

Stalinski, S. M. and Schellenberg, E. G. (2012). Music cognition: A developmental perspective. *Topics in Cognitive Science*, 4(4):485–497.

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science*, 32(2):418–444.

Trehub, S. E. (1987). Infants' perception of musical patterns. *Perception & Psychophysics*, 41(6):635–641.

Vanthornhout, J., Decruy, L., and Francart, T. (2019). Effect of task and attention on neural tracking of speech. *Frontiers in Neuroscience*, 13.

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., and Tervaniemi, M. (2012). The sound of music: differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia*, 50(7):1432–1443.

Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., and Chao, L. S. (2019). Learning deep transformer models for machine translation. *arXiv preprint arXiv:1906.01787*.

Zuk, N. J., Teoh, E. S., and Lalor, E. C. (2020). Eeg-based classification of natural sounds reveals specialized responses to speech and music. *NeuroImage*, 210:116558.

Appendices

.1 Supplemental Figures

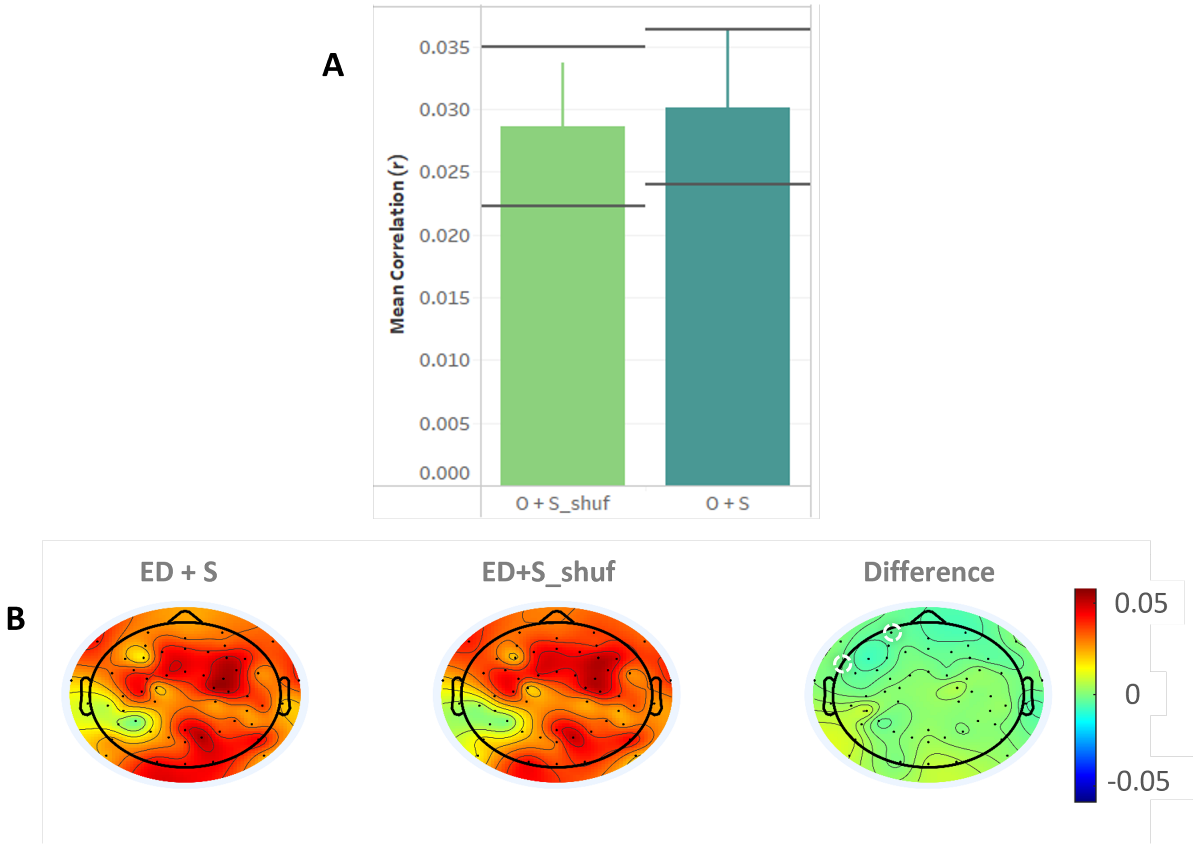

Figure 2: (**A**) Average correlation (r) of reconstructed EEG signal across all electrodes for TRF models trained with Note Onset Times + Randomly Shuffled Surprise (O + S_shuf) and Note Onset Times + Surprise (O + S). The difference was not statistically significant (p = 0.12, permutation test). (**B**) Average correlation (r) of reconstructed EEG signal for each individual electrode. Two significant electrodes from initial t-test (circled in white in Difference topography plot). After running FDR correction, no significant electrodes were found.

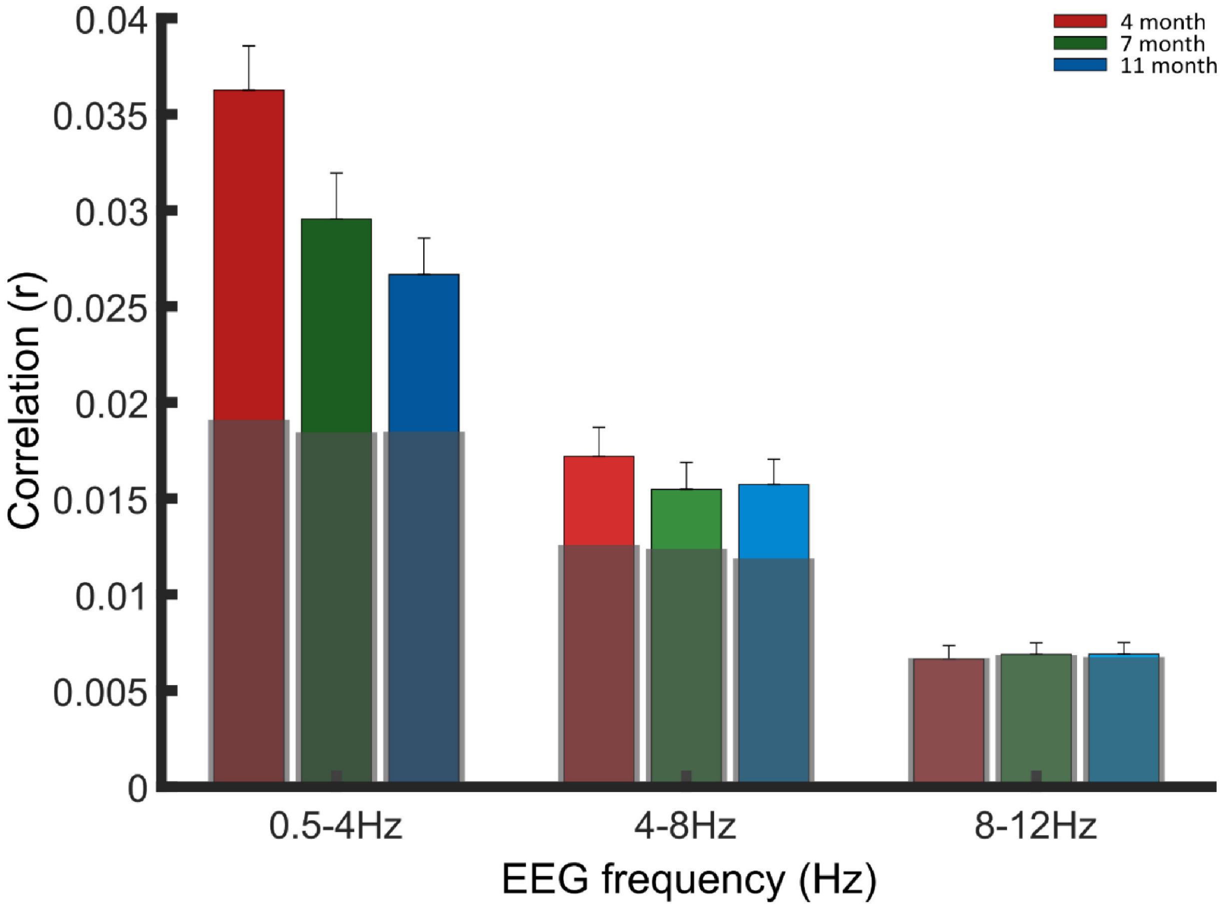

Figure 3: Average correlation (r) for reconstructed EEG signal from TRF model regressed on the nursery rhyme stimulus envelopes across 4, 7 and 11 months of age. (Attaheri et al., 2022)

.2 Security and Privacy Considerations

This work was completed as part of the CS7NS5: Security and Privacy module, where the intention was that it could be included in students' dissertations. It was adapted slightly from the original submitted work for inclusion here.

While the offline nature of this research is quite closed to outside threats, there are still various security and privacy considerations which needed to be considered for this specific project, and for similar ones in this area.

In this project, analysis was performed performed on adults' and infants' EEG responses as they listened to nursery rhymes. Before this could be performed, the sound files for the nursery rhymes had to be analysed so that relevant features could be correlated and regressed with the EEG responses. The files used for this dissertation are in the WAV file format. While they do not contain executables, it has been shown before that WAV files can contain hidden malware. Soni et al. (2019) had previously uncovered several attacks in which malware was inside an otherwise-clean looking WAV file. One of which was designed to mine cryptocurrency using the target machine, as well as one which could run a reverse shell into the machine. Guardicore Labs (2021) detailed another of such stenographic attacks in which malware which mined the Monero cryptocurrency was hiding inside a WAV file. This made use of the NSA's EternalBlue (Centre for Internet Security, 2019) vulnerability on Windows 7 to propagate through a network. This highlights the importance to researchers and organisations of installing the latest OS security patches, as well as the risks of running End-of-Life software or operating systems.

The audio files used in this dissertation came directly from the head of the original study. These were sent to my supervisor, who in turn sent them to me. This flow of information should be trustable. A more worrisome scenario would be a fake researcher appearing in good-faith and posting audio files containing malware. While this was not an issue in this dissertation, it is certainly something that future works, particularly when using publicly available data, should consider. Any initiatives and platforms which intend to openly host and share data should have procedures in place to scan and find hidden malware.

In terms of privacy, it is important to have data anonymised and not traceable back to the subjects. The EEG recordings used in this dissertation are simply numbered i.e. Subject 1, Subject 2, etc. As far as this particular study is concerned, we are blind to the real names of the subjects. The only way to directly link this back to a real person would be by cross referencing these names with the identity keys stored in the institution where the data was originally collected i.e. Cambridge University.

In this instance, even if the EEG and associated subject names were to leak/be stolen, they do not reveal anything meaningful about the subjects. It is merely neural responses from adults and infants listening to someone singing nursery rhymes. Nothing malicious can currently be inferred from this. Nonetheless, this could change with new research e.g. if it was discovered that such recordings gave an insight into an adults mental capacity, or how a infant would cognitively develop.

There are scenarios involving data from EEG or from neural imaging techniques such as fMRI or MRI where these anonymisation and data protection aspects could have much higher-stakes. These techniques could be used to measure the brain activity of subjects with neurological disorders e.g. schizophrenia. This can be by design i.e. a study of these subgroups specifically, or the data could have been collected for diagnostic purposes, perhaps even from a medical setting. A privacy breach here would amount to leaking sensitive medical information about the subject. Work in the field such as that by Agarwal et al. (2019), on privacy-preserving machine learning techniques for EEG would be worth investigating for these use cases.

It is possible that data or metadata contained within the EEG data or audio stimuli could reveal some privacy eroding information e.g. geographical location, timestamps. In the EEG files used here, there are no timestamps as to when they were created. There is one piece of metadata in the EEG files called 'deviceName' which gives the name of the device used to make the EEG recordings: "EGI, GES 300 amplifier, Geodesic Sensor Net". Similarly, the WAV files were also inspected. It is possible to encode the original recording time of a WAV file in it's metadata. This information was not present in the files used for this dissertation, but is something worth being aware of.