1-13-2023 9:30 AM

Perceptual benefits from long-term exposure to naturalistic sound patterns

Bruno A. Mesquita, *The University of Western Ontario*

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

Abstract

Our brains are proficient in learning recurring structures in the environment, in order to optimize perceptual inferences based on relevant information in a stochastic input. Sensory information is multi-dimensional, and the relationship between sound dimensions may be, in itself, a source of information. Many sounds in our environment covary dynamically, and these covariances may be learned, and therefore shape our perception, through exposure to them in our natural environment. In the present study we investigate how natural (long term), and experimental (short term), learning of statistical regularities in sounds may shape our ability to categorize them (Experiment 1) and to perceptually segregate them more easily from target speech (Experiment 2). Our results indicate that sounds that obey naturalistic pitch-speed relationships are more easily categorized than those that violate these expectations. However, these benefits did not translate into greater segregability of these naturalistic patterns from speech, although my method may have not been sufficiently sensitive to such effects. These findings highlight the ways in which long-term life experience may influence our auditory perception.

Keywords

Perceptual organization, Statistical learning, Auditory categorization, Bayesian perception, Multilevel modelling,

Summary for Lay Audience

We live in a very complex sensory world. All around us, we are constantly exposed to a variety of sights and sounds that compete for our attention. And yet, even if we don't notice it, our brains are excellent at identifying patterns to help us better make sense of our environment: Light usually comes from above, so dark patches in the ground are often shadows; and if you hear loud thunder, the storm is probably pretty close! Not only that, but many properties of a stimulus may vary together (or covary) over time. For instance, many sounds in our environment seem to show a positive covariance between pitch and speed: As machines 'power up' and their parts move faster, they sound higher pitched, and faster speech is usually higher pitched. The relationship between these sound properties is quite strong and can even result in illusory effects where sounds that are played at a higher pitch tend to also sound 'faster' even when that is not physically true. Learning these covariances may be a valuable tool for providing extra 'redundancy' in environmental information, allowing us to infer additional information about sounds even in noisy or ambiguous listening conditions.

My project aims to investigate how sound patterns that are learned over time influence our perception. More specifically, here we look at how this long-term familiarity with positive pitch-speed covariances may result in: (1) more accurate and faster categorization of sounds that obey these rules; and (2) more effective segregation of noise that matches these rules from target sounds, allowing us to better pay attention to more relevant information.

This project will ultimately contribute to better understanding of the ways in which long-term life experiences may shape our perception of the world around.

Co-Authorship Statement

The research done throughout this project will be submitted for publication with Bruno Mesquita as the first author. The co-authors will then be as follows: Björn Herrmann, Casey L. Roark, Ingrid S. Johnsrude. Björn Herrmann contributed with the scripts for stimuli generation, Casey L. Roark contributed with the design of the experiment and interpretation of results, and Ingrid S. Johnsrude was the principal investigator of this project, providing funding and counsel throughout all steps of the project.

Acknowledgments

I'd like to first of all thank my supervisor Ingrid, whose help along the way was invaluable not only through her wise mentorship, but also in her warmth and kindness.

I thank all of those in the CoNCHLab, who beyond simply co-workers, are also valuable friends in my daily life.

I also extend my most sincere gratitude and love to my family, and especially my parents, Flavio, and Debora, who were always there to support and encourage me in my life and passions.

And finally, to all those who are special to me, both near and far. From the new friends that I met here in Canada, as well as to those who have been with me throughout many years, with special mention to Augusto and Bruno (not to be confused with me), and my amazing partner Bruna (also not to be confused with me).

Table of Contents

List of Figures

List of Appendices

Chapter 1

1 Introduction

As we venture through our daily lives, we are constantly immersed in rich sensory environments. And yet, we are impressively proficient at deconstructing the complex input we receive from all around us into neatly organized sensory objects. When I look though my desk in front of me, I don't see a solid mass of colors and luminance; instead, I can quite clearly identify my monitor, a water bottle, and quite a few other objects (perhaps more than there should be!). Although this process tends to be more intuitive when thinking about visual information, our auditory system is perhaps even more impressive in its capacity to parse information from a complex input of competing sources. To highlight the impressiveness of this feat, Plack (2018) compares the accomplishments of our auditory system to that of a person who can determine how many swimmers there are in a lake, and which stroke they are each using, only by the pattern of ripples in the water that arrive at the shore.

To make matters even more complicated, the output from the various components of our sensory landscapes may interact in intricately complex manners, resulting in a stochastic combination beyond the sum of its parts: The light from my monitor may bounce off my water bottle, and the ripples produced by the swimmers may collide before reaching the shore. Fortunately for us, our brains are extremely sensitive to environmental statistics, allowing us to exploit redundancies and contextual cues in both prior and current experiences to aid in our understanding of our sensory scene: I usually keep my water bottle to my left, and swimmers using a butterfly stroke tend to produce larger splashes in the water.

Understanding how our brain captures and makes use of statistical information in the environment is an important topic in Psychology and Neuroscience (Batterink et al., 2019, Turk-Browne., 2012), with extensive implications for a better understanding of sensory-processing mechanisms, and of memory formation and consolidation (Henin et al., 2021; Conway, 2020; Schapiro et al., 2016). Most of the literature on this topic,

however, focuses on experiments that evaluate how the capacity of humans to learn short-term regularities results in improvement of performance on behavioral tasks (Turk-Browne., 2012). However, there are many types of regularities that extend to a much larger scale, being present for our entire lives. The present work aims to investigate the relationships between these short-term regularities and those potentially learned over a longer term (like a whole lifetime), and how both may aid in our ability to organize auditory sensory information.

1.1 Auditory Scene Analysis and Perceptual Organization

To adequately process incoming auditory information, our brains must first be able to perceptually organize this input into different streams of sounds from different sources in the environment, in a process known as 'Auditory Scene Analysis' (ASA) (Bregman, 1990). Picture yourself in a crowded restaurant, seated across from a friend and having a casual conversation. At this moment, you are receiving auditory information from all around you, with the babble of multiple people speaking, cutlery clinking on plates, and music in the background. And yet, you can both identify discrete events in this input, such a glass breaking, and group the sequence of sounds that comprise your friend's speech into a meaningful stream of words. These processes of streaming discrete and sequential sounds are respectively defined by Bregman (1990) as "simultaneous" and "sequential" grouping.

Much of the literature on ASA has focused on the classical 'ABA' or 'Horse-Morse' paradigm, as a simplified example of sequential grouping (Carlyon, 2004; van Noorden, 1975). In this design, a repeating cycle of pure tones is formed by interleaving two isochronous sequences of tone pips one of lower (A), and one of higher (B) frequency. Depending on manipulations made to these tones and their presentation patterns, listeners may shift between perceiving them as a singular stream resembling a 'galloping' sound or two concurrent streams that are reminiscent of morse code. The most typical manipulation of the tones is done by increasing the speed of the alternation of tones (Carlyon, 2004; van Noorden, 1975). Still, studies have also been done manipulating other cues that have been shown to be influential in sequential grouping, such as pitch differences between the two tones (Grimault, 2000), their timbre (Cusak & Roberts,

2000, and amplitude modulation (Dolležal et al., 2012). This phenomenon can be considered analogous to that observed by Gestalt psychologists in vision (Bregman, 1994), where the location of sensory components leads to their organization into distinct clusters. Take this assortment of circles for example:

OOO OOO

Although there are multiple circles present, we would tend to perceptually identify two 'clusters' of circles because of their physical distribution in space. Although in this case we are dealing with a physical distance between stimuli, this process extends to any form of distance in a 'perceptual space'. This 'perceptual distance' is defined as *d*, and in the case of the ABA paradigm is defined by the differences between A and B across any of several acoustic dimensions (frequency, time, spatial location…). Not all acoustic dimensions have the same perceptual weight on grouping, however, and so we also say that *d* is a *weighted* combination of differences between A and B.

Although the formation of sensory streams of information is a crucial concept for the perceptual organization of sound, another important concept the definition of an Auditory Perceptual Object. While an 'auditory stream' refers to a phenomenological unit of sound organization that is primarily characterized by its separability from other components, an Auditory Perceptual Object is a predictive representation, constructed from feature regularities extracted from auditory sensory input (Winkler, 2009; also see Griffiths & Warren, 2004). The emphasis of this definition as a pattern with predictable components expands upon the original concept of an auditory stream, highlighting the importance of top-down processes that guide grouping decisions through contextual information. Incorporating this concept will aid us in bridging the literature discussed in this section with the rest of the work.

1.2 Statistical Learning

Our brains are proficient at capturing regularities in the environment, making use of the recurrent nature of certain patterns in space and time to highlight important information

in a stochastic input. Studies investigating the extraction of stimulus statistics are commonly associated with the concept of Statistical Learning (SL; Saffran et al., 1996). In cognitive neuroscience and psychology, SL refers to the extraction of regularities in the environment over space and/or time (Turk-Browne, 2012). This concept was first introduced by Saffran et al. (1996) in their work on infant language acquisition, where 8-month-old infants were shown to be able to extract word boundaries from a continuous stream of speech based solely on the statistical relationships between neighboring sounds.

Over the past two decades, the field of SL has been expanded through a variety of experimental designs, across many sensory modalities, from vision (Seriés & Seitz, 2013; Bertels et al. 2012; Turk-Browne et al., 2009), to touch (Conway & Christiansen, 2005), to spatial orientation (Graves et al., 2022; Graves et al., 2020). In auditory neuroscience, it has been observed in studies involving both artificial, highly controlled sounds (Bianco et al., 2020; Woods & McDermott, 2018) as well as more naturalistic stimuli such as speech (Lehet & Holt, 2020; Stilp & Assgari, 2019; Lehet & Holt, 2017) and music (Pearce, 2018). An important aspect of SL that makes it such a crucial process for perception as whole is how it allows the brain to extract sensory objects from the combined undifferentiated input of the environment, overcoming variability through pooling of sensory data into statistical summaries (McWalter & McDermott, 2018). The concept of an auditory perceptual object discussed in the previous section, is relevant here since it is possible that SL may directly aid in the segregation of these predictive representations from the environment by facilitating isolating objects in a continuous stream of information. Together, these processes help us both extract important information from sensory input, and group it into organized perceptual objects.

According to Turk-Browne and colleagues (2012), statistical learning is defined by three criteria: (1) Can operate over undifferentiated input (streams of information without clear boundaries between stimuli), (2) occurs incidentally as a by-product of perception without intention or awareness of the subject and (3) is concerned with extracting how particular features and objects co-occur, which results in knowledge about specific stimuli. Although the present work owes much of its inspiration and theoretical background to this field of research, it is not designed to fulfill these conceptual criteria.

Therefore, although the extraction and learning of environmental and stimulus statistics will be often discussed as '*learning* of statistics', this is not synonymous with the classical concept of 'Statistical Learning'.

Although much of the discussion surrounding extraction and learning of statistical regularities in the environment of these classical SL studies is particularly focused on how our perceptual systems respond to *short-term* regularities, these are intended to simulate *long-term* regularities that may be learned a similar way.

1.3 Learning of long-term environmental regularities

Even though complex and inevitably stochastic, our acoustic environment is by no means *random*. Indeed, many properties of our sensory landscape tend to respect distinct patterns, due to the laws of physics that govern our world. For instance, when we notice something about to roll off a table, our immediate reaction is to reach *down* to try to grasp it mid-air. It would be quite a striking scene to see someone grasp *above* their heads in response to a pen falling off a desk.

It seems perfectly reasonable, therefore, that in a world filled with patterns, the ability to extract and somehow store these patterns to improve both speed and accuracy of perception would provide a distinct evolutionary advantage. The logical next step to this questioning would be to determine whether sensitivity to these patterns is a result of experience or a genetic predisposition. As is often seen in matters of the famous question of nature versus nurture, we have evidence for both. For instance, human infants have been shown to prefer listening to speech as opposed to samples of synthetic sine-wave analogs of speech (Vouloumanos & Werker, 2004, 2007), warbled tones (Samples & Franklin, 1978), or filtered speech (Spence & DeCasper, 1987). A study by Voulomanos and Werker (2007) demonstrated that this bias of human infants for speech could be present even in neonates, who adjusted their high amplitude sucking to preferentially listen to speech, compared to highly controlled non-speech analogues. These studies could potentially indicate that our brains are particularly tuned to certain preferred, expected, patterns –such as properties characteristic of speech –from birth.

On the other hand, the evidence for our sensory perception being shaped by experience is also plentiful, with much of it coming from studies investigating language learning and perceptual differences of native versus foreign speakers of a language. Speech is complex and highly variable, being comprised of multiple interacting acoustic dimensions that together define phonetic categories. Indeed, listeners are known to be more sensitive to statistical patterns within their native language compared to those found in an unfamiliar language (Maye et al., 2008, Maye, Werker & Gerken, 2002). Furthermore, the perceptual benefits related to long-term familiarity with the target talker are also well known in the literature (Johnsrude et al., 2013; Barker & Newman, 2004; Magnuson, Yamada & Nusbaum, 1995).

These experience-related benefits, however, do not only reflect long-term experience, but can also arise through short-term exposure in experimental settings. Listeners can adapt to novel speech statistics that violate patterns that reflect the phonetic system of their native language, providing benefits to phoneme categorization (Liu and Holt, 2016; Idemaru and Holt, 2011), intelligibility (Bradlow & Bent, 2008), and decreased listening effort (Brown et al., 2020). Similarly, previous works in the literature have also highlighted how the perceptual benefits related to voice familiarity can also arise through short-term exposure (Holmes et al., 2020; Kreitewolf et al., 2017).

A different way to look at the topic of learning of long-term statistics in through the perspective of Bayesian frameworks for human perception. These frameworks propose that sensory input is often ambiguous, and that perception is a process of unconscious inference, in which prior knowledge is used to resolve this ambiguity (Helmholtz 1867; R.L. Gregory, 1963; Skoe et al., 2015; Kersten et al., 2004; Mamassian et al., 2002). This knowledge, in the form of probabilistic Bayesian priors, is thought to be constantly integrated with sensory information in order to effectively perceive a dynamic sensory world. A classic example of how priors may shape our perception can be seen in Figure 1, where due to our consistent experience with an environment where light tends to come from above (Sun, J. & Perona, 1998), by manipulating the position of the shading in the circle, we can create a visual illusion that shifts our perception between concave and convex shapes even on a two-dimensional image (Adams et al., 2004).

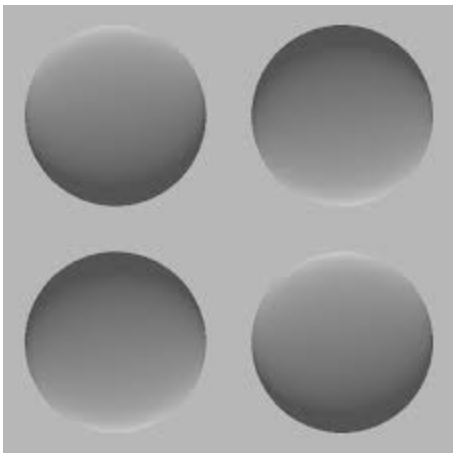

Figure 1: Example of the 'light-from-above' prior, where patches that are brighter at the top tend to be seen as convex, while those that are brighter below tend to be seen as concave. (Adams et al., 2004).

Taken together, we now have discussed how our brains act to perceptually organize an auditory scene into objects, and how this process can be aided by learning statistical regularities that can range both from short-term novel information presented in an experimental setting to long-term regularities learned throughout an individual's lifetime. In the next section we will discuss how these concepts relate to interactions between components of multi-dimensional stimuli.

1.4 How we learn what to expect

Even within a single sensory domain, stimulus 'objects' consist of many different dimensions that may change dynamically, and somewhat independently (Garner, 1976; Idemaru & Holt, 2011.). Auditory perception depends on a variety of acoustic properties such as envelope and spectral composition, which in turn give rise to perceptual features of sound such as pitch and timbre. Nevertheless, important environmental information is not only carried by individual dimensional properties themselves, but by how they relate to each other.

Many sounds in our environment may display patterns of covariance where certain acoustic dimensions tend to change together in a similar fashion. It could be expected, therefore, that our auditory system has several priors that relate to patterns of change in acoustic dimensions of naturalistic stimuli (e.g., speech, music, animal calls, and mechanical sounds). These patterns would be related to redundant attributes of sounds: sounds created by natural/real structures, such as musical instruments and vocal tracts, may have an inherent coherence between certain acoustic dimensions in accordance with physical laws governing sound-producing sources (Stilp et al., 2010).

For instance, music and speech tend to be perceived as faster (Collier & Hubbard, 2001; Boltz, 1998; Bond & Feldstein, 1982) and louder (Neuhoff, 2004; Neuhoff et al., 1999) when either pitch or sound intensity increases. The degree to which these covariances are learned, and shape our perceptions, is probably related to our exposure to them in our environment. For example, sounds that have positive covariance in amplitude modulation rate and pitch (Black, 1961; Broze & Huron, 2013), and between intensity and pitch (Neuhoff, 2004), seem to be more common in our natural acoustic environment than sounds with negative covariance. For instance, as machines such as a helicopter 'power-up' they produce higher pitched sounds as they spin faster. These expectations can even produce illusory effects in perception, where increases in pitch may lead to higher perceived intensity (Neuhoff, 2004; Neuhoff et al., 1999) or speed (Hermann et al., 2020; Hermann & Johnsrude, 2018). Although this illusory effect on perception may initially seem disadvantageous, this mechanism could be a result of an adaptive process of perception. Environmental information is often redundant, and covariance in stimulus dimensions may aid perceptual decisions, especially in noisy and ambiguous listening conditions (Hermann & Johnsrude, 2018).

1.5 Objectives

As I have highlighted so far, previous literature has demonstrated that our brains capture short-term statistical information in auditory stimuli. However, our life experience is also

filled with recurrent structures in sounds that may also enable learning of long-term regularities in how sounds are produced and propagate in our natural acoustic environments. In this regard, multiple works in the past have highlighted how listeners are sensitive to statistics of their native language (Hillenbrand et al., 2000; Dorman et al., 1977; Whalen et al., 1993), but others still have shown how listeners may also adapt to novel regularities in speech with a foreign or artificial accent (Liu and Holt, 2015; Idemaru & Holt, 2011). Nevertheless, fewer studies have looked into the effects of long-term expectations on more fundamental properties of simpler stimuli. A recent work by Roark and Holt (2022), for instance, showed that participants had better categorization performance when distinguishing between stimuli that conformed to perceptual priors thought to derive from the shared neural encoding of the two informative acoustic dimensions for the categorization boundaries. These effects were robust and resilient even to short-term passive exposure to sounds that violated these prior expectations. It is unclear, however, if similarly robust effects would also be observed for prior expectations derived from long-term life experience with the relationship between the informative dimensions, nor if these perceptual benefits extend to aspects of perceptual organization of this auditory information. Finally, previous works have mainly investigated the learning of time-invariant properties, as well as correlations between stable, unchanging properties. However, as we have mentioned in previous sections, patterns of dynamic change *over time* may very well provide important cues for our acoustic perception.

The present work aims to investigate how long-term priors and short-term learning of statistical regularities may shape auditory perceptual organization and categorization for dynamic features of sound. Here, I present two studies. The first experiment is designed to test whether long-term perceptual priors related to how sounds change over time may provide benefits in perceptual categorization of auditory stimuli. We employ stimuli whose categories are defined by covariance relationships in changes over time in the dimensions of amplitude modulation (AM; that is, how ‘fast’ stimuli are modulated over time) and carrier frequency modulation (CFM; how stimuli shift from high to low or low to high over time) thus, we create stimuli that are comprised of mostly identical acoustic properties, only differing by the direction of change in these dimensions over time.

I will: (1) Investigate how perceptual priors may act over stimuli with dimensions that are difficult to attend selectively and that differ solely in their *covariance* patterns; (2) Test priors that are, presumably, primarily related to learned relationships through daily exposure in our environment (structures tend to produce higher pitched sounds as they 'speed up'). Indeed, during piloting of our categories, participants commented on mnemonic devices used to aid them in the categorization task by saying that a certain category sounded like 'a helicopter speeding up' or 'a machine powering down' suggesting that these covariance rules do tap into empirical experiences of the individuals to some extent.

The second experiment will investigate whether long-term perceptual priors may provide benefits to perceptual organization of sounds, and whether learning of short-term statistical regularities may conflict with these long-term priors. This experiment introduces a novel design wherein participants will first go through active exposure where they will be trained only on sounds that either conform to, or violate, the same long-term priors tested on the first experiment, followed by testing of their ability to segregate target speech from maskers that belong to both categories. Previous work done by our research group has shown that learned acoustic cues associated with familiarity with voices can be exploited by listeners to both better segregate this familiar voice from competing masker speech as well as to better ignore it when attending to novel target speech targets (Johnsrude et al., 2013). Similarly, we hypothesize that sounds that are congruent with long-term perceptual priors will provide benefits to perceptual organization, and therefore be easier to segregate from the target speech. In addition, as has also been previously observed in the literature regarding adaptation to novel sound statistics (Liu and Holt, 2015; Idemaru & Holt, 2011), we also expect that short-term experience with sounds incongruent with prior expectations during a preceding familiarization task will result in more efficient segregation of this category relative to participants who were not familiarized with these sounds beforehand.

Chapter 2

2 Experiment 1

The first experiment aimed at investigating the behavioral effects of long-term priors on how sounds change over time in a categorization task.

2.1 Methods

2.1.1 Participants

Participants were recruited from Amazon's Mechanical Turk online participant pool (https://www.mturk.com) using the premium Cloud Research tool for sourcing participants (Litman et al., 2017). Data were collected from individuals ages 18-35 residing in the United States who were native English speakers with normal hearing.

In order to further guarantee that participants recruited would comply with these criteria, only CloudResearch approved participants were used for this study, which refines recruited participants to a select group who have passed a series of attention and engagement measures. Making use of such filtering tools by CloudResearch has been shown to provide reliable and high-quality data (Eyal et al. 2021).

The study was approved by Western University's Non-Medical Research Ethics Board (Project ID 112574; Appendix B).

In total, data from 192 participants were collected for Experiment 1.

2.1.2 Stimuli

Stimuli were generated and root mean-square (RMS) normalized via MATLAB (MathWorks, Inc., Natick, MA, USA) using custom functions. Stimuli consist of complex AM sounds of 150 components with 3s duration. Sounds varied in AM rate (AMr), carrier frequency (CFM), and spectral composition (timbre), with the first two jointly signaling stimulus category while timbre is an orthogonal dimension irrelevant to the categorization process, serving to increase task difficulty. Stimuli consisted of sounds

with amplitude modulation rate and frequency of components linearly increasing or decreasing over time (Fig 2). In this design, AM rate and frequency can either have positive covariance (with both increasing or decreasing simultaneously), or negative covariance (when one increases, the other decreases). The manipulation of the carrier frequency range of the complex tones was orthogonal to the categorization task and served to increase stimuli variety and task difficulty by varying the third dimension of timbre across stimuli.

Stimuli were generated by permuting across modulation and frequency range parameters (Figure 3). Amplitude was modulated at a depth of 0.7, varying over time with a minimum rate of 2 Hz, while the highest value reached throughout the stimulus varied between equally spaced steps on a range from 8 Hz to 12 Hz. Carrier frequency manipulation was proportional to starting values, with the base frequency values of components being the lowest possible point, and the highest point varying across stimuli in equally spaced values in the range from 1.5 to 2 times the base value. The base frequency range of components had a minimum value of 50 to 4000 and a maximum value of 40 to 4800 and was also varied by equally spaced steps along this range.

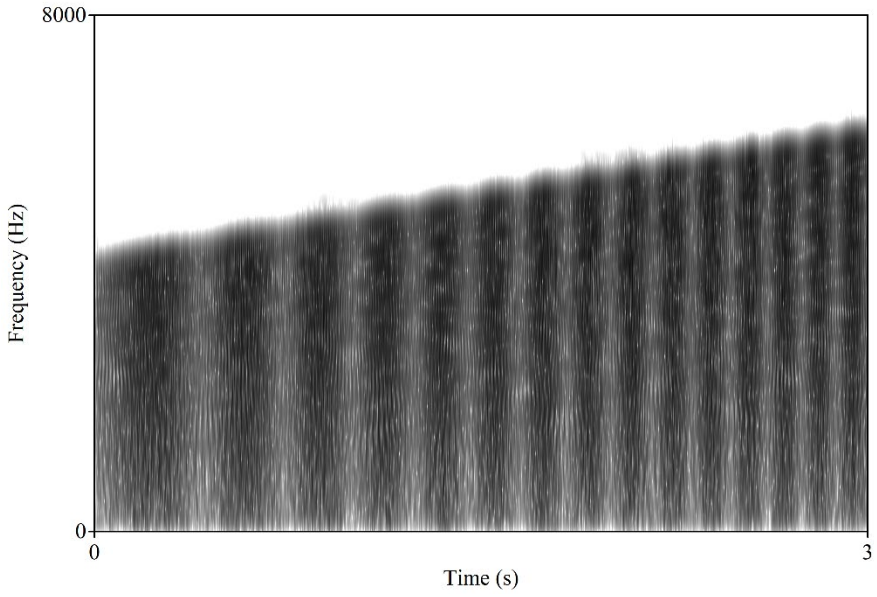

Figure 2: Spectrogram for an exemplar of the pCO category of stimuli.

In total, 192 stimuli were generated for Experiment 1 and 256 for Experiment 2. These could be classified into one of two categories, further divided into four total subcategories depending on how the relationships between AM rate and frequency dimensions vary over time:

- pCO_I: Increase in AM and increase in frequency over time.
- pCO_II: Decrease in AM and decrease in frequency over time.
- nCO_I: Decrease in AM, and increase in frequency over time
- nCO_II: Increase in AM and decrease in frequency over time.

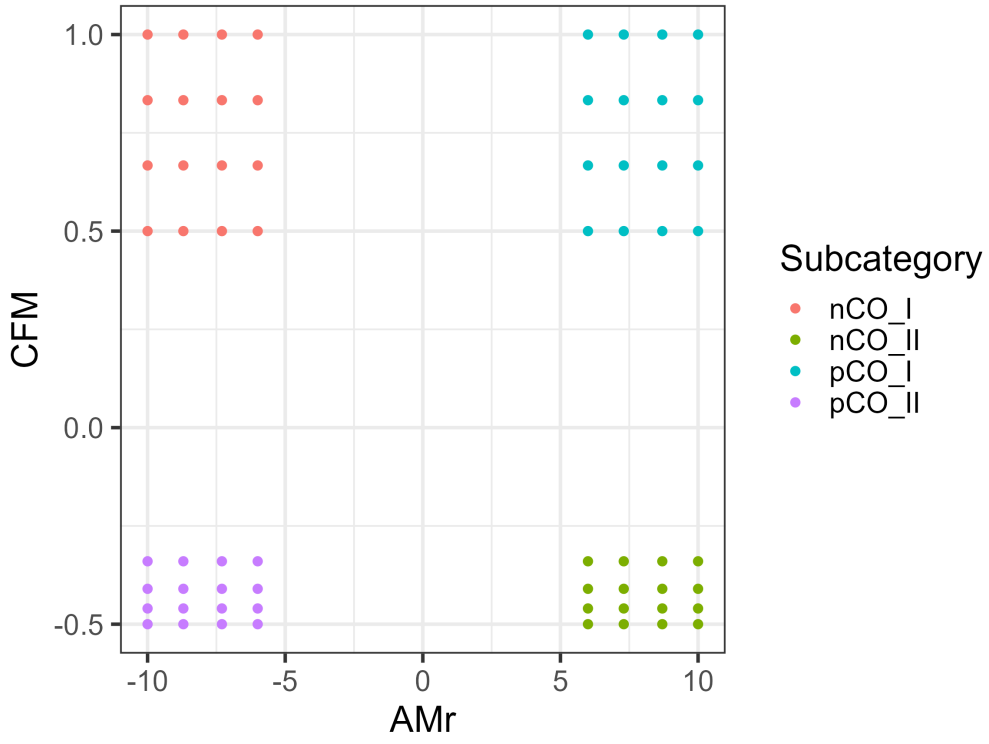

Figure 3: Distribution of generated stimuli across the two informative dimensions. Changes in timbre are not present here and would further result in more variations of stimuli. AMr = Amplitude modulation rate (Hz change over stimulus); CFM = Carrier frequency modulation (as proportion of initial value).

2.2 Procedure

Participants were instructed to wear headphones, and experiments were preceded by a screening procedure as designed by Chait et al. (2021) to ensure these instructions are followed, and the equipment is functioning adequately.

Participants took part in a categorization task structured in the form of a game where the objective is to correctly identify which "alien" is about to appear, based on the sounds they produce. There are four different aliens, distinguishable by shape and color (Fig 3), and each alien is pseudo-randomly assigned to a category based on sound statistics. Sound stimuli are designed in a way to signal alien categories based on covariances in the sound dimensions of AM rate and pure-tone frequency change over time.

The categorization task consisted of 192 trials in total, which were divided in eight blocks (24 trials per block). Each block was further subdivided into 6 smaller four-trial 'blocklets' consisting of one exemplar of each sub-category. Both block and 'blocklet' order were randomized across participants.

Each trial consisted of three stages: First, a sound clip was played. Next, a prompt appeared for participants to indicate an alien category by pressing one of four keys (<Q>, <W>, <E>, <R>). Finally, the corresponding alien appeared on the screen, allowing the participant to know if they responded correctly. Participants were instructed to respond as quickly as possible while still being accurate. Through this design, learning of the relationships between category and sound can be measured over the entire task.

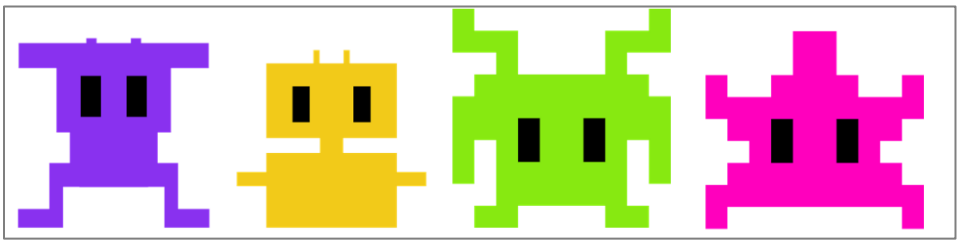

Figure 4: Four 'Aliens' used during the categorization tasks.

During four of the eight experimental blocks, one catch trial was randomly added to the block. During these catch trials, instead of the usual stimuli, participants would hear a pre-recorded computer-generated voice asking them to press the <SPACEBAR> instead of the usual response keys. Participants who failed to accurately respond in more than one of the catch trials had their data excluded from further analysis.

Following the experimental task, participants were forwarded to an online survey (Qualtrics, Provo, UT, https://www.qualtrics.com) where they reported on their qualitative experience with the task as well as providing demographical information, as well as reports on their experience with the task.

2.3 Data analysis

2.3.1 Pre-Processing

Participants who self-reported hearing or neurological issues had their data excluded from any statistical analysis, as well as participants who reported technical difficulties during the experimental task. Furthermore, data from participants who failed more than half of the headphone screening trials, or who reported not wearing headphones during the experiment were excluded from the statistical analysis (excluded N = 53 for Expt 1).

Participants who missed two or more of the catch trials were excluded from further data analysis (N = 2). For analysis of the reaction time data, we further excluded trials based on the following criteria in the following order: (1) Trials in which participant's response was incorrect (40%); (2) Reaction times that diverged more than three standard deviations from the mean of their category (0.52%) as recommended by Berger & Kiefer (2021). After this filtering process, 59.44% of our trials remained for subsequent analysis.

This resulted in a total of 137 participants for Experiment 1. Which was confirmed a sufficient sample size for testing our hypothesis based on a power analysis.

2.3.2 Analytical approach

Due to the data being structured in the form of repeated measures contained within the same individual over time, a multi-level modeling approach was appropriate. The use of multilevel models is advantageous when compared to traditional ANOVA frameworks that would employ a repeated-measures ANOVA as it allows for (i) higher sensitivity for the detection of effects due to the use of trial-level outcomes, and (ii) analysis of data that does not follow a normal distribution.

Furthermore, in order to obtain better resolution for the detection of effects, we elected to create models using an outcome variable at the level of trial. This took the form of the binary outcome of correct/incorrect alien categorization,

All analyses and manipulations were run using R version 4.1.2 (R Core Team, 2021), and all models were built using the "lme4" package version 1.30.1 (Bates et al., 2015). Models were fit using maximum likelihood estimation based on the Laplace approximation and the bound optimization by quadratic approximation (BOBYQA) algorithm (Powell, 2009). Model fit and model comparison was assessed by comparing differences in deviance statistics between models, as well as via likelihood ratio tests using the "anova" function of the "stats" package (R Core Team, 2021).

Inspection of the effect of predictor variables on the model was done via a combination of (i) chi-square likelihood ratio comparisons done via the "drop1" function of the "lme4" package (ii) Pseudo-R^2 estimations using the "MuMIn" package version "1.46.0" (Bartoń, 2009) based on Nakagawa et al.'s (2017) delta method (iii) profiled confidence intervals obtained by using the "confint.merMod" function of the "lme4" package. When not possible to compute profile confidence intervals, or when the Likelihood Ratio Test for the effect had a p-value too close to the alpha cut-off level, we computed bootstrapped confidence intervals instead, which can be more reliable at the cost of being much more computationally intensive. Post hoc analyses were done

through pairwise comparisons of estimated marginal means (Searle et al., 1980) by using the “emmeans” function from the “emmeans” package Version 1.8.1 (Russell V. Lenth, 2022).

2.3.3 Model building approach

For all models, we allowed the intercepts of fixed effects to vary by participant. Following the procedures of Barr et al. (2012) for determining the inclusion of random slopes in the final model, random slopes were added for each of the fixed effects in the model but were dropped in the final reported model in case they were shown to be not significant through a likelihood ratio test.

Throughout our analysis, we built both models containing solely main effects as well as models containing interaction terms. Unless otherwise stated, the reported main effects were appraised by inspection of the main-effects models while only interactions are reported from interaction models. In all interaction models, we still included constitutive terms for the interactions as recommended by Thomas Brambor et al. (2006).

2.3.3.1 Categorization Accuracy

In order to test the effects of learning and stimulus category on categorization performance, we built logistic regression models with the binary outcome variable of correct categorization on a given trial as the predicted variable of interest. Our models included the predictor variables of time, as defined by the experimental block in the task, and stimulus category, with intercepts being allowed to vary by participant. Only the random slopes for experimental block were shown to be significant, and therefore the random slopes of stimulus category were dropped in the final model.

The second model we built aimed at investigating the effects on performance of the direction of change in individual dimensions. To this end, we dissected the stimulus category variable into two new categorical variables that informed the direction of change for AMr and CFM (AMr up/down; CFM up/down). These were included in a model as fixed effects and allowed to vary both in slope and intercept with respect to participant.

In addition to the main effects models, we also built an interaction model with the goal of investigating the effect of the interaction between time and stimulus category. This was done in order to assess whether the differences in performance towards stimulus category would change over the course of the experiment.

2.3.3.2 Reaction time

The models aiming at investigating the effects of stimulus category on reaction time of response were built using a gamma distribution, which we initially attempted to fit using the identity link function, as recommended by Lo & Andrews (2015), in order to avoid log-transforming the reaction-time data and preserve the variability in responses. However, this resulted in non-convergence of the model, and we resorted to employing a log link function instead. Similarly to what was done for models of categorization accuracy, we included the fixed effects of time and stimulus category, allowed intercepts to vary across participants, and included random slopes for the time variable.

2.4 Results

2.4.1 Category Learning

We observed a significant effect of stimulus category on categorization accuracy (Figure 5). The summary of the output of the three multilevel models built for this step can be seen on Appendix A - Table 1.

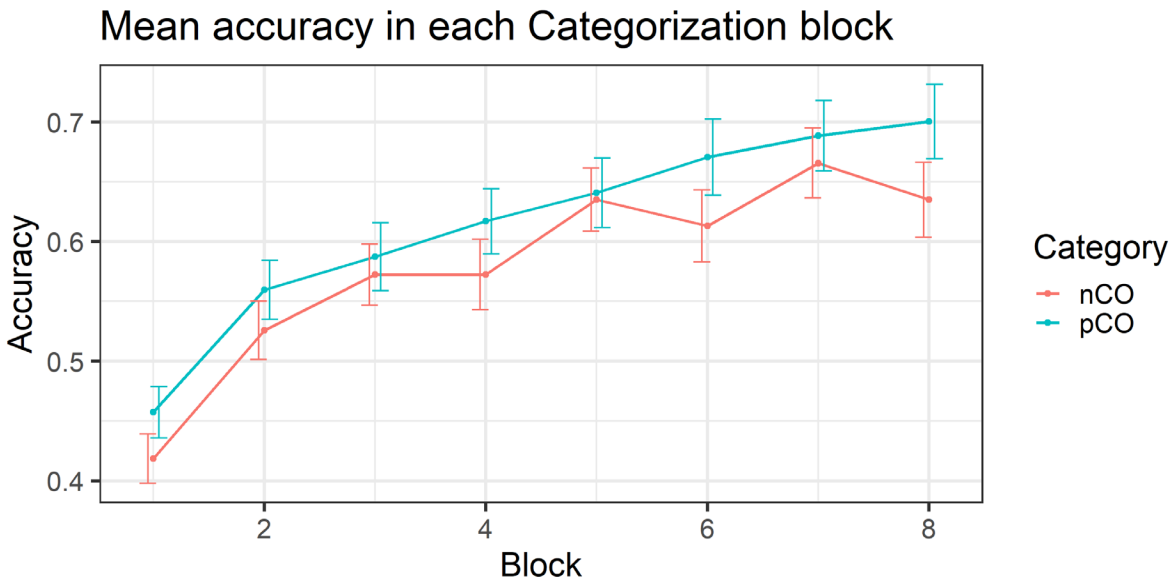

Figure 5: Mean accuracy in each categorization block during Expt 1 as a function of time (x axis) and stimulus category (line color). Error bars represent SE.

In the first model, we observed an effect of both stimulus category, as well as time on categorization accuracy. These effects were both confirmed significant through a likelihood ratio test comparing models where the predictor variables for time ($\chi 2$ (1) = 47.282, $p < .0001$) and category ($\chi 2$ (1) = 17,439, $p < .0001$) were individually dropped. There was a strong unconditional effect of time on the predicted probability of a participant accurately categorizing the sound on any given trial (b = 0.212, 95% CI [0.151, 0.270]). On average, the effect for stimulus category appears to be consistent, with positive covariances (pCO) being more accurately categorized than negative covariances (nCO), although variability between participants is considerable and so 95% confidence intervals are wide (b= 0.225, 95% CI [0.116, 0.344], SE = 0.052, z = -4.305, $p < .0001$) This means that, on average, trials in which participants categorized pCO stimuli had 25% higher odds of being accurately categorized. The variables included in the model accounted for 34% of the variance (R^2) across participants in categorization performance.

The interaction model showed that the interaction between time and category is on average greater for the positive covariance stimuli than negative covariance (Figure 6). However, the interaction effects are small (b = 0.040), 95% CI [0.006, 0.3071], although addition of the interaction term does meaningfully improve the model (explaining 38% of the variance compared to the 34% explained by the original model) according to a likelihood ratio test ($\chi 2$ (1) = 5.8179, p =.015). These results indicate that on average, for each successive block, the odds of participants more accurately categorizing pCO stimuli than nCO stimuli increased by 4%. When dissecting the category predictor variable in the second model, we observed no effect of CFM (b= 0.015, 95% CI [-0.124, 0.144]) nor AMr (b= 0.073, 95% CI [-0.026, 0.188]) modulation direction, and these predictor variables were not significant in a likelihood ratio test (CFM Direction: $\chi 2$ (1) = 0.059, p =.807; AMr Direction: $\chi 2$ (1) = 1.967, p =.161).

2.4.2 Reaction times

We observed a small effect of category on reaction time (RT, Figure 6). Results from the reaction time models are summarized in Appendix A - Table 2.

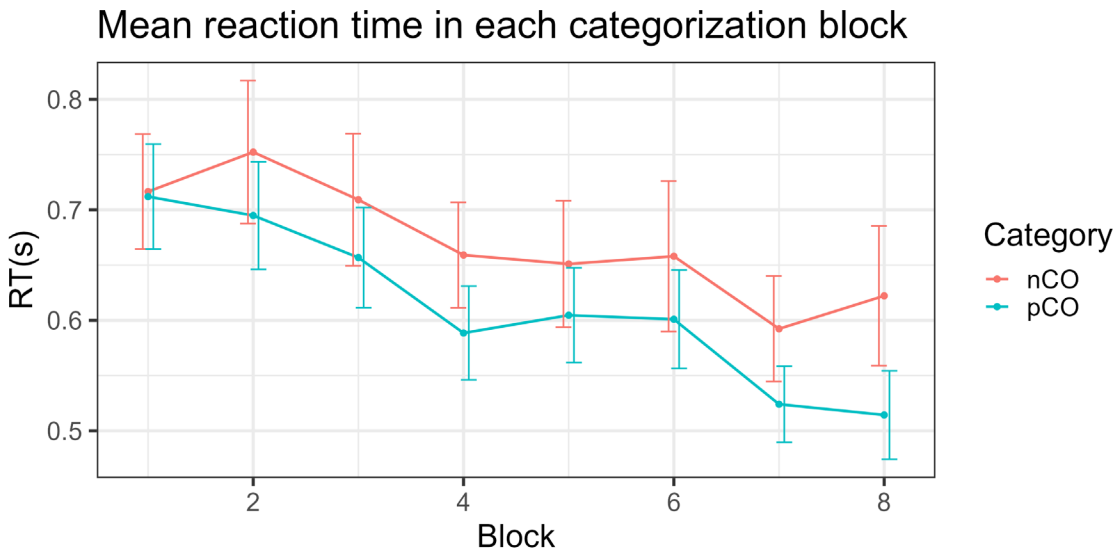

Figure 6:Mean RT in each categorization block during Expt 1 as a function of time (x axis) and stimulus category (line color). Error bars represent SE.

The unconditional main category model shows an average minor reduction in reaction times when categorizing pCO stimuli, although the CIs for this predictor coefficient border zero. (b = -0.070, 95% CI = [-0.132, -0.09]). Furthermore, the inclusion of the categorical predictor meaningfully improving the model ($\chi 2$ (1) = 8.8558, p =.003). Experimental block was also a significant predictor of RT reduction throughout the experiment (b = -0.042, 95% CI = [-0.064, -0.021]; $\chi 2$ (1) = 16.2663, p <.001).

On the interaction model, observed a marginally significant effect of the inclusion of the interaction between time and category on a likelihood ratio test ($\chi 2$ (1) = 1.8348, p =.176) where increases in block would be correlated with better performance towards pCO categorization speed. However, these results were not significant, and the CIs for the coefficient of the interaction included zero (b = -0.013, 95% CI = [-0.034, 0.007]).

On the dissected model, we observed a significant effect of AMr direction of reaction time change ($\chi 2$ (1) = 12.3678, p =.0004), where increases in AM over time were associated with slower reaction times (b = 0.099, 95% CI = [0.030, 0.159]). This effect was not seen for CFM direction, and the inclusion of this predictor did not meaningfully improve the model ($\chi 2$ (1) = 2.0119, p =.156).

Chapter 3

3 Experiment 2

The second experiment was aimed at investigating the effects of long-term experience on perceptual organization of sounds, as well as if these effects could be enhanced or overruled due to short-term training with sounds that either conform or violate these priors. The experiment was therefore divided into two main tasks.

The first task consisted of a familiarization task that was presented as a simplified version of the categorization task done during Experiment 1, with only two stimuli categories present instead of four.

The second task was a speech in noise task where participants reported words from the 5-word BUG sentences masked by novel complex AM sounds following the pattern of those used during experiment one and the first part of Experiment 2.

3.1 Methods

3.1.1 Participants

Participant recruitment followed the same procedures as for Experiment 1. In total, 189 participants were collected for Experiment 2.

3.1.2 Stimuli

3.1.2.1 Experimental stimuli

Stimuli generation for the amplitude modulated broadband noise stimuli used in this experiment followed the same procedures as for Experiment 1, with the exception that we generated more of them overall, for a total of 256 unique stimuli.

3.1.2.2 Speech stimuli

The speech stimuli used in Experiment 2 consisted of sentences constructed from the “Boston University Gerald” speech corpus (BUG; Kidd, Best, & Mason, 2008), which

adhere to the form “<Name> <past tense verb> <number> <adjective> <noun>” (e.g., “Bob bought five green bags”) and where all words are monosyllables. Words were spoken in citation form, without articulation between them, allowing for permuting across all pre-recorded exemplars to form full sentences. While the full corpus contains around 24 different speakers (12 stereotypically female voices, and 12 stereotypically male voices), only one female and one male voice were used during this experiment. Sentences were created by RMS normalizing and applying a low-pass filter at 4kHz with a 24dB roll off to the individual pre-recorded words for these speakers, and then semi-randomly permuting across the 8 words for each of the five components of the sentence, to have each speaker and word be equally likely to appear, with no sentence being repeated throughout the experiment. As part of the sentence construction process, inter-word interval was fixed at 100ms, and sentences were zero-padded at start and finish to have each audio clip be exactly 3s long.

This process was repeated for eight sentence sets divided into belonging to one of four main sets (A, B, C, and D). For each main set, half of the sentences were RMS normalized at an SNR of -3dB in relation to the ‘alien’ sounds (‘easy’ SNR) while the other half were normalized at a level of -7dB (‘hard’ SNR). This process was done two times, to counterbalance across the sentence-SNR pair by having each sentence exist as an easy and hard SNR version.

3.1.3 Procedure

3.1.3.1 Familiarization task

Stimuli for the familiarization task were generated in the same way as those used in Experiment 1, following the methods described in section 2.2.1. The familiarization task consisted of 128 trials divided into four blocks, which were further divided into two practice and two validation blocks which were assigned to one of two stimuli sets each. The two stimulus sets consisted of 32 exemplars of either solely pCO stimuli or nCO stimuli. Each exemplar was presented twice within its own stimulus set.

The task structure was nearly identical to the categorization task in Experiment 1, with the exception that each participant would either only categorize stimuli as belonging to one of two pCO sub-categories or between the two nCO sub-categories. Participants were randomly assigned to one of the two familiarization conditions, with stimulus sets counterbalanced between practice or validation blocks.

The separation between training and validation blocks was done to establish a criterion of learning so that participants who performed significantly worse than others within their familiarization group during the validation blocks would be excluded from further data analysis. This allows us to be more confident of the active engagement of participants with the task when accessing the effects of short-term familiarization on perceptual segregation during the speech matrix task.

3.1.3.2 Speech matrix task

Participants were instructed to disregard the masker sounds and concentrate on the words being spoken by the male and female voices. On any given trial, participants could click a button on the screen to play both the target speech and masker sounds simultaneously. After listening to the audio clip, participants were then prompted to click buttons on the screen to report the words they heard during the trial. Buttons were arranged in the form of a 5x8 matrix were each of the five columns represented one of the sentence components as following the form “<Name> <verb> <number> <adjective> <noun>” (Figure 7). The eight rows in the matrix consisted of the options for each of the words that could be spoken in each sentence. The buttons lit up in order, so as to have participants be questioned for each of the sentence components in order (First the possible names heard would light up, then the verbs, and so on...). Pilot studies were run to ensure that the masking effect would provide suitable challenge during the speech-in-noise task while being off floor and ceiling in performance.

Similarly to Experiment 1, following the experimental task, participants were forwarded to an online survey where they reported on their qualitative experience with the task and provided demographic information.

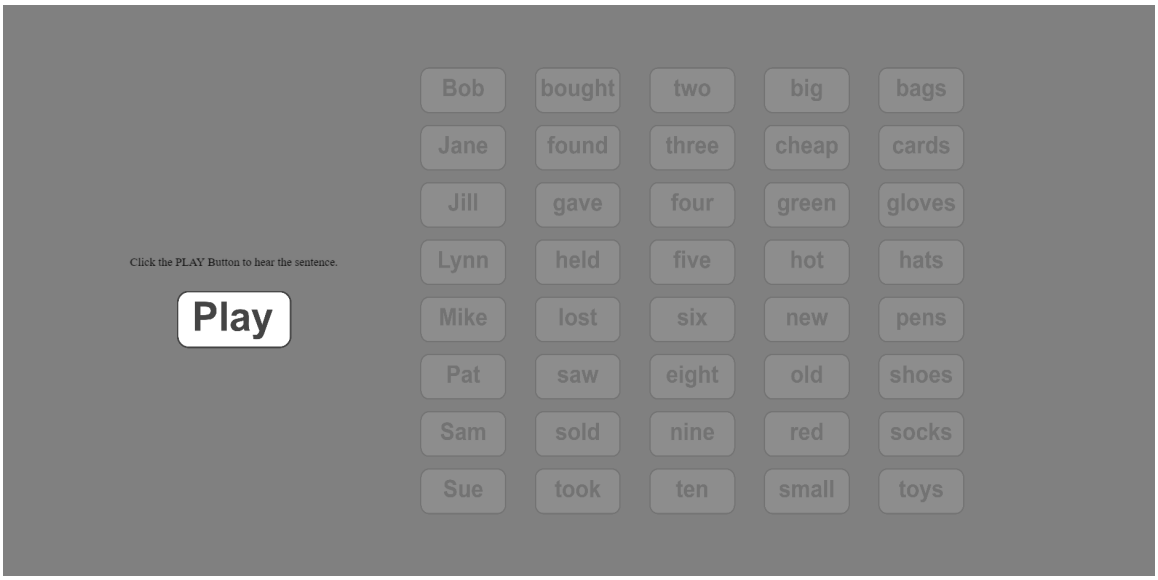

Figure 7: Screen capture showing the speech matrix design where participants input the words heard during the trial. Buttons lit up in sequence starting from the 'Play' button to hear the sentence and masker and then followed by each component of the sentence.

3.2 Data Analysis

3.2.1 Pre-processing

Participants who self-reported hearing or neurological issues had their data excluded from any statistical analysis, as well as participants who reported technical difficulties during the experimental task (Excluded N = 60). Furthermore, data from participants whose categorization performance during the validation block of the first task diverged by more than three standard deviations from the mean of their group were excluded from further analysis both for the first and second tasks (N = 4).

This resulted in a total of 125 participants for Experiment 2.

3.2.2 Analytical approach

Data analysis procedures followed a multilevel modeling approach as in Experiment 1.

For analyzing the speech matrix task data, we aggregated across trials with the same conditions under a block to compute an outcome variable of proportion of correctly identified words in a block. We chose to exclude the first and last words in the sentence as those would fall during moments most variable modulation, both in frequency and amplitude, and therefore any potential alterations to the masking effects of the stimulus from the psychoacoustic characteristics of it would be at their highest. By looking only at the three middle words we observe the changes in intelligibility while manipulation in both dimensions is around its 'middle point'.

3.2.3 Model building approach

3.2.3.1 Categorization accuracy

Models for comparing categorization performance across the two familiarization groups were similar to those used to evaluate categorization performance in Experiment 1, but the interpretation of the output differed somewhat since each participant group was trained on only one of the stimulus categories. As for in the first experiment, only the random slopes for Category were shown to be significant and therefore included in the final model.

Furthermore, as this task was divided into training and validations blocks which contained different stimulus sets, with the latter serving to assess learning, we opted to include only the two validation blocks in the analysis of these models. This would allow us to compare participant's performance once the rules for categorization have been learned.

3.2.3.2 Intelligibility scores

Intelligibility was measured by computing the proportion of correct word identifications within a block, which was fit as the dependent variable for multilevel binomial models. Predictor variables included in these models were the experimental block, masker

category, speaker, familiarization group the participant belonged to and the SNR of the trial in the form of a categorical variable distinguishing easy (-3dB) from hard (-7dB) trials.

Our next step was to build an interaction model where we included an interaction term between the stimulus category, SNR, and familiarization group. This three-way interaction was included to investigate if short-term learning with a specific stimulus category during the familiarization task would result in different intelligibility performances depending on masker category (i.e., higher intelligibility when masker matched familiarization stimuli), as well as to model if these improvements would only meaningfully be manifested during trials with more difficult SNR. Similarly, to our approach in Experiment 1. In a separate analysis, we modelled the effects of stimulus-change direction (up or down) on performance.

Finally, since our masker stimuli varied over time, with the extreme points of both manipulated dimensions occurring during the first and last words of the sentence, we conducted separate two-way ANOVAs to investigate the effect of both acoustic dimension manipulations on the first and last words of a sentence. This was done in order to assess if any effects observed on modulation direction were due to their change over time or the co-occurrence of the peaks in modulation with specific portions of the sentence.

3.3 Results

3.3.1 Familiarization task

Categorization performance in the familiarization task was consistent with what was observed in Experiment 1, showing increased performance for pCO stimuli, even with performance differences towards stimulus categorization being now investigated through a between-subjects design (Figure 8). The summary of results for the familiarization task models can be seen on Appendix A – Table 3.

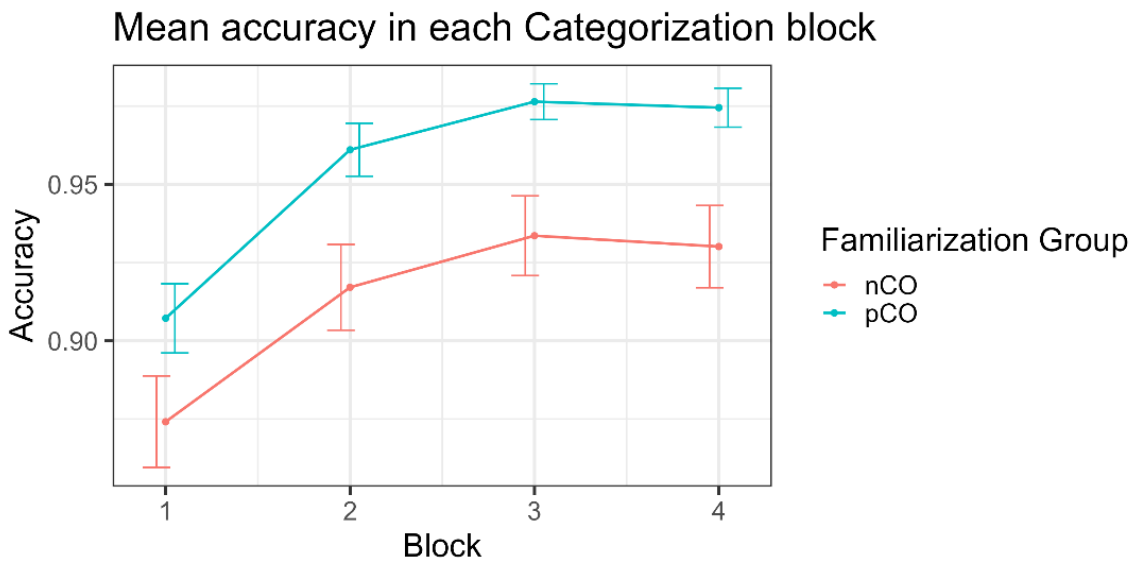

Figure 8: Mean accuracy in each categorization block during the familiarization task of Expt 2 as a function of time (x axis) and familiarization group (line color). Error bars represent SE. Blocks 1 and 2 of each participant are the training blocks, while block

In accordance with what was observed in Experiment 1, the first model indicates that participants who categorized pCO stimuli had an on average better unconditional performance on the validation blocks than those who categorized nCO sounds during familiarization (b = 1.343, 95% CI = [0.570, 2.247], SE = 0.39, z = -3.447, p = .0006). This effect was also confirmed through a likelihood ratio test ($\chi 2$ (1) = 12.3135, p =.0004). The time predictor variable did not show any significant effect with respect to differences between the two validation blocks and did not significantly improve the explanatory power of the model ($\chi 2$ (1) = 0.425, p =.51), this is likely due to this analysis only containing the two validation blocks after the first two of initial categorization learning. Consequently, the addition of the interaction term between time and category in the second model also did not further meaningfully improve the model ($\chi 2$ (1) = 0.008, p =.93).

3.3.2 Speech matrix task

Contrary to what was initially expected, we did not observe significant effects of stimulus category nor familiarization group on intelligibility performance (Figure 9). The summary output of models for this task can be seen on Appendix A- Table 4.

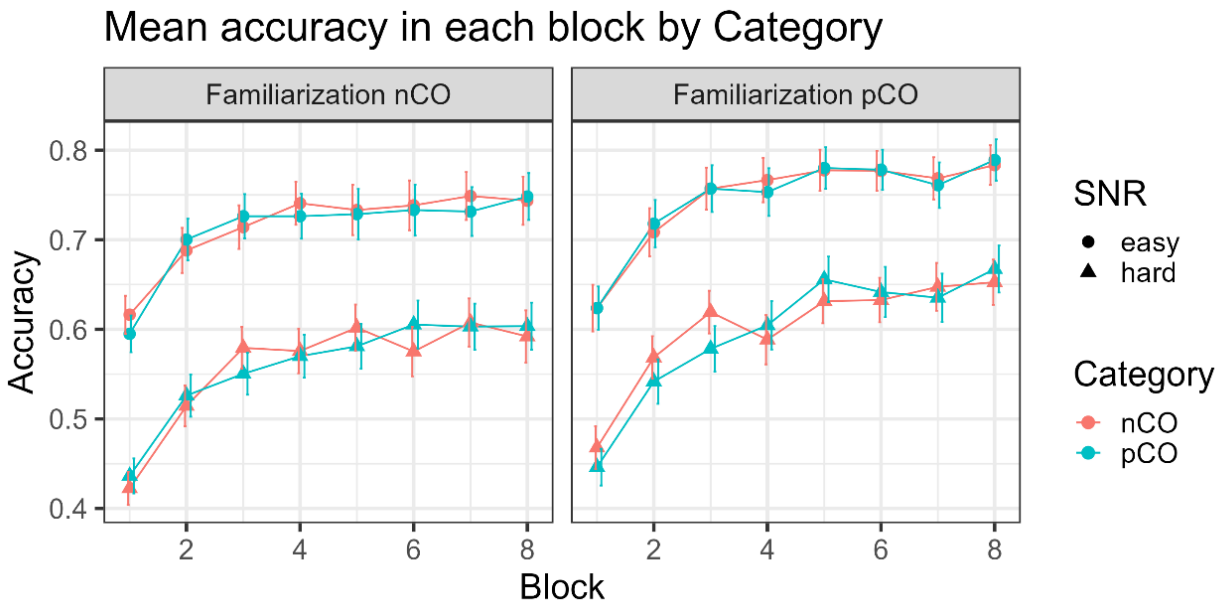

Figure 9: Mean accuracy (% of middle words accurately reported) in each block during the speech matrix task as a function of time (x-axis), stimulus category (line color), and trial SNR (symbols), faceted by each familiarization group. Error bars represent SE.

Through likelihood ratio tests, we confirmed the significance of the fixed effects of Block ($\chi2$ (1) = 104.021, p <.0001), SNR ($\chi2$ (1) = 315.498, p <.0001), and talker ($\chi2$ (1) = 70.4122, p <.0001). There was significant observed effect for the effect of time in the form of mean performance increasing alongside experimental blocks (b = 0.113, 95% CI = [0.095, 0.131]). Trials in which sentences were presented at an easy SNR (b = -0.834, 95% CI = [-0.880, -0.788], SE = 0.0234, z = 35.677, p < .0001) or spoken by a stereotypically male voice (b = 0.590, 95% CI = [0.536, 0.645], SE = 0.0277, z = -

21.324, p < .0001) resulted on higher odds of accurately identifying words. The masker category did not exhibit significant unconditional effects on intelligibility scores and was not a meaningful predictor of performance ($\chi 2$ (1) = 0.630, p = .4311). Finally, the familiarization group did not unconditionally show differences in performance and was not a significant predictor in the model ($\chi 2$ (1) = 1.625, p =.2024).

The interaction model investigating the effect of a three-way interaction between Category, Familiarization group, and SNR showed that this interaction was not significant. Including the interaction between the familiarization group and category (b = 0.024, 95% CI = [-0.061, 0.110]). Furthermore, the inclusion of the three-way interaction term (familiarization group, category, SNR) did not meaningfully improve the model ($\chi 2$ (1) = 0.027, p =.8697).

The third model shows the effect of the dissected stimulus dimensions on intelligibility scores. Contrary to what was observed for the categorization tasks, there was a strong effect of CFM direction on performance, with trials where CFM increased over time being associated with higher odds of accurate identification (b= 0.138, 95% CI = [0.110,0.166]) which was further confirmed by a likelihood ratio test ($\chi 2$ (1) = 69.7177, p <.0001). However, AMr direction exhibited the opposite effect, where increases over time were associated with lower performance, although with a smaller effect than that observed for CFM (b = -0.040, 95% CI = [-0.068, -0.011], $\chi 2$ (1) = 7.4988, p =.006).

Finally, we wanted to investigate if this effect of individual dimensions on intelligibility stems from the effect of the change in the dimension over time or due to an effect of the greater amount of modulation in a certain dimension occurring at the beginning or end of a sentence. To this end, we filtered data from the easy SNR trials, expecting any effects of the modulation to be more evident on the harder trials, and averaged participant's accuracy across all blocks in order to conduct two-way ANOVAs looking at the effects of AM direction and CFM direction on accuracy for each word position.

Looking at mean accuracy for the first word in the sentence, we observed a significant effect for AMr Direction (F (1,544) = 14.098, p = .002), increases over time (and consequently, modulation being at its lowest during the first word) were associated with

lower performance (mean difference = -0.055, 95% CI = [-0.083, -0.026]). Alternatively, CFM Direction was also shown to significantly affect mean accuracy ($F(1,544) = 20.966$, $p < .0001$), with increases over time associated with higher performance (mean difference = 0.067, 95% CI = [0.038, 0.095]). However, there was no significant effect of an interaction between these two dimensions ($F(1,544) = 0.772$, $p = .38$), further confirming that covariance relationships did not influence performance.

Results comparing performance on the last word of the sentence showed no significant effect of either AMr ($F(1,544) = 0.094$, $p = .759$) nor CFM ($F(1,544) = 2.144$, $p = .144$) direction of mean accuracy. However, although not statistically significant, the effect of both manipulations was trending in the opposite direction as that observed for the first word.

Chapter 4

4 Discussion

The current project investigated the effect of prior experience with short and long-term covariance patterns of stimulus statistics on perception. In Experiment 1, participants grouped stimuli into categories defined by the relationship between the two dimensions of AMr and CFM. Because of long-term prior knowledge, we predicted that people would categorize pCO sounds (both dimensions increasing or decreasing) more accurately than nCO sounds. In Experiment 2, participants were first familiarized with either pCO or nCO exemplars, and then reported words they heard from target speech masked by novel pCO and nCO exemplars. We predicted that people would be more successful at reporting words masked by sounds with a familiar covariance structure.

Overall, results from the categorization tasks were congruent with our hypothesis that sounds that present pitch-speed relationships consistent with naturalistic sound production and propagation would be more easily categorized. Participants were better at categorizing when presented stimuli presented positive covariance between AMr and CFM. This result was consistent across both Experiment 1 and the familiarization task in Experiment 2. Contrary to what we expected, whether maskers were learned or not did not change the effectiveness of the masker on the Speech Matrix task. This was true both for the effect on short-term familiarity with one of the covariance categories, as well as for an overall lack of effect for the maskers congruent with naturalistic covariance patterns.

4.1 Categorization performance was improved by positive covariance stimuli

Consistent with my predictions, participants performed more accurately on the categorization task in trials where the amplitude modulation rate and carrier frequency covaried positively. These improvements were apparent as a within-subjects effect in Experiment 1, as well as a between-subjects effect in Experiment 2. Reaction time

analysis in Experiment 1 showed that there was a minor effect of stimulus category on reaction time, although these effects were very small.

We also built models where both AMr and CFM were included as separate predictors to assess if any of these dimensions were more meaningfully informative than the other. In the models that dissected the effects of each informative dimension separately, we observed a significant individual effect of AMr on reaction time reduction, which may indicate that changes in this dimension may be more immediately apparent to participants. However, we did not observe an effect of either AMr or CFM on categorization accuracy. This further assures us that any improvements to categorization performance were generally consistent across both subcategories of pCO stimuli, and that the main informative criteria used by participants for categorization decisions were related to the interaction between the two dimensions.

Here we argue that these results are a consequence of a greater familiarity with stimuli that obey positive covariance patterns. This familiarity might be derived from life experience with naturalistic sound sources that obey consistent rules in sound production. Other authors have previously conjectured that changes in auditory dimensions may correlate in naturalistic sound sources. A well-documented example of this interaction is the relationship between the dimensions of pitch and loudness. Neuhoff (2004) argues, for instance, that many examples of sounds in our auditory environment such as speech, animal calls, and machinery noise covary in frequency and intensity, that is, as sounds increase in pitch, they also get louder. These relationships can be quite strong, and even result in illusory effects of perception where in experimental designs where only one of these dimensions is changed, the other is also perceived to increase or decrease accordingly (Neuhoff et al., 1999). Of greater relevance to the present study, we can also find examples of the interaction between the subjective perception of ‘speed’ and pitch of a sound, where listener’s tend to perceive a sounds modulation rate as slowing down when carrier frequency decreases (Hermann et al., 2020; Hermann & Johnsrude, 2018) Similar illusory changes have also been observed in the perception of musical tempo when the pitch of a musical composition is manipulated (Boltz, 2011).

Interestingly, music itself provides a valuable window through which to observe patterns and correlations between acoustic dimensions that may be recurrent in our environment. Music, much like language, may also act as a communication system (Aiello, 1994), and the structures of our communication may be correlated with our experiences in our environment. Indeed, both our perception and production of music and the arts may be highly influenced by our emotional and subjective judgments of their meaning. For instance, in both music and speech, lower energy states may be associated with lower pitch frequencies, loudness, and a slower tempo (Black, 1961), and tonal sequences that are higher and ascending in pitch are rated by listeners as happier, brighter, and faster than those that were lower and/or descending in pitch (Collier & Hubbard, 2001). These subjective attributions are also observed not only in how these compositions are perceived, but also produced, with musicians who were asked to perform a same musical piece with different emotional tones interpreting 'happy' performances with higher octaves and faster tempi (Gabrielsson, 1995). Interestingly, these associations seem to be quite consistent in the greater sum of Western music, with Broze and Huron (2013) conducting multiple approaches to an examination of Western musical compositions that suggested a distinct relationship between pitch and speed. The authors further highlight multiple possible causes for this observation, including associations derived from the physics of sound production, the human movements involved in vocal and instrumental performances, and our sensory and perceptual limitations and associations of the interaction between these dimensions (Broze & Huron, 2013).

The present work expands upon this discussion by further providing experimental evidence that this pitch-speed relationship of sounds is indeed somewhat special to our perceptual systems, and that these prior expectations may be exploited to aid in perceptual decisions in a categorization task. Indeed, stimuli that conformed to prior expectations that are congruent with naturalistic pitch-speed covariance patterns were often associated by participants during our debriefing questionnaire with real-world equivalents and examples (helicopters, machines powering up/down, and objects spinning....). More importantly, they were also more accurately categorized than those that were characterized by violations to said expectations. Our results are consistent with previous works in the literature that have shown that performance on a categorization

task can benefit from both short-term learning (Idemaru & Holt, 2015, Liu and Holt, 2014) and long-term prior experiences (Roark & Holt, 2020) with patterns of correlation between acoustic dimensions. Our design further distinguishes itself from others in the literature, however, by highlighting how these perceptual benefits may also arise in associations on how acoustic dimensions change over time, as the four simultaneous categories presented to participants could only be completely separated by further accounting by how they change throughout their presentation.

4.2 Masker category did not affect speech segregation

Contrary to our initial expectation that well learned maskers would be easier to segregate from the target speech, we did not observe a significant effect of either short or long-term familiarity with covariance categories on intelligibility performance during Experiment 2. In our main effects models, the stimulus category did not meaningfully improve the model performance, indicating that, unlike in Experiment 1, maskers that belonged to positive covariance categories were not better segregated from the targets, compared to negative-covariance maskers. Next, we turned to our full interaction model to investigate the effect of the masker category both at different levels of SNR, as well as in different familiarization groups. We expected to see a larger effect of the masker category on more challenging SNR trials, as well as an interaction between the familiarization group and category, which would indicate that short-term learning was able to influence performance. However, neither of these two interactions meaningfully improves predicted intelligibility performance.

On the other hand, our models to examine the effects of direction of movement in the two auditory dimensions indicated an individual effect for both AMr and CFM direction. We questioned if this would indicate a differential masking effect resulting from the change over time or merely the overlap of modulation levels and each word. To assess the origin of this effect, we also compared the effect of the direction of change in the two dimensions specifically for intelligibility of the first and last words in the sentence. Results for this comparison showed a significant effect of AMr and CFM direction only for the first word. Namely, an increase of AMr over time (and therefore with AM rate at its lowest during the first word) was associated with lower performance, while for CFM

we observed the opposite effect, with increases over time in this dimension being associated with higher performance. Although modulation direction for both dimensions was not significant for the last word in the sentence, possibly due to listener's having had more time to segregate the talker from the masker, or simply due to differences related to the spectral and temporal composition of the words themselves, the trend in performance was the inverse of that for the first word. That is, the periods of highest AMr trended towards better performance, while those of highest CFM trended for lower performance. This may indicate that potentially reduced masking effect arising from specific acoustic dimension properties may overrule potential benefits stemming from familiarization with the masker itself in our design.

The capacity for listeners to decompose complex auditory scenes in order to isolate and track a specific informative sound source is a widely researched topic. Commonly associated with the "cocktail party problem" (Cherry, 1953), investigating how listeners can selectively attend to one voice amidst other competing sounds is an important avenue for deciphering the acoustic cues used by our auditory system to organize information. Typically, research on segregating auditory streams is associated with identifying the physical acoustic characteristics that serve as cues for listeners to segregate the target source (Bregman, 1990; Darwin & Carlyon, 1995). Previous works in the literature have also highlighted the influence of experience and learning on these tasks, such as knowing the content of the target speech (Bregman, 1990), and familiarity with the talker (Holmes et al., 2020; Domingo et al., 2020; Johnsrude et al., 2013, Barker & Newman, 2004; Magnuson, Yamada & Nusbaum, 1995). However, the effects of experience with the masker, instead of the target, are still seldom explored. Our research group has previously demonstrated how familiarity with a speaker can provide knowledge that can be exploited to better ignore the highly familiar voice in order to attend to target speech from a stranger (Johnsrude et al., 2013). Here in this project, we expected similar intelligibility benefits to also arise from experience with the complex amplitude-modulated noise makers that conformed to prior expectations from life experience, however, no such effects were observed.

As we have previously conjectured, it is possible that physical acoustic properties inherent to amplitude-modulated makers may provide avenues for listeners to piece out information from target speech to a degree that obfuscates any possible benefits arising from experience in helping segregate the target. Previous works have shown how modulation may be an important cue for stream segregation (Dolležal et al., 2012), and Gustafsson & Arlinger (1993) have shown how listeners may make use of the dips or 'valleys' in modulation to capture information on the target speech. Although we were aware of these limitations when designing this experiment, we had expected that by reducing the modulation depth while also decreasing the SNR between target and masker, we would still be able to observe an effect of prior experience on performance, perhaps because participants would better predict when the modulation 'dips' would occur to aid in intelligibility. Nonetheless, it is possible that we inadvertently created a design that did not incentivize segregation of the masker.

4.3 Limitations and future directions

The stimuli used during this project were highly controlled computer-generated noise that although are ideal for investigating effects of experience on fundamental acoustic properties of sounds, are inevitably somewhat distant from the day-to-day acoustic landscape we are consistently immersed in. Future studies aiming to highlight covariance and/or statistical patterns on more naturalistic stimuli such as speech or animal sounds, may provide even greater evidence for the perceptual changes towards stimuli that conform to long-term expectations.

In addition, the lack of observable effects for both short and long-term experience on intelligibility performance in the speech matrix task also encourage both considerations for limitations with our experimental design as well as future approaches that seek to confirm our observations. This project aimed to simultaneously investigate both the prior expectations related to naturalistic pitch-speed relationships, as well as the perceptual benefits arising from these prior representations for both stimuli categorization as well as segregation, leaving us somewhat constrained on the manipulations that could be done to our stimuli. It is possible that the spectral overlap between target and masker resulting

from our usage of broad-band noise as maskers that near completely overlapped with the spectral composition of our target made it impossible for listeners to segregate masker from target during moments outside of the dips in modulation. The greater intelligibility during moments of higher amplitude modulation would suggest that the faster occurring dips might have allowed for more opportunity to segregate the target during moments of lower intensity of the masker. Similarly, better performance during moments of lower frequency modulation were associated with narrower spectral bands of masking which, although still had great overlap with the target speech, may have let more information leak through some of the highest frequency bands of the speech. Although we piloted many parameters for the maskers until settling on the ones we ultimately used to make sure performance was still off floor and ceiling during the task, it is possible that we have yet to find a 'sweet spot' that creates both a good masking effect while still allowing for sufficient acoustic cues that allow listeners to segregate the target speech. There are still many open questions about the way in which our brains can exploit experience-based cues with a masker to better segregate it from target sounds. Future research manipulating different acoustic dimensions of maskers that have potentially less interference with stimuli masking properties, for instance, may provide answers to these questions.

4.4 Conclusions

The present work found consistent benefits of positive pitch-speed covariance relationships on performance in a categorization task. This is consistent with previous works that have shown how prior expectations may influence performance in a categorization task (Roark & Holt 2020), as well as evidence for a strong relationship between pitch and speed for the human auditory system (Hermann, Augereau & Johnsrude, 2020; Hermann & Johnsrude, 2018). At the same time, it extends prior work to show that our auditory system is able to exploit these long-learned relationships on specific patterns of change over time within acoustic dimensions in order to aid perception. Still, contrary to what was initially expected, similar benefits were not observed regarding perceptual organization of sounds and facilitation on segregating familiar maskers from target speech.

The ways in which our brains can navigate and process information from such a rich and complex world around us is by no means a simple feat. Finding the threads that connect the patterns from which our intricately stochastic sensory landscape is formed is essential for being able to respond to our environment in an appropriate and timely manner. Thus, our ability to shape our perceptual systems in response to the statistics of the environment is a crucial component for how we interact with the world around us. Here we highlight that covariance structures between acoustic dimensions are one such source of valuable information that helps us adequately perceive and categorize auditory information.

References

Aiello, R., & Sloboda, J. A. (Eds.). (1994). *Musical perceptions*. *290*. https://psycnet.apa.org/fulltext/1994-97390-000.pdf

Allen, E. J., Moerel, M., Lage-Castellanos, A., De Martino, F., Formisano, E., & Oxenham, A. J. (2018). Encoding of natural timbre dimensions in human auditory cortex. *NeuroImage*, *166*, 60–70. https://doi.org/10.1016/j.neuroimage.2017.10.050

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2), B45–B53. https://doi.org/10.1016/j.cognition.2004.06.001

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Barton, K. (2009). MuMIn: multi-model inference. *Http://r-Forge. R-Project. Org/projects/mumin/*. https://ci.nii.ac.jp/naid/20001420752/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bertels, J., Boursain, E., Destrebecqz, A., & Gaillard, V. (2014). Visual statistical learning in children and young adults: how implicit? *Frontiers in Psychology*, *5*, 1541. https://doi.org/10.3389/fpsyg.2014.01541

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12420

Bianco, R., Harrison, P. M. C., Hu, M., Bolger, C., Picken, S., Pearce, M. T., & Chait, M. (2020). Long-term implicit memory for sequential auditory patterns in humans. *eLife*, *9*, e56073. https://doi.org/10.7554/eLife.56073

Black, J. W. (1961). Relationships among fundamental frequency, vocal sound pressure, and rate of speaking. *Language and Speech*, *4*(4), 196–199. https://psycnet.apa.org/fulltext/1963-06643-001.pdf

Boltz, M. G. (2011). Illusory Tempo Changes Due to Musical Characteristics. *Music Perception*, *28*(4), 367–386. https://doi.org/10.1525/mp.2011.28.4.367

Bond, R. N., & Feldstein, S. (11/1982). Acoustical correlates of the perception of speech rate: An experimental investigation. *Journal of Psycholinguistic Research*, *11*(6), 539–557. https://doi.org/10.1007/BF01067611

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association*, *14*(1), 63–82. http://www.jstor.org/stable/25791835

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. *773*. https://psycnet.apa.org/fulltext/1990-98046-000.pdf

Bregman, A. S., & McAdams, S. (02/1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. *The Journal of the Acoustical Society of America*, *95*(2), 1177–1178. https://doi.org/10.1121/1.408434

Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. J. (2020). Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Quarterly Journal of Experimental Psychology* , *73*(9), 1431–1443. https://doi.org/10.1177/1747021820916726

Broze, Y., & Huron, D. (2013). Is Higher Music Faster? Pitch–Speed Relationships in Western Compositions. *Music Perception*, *31*(1), 19–31. https://doi.org/10.1525/mp.2013.31.1.19

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, *8*(10), 465–471. https://doi.org/10.1016/j.tics.2004.08.008

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*, 975–979. https://doi.org/10.1121/1.1907229

Collier, W. G., & Hubbard, T. L. (23/2001). Musical Scales and Evaluations of Happiness and Awkwardness: Effects of Pitch, Direction, and Scale Mode. *The American Journal of Psychology*, *114*(3), 355. https://doi.org/10.2307/1423686

Conway, C. M., & Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, *62*(5), 1112–1120. https://doi.org/10.3758/bf03212092

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. *Hearing.*, *468*, 387–424. https://doi.org/10.1016/B978-012505626-7/50013-3

Denham, S., & Winkler, I. (2014). Auditory Perceptual Organization. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of Computational Neuroscience* (pp. 1–15). Springer New York. http://link.springer.com/10.1007/978-1-4614-7320-6_100-1

Dolležal, L.-V., Beutelmann, R., & Klump, G. M. (2012). Stream segregation in the perception of sinusoidally amplitude-modulated tones. *PloS One*, *7*(9), e43615. https://doi.org/10.1371/journal.pone.0043615

Domingo, Y., Holmes, E., & Johnsrude, I. S. (2020). The benefit to speech intelligibility of hearing a familiar voice. *Journal of Experimental Psychology. Applied*, *26*(2), 236–247. https://doi.org/10.1037/xap0000247

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. Perception & Psychophysics, 22(2), 109-122. doi:10.3758/BF03198744

Gabrielsson, A., & Lindström, E. (1995). Emotional expression in synthesizer and sentograph performance. *Psychomusicology: A Journal of Research in Music Cognition*, *14*(1-2), 94–116. https://doi.org/10.1037/h0094089

Garner, W. R. (1/1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*(1), 98–123. https://doi.org/10.1016/0010-0285(76)90006-2

Graves, K. N., Antony, J. W., & Turk-Browne, N. B. (09/2020). Finding the Pattern: On-Line Extraction of Spatial Structure During Virtual Navigation. *Psychological Science*, *31*(9), 1183–1190. https://doi.org/10.1177/0956797620948828

Graves, K. N., Sherman, B. E., Huberdeau, D., Damisah, E., Quraishi, I. H., & Turk-Browne, N. B. (2022). Remembering the pattern: A longitudinal case study on statistical learning in spatial navigation and memory consolidation. *Neuropsychologia*, *174*, 108341. https://doi.org/10.1016/j.neuropsychologia.2022.108341

Gregory, R. L. (1963). DISTORTION OF VISUAL SPACE AS INAPPROPRIATE CONSTANCY SCALING. *Nature*, *199*, 678–680. https://doi.org/10.1038/199678a0

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews. Neuroscience*, *5*(11), 887–892. https://doi.org/10.1038/nrn1538

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2000). Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *The Journal of the Acoustical Society of America*, *108*(1), 263–271. https://doi.org/10.1121/1.429462

Gustafsson, H. A., & Arlinger, S. D. (1994). Masking of speech by amplitude-modulated noise. *The Journal of the Acoustical Society of America*, *95*(1), 518–529. https://doi.org/10.1121/1.408346

Herrmann, B., Augereau, T., & Johnsrude, I. S. (2020). Neural Responses and Perceptual Sensitivity to Sound Depend on Sound-Level Statistics. *Scientific Reports*, *10*(1), 9571. https://doi.org/10.1038/s41598-020-66715-1

Herrmann, B., & Johnsrude, I. S. (2018). Neural Signatures of the Processing of Temporal Patterns in Sound. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(24), 5466–5477. https://doi.org/10.1523/JNEUROSCI.0346-18.2018

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013–3022. https://doi.org/10.1121/1.1323463

Holmes, E., To, G., & Johnsrude, I. (2020). *How do voices become familiar? Speech intelligibility and voice recognition are differentially sensitive to voice training*. PsyArXiv. https://osf.io/bm2uq

Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(6), 2014–2026. https://doi.org/10.1523/JNEUROSCI.1779-15.2016

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(6), 1939–1956. https://doi.org/10.1037/a0025641

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (10/2013). Swinging at a Cocktail Party: Voice Familiarity Aids Speech Perception in the Presence of a Competing Voice. *Psychological Science*, *24*(10), 1995–2004. https://doi.org/10.1177/0956797613482467

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, *55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Kreitewolf, J., Mathias, S. R., & von Kriegstein, K. (2017). Implicit Talker Training Improves Comprehension of Auditory Speech in Noise. *Frontiers in Psychology*, *8*, 1584. https://doi.org/10.3389/fpsyg.2017.01584

L. P. A. S. Noorden, V. (n.d.). *Temporal coherence in the perception of tone sequences*. Retrieved October 30, 2022, from https://pure.tue.nl/ws/files/3389175/152538.pdf

Lehet, M., & Holt, L. L. (04/2017). Dimension-Based Statistical Learning Affects Both Speech Perception and Production. *Cognitive Science*, *41*, 885–912. https://doi.org/10.1111/cogs.12413

Lehet, M., & Holt, L. L. (09/2020). Nevertheless, it persists: Dimension-based statistical learning and normalization of speech impact different levels of perceptual processing. *Cognition*, *202*, 104328. https://doi.org/10.1016/j.cognition.2020.104328

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

Magnuson, J. S., & Yamada, R. A. (1995). The effects of familiarity with a voice on speech perception. *Of the 1995 …*.

Mamassian, P., Landy, M., & Maloney, L. (2002). Bayesian modeling of visual perception. In *Probabilistic models of the brain: Perception and neural function* (pp. 13–36). MIT Press. https://nyuscholars.nyu.edu/en/publications/bayesian-modeling-of-visual-perception

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134. https://doi.org/10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. https://doi.org/10.1016/s0010-0277(01)00157-3

McWalter, R., & McDermott, J. H. (05/2018). Adaptive and Selective Time Averaging of Auditory Scenes. *Current Biology: CB*, *28*(9), 1405–1418.e10. https://doi.org/10.1016/j.cub.2018.03.049

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society, Interface / the Royal Society*, *14*(134). https://doi.org/10.1098/rsif.2017.0213

Neuhoff, J. G. (2004). Ecological psychoacoustics: Introduction and history. In *Ecological psychoacoustics* (pp. 1–13). Brill. https://brill.com/downloadpdf/book/9780080477442/B9780080477442_s004.pdf

Neuhoff, J. G., McBeath, M. K., & Wanzie, W. C. (1999). Dynamic frequency change influences loudness perception: A central, analytic process. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(4), 1050–1059. https://doi.org/10.1037/0096-1523.25.4.1050

Pearce, M. T. (07/2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation: Enculturation: statistical learning and prediction. *Annals of the New York Academy of Sciences*, *1423*(1), 378–395. https://doi.org/10.1111/nyas.13654

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Plack, C. J. (2018). *The sense of hearing: Second edition* (3rd ed.). Routledge.

Powell, M. J. D. (2009). *The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives*. http://dx.doi.org/

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roark, C. L., & Holt, L. L. (2019). Perceptual dimensions influence auditory category learning. *Attention, Perception & Psychophysics*, *81*(4), 912–926. https://doi.org/10.3758/s13414-019-01688-6

Roark, C. L., & Holt, L. L. (2022). Long-term priors constrain category learning in the context of short-term statistical regularities. *Psychonomic Bulletin & Review*, *29*(5), 1925–1937. https://doi.org/10.3758/s13423-022-02114-z

Russell V. Lenth (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.1-1. https://CRAN.R-project.org/package=emmeans

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Samples, J. M., & Franklin, B. (1978). Behavioral responses in 7 to 9 month old infants to speech and non-speech stimuli. *The Journal of Auditory Research*, *18*(2), 115–123. https://www.ncbi.nlm.nih.gov/pubmed/753823

Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. *The American Statistician*, *34*(4), 216–221. https://doi.org/10.2307/2684063

Seriès, P., & Seitz, A. R. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, *7*, 668. https://doi.org/10.3389/fnhum.2013.00668

Spence, M. J., & DeCasper, A. J. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior & Development*, *10*(2), 133–142. https://doi.org/10.1016/0163-6383(87)90028-2

Stilp, C. E., & Assgari, A. A. (8/2019). Natural speech statistics shift phoneme categorization. *Attention, Perception & Psychophysics*, *81*(6), 2037–2052. https://doi.org/10.3758/s13414-018-01659-3

Stilp, C. E., Rogers, T. T., & Kluender, K. R. (2010). Rapid efficient coding of correlated complex acoustic properties. *Proceedings of the National Academy of Sciences*, *107*(50), 21914–21919. https://doi.org/10.1073/pnas.1009020107

Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience*, *1*(3), 183–184. https://doi.org/10.1038/630

Turk-Browne, N. B. (2012). Statistical learning and its consequences. *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation*, *59*, 117–146. https://doi.org/10.1007/978-1-4614-4794-8_6

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945. https://doi.org/10.1162/jocn.2009.21131

von Helmholtz, H. (1924). *Handbuch der Physiologischen Optik. Leipzig: Voss. (English tranl. 1924 JPC Southall as Treatise on Physiological Optics)* (J. P. C. Southall (Ed.); Vol. 1, p. 3). https://doi.org/10.1037/13536-000

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276. https://doi.org/10.1111/j.1467-7687.2004.00345.x

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). FO gives voicing information even with unambiguous voice onset times. The Journal of the Acoustical Society of America, 93 (Pt. 1), 2152–2159. http://dx.doi.org/10.1121/1.406678

Woods, K. J. P., & McDermott, J. H. (2018). Schema learning for the cocktail party problem. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), E3313–E3322. https://doi.org/10.1073/pnas.1801614115

Appendix

Appendix A: Multilevel Model Summary Tables

In this table and in all following under this appendix, numbers represent beta values, with the values between parentheses corresponding to 95% CI.

Table 1

Expt 1 Categorization Models

	Dependent variable:		
	Categorization Outcome		
	Unconditional	Interaction	Dissected
Time(Block)	0.212	0.193	0.213
	(0.151, 0.270)	(0.140, 0.252)	(0.164, 0.272)
pCO Category	0.225	0.066	
	(0.116, 0.344)	(-0.110, 0.238)	
Time x pCO Category		0.040	
		(0.006, 0.071)	
AMr Increase			0.073
			(-0.026, 0.188)
FMr Increase			0.015
			(-0.124, 0.144)
Constant	-0.426	-0.349	-0.360
	(-0.600, -0.291)	(-0.499, -0.176)	(-0.518, -0.197)
Conditional R2	0.336	0.338	0.34
Observations	16,128	16,128	16,128
Log Likelihood	-9,086.817	-9,083.908	-9,075.694
Akaike Inf. Crit.	18,191.630	18,187.810	18,179.390
Bayesian Inf. Crit.	18,260.830	18,264.700	18,287.030

Table 2

Expt 1 Reaction Time Models

	Dependent variable:		
	Reaction Time (in log seconds)		
	Unconditional	Interaction	Dissected
Time(Block)	-0.043***	-0.039***	-0.043***
	(-0.064, -0.021)	(-0.063, -0.015)	(-0.066, -0.022)
pCO Category	-0.028*	0.008	
	(-0.081, 0.017)	(-0.110, 0.120)	
Time x pCO Category		-0.007	
		(-0.028, 0.016)	
AMr Increase			0.096***
			(0.045, 0.144)
FMr Increase			0.010
			(-0.040, 0.059)
Constant	-0.508***	-0.526***	-0.576***
	(-0.664, -0.384)	(-0.681, -0.391)	(-0.721, -0.450)
Conditional R2	0.148	0.148	0.151
Observations	9,185	9,185	9,185
Log Likelihood	196.319	197.237	224.390
Akaike Inf. Crit.	-378.639	-378.474	-432.780
Bayesian Inf. Crit.	-328.762	-321.471	-375.777

Table 3

Expt 2 Categorization Models

	Dependent variable:	
	Categorization Outcome	
	Unconditional	Interaction
Time(Block)	-0.071	-0.065
	(-0.305, 0.145)	(-0.337, 0.189)
pCO Category	1.343	1.418
	(0.570, 2.247)	(-0.444, 3.466)
Time x pCO Category		-0.021
		(-0.539, 0.518)
Constant	3.991	3.971
	(3.106, 5.074)	(3.012, 5.227)
Conditional R2	0.145	0.145
Observations	8,768	8,768
Log Likelihood	-1,349.863	-1,349.859
Akaike Inf. Crit.	2,707.726	2,709.718
Bayesian Inf. Crit.	2,736.042	2,745.113

Table 4

Expt 2 Inteligibility Models

	Dependent variable:		
	Inteligibility Score		
	Unconditional	Interaction	Dissected
Time(Block)	0.113	0.113	0.113
	(0.095, 0.131)	(0.095, 0.131)	(0.095, 0.131)
pCO Category	-0.011	-0.026	
	(-0.040, 0.017)	(-0.084, 0.032)	
AMr Increase			-0.040
			(-0.068, -0.011)
CFM Increase			0.138
			(0.110, 0.166)
Familiarization pCO	0.130	0.130	0.130
	(-0.070, 0.330)	(-0.130, 0.390)	(-0.070, 0.330)
Hard SNR	-0.834	-0.846	-0.835
	(-0.880, -0.788)	(-0.918, -0.773)	(-0.881, -0.789)
Talker Male	0.590	0.590	0.591
	(0.536, 0.645)	(0.536, 0.645)	(0.536, 0.646)
pCO Category:Familiarization pCO		0.024	
		(-0.061, 0.110)	
pCO Category:Hard SNR		0.030	
		(-0.048, 0.108)	
pCO Category:Familiarization pCO:SNR Hard		0.019	
		(-0.084, 0.122)	
CategorypCO:FamiliarizationpCO:SNRhard		-0.052	
		(-0.166, 0.062)	
Constant	0.341	0.342	0.287
	(0.176, 0.506)	(0.159, 0.526)	(0.121, 0.453)
Conditional R2	0.22	0.22	0.22
Observations	8,768	8,768	17,536
Log Likelihood	-16,676.010	-16,675.560	-26,153.780
Akaike Inf. Crit.	33,384.010	33,391.130	52,341.560
Bayesian Inf. Crit.	33,497.280	33,532.700	52,473.680

Appendix B: Ethics approval

Western Research

Date: 4 March 2019

To: Dr. Ingrid Johnsrude

Project ID: 112574

Study Title: Studies in auditory category learning, memory, and generalization

Application Type: NMREB Initial Application

Review Type: Delegated

Full Board Reporting Date: April 5 2019

Date Approval Issued: 04/Mar/2019

REB Approval Expiry Date: 04/Mar/2020

Dear Dr. Ingrid Johnsrude

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
2_AudCalib	Other Data Collection Instruments	07/Dec/2018	1
3_PrimaryTask	Other Data Collection Instruments	07/Dec/2018	1
4_AudCogAbilities	Other Data Collection Instruments	07/Dec/2018	1
5_Questionnaires	Online Survey	17/Dec/2018	1
5_Questionnaires	Paper Survey	07/Dec/2018	1
AP_EmailDatabase	Recruitment Materials	26/Feb/2019	2
AP_EmailFollowup	Recruitment Materials	26/Feb/2019	2
AP_WebsiteRecruitment	Recruitment Materials	26/Feb/2019	2
APRecruitment112574	Recruitment Materials	26/Feb/2019	2
Consent_Online	Implied Consent/Assent	26/Feb/2019	2
EmailRecruitment	Recruitment Materials	26/Feb/2019	2
General_EmailFollowup	Recruitment Materials	26/Feb/2019	2
General_WebsiteRecruitment	Recruitment Materials	26/Feb/2019	2
GeneralRecruitment112574	Recruitment Materials	26/Feb/2019	2
LOI-C_InPerson	Written Consent/Assent	26/Feb/2019	2
LOI-C_Online	Implied Consent/Assent	26/Feb/2019	2
MTurk_Recruitment	Recruitment Materials	26/Feb/2019	2
Musician_EmailDatabase	Recruitment Materials	26/Feb/2019	2
Musician_EmailFollowup	Recruitment Materials	26/Feb/2019	2
Musician_WebsiteRecruitment	Recruitment Materials	26/Feb/2019	2
MusicianRecruitment112574	Recruitment Materials	26/Feb/2019	2
SONA_Recruitment	Recruitment Materials	26/Feb/2019	2

Curriculum Vitae

Bruno Mesquita

Education:

- **Master's degree in Neuroscience**

University of Western Ontario, London, Canada

January 2021 - present

- **Bachelor's degree in Biological Sciences**

University of São Paulo, Ribeirão Preto, Brazil

February 2015 - July 2020

- **Licentiate degree in Biological Sciences**

University of São Paulo, Ribeirão Preto, Brazil

February 2015 - July 2020

Research Interests:

Neuroscience, Cognitive Neuroscience, Evolution, Speech Perception, Statistical Learning, Category Learning, Science Education.

Research Experience:

Graduate Researcher
ConCHLab
University of Western Ontario

JAN 2021 - Present

- Working under the supervision of Professor Ingrid Johnsrude.
- Thesis project: Investigating the role of learning of sound statistics on auditory category learning, and sound perception.
- Side project investigating the role of cognitive control on the processing of degraded speech.
- Collaboration with projects from other members of the lab.

Mitacs Research Internship
Équipe de recherche en psychologie cognitive
Université du Québec à Trois-Rivières

JULY 2019 - AUGUST 2019

- Worked under the supervision of Professor Isabelle Blanchette.
- Designed experimental protocol, collected and analyzed EEG data for a project entitled: Cognitive neuroscience of emotion-cognition interaction.

Research Internship
Laboratory of Epistemology and Didactic of Biology
University of São Paulo at Ribeirão Preto.

MARCH 2019 - JUNE 2020

- Working under the supervision of Professor Fernanda da Rocha Brando.
- Conducting primary and secondary source research for a project entitled "Seeding evolutionary knowledge in Brazilian public schools".

Research Internship
Laboratory of Investigation in Epilepsies.
University of São Paulo at Ribeirão Preto.

JUNE 2018 - JUNE 2020

- Working under the supervision of Professor João Pereira Leite.
- Developing and working on the project "Development of a within-subject design to investigate the neural encoding of degrees of control over aversive stimuli."

Technical Training
Laboratory of Investigation in Epilepsies.
University of São Paulo at Ribeirão Preto.

JUNE 2016 - DECEMBER 2017

- Worked under the supervision of Professor João Pereira Leite.
- Learned relevant theory on topics of neuroscience, neuropsychology, and neuropsychiatry.
- Learned relevant techniques for behavioral testing of rodents.

Published works:

- Marques, D. B., Rossignoli, M. T., de Avó Mesquita, B., Prizon, T., Zacharias, L. R., Ruggiero, R. N., & Leite, J. P. (n.d.). Decoding fear or safety and approach or avoidance by brain-wide network dynamics. BioRxiv. https://doi.org/10.1101/2022.10.13.511989

Conference Presentations:

Oral Presentations:

- Mesquita, B.A.,Herrmann, B.,Roark, C.L., Johnsrude, I.S. (2023,02) *Perceptual benefits from long-term exposure to naturalistic sound patterns.* ARO Midwinter Meeting 2023. Orlando, Fflorida, USA.
- Mesquita, B.A., Marques, D.B., Rossignoli, M.T., Ruggiero, R.N., Leite, J.P. (2019,10). *Development of a within-subject design to investigate the neural encoding of degrees of control over aversive stimuli.* Annual meeting of the Brazilian Society of Neurosciences and Behavioral Sciences. Campos do Jordão, São Paulo, Brazil.

Poster Presentations:

- Mesquita, B.A., Marques D.B, Rossignoli, M.T., Ruggiero, R.N., Leite, J.P. (2020, 10). *Development of a new animal model to investigate the neural encoding of degrees of control over aversive stimuli.* XII Forum on Neurobiology of Stress and International Symposium on Ethanol Research. Ribeirão Preto, São Paulo, Brazil.

- Mesquita, B.A., Marques, D.B., Rossignoli, M.T., Ruggiero, R.N., Leite, J.P. (2019,10). *Development of a within-subject design to investigate the neural encoding of degrees of control over aversive stimuli.* Annual meeting of the Brazilian Society of Neurosciences and Behavioral Sciences. Campos do Jordão, São Paulo, Brazil. (Awarded Honorable Mention)

- Mesquita, B.A., Marques, D.B., Rossignoli, M.T., Ruggiero, R.N., Leite, J.P. (2019,09). *Characterization of the ability to discriminate between controllable and uncontrollable aversive stimuli in rats.* International symposium of scientific initiation of the University of São Paulo, Ribeirão Preto, São Paulo, Brazil. (Awarded Honorable Mention)

Teaching Experience:

Graduate Teaching Assistant
Department of Psychology.
University of Western Ontario.

SEP 2021 - current

Institutional Program of Scholarships for Teaching Initiation (PIBID)
Laboratory of Education in Biology.
University of São Paulo at Ribeirão Preto.

MARCH 2017 - DECEMBER 2017

- Worked under the supervision of Professor Marcelo Motokane.
- Created and implemented didactic sequences on different topics of science education in Brazilian public schools.

Leadership, Organization, and Community Service:

Member of the Science Outreach Committee Society of Graduate Students - Western University	**2022 - Current**
Expositor and Organization "Bio na Rua" (Biology in the Streets)	**2017**
Expositor Semana Nacional do Cérebro (Brain Awareness Week)	**2016**
Students Representative University of São Paulo at Ribeirão Preto - Department of Biology	**2016/01 to 2017/01**
English teacher Public preparatory course for the Law School of Ribeirão Preto	**2016/01 to 2017/06**
Organization Annual week of bio studies	**2015,2016**
Member of the Student Union University of São Paulo at Ribeirão Preto - Department of Biology	**2015/03 to 2016/12**

Scholarships and Grants Awarded:

- **Western Graduate Research Scholarship (x3)**

  UWO

  2021 - 2023

- **Globalink Graduate Fellowship**

  Mitacs

  2021

- **Scientific initiation scholarship**

  São Paulo Research Foundation (FAPESP)

2019

- **Mitacs Research Internship**

  Mitacs

  2019

- **Scholarship for teaching initiation**

  Institutional Program of Scholarships for Teaching Initiation (PIBID)

  2017