Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

8-11-2023 11:00 AM

Temporal dynamics of natural sound categorization

Ali Tafakkor, *Western University*

Supervisor: Mohsenzadeh, Yalda., *The University of Western Ontario*
Co-Supervisor: Johnsrude, Ingrid S., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Neuroscience
© Ali Tafakkor 2023

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Computational Neuroscience Commons

Recommended Citation

Tafakkor, Ali, "Temporal dynamics of natural sound categorization" (2023). *Electronic Thesis and Dissertation Repository*. 9452.
https://ir.lib.uwo.ca/etd/9452

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

While extensive research has elucidated the brain's processing of semantics from speech sound waves and their mapping onto the auditory cortex, the temporal dynamics of how meaningful non-speech sounds are processed remain less examined. Understanding these dynamics is key to resolving the debate between cascaded and parallel hierarchical processing models, both plausible given the anatomical evidence. This study investigates how semantic category information from environmental sounds is processed in the temporal domain, using electroencephalography (EEG) collected from 25 participants and representational similarity analysis (RSA) along with models of acoustic and semantic information. We examined information extracted by the brain from 80 one-second natural sounds across four categories. The results revealed a cascaded temporal hierarchy of processing of information towards identifying the sound category, which supports the well established anatomical hierarchy. Low-level information is decodable at ~ 30 ms, and semantic information begins to emerge ~ 40 ms later. We conclude that basic information transforms to more complex information over time, while semantic representations are more stable over time than representations of acoustic information.

Keywords: temporal hierarchy, naturalistic sound, semantic category, EEG, RSA, MVPA, temporal generalization

Summary for lay audience

Our brains are constantly processing information from our surroundings to help us understand and interact with the world. One way we do this is through grouping similar objects, events, or ideas based on their shared characteristics or meanings. This process, called semantic categorization, is not just limited to what we see, but also extends to sounds we hear. Consider the non-speech sounds you hear daily, like a kettle's whistle. These sounds carry important 'semantics' or meanings that help us understand our environment. By studying how our brain processes these sounds, we can gain insights into our cognitive abilities and understand how they might be affected by aging or neurological disorders.

The processing of sounds in our brain involves multiple stages, from the initial reception of the sound waves in our ears to the complex functions carried out by the cortex, the outer layer of our brain. Studies have shown that there are multiple regions in the brain involved in processing sounds, and these regions are interconnected. Research has also shown that these different regions are responsible for processing different levels of information, from basic acoustic features to complex semantic meanings. The question is whether the complexity of the information processed also increases over time.

In a recent study, we used electroencephalography (EEG), a method that records electrical activity of the brain, to investigate how the brain processes semantic information from environmental sounds over time. We found that the brain first decodes basic information about the sound around 30 milliseconds after hearing it, and then starts to extract semantic information about 40 milliseconds later. This suggests that our brain transforms basic information into more complex information over time, which confirms our knowledge from the brain anatomy of auditory processing regions.

In simple terms, when we hear a sound, our brain quickly identifies its basic features and then takes a bit more time to figure out what it means, instead of these two stages happening simultaneously. This process helps us understand and respond to our environment effectively.

Acknowlegements

I extend my deepest gratitude to my supervisors, Dr. Yalda Mohsenzadeh and Dr. Ingrid Johnsrude, for their invaluable guidance and unwavering support throughout this thesis. I am equally thankful to my lab members for their enriching discussions and constructive feedback, which have been instrumental in shaping this work. I would also like to express my sincere appreciation to the participants of this study, whose cooperation was vital to the success of this research. My heartfelt thanks go to my family and friends for their constant encouragement and companionship. This thesis is a testament to their unwavering faith and support, and I am deeply indebted to each one of them for their contributions to my academic journey.

Contents

Abstract ii

Summary for lay audience iii

Acknowlegements iv

List of Figures viii

List of Tables ix

List of Appendices x

1 Introduction **1**

1.1 General Introduction 1

1.2 Research question 3

1.3 Auditory processing stages 4

1.3.1 Cochlea 4

1.3.2 Brainstem auditory-related areas 5

1.3.3 Forebrain auditory-related areas 6

1.4 Electroencephalography: a tool to track information flow in auditory pathways 7

1.5 Representational Similarity Analysis: a method for decoding semantic representations 9

1.6 Hierarchical processing of auditory semantic information 11

1.7 Current study 20

2 Methods **23**

2.1 Participants 23

2.2 Stimuli 23

2.3 Experiment design and procedure 24

2.3.1 Behavioral Pilot Experiment 24

Familiarization task 24

Recognition task 24

2.3.2 EEG Experiment 25

2.4 EEG acquisition and preprocessing 25

2.5 Multivariate pattern analysis (MVPA) 26

2.6 Temporal generalization analysis 27

2.7 Representational similarity analysis (RSA) 28

2.8 Models 30

2.8.1 Auditory processing models 30

2.8.2 Semantic models 30

2.8.3 Noise ceiling 32

2.9 Statistics 32

2.9.1 Cluster-correction permutation test 32

2.9.2 Onset and peak time identification 33

3 Results **34**

3.1 Behavioral results provide a refined model for category perception . . . 34

3.2 Decoding time series show a latency in emergence of category information compared to non-category information 37

3.3 Temporal generalization depicts category information is more sustained 40

3.4 Representational similarity analysis distinguishes levels of information . 41

4 Discussion **44**

4.1 Hierarchical auditory processing 45

4.2 Temporal hierarchy of auditory semantic processing 46

4.3 Limitations 49

4.4 Conclusion 51

References **52**

A Health sciences research ethics board approval letter **59**

Curriculum Vitae **61**

List of Figures

2.1 Experiments’ design 26
2.2 Multivariate pattern analysis (MVPA) 28
2.3 Representational similarity analysis (RSA) 31

3.1 Behavior experiment results 36
3.2 Decoding time series analysis results 39
3.3 Temporal generalization analysis results 41
3.4 Representational similarity analysis results 43

List of Tables

3.1 Decoding onset and peak time summary 39

3.2 Models correlation peak time comparison 43

List of Appendices

Appendix A 59

Chapter 1

Introduction

1.1 General Introduction

Semantic categorization, a cognitive process that involves grouping similar objects, events, concepts, or ideas based on their shared characteristics or meanings, facilitates inference and conceptualization of our environment. It plays a pivotal role in humans' survival and development as it underpins comprehension and learning, efficient information retrieval, communication and language understanding, decision-making, and problem-solving. Semantic categorization is a ubiquitous cognitive phenomenon that happens in all sensory modalities, including hearing. This phenomenon and processing of the semantic information that comes through the auditory sense is mostly and extensively studied in the context of speech and language. However, semantic information also exists in and can be inferred from non-linguistic auditory stimuli. Semantics inferred from non-speech sounds can have a significant role in how humans make sense of their environment and in human communication. For example, consider the non-speech sounds of morning birdsong and a kettle whistling. Birdsong at dawn signals the beginning of a new day, serving as a natural alarm. Meanwhile, the whistle of a kettle conveys that water has reached boiling point, preparing us for morning routines like making tea or coffee. These auditory cues provide crucial semantic information, subtly guiding our actions at the start of each day, showcasing how non-speech sounds play a fundamental role in our daily lives by communicating important concepts. Investigating how the brain accomplishes semantic classification of natural sounds may deepen our understanding of broader cognitive abilities involving semantic categorization such as attention, memory, and decision-making. Furthermore, it may also provide us with valuable approaches to examine how these mechanisms may be affected by aging or neurological disorders. Hence, the question of how the brain processes semantic data from natural, non-speech auditory stimuli presents a compelling area of study.

Processing of sounds consists of multiple stages from transduction of acoustic waves to neural codes in the cochlea to complex functions carried out by the cortex. The auditory cortex is defined as the cortical areas that receive their primary input from auditory related subcortical regions. Studies on mammalian species, including non-human primates, have shown there are multiple auditory regions in the brain with lateral connections. Anatomical and functional organization of the auditory-related areas has been quite well studied in non-human mammals (Hackett et al., 2001; Kaas et al., 1999; Hackett et al., 1998; Kaas and Hackett, 2000; Hackett, 2011; Carrasco and Lomber, 2009). However, in humans, understanding of the functional organization generally relies on functional imaging techniques, and remains to be studied in more detail. These functional imaging techniques have been used to determine areas responsive to certain types of information or responsible for certain processes, allowing us to investigate the functional roles and organization of auditory-related brain regions. For instance, correspondence between the auditory stages of processing and the auditory evoked responses recorded by electroencephalogram (EEG) has been established based on different latencies for the signal to travel through the auditory pathway. Auditory evoked potentials (AEPs) have been studied for over fifty years and have been employed as standard procedures in clinical practice. However, AEPs are not highly informative about the functional role of processing stages due to methodological limitations including frequent reliance upon very simple stimuli and inter-individual variability.

Moreover, studies have shown that the human brain processes auditory semantics in a hierarchy of regions in the auditory cortex (Binder et al., 2000; Davis and Johnsrude, 2003; Okada et al., 2010; DeWitt and Rauschecker, 2012; Yi et al., 2019). This step-by-step processing scheme has been mostly shown by speech studies but has also been demonstrated by studies using non-speech sounds (Kell et al., 2018; Lowe et al., 2021; Benner et al., 2023; Giordano et al., 2023). These studies have primarily relied upon functional Magnetic Resonance Imaging (fMRI), which offers the highest spatial resolution to indirectly infer neural activity non-invasively. Using this imaging technique, brain regions have been identified as responsible for processing distinct levels of information from low-level, acoustic features to high-level, semantic features of sound (Binder et al., 2000; Davis and Johnsrude, 2003; Okada et al., 2010; DeWitt and Rauschecker, 2012; Yi et al., 2019; Kell et al., 2018; Lowe et al., 2021; Benner et al., 2023; Giordano et al., 2023). Additionally, it has been shown that those functionally distinct regions are organized in a specific order resembling a hierarchy. However, hierarchical processing implies a temporal hierarchy as well as a spatial. All brain regions within this hierarchy receive direct thalamocortical projections (Hackett, 2011). While it might be anticipated that the levels of information processed increase in complexity over time, forming a cascaded hierarchy, these parallel thalamocortical connections potentially allow a parallel hierarchy as well. Consequently, it is imperative to identify the flow of information within this hierarchy. However, the use of fMRI does not provide the necessary temporal resolution to investigate this, hence imaging methods with high temporal resolution should be employed to validate the presence of a cascaded hierarchy.

1.2 Research question

The main question of this study is whether low-level acoustic information and higher-level semantic information are decodable from brain activity at different latencies, with the former earlier than the latter. In that case, we will have more evidence in favor of a hierarchical functional organization of auditory-related brain regions, and it will shed light on information flow in the brain when processing semantic information. Therefore, we aim to investigate the brain representations of sounds through time and to examine the nature of these representations and their temporal dynamics.

1.3 Auditory processing stages

Essentially, sound is a mechanical vibration propagating in a medium such as air as a pressure wave. Alternating patterns of high- and low-pressure air carry energy away from the sound source to be captured by the external ear. The pinna reflects and focuses the sound into the ear canal, which culminates in translational movement of the eardrum (tympanic membrane). The eardrum vibrates as the sound waves hit it, and these vibrations are transmitted through the three small bones of the middle ear, called the ossicles (malleus, incus, stapes), to the cochlea in the inner ear.

1.3.1 Cochlea

The cochlea transduces the mechanical vibrations of sound to electrical signals interpretable by the brain. The cochlea resembles a small snail-like structure whose diameter shrinks as it winds around a bony core, and the whole structure is embedded in the temporal bone. The cochlea is divided into three fluid-filled parallel chambers, and the vibrations of the eardrum transmit to the fluids in the cochlear chambers through the ossicles and the oval window (a window at the base of the top chamber). The top and bottom chambers only connect at the apex of the cochlea (helicotrema). The pressure changes induced by the movement of stapes propagate through the top chamber's fluid and eventually travels to the bottom chamber which causes a traveling wave to move along the basilar membrane (the membrane separating middle and bottom chambers). As the traveling wave moves toward the cochlear apex its amplitude grows to a maximum and then plummets. The basilar membrane varies in stiffness from base to apex, with the stiffest part at the base and the floppiest part at the apex, and the cochlear chambers get smaller from the base toward the apex. These variations of the physical attributes of the cochlea and basilar membrane along their length makes the position where the traveling wave reaches its maximum amplitude depend on the sound's frequency. The basilar membrane responds best to the highest and lowest audible frequencies (20 kHz and 20 Hz) at its base and apex respectively, and intermediate frequencies are represented along the basilar membrane forming a continuous logarithmic tonotopic map. Auditory transduction happens in the organ of Corti that lies along the middle chamber attached to the basilar membrane. The organ of Corti consists of different cell types, among which are the inner and outer hair cells. The inner hair cells rapidly transform mechanical displacement of the basilar membrane to a voltage change, and they form synapses to sensory neurons. Finally, the bipolar neurons with their cell bodies in the spiral ganglion take information from cochlea toward brain in the tonotopically mapped cochlear nerve.

1.3.2 Brainstem auditory-related areas

The auditory neural pathways begin at the ear and extend to the cerebral cortex. The information is first conveyed from the spiral ganglion to various neurons in the cochlear nuclei within the brainstem. These neurons then transport the information via different paths to the brainstem and midbrain. Some connections lead directly to the contralateral inferior colliculus, while others pass through more synaptic stages. From the bilateral inferior colliculi, sound data is relayed to the ipsilateral superior colliculus (helping in orientation towards sounds) and the ipsilateral thalamus (the gateway to the cerebral cortex's auditory areas). Notably, these auditory pathways involve efferent feedback at multiple stages (Warr, 1992; Darrow et al., 2006; Kandel et al., 2021).

The cochlear nuclei exhibit tonotopic organization where fibers from the cochlea's apex to base (carrying low- to high-frequency information) terminate ventrodorsally in the ventral and dorsal cochlear nuclei. Additionally, every cochlear nerve fiber innervates multiple areas within the cochlear nuclei, contacting different neuron types with unique projection patterns to higher contralateral auditory centers such as superior olivary complex, nucleus of the lateral lemniscus, and inferior colliculus. Consequently, the auditory pathway consists of at least four parallel ascending pathways, each extracting different acoustic information.

Neurons within the superior olivary complex use the neural activity from the bilateral cochlear nuclei to pinpoint the sources of sounds. They have separate circuits specifically for detecting interaural time and intensity differences respectively in Medial and Lateral Superior Olive (MSO and LSO), which then project this information to the inferior colliculi. The Superior Olivary Complex also provides feedback to the cochlea, enabling a dynamic response to auditory stimuli.

The inferior colliculus is the major auditory hub which organizes and refines all ascending auditory signals. It is subdivided into the central nucleus, dorsal cortex, and external cortex. The central nucleus contains neurons that carry information about sound source locations. These neurons receive a broader spectral range of inputs than at previous stages in the auditory pathway. However, inhibitory processes help narrow the responses of excitatory neurons. The inferior colliculus plays a pivotal role in transmitting auditory information to the cerebral cortex, with both ascending and descending pathways involving the medial geniculate body of the thalamus. Neurons at higher stages are sensitive to progressively more complex features along this ascending auditory pathway, allowing for multidimensional responses to sound based on aspects like frequency, bandwidth, modulation frequency, intensity, and location.

1.3.3 Forebrain auditory-related areas

Neurons in the inferior colliculus send signals to the superior colliculus and the medial geniculate nucleus (MGN) in the thalamus. The MGN then projects to the cortex via core and belt pathways. The MGN is an intermediate processing region for sensory information and is thought to participate in attention processes and the integration of auditory data with other sensory inputs. Further, there are also structurally and functionally significant connections from the cortex back to the MGN (Kandel et al., 2021).

The auditory cortex comprises the cortical regions that receive thalamocortical inputs from the MGN. Mammalian auditory cortex comprises several different areas characterized by differing neural architecture and cellular structure, and is mainly classified into three regions: core, belt, and para-belt which lie in different locations and can be further divided into smaller subregions (Hackett, 2011; Kaas and Hackett, 2000). The core receives most of its input from the MGN, it contains the primary auditory cortex (PAC) or A1, and neurons' organization shows a strong tonotopic map (Hackett, 2011; Kaas and Hackett, 2000). The PAC of human occupies the Heschl's gyrus of the temporal lobe. The belt and para-belt regions which surround core areas also receive inputs from MGN but less densely, and there are substantial and significant connections among these regions (Hackett, 2011; Kaas and Hackett, 2000). This auditory cortical organization seems to mirror the division of labor observed in other sensory systems, like the visual and somatosensory cortices, which segregate information related to object location and identification. Similarly, in the auditory cortex, different areas project to distinct regions of the temporal and frontal lobes (Hackett, 2011; Kaas and Hackett, 2000), suggesting a division of spatial and nonspatial processing (Rauschecker and Tian, 2000; Alain et al., 2001; Arnott et al., 2004). However, this division may not be as clear-cut, as neurons with broad spatial responsiveness are found throughout various areas (Hackett, 2011; Kaas and Hackett, 2000). Whereas the anatomical distinctions between these regions hint at possible functional differences, further experiments are necessary for confirmation.

1.4 Electroencephalography: a tool to track information flow in auditory pathways

Functional imaging has a pivotal role in identifying brain processes including those of the auditory system. An accessible and popular technique has been electroencephalography (EEG), a non-invasive technique that records electrical activity of the brain. Electrodes placed on the scalp detect micro voltage changes as fast as every millisecond or less, resulting from ionic current flows within the neurons (Niedermeyer and da Silva, 2005). The EEG signal, which originates from the synchronous activity of thousands to millions of neurons, provides a measure of the post-synaptic potentials (PSPs) in the cortical pyramidal cells (Woodman, 2010).

EEG has been used to follow the information flow through the auditory system. Auditory evoked potentials (AEPs), the electrical potentials recorded from the nervous system in response to auditory stimuli, are popular in both basic and clinical research. AEPs are grouped into three main categories: auditory brainstem responses (ABRs), mid-latency responses (MLRs), and long-latency responses (LLRs). ABRs, MLRs, and LLRs occur within the first 10 milliseconds, 10 to 50 milliseconds, and beyond 50 milliseconds, respectively, after an impulsive auditory stimulus, or after sound onset (Picton, 2010). Since the electrical signal propagation through auditory pathway has a limited speed, each of these groups of potentials corresponds to distinct stages in the auditory pathway, reflecting the sequential activation of neural structures from the auditory nerve to the auditory cortex. To identify the specific AEPs, the EEG data is time-locked to the onset of the auditory stimulus and averaged across multiple trials to enhance the signal-to-noise ratio, allowing for the isolation of the potentials of interest (Luck, 2014). ABRs correspond to the activation of the auditory nerve and brainstem structures, including the cochlear nuclei and the superior olivary complex. MLRs represent the activity of the thalamocortical radiations and PAC, whereas LLRs are associated with the secondary auditory cortex and other higher-level auditory processing centers in the brain (Näätänen and Picton, 1987; Picton, 2010). These potentials provide valuable information about the functional integrity of the auditory pathway and can aid in the diagnosis of various auditory disorders, as well as contribute to our understanding of auditory perception and cognition. Nevertheless, AEPs are limited for interpreting the type of processing happening in the brain, particularly given their predominant reliance on simplistic stimuli such as clicks or brief tones. AEPs capture the collective neural activity, consequently masking the intricate interactions among brain regions. As a result, we are primarily presented with an averaged response, limiting our insights into individualized or localized neural processes. Although, AEPs' association with hierarchy of auditory processing in the auditory system is established, they are not as helpful to track information flow in response to complex stimuli.

1.5 Representational Similarity Analysis: a method for decoding semantic representations

The ongoing endeavor of neuroscience to correlate behavior, neural activity, and models has seen considerable advances due to numerous technological and methodological innovations. An increasingly prevalent method is Representational Similarity Analysis (RSA) (Kriegeskorte et al., 2008). RSA offers a framework that can bridge the gap between neural and behavioral data and models, thus fostering a more comprehensive understanding of the brain's functional organization. RSA is a multivariate approach and captures the complex, high-dimensional nature of neural representation by comparing patterns of neural activation in response to diverse stimuli or tasks (Kriegeskorte et al., 2008). Multivariate pattern analysis (MVPA) is sensitive to the pattern of activity across units (e.g., neurons, channels, voxels, etc.) rather than to the average amplitude of the response. RSA transforms multivariate data into a common space, namely a dissimilarity or similarity space. This transformation is achieved by constructing representational dissimilarity matrices (RDMs), which capture the dissimilarity between patterns of brain activation across multiple conditions (Kriegeskorte et al., 2008). Each element of an RDM represents the dissimilarity between two condition-evoked activity patterns. By mapping activities and high-dimensional neural representations to a common space, RSA allows comparing representations in various brain regions, tasks, subjects, species, models or even neuroimaging techniques (Kriegeskorte and Kievit, 2013). This flexibility makes RSA an ideal tool for connecting observations at multiple levels, from cellular activity to behavior, empirical neuroimaging data, and computational models (Kriegeskorte and Douglas, 2018; Cichy et al., 2016; Cichy and Pantazis, 2017). By aligning computational models with complex brain activity, RSA can provide insight into underlying cognitive and neural processes.

Several studies have successfully employed RSA in a variety of contexts, highlighting its versatility and efficacy. For instance, a study by Kriegeskorte et al. (2008) utilized RSA to compare representations in human ventral-temporal cortex and a computational model of object recognition. They demonstrated a remarkable similarity between the artificial model and the human brain, providing strong evidence for a hierarchy of increasingly complex features in the brain's representation of objects. Connolly et al. (2012) employed fMRI and RSA to investigate neural representations of fine-grained categories within the animate domain (six animal species across three biological classes: primates, birds, and insects). The findings suggested that multivariate patterns of brain activity within the ventral object vision cortex significantly correlate with behavioral judgments of biological similarity. Furthermore, these brain activity patterns were found to span a continuum in abstract representational space, with primates and insects at opposite ends. The results reveal the complexity of category-specific representation in the human visual system, implying that both specific and general visual processing regions contribute to these representations.

Mohsenzadeh et al. (2018) aimed to disentangle feedforward and feedback neuronal processes involved in human visual recognition. To achieve this, the authors employed ultra-rapid serial visual presentation to suppress sustained activity, which enabled them to resolve two distinct stages of processing using multivariate pattern classification on magnetoencephalography (MEG) data (similar in temporal resolution to EEG). The first stage of processing represented the rapid bottom-up cascade, which terminated earlier as the rate of visual stimulus presentation increased. The second stage signified the emergence of category information, with a peak latency that shifted later with progressively faster stimulus presentations, indicative of time-consuming recurrent processing. By integrating MEG with functional fMRI using RSA, the authors localized recurrent signals in the early visual cortex. The findings separated an initial bottom-up processing stage from a subsequent feedback processing and uncovered the neural signature of increased recurrent processing demands under challenging viewing conditions.

1.6 Hierarchical processing of auditory semantic information

The auditory cortex (AC) is known to carry out many complex processes on auditory input to extract information that in turn affect behavior, perception, attention, and further high-level cognitive functions, such as semantic categorization (Grossberg, 2020). However, how inputs to auditory cortex are modulated and transformed to semantic information is not understood. Most studies in this area of research focus on a certain type of auditory input, usually speech sounds (Davis and Johnsrude, 2003; Okada et al., 2010; DeWitt and Rauschecker, 2012; Yi et al., 2019), since speech is profoundly important in human communication, and it is rich in semantics. Nonetheless, context and semantics can also be inferred from natural non-speech sounds. As functional organization of the cortex can be highly task dependent, transformation of natural non-speech sounds needs to be separately investigated.

Several lines of converging empirical evidence and theoretical rationale suggest that auditory and other cortical regions are organized hierarchically for auditory perception. Hierarchical processing posits a multi-tiered network structure within the brain, wherein simpler sensory representations are formed in early processing stages, and these are then progressively integrated and abstracted to form complex representations in higher-level areas (Felleman and Van Essen, 1991). This model has been substantiated through a wealth of neuroimaging studies, demonstrating its prevalence across various sensory modalities, including vision, auditory, and somatosensory processing (Hubel and Wiesel, 1962; Bizley and Cohen, 2013; Kaas et al., 1999; Hackett et al., 1998; Kaas and Hackett, 2000; Hackett, 2011; Rauschecker and Scott, 2009; Rauschecker and Tian, 2000; Mountcastle, 1997). Hierarchical processing aligns well with neuroanatomical observations, as the connections between the brain regions are largely organized in a hierarchical manner (Felleman and Van Essen, 1991). Moreover, the viability of hierarchical processing is also supported by computational models of brain function (Serre et al., 2007), which have provided insights into how this scheme could support a range of cognitive functions, including object recognition and spatial navigation. Finally, hierarchical processing has been instrumental in the development of deep learning models in artificial intelligence, inspired by the layered structure and the function of the cerebral cortex (LeCun et al., 2015). The success of these models in tasks related to sensory processing, such as image and speech recognition (He et al., 2015; Li et al., 2019), further corroborates the plausibility of hierarchical processing in the brain.

We reviewed the anatomical organization of auditory related areas of the brain from the peripheral auditory system to the auditory cortex. A great portion of this information arises from the neuroanatomical and cytoarchitectural investigations of non-human primates and other mammals (Hackett et al., 2001; Kaas et al., 1999; Hackett et al., 1998; Kaas and Hackett, 2000; Hackett, 2011). Although, the functional organization of the cortex needs to be investigated using functional imaging techniques for several reasons: (a) the neuroanatomical evidence is not as strongly established in humans; (b) distinct regions characterized with different neuroanatomy may not necessarily be functionally distinct; (c) the human auditory cortex is thought to be organized differently to facilitate humans' unique capabilities in speech recognition and music (Norman-Haignere et al., 2019).

Functional Magnetic Resonance Imaging (fMRI), a non-invasive neuroimaging technique, has been the most popular tool in these studies. fMRI leverages the principles of nuclear magnetic resonance to construct detailed maps of brain activity. It focuses on the Blood Oxygen Level Dependent (BOLD) contrast, which exploits the different properties of paramagnetic deoxygenated and diamagnetic oxygenated hemoglobin that distort external magnetic field differently. When a particular brain region is engaged, local cerebral blood flow is increased to supply the heightened demand for glucose and oxygen, resulting in more diamagnetic oxygenated hemoglobin. This results in a stronger

MR signal that is measured by the fMRI scanner over time and across different brain regions (Hillman, 2014), enabling it to capture and illustrate the spatial localization of brain activity with high resolution as precise as 1 mm.

Binder et al. (2000) conducted a study where participants were exposed to various auditory stimuli while their brain activity was recorded by fMRI. The study found that different regions within the human lateral temporal cortex are responsible for processing different types of sounds. Dorsal areas, specifically the planum temporale (PT) and dorsolateral superior temporal gyrus (STG), were more responsive to frequency-modulated tones, indicating their role in processing simple auditory information. On the other hand, regions in the superior temporal sulcus (STS) showed a greater response to speech stimuli, suggesting their involvement in processing complex speech sounds. This research highlighted the spatial hierarchy of sound processing in the lateral temporal cortex, with distinct neural mechanisms for different types of auditory stimuli.

Davis and Johnsrude (2003) also used fMRI to investigate the brain's functional organization in understanding spoken language. They degraded 190 English sentences in three ways: ”segmented” speech, ”vocoded” speech, and ”speech in noise”, each producing significantly different acoustic outcomes while preserving core attributes of the speech. Degradations were applied in different strengths rendering different speech intelligibility levels which were measured by a word report task. A signal-correlated noise baseline was also generated, which carried no linguistic information.

The authors correlated neural activity with behavioral performance and identified regions sensitive to speech intelligibility. They also examined these regions for their sensitivity to the acoustic form of the stimuli. Regions close to PAC were associated with sound-based processes, such as phoneme discrimination and acoustic-phonetic analysis as activity in those regions depended on the type of degradation. In contrast, higher-level processes like semantic processing and syntactic analysis were linked to activity in the other intelligibility-responsive regions: inferior frontal gyrus (IFG) and posterior middle temporal gyrus (pMTG) since they were insensitive to this acoustic variability.

Furthermore, a correlation was reported between individual differences in performance and brain activity. Participants with better performance on sound-based tasks showed greater activity in the left STG and PT, while those with better performance on semantic tasks showed greater activity in the left IFG. The findings supported a hierarchical model of spoken language comprehension, where speech signals are initially processed by sound-based mechanisms before being integrated into higher-level meaning representations.

Okada et al. (2010) used fMRI and a multivariate pattern analysis to assess sensitivity of auditory cortex to acoustic variation within intelligible versus unintelligible speech. The researchers used two types of intelligible but acoustically different stimuli (clear speech and noise-vocoded speech) and two types of unintelligible stimuli (spectrally rotated versions of the clear and vocoded speech).

The results showed that core auditory regions on the dorsal plane of the STG exhibited high levels of sensitivity to acoustic features. In contrast, downstream auditory regions in both anterior STS and posterior superior temporal sulcus (pSTS) bilaterally showed greater sensitivity to whether speech was intelligible or not and less sensitivity to acoustic variation (acoustic invariance). Acoustic invariance was most pronounced in more pSTS regions of both hemispheres.

Okada et al. (2010) employed a multivariate pattern analysis which is potentially capable to detect effects that are not detectable by the univariate method used by Davis and Johnsrude (2003). In their review, Peelle et al. (2010) highlighted that previous studies (Scott et al., 2000, 2006; Davis and Johnsrude, 2003), suggested that the anterior temporal regions of the brain are largely ”acoustically invariant”. However, Okada et al. (2010) found evidence that contradicts this view. Their findings suggest that the anterior temporal regions do show sensitivity to the acoustic properties of speech. Despite the contradiction of results with previous research pointed out by Peelle et al. (2010) the findings still provide evidence for a spatial hierarchical organization of the human auditory cortex. The core auditory regions, which are more sensitive to acoustic features, represent the lower levels of the hierarchy, while the downstream regions, which are more sensitive to the intelligibility of speech and less sensitive to acoustic variation, represent the higher levels.

The mentioned studies investigated functional organization of AC by manipulating the stimulus to identify functionally distinct brain regions. The human brain has evolved to proficiently interpret natural sounds; thus, to comprehensively decipher these mechanisms, naturalistic stimuli may offer significant insights. As stated, task and stimuli may greatly modulate the processing scheme and identified hierarchy. Hence, using naturalistic stimuli in neuroscience has great significance as it provides a comprehensive understanding of the brain's functionality in realistic scenarios. It allows for the study of intricate neural processes shaped by daily sensory input, which can be oversimplified in controlled stimuli. This approach also illuminates how the brain has evolutionarily adapted to process information within natural environments. In essence, the use of naturalistic stimuli enhances the ecological validity of findings, bridging the gap between lab-based research and the complexities of real-world experiences (Miller et al., 2022).

Using naturalistic stimuli in traditional evoked response approaches is not feasible as controlling all aspects of stimuli is unattainable. Consequently, further methodological approaches must be taken which can differentiate the features of interest. For example, a recent study by Kell et al. (2018) approached the investigation of functional organization in auditory cortex by employing a novel computational model. This model, a deep neural network (DNN), was meticulously designed and trained to perform auditory tasks that mirror real-world scenarios. These tasks included the identification of words and musical genres from raw naturalistic sound waveforms. The network was optimized using extensive sets of labeled data, reflecting the authors' hypothesis that the constraints imposed by everyday recognition tasks could lead to the development of representational transformations that mimic those found in the human brain.

The performance of this neural network was then juxtaposed with the performance of human listeners across a diverse range of conditions. The network not only demonstrated proficiency in recognizing words and musical genres that were on par with human listeners, but it also exhibited error patterns that closely mirrored those of humans. This suggests that the network was able to replicate key aspects of human auditory behavior.

Further, the authors used the features of the network to predict fMRI voxel responses throughout the auditory cortex. By analyzing the responses of different layers of the network to auditory stimuli, the authors found that intermediate layers of the model best explained responses in the PAC, while deeper layers best explained responses in non-primary areas. This provided compelling evidence of a hierarchical organization in the human auditory cortex, with different levels of the hierarchy responsible for different aspects of auditory processing.

The study by Kell et al. (2018) provided an innovative approach to understanding the human auditory system. By developing a task-optimized neural network that replicates human auditory behavior and predicts brain responses, the authors shed light on the hierarchical organization of the auditory cortex.

Recently, Giordano et al. (2023) delved into a more comprehensive investigation of the computational dynamics implicated in the recognition of natural sounds. The authors employed a model comparison framework to pinpoint intermediate acoustic-to-semantic representations that bridge behavioral and neural responses to natural sounds. They utilized a dataset examining behavioral and auditory cortical responses (fMRI) to natural sounds and determined the predictability of these responses by computational models of sound representation. The models stemmed from three distinct classes: acoustic-processing models, semantic-processing models, and DNNs including the Kell et al.

(2018) model.

Their findings showed that semantic-processing models surpass acoustic-processing models in terms of their ability to predict both behavioral and neural responses. This leads to an inference that semantic representations hold a significant role in the recognition of natural sounds. Additionally, the study substantiates that spectrotemporal modulations have the capacity to forecast early auditory cortex (Heschl's gyrus) responses, while auditory dimensions (for example, loudness and periodicity) can anticipate STG responses and perceived dissimilarity. The DNN models parallel acoustic models in predicting responses in Heschl's gyrus, but they excel beyond all competitive models when forecasting both STG responses and perceived dissimilarity. Such evidence implies that these types of models could prove particularly beneficial for comprehending how complex sounds undergo processing in the brain.

Collectively, Binder et al. (2000), Davis and Johnsrude (2003), Okada et al. (2010), Kell et al. (2018), and Giordano et al. (2023) have demonstrated a spatial hierarchical organization for semantic processing of sounds. However, a spatial hierarchy alone cannot evince hierarchical processing since distinct regions of AC receive parallel projections from thalamus. Hence, to confirm a hierarchical scheme a temporal hierarchy must be validated. To investigate a temporal hierarchy a higher temporal resolution is required than what fMRI offers, thus either MEG or EEG, the only non-invasive high temporal resolution methods, should be employed.

Benner et al. (2023) investigated the organization of the human AC, drawing parallels with the core-belt-parabelt organization observed in nonhuman primates. The study employs a robust methodology, utilizing both fMRI and MEG to achieve spatial and temporal segregation of human auditory responses. This approach was applied to a cohort of musicians.

They identified individual fMRI activations induced by different sampled instrumental and synthesized complex harmonic tones within four distinct regions of the AC: the medial Heschl's gyrus (HG), lateral HG, anterior STG, and PT. The same 23 musicians who participated in the fMRI experiment also underwent MEG with similar acoustic stimuli. The spatial coordinates of BOLD activations in the four ROIs in the AC were used as seeding points for source modeling. In each hemisphere, four dipoles were used, one for each ROI. To distinguish between early, middle, and late MEG components (P30 / P1–N1 / P2 & P2a), the time windows for source modeling were adjusted to 20–50, 50–120, and 120–250 ms, respectively. Each ROI yielded a distinct average source waveform. By measuring the onset and peak latencies of the evoked responses (in milliseconds after stimulus onset), a chronotopic order of responses was observed in the four ROIs. The earliest primary response (P30) was localized to the medial HG (ROI 1) with an average onset latency of approximately 20 ms. The secondary response (P1-N1 complex) was localized to the lateral HG (ROI 2) with an onset latency of approximately 28 ms. The later P2 and P2a responses were localized to the anterior STG and PT, respectively, with onset latencies of approximately 51 and 57 ms (ROIs 3 and 4). The average latencies of the first peak were approximately 34, 63, 87, and 90 ms for ROIs 1–4, respectively.

These findings, therefore, provided compelling evidence for a temporal hierarchy within the human AC. The study's innovative use of combined MEG and fMRI measurements allowed the demonstration of this characteristic temporal hierarchy, thereby reflecting the serial processing predictions derived from nonhuman studies. However, as mentioned functional organization is highly task dependent and processing semantic information needs to be independently investigated.

Lowe et al. (2021) have conducted the only study focusing on exploring the spatiotemporal hierarchy of semantic processing from naturalistic non-speech sounds. Using fMRI and MEG with 16 participants, the researchers examined the neural responses to a set of naturalistic sounds, such as voices, animals, objects, and scenes. These sounds were presented in an event related design, and each run contained all sounds plus null and oddball trials that participants were instructed to detect.

The authors utilized MVPA on MEG data in trial epochs spanning -200 to +3000 ms relative to the stimulus onset, generating MEG RDMs through time for each participant. For fMRI, RDMs were computed for each voxel using a whole-brain searchlight approach per participant. The group-averaged MEG RDMs were then correlated with subject-specific fMRI RDMs, resulting in a sequence of whole-brain maps showing MEG-fMRI correspondences across time. Significant fusion correlations were seen starting 55-60 ms post-stimulus onset in Heschl's Gyrus and the superior temporal gyrus (STG), progressing to pre-frontal, ventral occipitotemporal, and medial regions by around 130-140 ms. This progression suggests an orderly spatiotemporal procession of responses from early sensory cortices to higher-level and extratemporal regions.

This progression was further quantified by employing an ROI analysis. Various primary and nonprimary auditory anatomical ROIs were identified, including the PAC, PT, and planum polare (PP). From these, fMRI BOLD responses were extracted and an RDM was computed. The results revealed significant clusters in all ROIs, except for the early visual cortex (EVC), with peak latencies of approximately 115 ms post-stimulus onset in early primary and nonprimary auditory ROIs. Higher cortical regions displayed a later onset of sound-evoked neural activity, reinforcing the forward signal propagation.

They determined representational levels using two extremities: cochleagram RDM to model frequency-based similarity structure across stimuli, and category RDM to model generalized high-level semantic category selectivity across all four categories. The dominance of acoustic or semantic properties in neural representations was quantified as Semantic Dominance (SD). Both cochleagram and category models correlated significantly with the whole-brain MEG signal, reaching significance at similar onset times of around 80ms. Notably, the period of significant cochleagram model correlation was shorter and peaked earlier. Semantic Dominance was significantly negative early on but became significantly positive later, indicating a temporal progression where neural response patterns are more identifiable by semantic category than by their spectra after about 200ms. These results provided strong evidence of a systematic progression from initial acoustic representation to a semantic, category-specific neural coding process.

Benner et al. (2023) and Lowe et al. (2021)'s investigations of temporal dynamics in auditory semantic processing aligned with the spatial hierarchy demonstrated previously (Binder et al., 2000; Davis and Johnsrude, 2003; Okada et al., 2010; Kell et al., 2018; Giordano et al., 2023) and provided more evidence supporting a core-belt-parabelt hierarchy in human AC. However, their findings have yet to be subjected to replication attempts.

1.7 Current study

In the current study, we aimed to investigate the temporal dynamics of processing semantic information from naturalistic sounds. More specifically, we intended to confirm a temporal hierarchy in processing semantic information to complement the previously established spatial hierarchy and provide evidence for a hierarchical organization scheme in auditory cortex. Thus, we replicated and extended prior work in this area by Lowe et al. (2021).

To better understand the dynamics of natural sound representation, we conducted an EEG experiment. This method was selected due to its capacity for high temporal resolution. The stimulus set was deliberately chosen, comprised of four naturalistic sound categories: animals, people, objects, and scenes. Each category included twenty, one-second-long sounds, thereby ensuring a diverse and representative dataset. To ensure clarity of the selected sounds, a behavioral pilot study was performed. This not only affirmed the quality of our stimuli but also allowed us to refine our category assignments. The caveat being that the categories we ascribed might not be the sole interpretation of the sounds. The EEG experiment was conducted with a sample size of twenty-five participants, a number deemed sufficient for this initial exploratory study. Stimuli were presented in an event related design and participants were asked to do an oddball detection task identifying a white noise sound while hearing the whole stimulus set in each run. This task ensured vigilance of participants while not modulating the natural sound representations.

Our analysis strategy entailed an MVPA approach, employing RSA. The EEG signals provided nuanced sound representations in small time intervals of 2 ms, ranging from -200 to 1300 ms relative to the stimulus onset. To discern the dissimilarities at each time sample, we decoded the activity patterns associated with each sound (across channels) from activity patterns of another sound using a linear classifier. By performing this classification for all pairs of sounds, we obtained a dissimilarity pattern over the whole stimulus set at each time sample. Given our hypothesis of a non-linear transformation of representations over time depicting the semantic categorization, we employed linear discrimination. This method allowed us to assess the represented information at different times without intermixing information from disparate timeframes.

To evaluate the contribution of different information levels over time, we compared the dissimilarity patterns or RDMs extracted from the EEG to those derived from computational models. The models we utilized contained a broad range from low-level acoustic features to high-level semantics, including cochleagrams and category models, among others. Finally, a temporal generalization analysis was conducted to scrutinize the generalizability and stationarity of the low- and high-level information.

Through these methods we first replicated the temporal hierarchy from Lowe et al. (2021) using a different complementary imaging method and stimulus set. Besides, we extended their result and delved deeper into temporal dynamics using other computational models than the two extremists (cochleagram and category models) and performing temporal generalization analysis. Further, we addressed some potential shortcomings of their study. For instance, Lowe et al. (2021) did not control the role of mental imagery of sounds in their study, which potentially poses a limitation in their inference. Given these considerations, the current investigation is primarily motivated by two principal objectives. First, it seeks to validate Lowe et al. (2021) through an attempt to replicate the presence of a temporal hierarchy. Second, the current study aims to expand upon their original work by incorporating a more comprehensive analysis involving various models, examining additional facets of temporal dynamics such as temporal generalization. Furthermore, the present study utilizes EEG, not only because of its high temporal resolution but also due to pervasive usage and availability in research institutions. This approach offers a rich source of data that might further enhance our understanding of the temporal hierarchy in the semantic processing of natural sound.

The subsequent chapters of this thesis will sequentially comprise a method chapter detailing the study undertaken, a results chapter presenting the findings of the study, and a discussion chapter pertaining to the study's implications, contribution, and significance.

Chapter 2

Methods

2.1 Participants

Thirty-four participants who self-reported normal hearing and no neurological or psychiatric disorders were recruited. Sixteen participants (9 female, age: mean ± s.d. = 22.9 ± 10.2 years) completed the behavioral experiment, and 25 participants (15 female, age: mean ± s.d. = 22.1 ± 8.6 years), five of whom had completed the behavioral experiment, participated in the EEG experiment. Both experiments were approved by The University of Western Ontario Health Sciences Research Ethics Board (HSREB, Study ID: 120730) and all participants provided written informed consent for each of the experiments in which they participated.

2.2 Stimuli

We selected 20 naturalistic sound clips from each of four categories: Animals, People, Objects, and Scenes (80 in total). The sounds were root-mean-squared normalized, resampled to 48 kHz, and trimmed to one –second in length. We identified the sound onsets by thresholding the signals and used the right channel of the sound files to send a step function marking the onsets for synchronization. The left channel was used to feed participants' headphones: sounds were presented diotically. We presented the stimuli to participants using Psychtoolbox-3 in MATLAB and through a Steinberg UR22C USB audio interface.

2.3 Experiment design and procedure

2.3.1 Behavioral Pilot Experiment

The behavioral pilot experiment aimed to confirm participants' ability to identify each sound's category and to determine if familiarization with the sounds biases participants' categorization. The behavioral experiment included two tasks: one for familiarization and one for recognition. Only half of the participants completed the first task, while all participants completed the latter.

Familiarization task

During the familiarization task, participants listened to the sounds while accompanied by a written description on the screen. The goal was to ensure participants have a clear and correct understanding of the stimuli. Participants were instructed not to memorize the sounds or descriptions but to focus on comprehending the stimuli to diminish ambiguity. They could repeat a sound as many times as they desired, move forward to the next, or replay the previous sound. After the participants felt confident in understanding the sounds, they proceeded to the next task (see Figure 2.1 A).

Recognition task

We asked participants to recognize the category (with four options) for each of the 80 sounds as they were played in a random order, with an inter-trial interval (ITI) of one second. Participants responded to each sound by pressing one of four keyboard buttons corresponding to the four categories: Animals, People, Objects, Scenes, the correspondence between category and key was shown on the screen. In addition, an "hourglass" bar on the screen indicated the time participants had to respond to each sound (2 seconds from the beginning of the sound). Two runs of the task were completed by each participant: in each run they heard all 80 sounds (see Figure 2.1 B).

2.3.2 EEG Experiment

EEG was used to measure brain activity in response to hearing the sounds. As we were not interested in a task-driven or top-down modulated representation of the sounds, we employed a target detection task. In this task, all 80 sounds were played in a random order in each run while a one-second white noise (oddball) was randomly played in between the sounds 20 times (two-second ITI). This task ensures participants' attend to the stimuli, but it does not require inference about the sounds other than whether or not they are the target white noise. All participants were familiarized with the stimuli by doing the familiarization task from the behavioral pilot experiment prior to recording in the EEG session (Figure 2.1 A). Participants were asked to press a button when they heard the oddball noise, to keep their eyes open, and to fixate on the center of the screen as they performed 12 runs of this task (Figure 2.1 C). Thus, we had 12 repetitions for each sound and in turn 12 samples of brain activity in response to each sound trough time, and oddball trials were excluded from the analysis.

2.4 EEG acquisition and preprocessing

EEG signals were collected from the participants while they were performing the target detection task using a BioSemi ActiveTwo system, 1020 64-electrode cap, and a sampling rate of 2048 Hz. The sound onset events were precisely marked on the EEG signals synchronized with what participants heard through their headphones. Preprocessing used the Brainstorm toolbox (Tadel et al., 2011), and included following steps: 0.5-30 Hz bandpass filtering, identifying and removing bad channels, Independent Components Analysis (ICA) and removing eye movement and blink artifacts, manually rejecting bad segments of the signal, down-sampling to 512 Hz, removing baseline and epoching data from –200 to 1300 ms relative to each sound onset event (769 timepoints). Bad channels were identified as outliers in the power spectral density and time domain, and bad segments were segments with excessive amounts of muscle noise or movement artifact assessed by monitoring the signal. These steps yield trials, with labels corre-

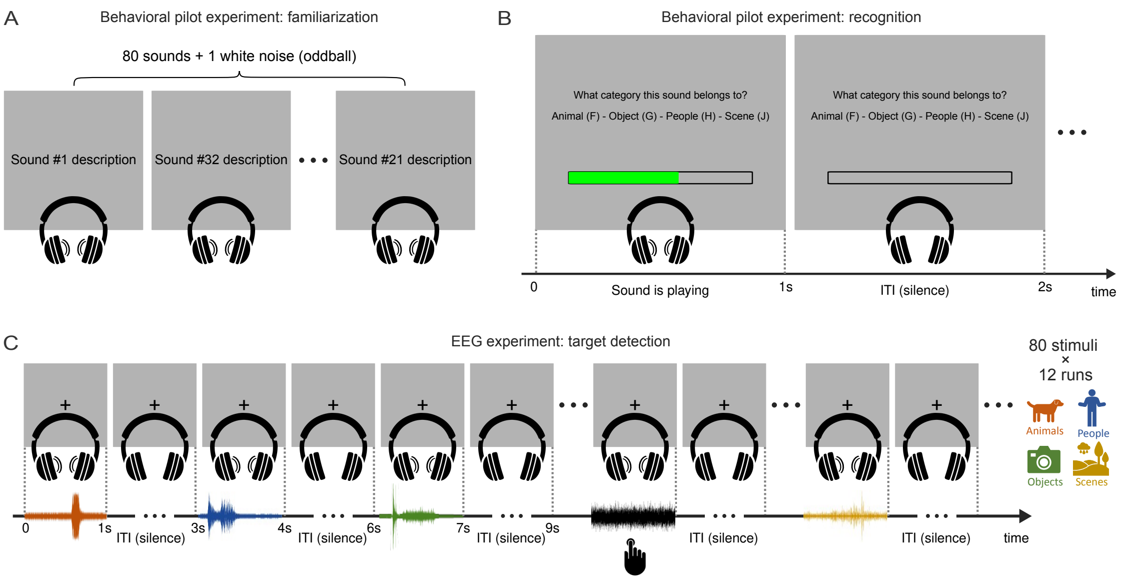
A
Behavioral pilot experiment: familiarization
80 sounds + 1 white noise (oddball)
Sound #1 description
Sound #32 description
• • •
Sound #21 description
B
Behavioral pilot experiment: recognition
What category this sound belongs to?
Animal (F) - Object (G) - People (H) - Scene (J)
What category this sound belongs to?
Animal (F) - Object (G) - People (H) - Scene (J)
• • •
0
Sound is playing
1s
ITI (silence)
2s
time
C
EEG experiment: target detection
80 stimuli
×
12 runs
Animals
People
Objects
Scenes
0
1s
ITI (silence)
3s
4s
ITI (silence)
6s
7s
ITI (silence)
9s
ITI (silence)
ITI (silence)
time

Figure 2.1: Experiments' design. (**A**) behavioral session's first task was familiarization, where participants freely explored the stimulus set while a description on the screen accompanied each sound, half of participants were asked to do this task. (**B**) Participants were asked to categorize 80 sounds into four categories: animals, people, objects, scenes. The sounds were played randomly with a one-second ITI. Participants responded by pressing one of four corresponding keyboard buttons. An on-screen 'hourglass' bar indicated the 2-second response time for each sound. Each participant completed two runs of this task, hearing all 80 sounds in each run. (**C**) In the EEG session participants were first familiarized with the stimuli, doing the familiarization task in A. In the target detection task, 80 sounds were played randomly in each of the 12 runs with a two-second ITI. A one-second white noise (oddball) was interspersed randomly 20 times to ensure participants' attention. Participants were asked to press a button when they heard the oddball noise, keep their eyes open, and fixate on the screen's center.

sponding to each sound, that capture the participant's brain activity in response to the sound. Each sound had 12 repetitions in the experiment but trials that overlapped with bad segments were excluded from the analysis.

2.5 Multivariate pattern analysis (MVPA)

A multivariate pattern analysis (MVPA) approach was taken to uncover the information represented in the brain, reflected in EEG signals through time within each participant. At timepoint t we extracted the activity pattern—the pattern of voltage levels across channels—across all repetitions of each sound (see Figure 2.2 A). Subsequently, we conducted a pairwise classification on the activity patterns for each pair of sounds, meaning a classifier was trained and tested to discriminate between every possible pair of sounds. Specifically, all available activity patterns (across 64 channels at timepoint t) for a particular sound were randomly grouped and averaged into five folds (pseudo-trials) to get more robust response patterns (Guggenmos et al., 2018). A linear Support Vector Machine (SVM) was used to discriminate between patterns of one sound and the other, using the LIBSVM implementation (Chang and Lin, 2011) with default parameters for the classification. The pairwise classification was leave-one-out cross validated (selecting one pseudo-trial from each sound in the pair for test and exclude from training), and it was repeated 100 times with different random folding of activity patterns and the decoding accuracy was averaged. The averaged decoding accuracy yields a measure of dissimilarity of the sounds' activity patterns at time t and was computed for all 769 timepoints (see Figure 2.2 B).

2.6 Temporal generalization analysis

To test the stationarity of information's representations through time, we performed temporal generalization analysis (King and Dehaene, 2014) on the EEG time-series. In the MVPA described in the previous section, SVM classifiers were trained on folded activity patterns of a timepoint t and tested on a different fold of activity pattern from the same timepoint. However, in temporal generalization analysis, the classifiers trained on timepoint t are tested on all timepoints including t. This analysis assesses if the discriminating information represented in the activity patterns at a certain time (learned by the classifier) is present in or generalizable to other timepoints. The analysis yields a decoding accuracy for every pair of test-train timepoints and for all pairs of sounds. If the decoding accuracy is high for t and t's which are close but it declines rapidly for further t's, where t and t' are train and test timepoints respectively, it means the discriminating information at t are not useful at t' when they are not close meaning representations at t are transient and change quickly. Conversely, if the decoding accuracy does not decline rapidly but slowly as t and t' get farther apart, it means the representations at that

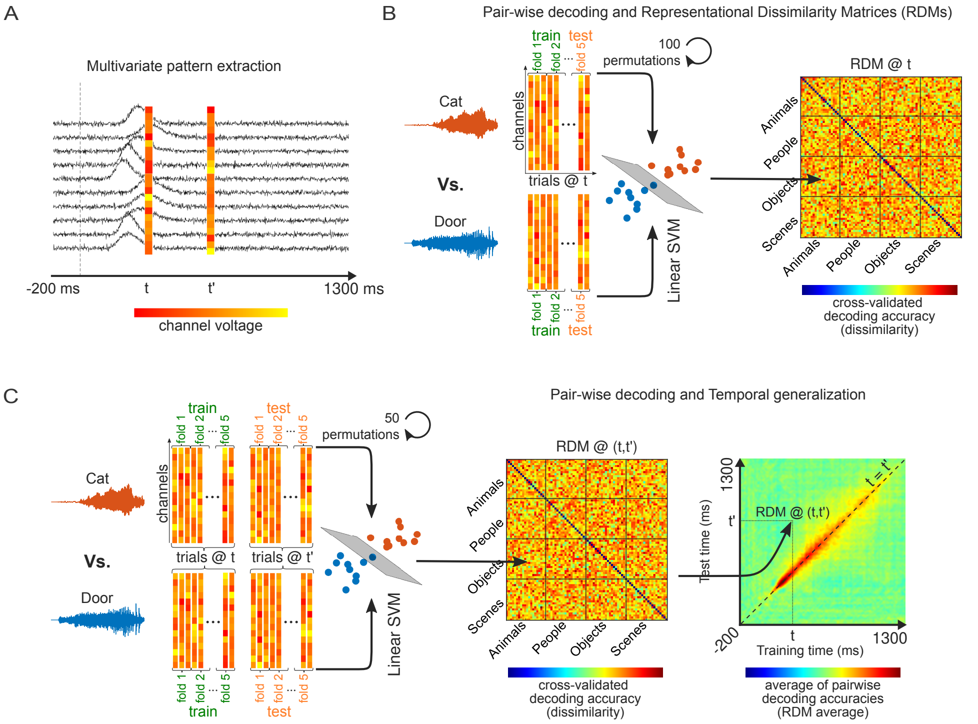

Figure 2.2: Multivariate pattern analysis (MVPA). (**A**) EEG signals were epoched from -200 to 1300 ms relative to each sound onset, oddball sound excluded. (**B**) At each timepoint (769 in total) activity patterns for all sound pairs were extracted and decoded by a leave-one-out cross-validated linear SVM within each participant. The decoding accuracies were used as dissimilarity measure to form a RDM at each time point. Activity patterns were subgrouped and averaged with different permutations (100). (**C**) Temporal generalization analysis was performed similar to MVPA, however classifiers trained at a timepoint t were tested on all other timepoints t' yielding RDMs for each (t, t') to test generalizability of representations across time.

time point are more sustained, and probably t and t' belong to one stage of processing. Hence, how stationary, or transient are the representations can be determined. Temporal generalization analysis is computationally intensive; thus, the EEG time-series were down-sampled by a factor of 4 (to 128Hz). Additionally, the number of permutations for random folding was 50 iterations (see Figure 2.2 C).

2.7 Representational similarity analysis (RSA)

To quantify the sound's representation at timepoint t a Representational Dissimilarity Matrix (RDM) was formed. An RDM is a symmetric matrix where each row and column correspond to a sound and each entry denotes the dissimilarity between activity patterns (voltage pattern over 64 channels) of its row and column corresponding sounds. Hence

RDMs (Representational Dissimilarity Matrices) were generated for all timepoints using all pairwise decoding accuracies at their corresponding timepoints. RDMs provide a method for comparing neural representations (EEG) with representations derived from cognitive theory or computational models. Different models were used to predict representations at different perceptual levels in the analysis, ranging from acoustic models to semantic models, as discussed in the next section (see Figure 2.3 A). We derived RDMs from the models' representations and compared them to RDMs from EEG across all timepoints. Spearman's rank-correlation (ρ) was used to compare EEG and model RDMs (see Figure 2.3 B). Specifically, an estimate of the expected Spearman's ρ under random tiebreaking was used to remove the bias toward higher predictions for tied ranks:

$$\rho_a(\mathbf{x}, \mathbf{y}) = \mathbb{E}_{\substack{\tilde{\mathbf{a}}=\tilde{\mathbf{x}}-\frac{1}{n}\sum_{i=1}^{n} i, \tilde{\mathbf{x}}\sim Rae(\mathbf{x}) \\ \tilde{\mathbf{b}}=\tilde{\mathbf{y}}-\frac{1}{n}\sum_{i=1}^{n} i, \tilde{\mathbf{y}}\sim Rae(\mathbf{y})}} \left[\frac{\tilde{\mathbf{a}}^\top \tilde{\mathbf{b}}}{||\tilde{\mathbf{a}}||_2 ||\tilde{\mathbf{b}}||_2} \right] \tag{2.1}$$

$$= \frac{12}{n^3 - n} \mathbb{E}_{\tilde{\mathbf{a}}}[\tilde{\mathbf{a}}]^\top \mathbb{E}_{\tilde{\mathbf{b}}}[\tilde{\mathbf{b}}] \tag{2.2}$$

$$= \frac{12\mathbf{x}^\top \mathbf{y}}{n^3 - n} - \frac{3(n+1)}{n-1} \tag{2.3}$$

The ρ_a is fast to compute, hence it was used to determine the similarity of RDMs. RSA frequently employs rank correlation, specifically Spearman's rank correlation, due to its robustness against non-linearity and outliers. Spearman's correlation does not assume a linear relationship, and its non-parametric nature ensures flexibility, accommodating diverse data distributions and facilitating comparisons across varying measures or modalities in RSA. A high similarity between a model RDM and an EEG RDM at t indicates that the information represented in the model is also represented in the EEG activation of timepoint t.

2.8 Models

2.8.1 Auditory processing models

Two models were used to characterize two stages of auditory processing: a cochleagram model, and a spectrotemporal model.

The cochleagram model mimics the frequency decomposition that takes place in the cochlea due to different resonance frequencies of basilar membrane along its length. The model used in this study is the NSLtoolbox implementation of the cochleagram representation (Chi et al., 2005). This implementation models the peripheral auditory system including cochlear filtering, hair cell transduction, and the lateral inhibitory network in the cochlea. The model considers 128 overlapping bandpass filters which center frequencies cover 5.3 octaves uniformly spaced along logarithmic frequency axis. Euclidean distance between sounds' cochleagram were used to build RDMs for comparing to EEG RDMs. For time points before the stimulus onset ($t < 0$) and after one second ($t > 1$), we used full cochleagram of sounds, however for time points in between ($0 < t < 1$) only the part of sounds' cochleagram corresponding to $[0, t]$ interval were used to form the RDM corresponding to each time point (see Figure 2.3 A).

The spectrotemporal model mimics responses of higher auditory stages, in particular, the primary auditory cortex, implemented in the NSLtoolbox (Chi et al., 2005). This model estimates the spectral and temporal modulation of the cochleagram by a bank of filters selective to different spectrotemporal modulation rates (ranging from slow to fast) and scales (ranging from narrow to broad). Euclidean distance of sounds' spectrotemporal modulation features were used to form a fixed RDM through time (see Figure 2.3 A).

2.8.2 Semantic models

We used two different models to represent high-level semantic information. One model, the category model, only reflects the categorical differences among stimuli.

Hence, in this model, any between-category pair of sounds is considered maximally dissimilar, and any within-category pair is considered maximally similar (see Figure 2.3 A), forming an 80×80 matrix.

However, participants may not perceive the sounds as equally good exemplars of a particular category or might. Therefore, a behavioral model, based on the behavioral pilot experiment, was also used to account for perceptual similarities across category boundaries, and perceptual dissimilarities within categories. Confusion data from the behavioral recognition task was used to build this model: if two categories were confused more often in that task that means they are more likely to be similar (see Figure 2.3 A). To represent dissimilarities based on the confusion matrix, we calculated $1 -$ confusion matrix, made it symmetric, and then upscaled to an 80×80 dimension, ensuring a zero diagonal.

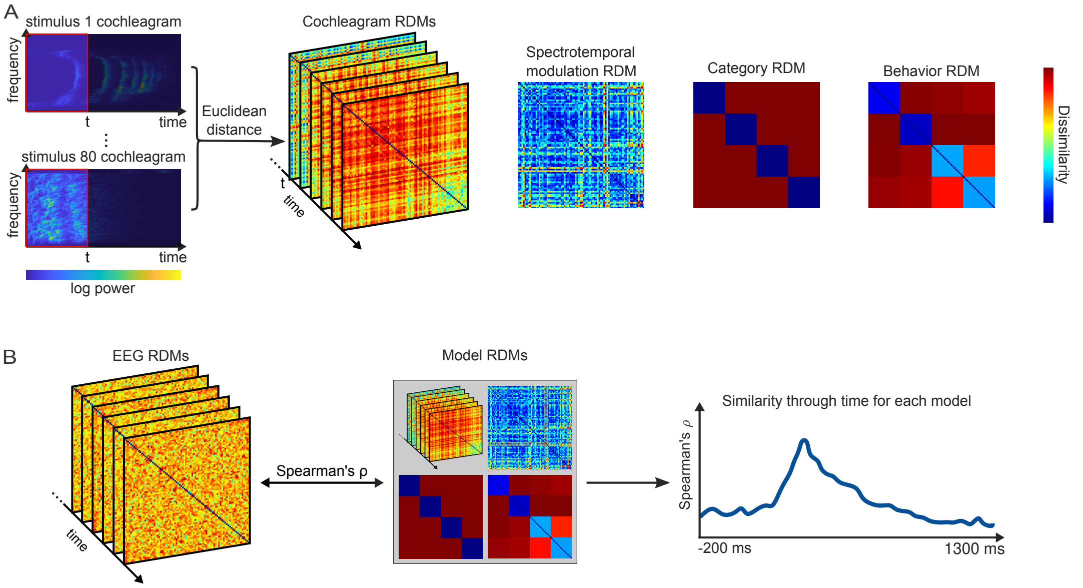

Figure 2.3: Representational similarity analysis. (**A**) Model RDMs of the analysis: cochleagram model built for each timepoint t only using the $[0,t]$ time window of the sounds' cochleagram (for $t < 0$ and $t > 1$ whole cochleagram was used), spectrotemporal modulation model based on the model of cortical responses implemented in NSLtoolbox (Chi et al., 2005) (using Euclidean distance), category model assuming sound belonging to different categories maximally dissimilar and vice versa, behavior model based on confusion matrix of behavioral experiment. (**B**) Spearman's correlation of model RDMs and EEG RDMs through time was estimated with random tiebreaking within each participant.

2.8.3 Noise ceiling

The noise ceiling provides an upper limit on the correlation between models and observed RDMs, constrained by the inherent noise in the observed data Nili et al. (2014). Essentially, it offers a measure of the maximum achievable correlation given the noise level in the data, thereby providing a benchmark for model performance. A model that approaches or reaches the noise ceiling is considered to be performing optimally given the noise in the data (Nili et al., 2014).

We computed within-participant noise ceiling of the EEG RDMs by correlating each individual participant's RDMs with the average of the other participants' RDMs, providing an estimate of the reliability of each participant's data. This process was iteratively performed for each participant, and the mean of these correlations formed the within-participant noise ceiling (Nili et al., 2014).

2.9 Statistics

2.9.1 Cluster-correction permutation test

To address the multiple-comparisons problem when testing whether decoding accuracies or RSA correlations were significant, we employed sign flip permutation test in combination with cluster correction. The null hypothesis of no effect (i.e., mean of the test statistic: decoding accuracy, correlation, etc.) across participants equals zero was considered. The cluster correction algorithm estimates the null distribution over time by sign flip permutation, a non-parametric statistical test. In multiple permutations (e.g., 10000), test statistics corresponding to each participant were sign flipped with a 50% probability, and then averaged to produce an observation of the null distribution. Applying a cluster defining threshold of $p < 0.001$, consecutive significant time points were considered as a cluster at each permutation. Each permutation sample yielded a maximum cluster size statistic, hence the distribution for maximum cluster size under the null distribution was estimated. Then a significance threshold (e.g., $p < 0.01$) was applied to get the significant time clusters in the original sample of participants.

2.9.2 Onset and peak time identification

To elucidate the onset and peak times in our time series data (average decoding, correlation, etc.), we employed a bootstrap resampling technique. This method entailed resampling our participant pool with replacement over several iterations (e.g., 10000) maintaining an equivalent sample size. In each iteration, the time series of randomly chosen participants were averaged, creating a resampled time series for further analysis. To facilitate a more accurate identification of the onset and peak times, we applied Gaussian smoothing to each resampled time series. Smoothing utilized a window size corresponding to 5% of the time series total length, or 39 samples (0.05×769). The standard deviation of the Gaussian window was calculated relative to the window size: $std = \frac{\text{window size -1}}{6}$. This smoothing procedure allowed us to mitigate minor fluctuations and enhance the estimation of onset and peak times. Following the smoothing process, we calculated the derivative of the smoothed time series. The onset of the time series was defined as the initial time point where the derivative exceeded 20% of its maximum value for a duration exceeding half the size of the smoothing window. This derivative-based onset detection methodology is more sensitive to the inflection points where the curve ascends.

To ascertain the peak time of the time series, we first identified the five most significant local peaks at the time points where the derivative crossed zero from positive to negative (i.e., local maxima). Among these peaks, we reported the earliest one as the peak time. Each iteration of the bootstrap procedure yielded an onset and a peak time, allowing us to form empirical distributions for both these variables. These distributions were subsequently utilized to define confidence intervals for the onset and peak times, and to statistically evaluate any significant differences.

Chapter 3

Results

3.1 Behavioral results provide a refined model for category perception

We selected auditory stimuli from four categorical divisions of natural sounds encompassing animals, people, objects, and scenes. Nevertheless, there exists a potential discrepancy between the designated categories and the actual perceptual interpretation of the participants. To validate the categories to which individual stimuli were assigned, a pilot behavioral experiment was conducted, in which participants assigned each stimulus to one of the four categories. Participants' proficiency in the sound recognition task demonstrated the ability to accurately recognize the category to which our stimuli belonged (see Figure 3.1 A). Despite the four categories differing in identification accuracy (animals: 95%, people 96%, objects: 78%, and scenes: 76%), the participants' overall performance in the sound recognition task was significantly above chance.

In assessing participant performance during the recognition task, we drew comparisons between two groups, each comprising eight participants; one group had been familiarized with the stimuli, while the other had not. The accuracy difference of the two groups difference was not statistically significant ($t(15) = 1.54, p = 0.16$, see Figure 3.1 B). Consequently, this outcome confirms familiarization does not undermine the naturalistic aspect of our stimuli perception by inducing a learning effect over the EEG experiment. To maintain consistency, all participants were familiarized with the stimuli at the beginning of the EEG experiment session.

Furthermore, the confusion matrix—a matrix depicting the frequency of misidentified sound categories—provides insightful data regarding the overlap and potential similarity between categories. Instances where stimuli of one category are regularly misinterpreted belonging to another particular category indicates a perceptual resemblance between those categories (see Figure 3.1 C). Thus, the confusion matrix provides a refined model for category similarity, alternative to assuming maximum dissimilarity between categories and complete similarity within a single category. We assume the information gathered from the behavior model would be beneficial in our examination of the representation of information in the brain across the post stimulation interval.

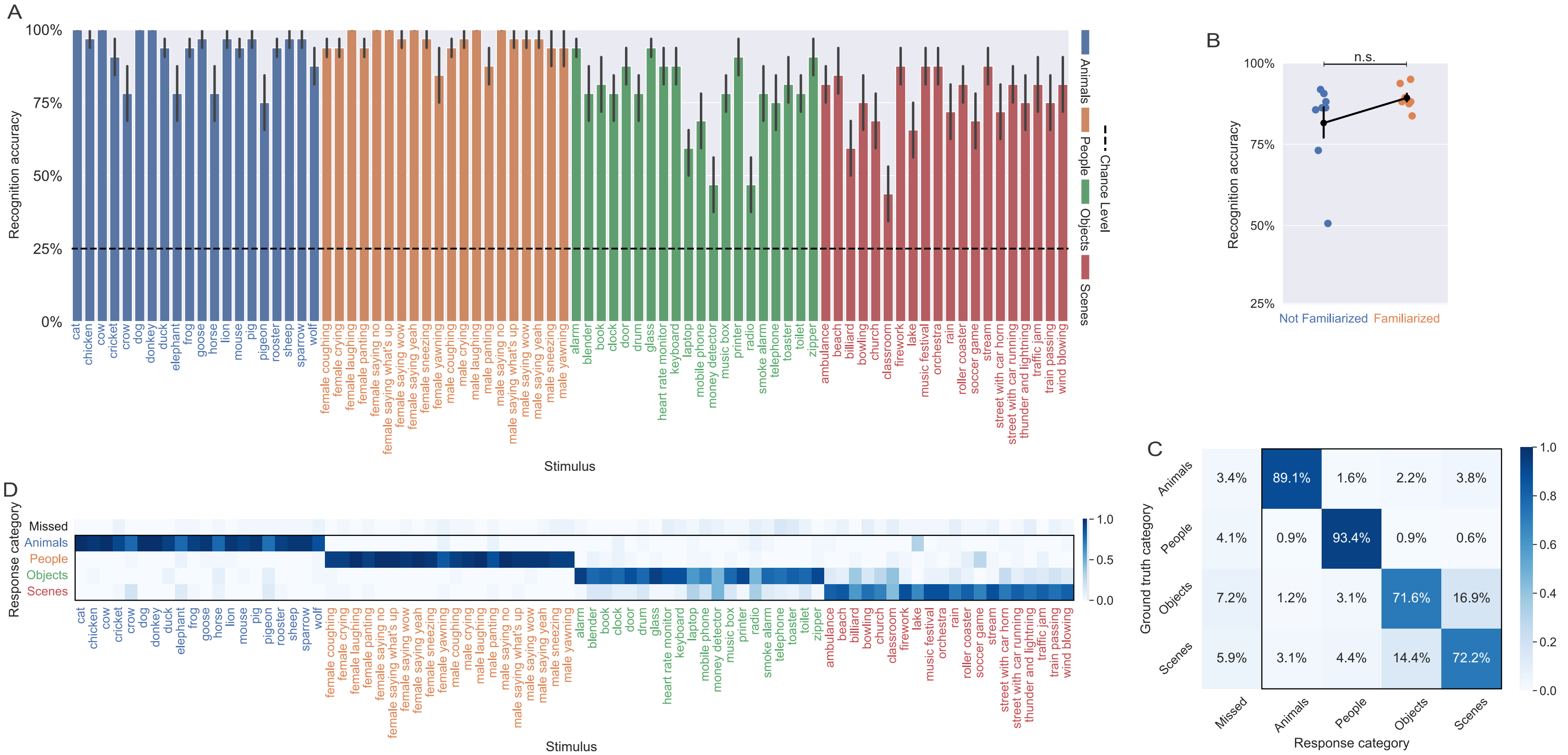

Figure 3.1: Behavior experiment results. (**A**) Performance accuracy of 16 participants over stimuli. (**B**) Performance accuracy of 8 familiarized and 8 not familiarized participants was compared in the recognition task showing non-significant difference ($t(15) = 1.54$, $p = 0.16$). (**C**) Average confusion matrix showing frequency of various responses (four categories or missed response) to a sound from certain category averaged across n=16 participants. Off diagonal elements in the outlined part indicate confusion between two categories, and diagonals are the accuracy of each category. Hence, the outlined part was used as a refined model of within and between category similarities as larger confusion between two categories implies greater similarity. (**D**) Confusion matrix for each sound showing individual sounds were correctly categorized with an accuracy significantly above chance. Confusion matrix shown in **C** is an average summary of this confusion matrix.

3.2 Decoding time series show a latency in emergence of category information compared to non-category information

A pairwise classification was conducted on all sound pairs (comprising 3160 pairs) across all 769 timepoints, spanning from -200 ms to 1300 ms in reference to the stimulus onset, individually for each participant. The average of all decoding time series (grand average) elucidates the extent of information that EEG multivariate patterns encapsulate that can be leveraged to linearly differentiate between sounds. The chance level for the averaged decoding is 50%. The statistical significance of averaged decoding accuracy (tested against 50% chance accuracy) was ascertained using a sign-flip permutation test with cluster correction (cluster defining threshold $p < 0.001$, significance threshold $p < 0.01$, 10000 permutations) to compensate for the multiple testing across timepoints (see Figure 3.2 A). Further, the onset and peak time of the average decoding was identified by bootstrapping all 25 participants with substitution for 10,000 permutations yielding 10000 onsets and peak times for each selected subset of participants. Onset and peak were defined using the method explained in the Methods section. The results revealed a mean onset for average decoding accuracy at 39 ms (95% CI [21, 60]) and a mean peak time at 185 ms (95% CI [175, 198]) (see Figure 3.2 B). This implies that discernible information is extracted by the brain and mirrored in the EEG signal as promptly as 39 ms post stimulus onset. It is plausible that this initial information primarily includes lower-level details such as cochleagram representations of the sound, a hypothesis supported by the alignment with the timings of auditory evoked potentials (Hari and Salmelin, 1997; Näätänen and Picton, 1987). Moreover, the peak time of decoding accuracy corresponds to the moment when the EEG signal embodies the most abundant information about the sounds, including semantic information.

For distinguishing between high-level, category information, and low-level, non-category information, a common RSA strategy involves comparing the average decoding accuracy of within-category and between-category sound pairs. The average of within-category pairwise decodings reflects the non-category information, given the absence of category-discriminating information in these decoding pairs. Conversely, the average of between-category pairwise decodings relies on both category and non-category information for sound discrimination. Thus, the difference between the between- and within-category averages illustrates the contribution of category information to decoding accuracy. Figure 3.2 C presents the within-category average and the difference of between- and within-category averages (the difference curve). For the within-category average, the chance level is 50%, whereas for the difference curve, it is set at 0%, given that it represents the difference between two decoding accuracies with the same chance level. The statistical significance of both time series was tested against the corresponding chance level employing sign-flip permutation test with cluster correction (cluster defining threshold $p < 0.001$, significance threshold $p < 0.01$, 10000 permutations, see Figure 3.2 C). Analogous to the grand average, the confidence intervals of the onset and peak times of the decoding curves were determined by bootstrapping (10000 permutations). The onsets of the two curves were statistically compared using their empirical distributions from bootstrapping (see Figure 3.2 D), and the mean onset of the difference curve was found to be 36 ms (95% CI [2, 82]) later than the within-category average onset ($p < 0.05$) while peak times were not significantly different. The delayed onset of the difference curve, symbolizing the contribution of category information, suggests that category information emerges after non-category information. Table3.2 summarizes mean and confidence interval for onsets and peaks of all time series.

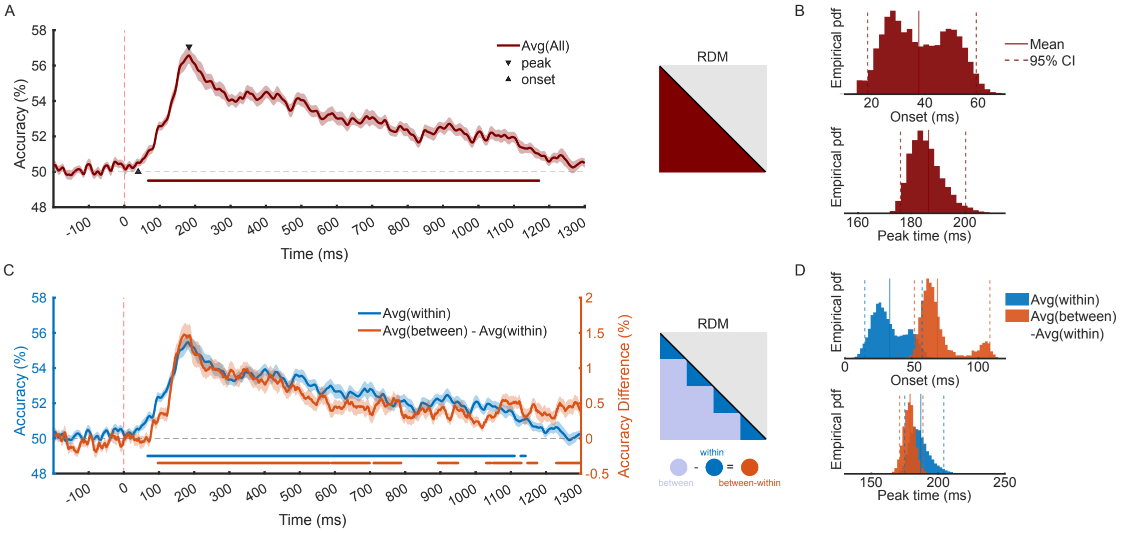

Figure 3.2: Decoding time series analysis results. (**A**) Average of all pairwise decodings (Avg(all)) over time is shown. The averaged part of the RDM (3160 pairwise classifications) is depicted on the right side of the time series plot. (**C**) Average of within-category decodings (Avg(within)) is compared to difference of between- and within-category averaged decodings (Avg(between)-Avg(within)). The averaged parts of the RDM are depicted on the right side of the time series plot, within-category pairs, between-category pairs, and the difference of their averages are shown in blue, light purple, and orange. (**B** and **D**) Empirical distribution for peak times and onsets of time series (color-coded accordingly) were calculated by bootstrapping 25 participants with substitution for 10000 times. Mean and confidence interval (annotated on the distributions by solid and dashed lines respectively) for onsets and peak times of all time series in this figure are summarized in Table 3.1. (**A** and **C**) The time series are averaged across 25 participants. The color-coded shaded areas show the standard error of the mean across participants. Stimulus onset and chance level accuracy are marked as red vertical and gray and horizontal dashed lines, respectively. The color-coded horizontal solid line indicates the time intervals where corresponding time series are significantly above chance across participants, tested by sign-flip permutation with cluster correction (cluster defining threshold $p < 0.001$, significance threshold $p < 0.01$, 10000 permutations).

Table 3.1: Onset and peak times of the time series were statistically identified through bootstrapping 25 participants with substitution in 10000 permutations. The mean and 95% confidence interval (CI) of the mean were calculated using the empirical distributions for onsets and peak times.

Decoding	Onset mean	Onset CI_{95}	Peak time mean	Peak time CI_{95}
Avg (All)	39 ms	[21 ms, 60 ms]	185 ms	[175 ms, 198 ms]
Avg (within)	33 ms	[15 ms, 58 ms]	187 ms	[175 ms, 204 ms]
Avg (between) - Avg (within)	69 ms	[52 ms, 109 ms]	179 ms	[171 ms, 189 ms]

3.3 Temporal generalization depicts category information is more sustained

The category information was decodable later than the decoding of non-category information. However, the temporal dynamics of these processes are not solely defined by the latency of these different processing levels. A critical aspect of understanding these temporal dynamics involves determining the duration for which different levels of representations are retained by the brain. We evaluated the generalizability of information represented at a particular timepoint to other earlier and later timepoints in order to answer this question. The results from the temporal generalization approach provide insights into the persistence of individual representation levels and offer the means to dissect distinct processing stages.

We calculated the average decoding accuracy for within- and between-category pairs and compared the within-category average with between- and within-category averages' difference. This approach was used to differentiate between low- and high-level information. As depicted in Figure 3.3 A, the decoding accuracy peaked close to the diagonal line (where t equals t') for the within-category average. This signifies that the representations of non-category information are transient in nature and are not held for a long period.

However, Figure 3.3 B reveals that the discrepancy between the average of within- and between-category decoding accuracies, which indirectly captures the influence of category information, was still statistically significant within the larger zones surrounding the diagonal line. This finding suggests that category information is indeed more generalizable and is consistently represented.

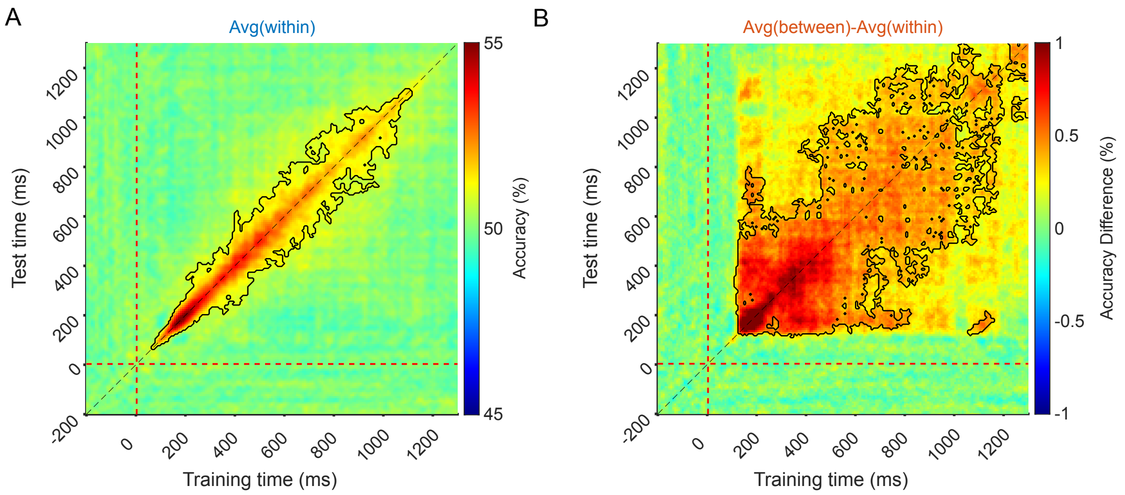

Figure 3.3: Temporal generalization analysis results. (**A**) Average of within-category decodings at each train and test timepoint (t,t') pair. (**B**) Difference of between- and within-category averaged decodings (Avg(between)-Avg(within)). Heatmaps are averaged across 25 participants. The black outlines show significant 2d clusters across participants (cluster defining threshold $p < 0.01$, significance threshold $p < 0.05$, 10000 permutations) tested by sign-flip permutation. Diagonal dashed line indicates $t = t'$.

3.4 Representational similarity analysis distinguishes levels of information

What is the nature of this non-category information? Although category and non-category information could be decoded from EEG data at different latencies, the precise nature of the information in the signals is not clear. To discern the nature of information represented across various temporal intervals, we turned to representational similarity analysis (RSA) and computational models. Two models, cochleagram and spectrotemporal modulation, were used to illustrate low-level information, while category and behavior (with adjustments to category dissimilarities) were chosen to depict high-level semantics. The measure of similarity between models and EEG data over time was evaluated via Spearman's correlation, with corrections for tied ranks. Moreover, the noise ceiling was assessed as the correlation of participants' RDMs in a leave-one-out cross-validation (comparing each participant's RDMs with the average of others), which unfolds the noise level in the data by showing how well the data correlates with itself (see Figure 3.4 A). Finally, we pinpointed the peak time of similarity for each model by bootstrapping participants with substitution (see Figure 3.4 B).

The cochleagram model was the earliest to exhibit correlation with EEG representations (mean peak time of 126 ms), and for a brief duration, it highly synchronized with the noise ceiling. This outcome suggests that cochleagram model was the best model for EEG representations before 100 ms. This aligned with expectations, given that cochleagram representations model the initial stage in the auditory processing hierarchy. Subsequently, the spectrotemporal modulation model demonstrated the highest degree of similarity with the EEG data (mean peak time of 167 ms), thereby reinforcing its role in modeling sound representations in the primary auditory cortex (Chi et al., 2005; Norman-Haignere and McDermott, 2018). Semantic models, on the other hand, showed a significantly delayed peak ($p < 0.05$) compared to each low-level model (mean peak times of 183 ms and 203 ms for category and behavior models respectively). Moreover, the behavioral model exhibited a larger peak than the category model (see Figure 3.4 A), suggesting a more accurate model of represented information. The peak similarity times were markedly distinct for each pair of models, except for the cochleagram and spectrotemporal models (see Table 3.2). This was tested using the empirical distributions for the peak times of the models. Furthermore, the models' peak times were arranged in ascending order based on their complexity, confirming the hierarchical structure of temporal processing in the auditory semantic information of sounds (see Figure 3.4 B). This arrangement of models provides more credibility to the concept of temporal hierarchy in the semantic processing of sounds.

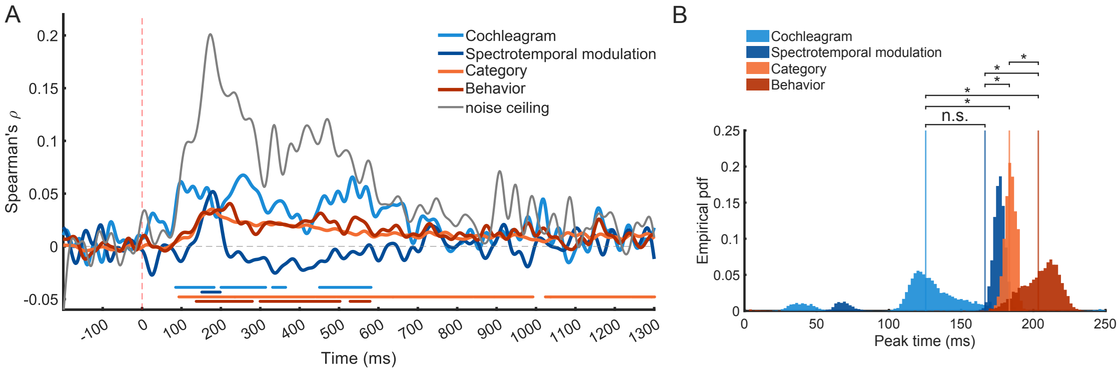

Figure 3.4: Representational similarity analysis results. (**A**) Spearman's correlation of model RDMs and EEG RDMs through time averaged across 25 participants. The color-coded horizontal solid line indicates the time intervals where corresponding correlations are significantly above zero across participants, tested by sign-flip permutation with cluster correction (cluster defining threshold $p < 0.001$, significance threshold $p < 0.01$, 10000 permutations). (**B**) Empirical distributions for peak time of model correlations shown in A identified by bootstrapping 25 participants for 10000 iterations. Asterisks indicate $p < 0.05$ significance of paired test on peak time of model i being earlier than model i+1. P-values are summarized in Table 3.2.

Table 3.2: Peak times of model similarity with EEG representations were statistically identified through bootstrapping 25 participants with substitution in 10000 permutations and tested against one another. Peak time for all models were tested against each other using the empirical distribution from bootstrap analysis. The p-values for these paired-sample tests are reported in this table. Asterisks indicate significant p-values ($p < 0.05$, one-tail).

Peak time comparison			p-value
Spectrotemporal modulation	>	Cochleagram	0.1180
Category	>	Cochleagram	0.0377^*
Category	>	Spectrotemporal modulation	0.0381^*
Behavior	>	Cochleagram	0.0413^*
Behavior	>	Spectrotemporal modulation	0.0247^*
Behavior	>	Category	0.0424^*

Chapter 4

Discussion

Our pilot behavioral experiment demonstrated that participants can accurately recognize the category of sounds within our stimulus set, and gave us empirical data for a model of category dissimilarity based in human perception. In the conducted EEG experiment, participants were subjected to stimuli that acted as ”distractors” in relation to the oddball white noise to which they were required to respond. The design of this experiment was intended to enable the capture of passive sound representations. However, this task might have potentially allowed participants to disregard the other sounds, which could potentially attenuate the decoding outcomes. Despite this plausible constraint, our results demonstrated that sound decoding from the EEG signals significantly exceeded chance levels as early as 40 ms post-stimulus onset with peak decoding around 185 ms (95% CI [175, 198]). We managed to differentiate the cognitive levels of information (non-category, including acoustic, vs category) by averaging the decoding pairs within and between categories, subsequently comparing the within-category average and the difference of between- and within-category averages. The onset of the latter time series (mean=69 ms, 95% CI [52, 109]) was delayed by approximately 36 ms (95% CI [2, 82]) compared to onset of within-category average (mean=33 ms, 95% CI [15, 58]), indicating the later emergence of category information. Applying a similar logic to discriminate information in the temporal generalization analysis, we found that category information exhibits more sustained representations. We further explored the distinction of representation levels through Representational Similarity Analysis (RSA). RSA confirmed the delayed emergence of category information compared to acoustic features. We assessed the correlation between model RDMs and EEG RDMs over time. Our observations indicated that the correlations peak at distinct times and in a specific sequence: the cochleagram model at 126 ms, the spectrotemporal modulation model at 167 ms, the category model at 183 ms, and the behavior model at 203 ms. The cochleagram was the first model to exhibit peak correlation with EEG data. This validated the cochleagram representation (low-level information) as the earliest coding of sounds in the auditory pathway and provided a sanity check. The spectrotemporal modulation model (a model for PAC) peaked later than the cochleagram model indicated by the empirical peak time distributions, however this lag was not significant. Both semantic models: category and behavior peak times was significantly later than peak times for cochleagram and spectrotemporal modulation. Interestingly, the behavior model demonstrated a higher maximum correlation than the simplistic category model. This indicates that the refined category model obtained from the behavioral experiment is a better model for brain representations captured by EEG data.

4.1 Hierarchical auditory processing

Auditory processing in nonhuman primates is characterized by distinct regions that respond to sound properties of increasing complexity (Bizley and Cohen, 2013; Kaas and Hackett, 2000; Kaas et al., 1999; Rauschecker and Scott, 2009; Rauschecker and Tian, 2000). This has been likened to the multistage hierarchy observed in the visual system (Felleman and Van Essen, 1991; Cichy et al., 2016). In humans, this hierarchical model has been primarily investigated in relation to speech features, from acoustic to semantic levels (Davis and Johnsrude, 2003; Okada et al., 2010; Peelle et al., 2010; Rauschecker and Scott, 2009; Chang et al., 2022).

However, the primary auditory cortex (PAC) in humans exhibits more complex processing compared to the primary visual cortex (V1). The anatomical structure of the auditory pathway is more parallel, with direct projections from the medial geniculate to both PAC and non-primary areas (Hackett, 2011; Kaas and Hackett, 2000). There is also a broad overlapping selectivity for complex sounds along the superior temporal plane (Bizley and Cohen, 2013). These findings challenge the strict serial hierarchical model, suggesting a more parallel approach to high-level auditory processing (Hamilton et al., 2021; Formisano et al., 2008; Staeren et al., 2009). It is noteworthy that the understanding of spatial organization within this potential hierarchy is relatively well-developed. This comprehension, however, does not extend equally to our grasp of the temporal aspects of auditory processing. The relative latencies involved in these processing stages remain less explored and understood (Lowe et al., 2021; Benner et al., 2023).

The primary objective of this study was to ascertain whether distinct temporal latencies exist between the emergence of low-level, acoustic-like, and high-level, semantic, information from auditory stimuli. Our findings provide compelling evidence in support of a cascaded hierarchy of processing, thereby contributing to the understanding of a hierarchical functional organization of auditory-related brain regions and the information flow in the brain when processing sounds.

4.2 Temporal hierarchy of auditory semantic processing

The temporal hierarchy that we observe is in line with propagation delays across the auditory system established by auditory evoked potential studies. The data from our study revealed that auditory information can be decoded from the raw EEG approximately 40 ms post-stimulus onset. The peak decoding occurs around 185 ms. This suggests that the brain begins processing auditory information rapidly after stimulus onset, with the highest level of decoding occurring within a relatively short time frame. A peak in decoding signifies the point in time at which the sounds are most distinguishable within the EEG data, indicating that the brain has effectively resolved the auditory information. The onset of decoding falls in the timing range for mid-latency responses (MLRs) (10 to 50 ms) that is associated with the activity of the thalamocortical radiations and primary auditory cortex (Näätänen and Picton, 1987; Picton, 2010).

Our analysis of the representation levels, achieved by averaging the decoding pairs within and between categories, demonstrated a delayed emergence of category information compared to non-category information. This delay of approximately 36 ms, indicates that high-level, semantic information from auditory stimuli emerges after low-level, acoustic-like information. These onset latencies for non-category (33 ms) and category (69 ms) information are aligned with average onset latencies reported by Benner et al. (2023) for medial Heschl's gyrus (HG) and lateral HG (34 ms and 63 ms, respectively). This finding elucidates the timecourse of processing stages in auditory-related brain regions, suggesting a sequential progression from low-level to high-level information processing. The onset for emergence of category information falls beyond 50 ms, in line with expectations for secondary auditory cortex and other higher-level auditory processing centers (late-latency responses) (Näätänen and Picton, 1987; Picton, 2010). Furthermore, our investigation into the temporal generalization of category information revealed a more stable representation over time. This stability suggests that once high-level, semantic information emerges, it remains consistent. The lingering of category representations versus transient representations of lower-level information provides more insight to viewing them as separate stages, further supporting the idea of a hierarchical information flow in the brain.

However, lower-level information might differ systematically between categories of sound. Thus, the difference between average of between-category and average of within-category pairs is not solely indicative of the category information contribution but also benefits from non-category information with a systematic difference between categories. For instance, sound spectral density, regarded as a low-level feature, could exhibit a higher concentration in high frequencies in animal sounds than human sounds. To account for this confounding factor and better distinguish information levels, we employed representational similarity analysis to scrutinize information levels as models, thereby gaining increased flexibility. RSA allowed us to delve deeper into the distinction of information levels and confirmed the delayed emergence of category information. The correlations between model RDMs (cochleagram, spectrotemporal modulation, category, and behavior) and EEG RDMs peaked at distinct time frames and in a specific sequence, with the cochleagram model peaking first (at 126 ms) and the behavior model peaking last (at 203 ms). This sequence of peak correlations provides additional evidence of the temporal hierarchy in neural representations of sounds, with low-level acoustic-like information resolved significantly earlier than high-level semantic information. The results replicate the previous study by Lowe et al. (2021) using MEG, which reported correlation peaks with model RDMs at 127 ms and 218 ms for cochleagram and category models, respectively.

The current study employed a different stimulus set compared to Lowe et al. (2021), who used shorter sounds (500 ms). Additionally, whereas Lowe et al. (2021) used a fixed cochleagram model over time, we utilized a time-varying cochleagram model, which calculates cochleagram features at each time point using the portion of the sound heard up to that moment, rather than using the entire sound waveform for all times. Further, we used a refined model of category similarity based on the behavioral experiment that was shown to better resemble the neural representations.

EEG and MEG are both sensitive to the postsynaptic potentials generated by cortical pyramidal neurons; however, they capture different facets of these potentials due to the divergent propagation of electrical and magnetic fields. EEG registers both tangential and radial components of these electrical potentials, making it sensitive to activity within both the sulci and gyri of the cortex (Nunez and Srinivasan, 2006; Ahlfors et al., 2010). In contrast, MEG predominantly detects the magnetic fields associated with the tangential components of intracellular currents, as radial components tend to self-cancel magnetically. Consequently, MEG is preferentially sensitive to neural activity in the sulci (Hämäläinen et al., 1993; Ahlfors et al., 2010). Conducting this study utilizing EEG was complementary to the previously tested MEG datasets (Lowe et al., 2021; Benner et al., 2023). Despite these methodological differences, the timings reported in our study align well with the timings reported by Lowe et al. (2021) and Benner et al. (2023). This alignment suggests that the observed temporal hierarchy is not merely a product of the stimulus but rather reflects a processing scheme employed by the brain.

4.3 Limitations

Multivariate pattern-information approaches like MVPA and RSA are advanced techniques that offer significant advantages for the exploration and understanding of complex neural processes. These techniques provide a more nuanced and comprehensive view of brain activity, which can be instrumental in elucidating the intricate mechanisms of the brain. However, the pairwise nature of classifications in MVPA, particularly when using SVMs, indeed presents certain limitations. Primarily, this approach does not generate a comprehensive distance measure that considers all stimuli simultaneously. Instead, it contrasts one stimulus against another at a time. This pairwise comparison could potentially restrict the scope of the analysis and the insights that can be derived from it, as it does not fully capture the multidimensional nature of the stimulus space. This, combined with the lack of adherence to the triangle inequalityin RDMs, can impose certain limitations on the inferability of RDMs. Essentially, the triangle inequality is a fundamental property of metric spaces, which posits that the distance between any two points should not exceed the sum of their distances to a third point. When this property is not observed, as in the case of RDMs formed by pairwise classification, the resulting space is non-Euclidean (Kriegeskorte and Kievit, 2013). This deviation from Euclidean geometry can complicate the interpretation of distances and may limit the applicability of numerous standard statistical techniques that are predicated on Euclidean distances.

Moreover, the absence of the triangle inequality can lead to inconsistencies in the scaling of distances. For instance, the dissimilarity between two highly similar conditions might be disproportionately large compared to the dissimilarity between two markedly different conditions, thereby distorting the perceived relationships between conditions (Kriegeskorte et al., 2008). This can negatively affect the correlation of EEG RDMs with correlation models. RDMs offer a powerful tool for comparing brain activity patterns across different conditions, tasks, or modalities, the lack of triangle inequality is a limitation that should be considered.

While SVMs offer robustness against overfitting and noise and are efficient in high-dimensional spaces, their interpretability can be challenging. The understanding of what decision boundaries signify in the context of the original data space (channel space) may not always be clear (Haufe et al., 2014). This lack of transparency can limit the interpretability of the results, which RSA addresses. Furthermore, the feature weights in SVM do not directly correlate with their importance, which could potentially obscure the understanding of the underlying patterns in the data, thus limiting the depth of the analysis (Davis and Poldrack, 2013). Nevertheless, interpreting channel contributions in decoding neural representations (presumably employing methods similar to searchlight (Kriegeskorte et al., 2006) or imposing sparse weight criteria on the model) proposes an interesting future direction to studies performing this type of analysis.

RSA enables the comparison of observed brain activity patterns with those predicted by computational or theoretical models. This capability provides a robust framework for testing specific hypotheses about the nature of the information represented in particular brain regions or at a particular time, thereby enhancing the validity of the research findings. While RSA is a potent tool for hypothesis testing, it necessitates robust a priori models or hypotheses. Consequently, its utility may be limited in exploratory analyses where the nature of the neural representations is less well defined. This limitation underscores the critical role of the models used in RSA. If the models are not well-formulated or if they do not accurately reflect the underlying neural processes, the results of the RSA may not provide meaningful or accurate insights into the neural representations of interest. For example, in our study, we demonstrated that utilizing a behavior-driven semantic model enhanced the correlation with EEG RDMs in comparison to the category model. However, it is important to note that the behavior model was derived from a pilot study where participants were instructed to categorize the stimuli into one of four categories (animals, people, objects, and scenes). This may have overlooked other potential categories that participants might have selected in an open-set task, thereby potentially limiting the explanatory power of our model in relation to neural representations. To address this limitation, a behavior task that allows participants to freely cluster the stimuli could be beneficial.

Auditory stimuli are conveyed over an extended period (one second in our study), contrasting with visual studies where the entire stimulus is presented in a brief period (less than 50 ms in most studies). Consequently, the timing of processed information identified by multivariate analysis and RSA could potentially be driven by the structure of the stimulus set, rather than being an inherent characteristic of the brain. Hence, studying the temporal dynamics of sound processing presents a significant challenge. This is due to the fact that naturalistic stimuli cannot be entirely controlled for their low-level features. Therefore, the importance of replicating results of such studies with different stimulus sets is signified.

4.4 Conclusion

In conclusion, our findings provide strong evidence in favor of different temporal latencies between the emergence of low-level, acoustic-like, and high-level, semantic, information from auditory stimuli. This contributes to our understanding of the functional organization of auditory-related brain regions and the temporal dynamics of information processing in the brain. The temporal dynamics reported in this study are consistent with, and complementary to, the spatial hierarchy of auditory processing established by previous studies (Binder et al., 2000; Davis and Johnsrude, 2003; Okada et al., 2010; Yi et al., 2019; Kell et al., 2018; Lowe et al., 2021; Benner et al., 2023; Giordano et al., 2023) and together suggest a hierarchical functional organization for auditory processing.

References

Ahlfors, S. P., Han, J., Belliveau, J. W., and Hämäläinen, M. S. Sensitivity of MEG and EEG to source orientation. *Brain Topogr.*, 23(3):227–232, September 2010.

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., and Grady, C. L. “what” and “where” in the human auditory system. *Proceedings of the National Academy of Sciences*, 98 (21):12301–12306, 2001.

Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. Assessing the auditory dual-pathway model in humans. *Neuroimage*, 22(1):401–408, May 2004.

Benner, J., Reinhardt, J., Christiner, M., Wengenroth, M., Stippich, C., Schneider, P., and Blatow, M. Temporal hierarchy of cortical responses reflects core-belt-parabelt organization of auditory cortex in musicians. *Cereb. Cortex*, 33(11):7044–7060, May 2023.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex*, 10(5):512–528, May 2000.

Bizley, J. K. and Cohen, Y. E. The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.*, 14(10):693–707, October 2013.

Carrasco, A. and Lomber, S. G. Evidence for hierarchical processing in cat auditory cortex: nonreciprocal influence of primary auditory cortex on the posterior auditory field. *J. Neurosci.*, 29(45):14323–14333, November 2009.

Chang, C.-C. and Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):1–27, May 2011.

Chang, C. H. C., Nastase, S. A., and Hasson, U. Information flow across the cortical timescale hierarchy during narrative construction. *Proc. Natl. Acad. Sci. U. S. A.*, 119 (51):e2209307119, December 2022.

Chi, T., Ru, P., and Shamma, S. A. Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.*, 118(2):887–906, August 2005.

Cichy, R. M. and Pantazis, D. Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space. *Neuroimage*, 158:441–454, September 2017.

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., and Oliva, A. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.*, 6:27755, June 2016.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., Abdi, H., and Haxby, J. V. The representation of biological classes in the human brain. *J. Neurosci.*, 32(8):2608–2618, February 2012.

Darrow, K. N., Maison, S. F., and Liberman, M. C. Cochlear efferent feedback balances interaural sensitivity. *Nat. Neurosci.*, 9(12):1474–1476, December 2006.

Davis, M. H. and Johnsrude, I. S. Hierarchical processing in spoken language comprehension. *J. Neurosci.*, 23(8):3423–3431, April 2003.

Davis, T. and Poldrack, R. A. Measuring neural representations with fMRI: practices and pitfalls. *Ann. N. Y. Acad. Sci.*, 1296:108–134, August 2013.

DeWitt, I. and Rauschecker, J. P. Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U. S. A.*, 109(8):E505–14, February 2012.

Felleman, D. J. and Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, 1(1):1–47, 1991.

Formisano, E., De Martino, F., Bonte, M., and Goebel, R. “who” is saying “what”? brain-based decoding of human voice and speech. *Science*, 322(5903):970–973, November 2008.

Giordano, B. L., Esposito, M., Valente, G., and Formisano, E. Intermediate acoustic-to-semantic representations link behavioral and neural responses to natural sounds. *Nat. Neurosci.*, 26(4):664–672, April 2023.

Grossberg, S. Developmental designs and adult functions of cortical maps in multiple modalities: Perception, attention, navigation, numbers, streaming, speech, and cognition. *Front. Neuroinform.*, 14:4, February 2020.

Guggenmos, M., Sterzer, P., and Cichy, R. M. Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *Neuroimage*, 173:434–447, June 2018.

Hackett, T. A., Stepniewska, I., and Kaas, J. H. Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J. Comp. Neurol.*, 394(4):475–495, May 1998.

Hackett, T. A., Preuss, T. M., and Kaas, J. H. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J. Comp. Neurol.*, 441(3):197–222, December 2001.

Hackett, T. A. Information flow in the auditory cortical network. *Hear. Res.*, 271(1-2): 133–146, January 2011.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.*, 65(2):413–497, April 1993.

Hamilton, L. S., Oganian, Y., Hall, J., and Chang, E. F. Parallel and distributed encoding of speech across human auditory cortex. *Cell*, 184(18):4626–4639.e13, September 2021.

Hari, R. and Salmelin, R. Human cortical oscillations: a neuromagnetic view through the skull. *Trends Neurosci.*, 20(1):44–49, January 1997.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87:96–110, February 2014.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. December 2015.

Hillman, E. M. C. Coupling mechanism and significance of the BOLD signal: a status report. *Annu. Rev. Neurosci.*, 37:161–181, 2014.

Hubel, D. H. and Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160(1):106–154, January 1962.

Kaas, J. H. and Hackett, T. A. Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U. S. A.*, 97(22):11793–11799, October 2000.

Kaas, J. H., Hackett, T. A., and Tramo, M. J. Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.*, 9(2):164–170, April 1999.

Kandel, E. R., Koester, J. D., Mack, S. H., and Siegelbaum, S. A. *Principles of Neural Science, Sixth Edition*. McGraw Hill Professional, April 2021.

Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V., and McDermott, J. H. A Task-Optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, 98(3): 630–644.e16, May 2018.

King, J.-R. and Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.*, 18(4):203–210, April 2014.

Kriegeskorte, N. and Douglas, P. K. Cognitive computational neuroscience. *Nat. Neurosci.*, 21(9):1148–1160, September 2018.

Kriegeskorte, N. and Kievit, R. A. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.*, 17(8):401–412, August 2013.

Kriegeskorte, N., Goebel, R., and Bandettini, P. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.*, 103(10):3863–3868, March 2006.

Kriegeskorte, N., Mur, M., and Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.*, 2:4, November 2008.

LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *Nature*, 521(7553):436–444, May 2015.

Li, J., Lavrukhin, V., Ginsburg, B., Leary, R., Kuchaiev, O., Cohen, J. M., Nguyen, H., and Gadde, R. T. Jasper: An End-to-End convolutional neural acoustic model. April 2019.

Lowe, M. X., Mohsenzadeh, Y., Lahner, B., Charest, I., Oliva, A., and Teng, S. Cochlea to categories: The spatiotemporal dynamics of semantic auditory representations. *Cogn. Neuropsychol.*, 38(7-8):468–489, 2021.

Luck, S. J. *An Introduction to the Event-Related Potential Technique, second edition*. MIT Press, May 2014.

Miller, C. T., Gire, D., Hoke, K., Huk, A. C., Kelley, D., Leopold, D. A., Smear, M. C., Theunissen, F., Yartsev, M., and Niell, C. M. Natural behavior is the language of the brain. *Curr. Biol.*, 32(10):R482–R493, May 2022.

Mohsenzadeh, Y., Qin, S., Cichy, R. M., and Pantazis, D. Ultra-Rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *Elife*, 7, June 2018.

Mountcastle, V. B. The columnar organization of the neocortex. *Brain*, 120 (Pt 4): 701–722, April 1997.

Näätänen, R. and Picton, T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24 (4):375–425, July 1987.

Niedermeyer, E. and da Silva, F. H. L. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2005.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. A toolbox for representational similarity analysis. *PLoS Comput. Biol.*, 10(4):e1003553, April 2014.

Norman-Haignere, S. V. and McDermott, J. H. Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLoS Biol.*, 16(12):e2005127, December 2018.

Norman-Haignere, S. V., Kanwisher, N., McDermott, J. H., and Conway, B. R. Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. *Nat. Neurosci.*, 22(7):1057–1060, July 2019.

Nunez, P. L. and Srinivasan, R. *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2006.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J. T., and Hickok, G. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb. Cortex*, 20(10):2486–2495, October 2010.

Peelle, J. E., Johnsrude, I. S., and Davis, M. H. Hierarchical processing for speech in human auditory cortex and beyond. *Front. Hum. Neurosci.*, 4:51, June 2010.

Picton, T. W. *Human Auditory Evoked Potentials*. Plural Publishing, September 2010.

Rauschecker, J. P. and Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.*, 12(6):718–724, June 2009.

Rauschecker, J. P. and Tian, B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences*, 97 (22):11800–11806, 2000.

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123 Pt 12(Pt 12):2400–2406, December 2000.

Scott, S. K., Rosen, S., Lang, H., and Wise, R. J. S. Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *J. Acoust. Soc. Am.*, 120(2):1075–1083, 2006.

Serre, T., Oliva, A., and Poggio, T. A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U. S. A.*, 104(15):6424–6429, April 2007.

Staeren, N., Renvall, H., De Martino, F., Goebel, R., and Formisano, E. Sound categories are represented as distributed patterns in the human auditory cortex. *Curr. Biol.*, 19(6):498–502, March 2009.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.*, 2011:879716, April 2011.

Warr, W. B. Organization of olivocochlear efferent systems in mammals. In Webster, D. B., Popper, A. N., and Fay, R. R., editors, *The Mammalian Auditory Pathway: Neuroanatomy*, pages 410–448. Springer New York, New York, NY, 1992.

Woodman, G. F. A brief introduction to the use of event-related potentials in studies of perception and attention. *Atten. Percept. Psychophys.*, 72(8):2031–2046, November 2010.

Yi, H. G., Leonard, M. K., and Chang, E. F. The encoding of speech sounds in the superior temporal gyrus. *Neuron*, 102(6):1096–1110, June 2019.

Appendix A

Health sciences research ethics board approval letter

Western
Research

Date: 15 June 2022

To: Dr Yalda Mohsenzadeh

Project ID: 120730

Review Reference: 2022-120730-67651

Study Title: Computational Models for Auditory Processing in the Brain

Application Type: HSREB Initial Application

Review Type: Delegated

Full Board Reporting Date:28/June/2022

Date Approval Issued: 15/Jun/2022 11:55

REB Approval Expiry Date: 15/Jun/2023

Dear Dr Yalda Mohsenzadeh

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the above mentioned study as described in the WREM application form, as of the HSREB Initial Approval Date noted above. This research study is to be conducted by the investigator noted above. **All other required institutional approvals and mandated training must also be obtained prior to the conduct of the study**.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
Other	Other Data Collection Instruments	06/Jun/2022	
Auditory Perception_Flyer	Recruitment Materials	26/May/2022	v2
Auditory Perception_ Email	Email Script	30/May/2022	
Auditory Perception_ LOI & Consent	Written Consent/Assent	08/Jun/2022	v3
Auditory Perception_Protocol	Protocol	14/Jun/2022	v 2

No deviations from, or changes to, the protocol or WREM application should be initiated without prior written approval of an appropriate amendment from Western HSREB , except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Electronically signed by:

Ms. Nicola Geoghegan-Morphet , Ethics Officer on behalf of Dr. Philip Jones, HSREB Chair, 15/Jun/2022 11:55

Reason: I am approving this document

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations, See

Curriculum Vitae

Name:	Ali Tafakkor
Post-Secondary Education and Degrees:	Sharif University of Technology, Tehran, Iran. B.Sc. Electrical Engineering Sept. 2016 - June 2021
	University of Western Ontario, London, Canada. M.Sc. Neuroscience Sept. 2021 - Aug. 2023
Honours and Awards:	Vector Research Grant, Summer 2023, Amount: 4000 CAD.
	Vector Scholarship in AI, Sept. 2021 - Aug. 2022, Amount: 17,500 CAD

Presentations:

- Tafakkor, A., Johnsrude I., Mohsenzadeh, Y., 2023. *Temporal Dynamics of Natural Sound Categorization.* OHBM 2023, Montréal, Canada.
- Tafakkor, A., Johnsrude I., Mohsenzadeh, Y., 2023. *Temporal Dynamics of Natural Sound Categorization.* NRD 2023, London, Canada.