The Influence of Expectation and Experience on Auditory Perception

A Thesis

submitted for the degree of

Doctor of Philosophy

in Physiology, Anatomy and Genetics

by

Katharina S. Bochtler

Department of Physiology, Anatomy and Genetics
Keble College, University of Oxford

Statement of Contributions

Chapter 2:

Animal behaviour: hardware installed by Kerry Walker (KW), maintained by Katharina Bochtler (KB). Software written by KB adapted from scripts written by KW. Behavioural training and testing performed by Katharina Bochtler (KB), Susan Spires (SS), Ellie Faulkner (EF), and Kerry Walker (KW).

Animal data analysis: preprocessing code was adapted from KW but analysis code was created exclusively by KB.

Human behaviour: software was based on existing experiments from Lori Holt (LH), Fred Dick (FD), and Sijia Zhao (SZ) but was substantially expanded and modified by KB. Experiment was performed under approved UCL ethics application fMRI_2022_002 with principal researcher FD. KB and KW are named additional applicants.

Human data analysis: completed exclusively by KB using original scripts.

Chapter 3:

Animal behaviour: hardware installed by KW, maintained by KB. Software written by KB adapted from scripts written by KW. Behavioural training and testing performed by Katharina Bochtler (KB), Susan Spires (SS), Ellie Faulkner (EF), and Kerry Walker (KW).

Analysis: preprocessing code was adapted from KW but analysis was performed using scripts written by KB.

Early results were submitted as part of an SBE-UKRI grant application.

Chapter 4:

Surgery: performed by KB, Hemant Kumar Srivastava (HKS), and Kyunghee Kim (KK).

Apparatus and Software: hardware installed by KB, HKS, Hong Jiang (HJ), August Pfliger (AP), and James Webb (JW). Software written by HJ based on specifications from KB.

Oddball task: training and testing performed by KB.

Vertical Sound Localisation task: training and testing performed by KB and Nikolas Scarcelli (NS).

Data analysis: pupil preprocessing pipeline written by Jan Willem de Gee (JWG) and implemented by KB. Ear tracking algorithm developed by J. Drew (JD) with training data and support provided by KB. Oddball preprocessing code built on code of AP with substantial development by KB. Oddball data analysis created by KB. Vertical sound localisation behavioural analysis created by KB. Ear-specific analysis for both tasks was done by JD. Analysis feedback provided by Matthew McGinley (MM).

Some results were submitted as part of an NIH RO1 grant application.

Abbreviations

STRF: spectrotemporal receptive field

RT: reaction time

PC: percent correct, used interchangeably with performance accuracy

2AFC: two-alternative forced choice task

SNR: signal-to-noise ratio

LME: linear mixed-effects model

GLME: generalised linear mixed-effects model

GNG: go/no-go

DCN: dorsal cochlear nucleus

HRTF: head-related transfer function

CI: confidence Interval

CONTENTS

THE INFLUENCE OF EXPECTATION AND EXPERIENCE ON AUDITORY PERCEPTION

Katharina S. Bochtler

Keble College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary, 2025

Listeners exploit statistical regularities in the environment to guide auditory attention. This thesis investigates how expectations based on stimulus frequency and spatial location influence auditory perception across species and tasks. The first two chapters focus on the probe signal effect, using paradigms in which frequency is orthogonal to the task. Humans and ferrets completed a tone-in-noise detection task, while a separate duration discrimination task was conducted in ferrets. Both tasks manipulated the probability of target frequencies. The results show that whether expected or unexpected frequencies are prioritised varies by species and task design. In some conditions, ferrets showed improved performance for low-probability probe tones, consistent with deviance detection, while in others they showed higher accuracy for expected tones. Some human results diverged from previous findings using two-alternative forced-choice designs, suggesting that task structure can shape attentional biases. Together, these findings indicate that frequency-based expectations influence behaviour, but that this influence is not uniform across tasks or species.

The third chapter examines spatial attention in mice using a vertical oddball paradigm combined with pupillometry and ear tracking. In addition to deviance detection, mice showed greater pupil dilations for sounds from above, suggesting a threat detection mechanism that was still present after pinnae removal. A subsequent go/no-go task showed that mice, contrary to popular belief, have strong vertical localisation abilities. After altering spectral cues by pinnae manipulation, mice could quickly relearn localisation to the level before manipulation.

Overall, the results demonstrate that expectations guide auditory attention in both frequency and spatial domains, but that the form and direction of this bias depend on species, task demands, and stimulus context. Together, these behavioural paradigms establish a foundation for future physiological investigations into neural mechanisms of statistics-driven and spatial auditory attention across species.

1 INTRODUCTION

1.1 AUDITORY SELECTIVE ATTENTION

Without much conscious thought, we use selective auditory attention every day. Whether it's catching up with a friend in a busy café or following a lecture while ignoring students typing their notes, we can focus our attention on the sounds that matter. Despite its relevance to daily life, our understanding of this mechanism is surprisingly limited. Selective auditory attention requires the brain to identify which dimensions of sensory signals are pertinent to a situation and sustain attention to these while potentially suppressing irrelevant dimensions (Bizley & Cohen, 2013; Woods & McDermott, 2018). In the café example, this means attending to the specific frequencies of your friend’s voice and disregarding the array of frequencies from background noise. This may seem like a straightforward skill, yet becomes more difficult with age and is impaired in several clinical pathologies, such as stroke, traumatic brain injury, and neurodevelopmental disorders, which can lead to social isolation, depression, cognitive dysfunction, poor educational attainment, and challenges in the workforce (Carlyon et al., 2001; Dai et al., 2018; Medalia et al., 1998; Robertson, 2001; Sohlberg & Mateer, 2001; Stevens et al., 2006). Without a better understanding of auditory attention, rehabilitation and treatment options remain limited, perpetuating the negative consequences of attentional deficits.

Traditional studies on selective auditory attention typically present stimuli that are meant to be attended to and others to be ignored (Brodbeck et al., 2020; Kurmanavičiūtė et al., 2023; Yao et al., 2019) or compare active task engagement versus passive listening (Alho et al., 2014; Otazu et al., 2009). However, the work presented in this thesis takes a different approach: auditory statistical learning. This type of learning describes the implicit ability to detect patterns and process regularities from sensory input. Originally pioneered in language research by Saffran et al. (1996) and Reber (1967), the literature has since shown that listeners quickly, sensitively, and passively build statistically-driven expectations of auditory stimuli (Armstrong et al., 2017; Frost et al., 2019; Santolin & Saffran, 2018). These expectations can affect perception (Holt, 2005; Southwell & Chait, 2018; Stilp et al., 2010) and may guide attentional allocation not only to high-probability frequency values, but also to more probable spatial locations or ecologically salient cues (Greenberg & Larkin, 1968; Mondor & Bregman, 1994; Scharf et al., 1987). What underlies statistics driven selective auditory attention remains debated, with two opposing models proposed (Press et al., 2020): expectation-based models suggest cortical enhancement of neural representations of the expected stimuli (de Lange et al., 2018; Kaiser et al., 2019) while deviance-based models suggest that unexpected stimuli should be perceptually prioritised (Kok et al., 2012; Summerfield & De Lange, 2014).

1.1.1 Auditory Object Perception and Processing

Although the precise definition of an auditory object is contested (Bizley & Cohen, 2013), it can be understood as a perceptual entity that attributes a sound to a specific source (Griffiths & Warren, 2004; Kubovy & Van Valkenburg, 2001; Shinn-Cunningham, 2008; Winkler et al., 2009). The formation of an auditory object necessitates the auditory system's ability to detect, extract, segregate, and group spectrotemporal patterns from the surrounding acoustic environment.

Auditory objects arise from the brain's ability to group and segregate spectrotemporal patterns, allowing listeners to distinguish relevant sources in complex acoustic environments (Bregman, 1990; Shinn-Cunningham, 2008; Winkler et al., 2009). Despite variability in acoustic features, these representations are stable and generalisable (Ding & Simon, 2012; Griffiths & Warren, 2004).

Initially, theories on auditory processing were influenced by findings from the visual field, leading to the hypothesis that two parallel pathways analyse the identity and location of objects within an auditory scene separately (Kaas & Hackett, 2000; Rauschecker, 1997; Romanski et al., 1999). There is evidence both supporting and challenging this hierarchical parallel model. On one hand, the ventral pathways are thought to handle object extraction, while the dorsal pathways are responsible for sound localisation. For instance, a study on rhesus monkeys found that different regions within the auditory cortex responded to vocalizations based on type or location, supporting the idea of functional specialization (Tian et al., 2001). Similarly, early human imaging studies and a meta-analysis of fMRI data showed distinct activation patterns for spatial and non-

spatial tasks, further supporting this model (Alain et al., 2001; Arnott et al., 2004; Maeder et al., 2001). However, evidence also indicates that both auditory object processing and spatial processing can occur in both pathways, suggesting a more integrated and complex model than the strictly parallel hierarchical one (Bizley et al., 2009; Miller & Recanzone, 2009; Obleser & Eisner, 2009; Rauschecker, 2012; Stecker & Middlebrooks, 2003).

A more nuanced understanding suggests that spatial and non-spatial information may work together to create stable perceptual representations necessary for guiding goal-directed behaviour. For example, spatial information can serve as a grouping cue, enabling the formation of auditory streams. A sequence of identical sound bursts presented from one location is perceived as a single source, while the same sequence from two different locations is perceived as two distinct sources with separate rhythms (Middlebrooks & Onsan, 2012). Furthermore, non-spatial information processed in the dorsal stream may assist in tasks such as target selection and the processing of dynamic auditory information, which are crucial for organizing the auditory scene (Obleser & Eisner, 2009; Rauschecker, 2011, 2012; Teki et al., 2011).

Attention is not required to detect changes in a stimulus feature. For example, in oddball paradigms where a rare sound is interposed into a stream of repeating standard sounds, deviance-detection mechanisms operate automatically without the subject needing to attend to the stimulus (Alain et al., 1994; Kujala et al., 2007; Näätänen & Picton, 1987; Picton et al., 2000). Such effects have been demonstrated even in passive listening conditions, where physiological responses to prediction error can be observed. Without the need for explicit attentional

engagement, the auditory cortex continuously monitors the environment and updates predictions about the current sound scene (Winkler et al., 2009). While attention is not required for streams to form, it plays a critical role in shaping perception, particularly in complex listening conditions with competing sources. In such environments, attentional selection is guided by behavioural goals and listener intent, influencing which streams are prioritised. (Cusack et al., 2004; Shinn-Cunningham, 2008).

There is likely a two-stage process where initially, an auditory scene is parsed into individual objects based on acoustical features, and then selective attention resolves competition between multiple sources by selecting the most salient or behaviourally relevant (Knudsen, 2007; Shinn-Cunningham & Best, 2008). Both bottom-up and top-down cues can be employed to direct attention (Shamma, 2008; Shinn-Cunningham & Best, 2008) and attention can operate on the level of auditory objects (Alain & Arnott, 2000; Desimone & Duncan, 1995; Zatorre et al., 1999). Attention can enhance the representation of attended sounds in the auditory system by increasing the neural response to these sounds while suppressing the response to unattended sounds. This modulation can be observed in early auditory cortex and extends to higher-order auditory areas (Atiani et al., 2009; Ding & Simon, 2012; Fritz et al., 2003; Niwa et al., 2012). In these higher-order areas, neural signals are thought to reflect the listener's perception of a particular auditory object. For example, Ding & Simon (2012) showed that the speech signal attended to by the listener, while ignoring another spectrotemporally overlapping speech signal, preferentially modulated neural activity in the posterior auditory cortex. Similarly, Mesgarani & Chang (2012)

showed that neural responses to unattended speech were suppressed in secondary, but not primary, auditory cortex. Attentional modulation is not mediated via a simple feedforward network but rather involves distinct activity patterns for spatial and non-spatial auditory attention (Hill & Miller, 2010; Lee et al., 2013). It is likely that, similar to visual processing, attentional feedback is sent via a top-down mechanism to early sensory areas (Buffalo et al., 2010). Understanding these mechanisms across species, may help reveal conserved principles of auditory attention and its neural implementation at the cellular level.

1.1.2 Statistics-driven Learning and Attention

Statistical learning refers to the brain's ability to extract regularities from the environment, often without explicit instruction or conscious awareness. It has been demonstrated across a range of cognitive domains and is now viewed as a central component of many theories of perceptual learning (Frost et al., 2019). The term gained prominence through work on speech segmentation in infants (Saffran, Aslin, et al., 1996), but the subsequent research has shown that statistical learning operates across different modalities. In the auditory domain, statistical learning enables the listener to form expectations about acoustic features based on prior experience, even when these features are not explicitly relevant to a behavioural task.

1.1.2.1 Role of Attention in Statistical Learning

The influence of attention on statistical learning remains understudied and poorly understood, especially when contrasted with the extensive attention literature in implicit learning (Frost et al., 2019). While foundational studies in the auditory domain suggest that statistical learning can proceed without overt attention (Saffran et al., 1997), later findings challenge this view. For instance, Toro et al. (2005) reported that speech segmentation is impaired when attention is not allocated towards it, implying a role for attention in auditory statistical learning. In the visual domain, Turk-Browne et al. (2005) found that statistical learning of shape sequences occurred only for stimuli in the attended colour stream, suggesting that attention gates access to learning. However, subsequent replication attempts by Musz et al. (2015) argued that statistical learning does not require selective attention, as they observed learning in both attended and unattended streams. They proposed that earlier failures to find learning in unattended conditions may have reflected task-specific factors, such as reduced encoding time or dual-task interference, rather than a general gating role for attention in statistical learning.

A related line of research in visual attention, the contextual cueing paradigm, may provide further clarity on how statistical learning can influence attentional allocation. The original contextual cueing task (e.g. Chun & Jiang, 1998) asked participants to search for a target (the letter ‘T’) within different spatial configurations full of distractor letters (‘L’). Half of these configurations were repeated, while the other half were novel. The contextual cueing effect was

observed when participants detected the target more quickly in the repeated configurations compared to the novel ones. While contextual cueing is not always explicitly labelled as statistical learning, it does involve the implicit learning of spatial contingencies and is similar to the task described by Fiser & Aslin (2001). This similarity allows contextual cueing to be considered a form of statistical learning (Goujon et al., 2015). The contextual cueing effect can be explained as learning the regularities of where the target letter is located, which in turn directs visual attention toward the right location, leading to faster response times. Thus, contextual cueing demonstrates how visual statistical learning can guide attentional allocation, even when learning is incidental. Consistent with findings in the statistical learning literature, this learning occurs without participants consciously recognizing the repeated configurations. Nevertheless, attention is systematically guided by the learned regularities.

Building on the insights from contextual cueing and statistical learning, it is now well established that statistical regularities can guide attentional allocation, even when acquired incidentally. Work by Wang & Theeuwes (2018a, 2018b, 2018c) demonstrates that attention is automatically drawn to irrelevant visual distractors, yet this capture is modulated by the statistical likelihood of the distractor's location. Importantly, such modulation occurs without conscious awareness, suggesting that statistical learning can operate independently of explicit recognition. When participants were instructed to ignore distractors, response times nonetheless increased if the distractor appeared in a high-probability location. Such effects are most pronounced when the distractor consistently appears in the same location, drawing attention due to learned

regularities rather than task relevance These findings provide a compelling parallel to the current investigation, which asks whether statistical regularities in irrelevant auditory features, such as frequency, can similarly shape attentional processes, even when those features are orthogonal to task demands.

1.1.3 Statistics-driven Auditory Attention: The Probe Signal Literature

The work presented in this thesis builds upon the human probe signal literature, which investigates how stimulus statistics shape selective auditory attention. Greenberg & Larkin (1968) introduced the probe signal task to test whether listeners form frequency-specific expectations that act as auditory filters. They likened their approach to measuring the frequency-response characteristics of an electronic filter: just as one probes a filter by presenting a range of input signals and observing its output, they assessed listeners' attentional filters by presenting tones of varying frequencies. In their task, participants were first exposed to repeated presentations of a standard tone (e.g. 1100 Hz) to build frequency expectation. Participants completed a two-alternative forced choice task, identifying which of two intervals contained a brief tone embedded in continuous wideband noise. Probe tones deviating from this standard were then introduced on a small proportion of randomly interleaved trials, with each block containing only one probe frequency (600, 900, 1300, or 1600 Hz). Detection performance was highest for the expected tone (the standard) and declined for

unexpected probe tones with increasing spectral distance, consistent with the idea of a frequency-tuned attentional filter (Figure 1). These findings were interpreted in terms of critical bands, that is, the frequency tuning width of an auditory nerve fibre in the inner ear, with probes outside a half-band range being poorly detected. Importantly, the effect emerged across 24 experimental sessions, under the assumption that frequency expectations build gradually with experience.

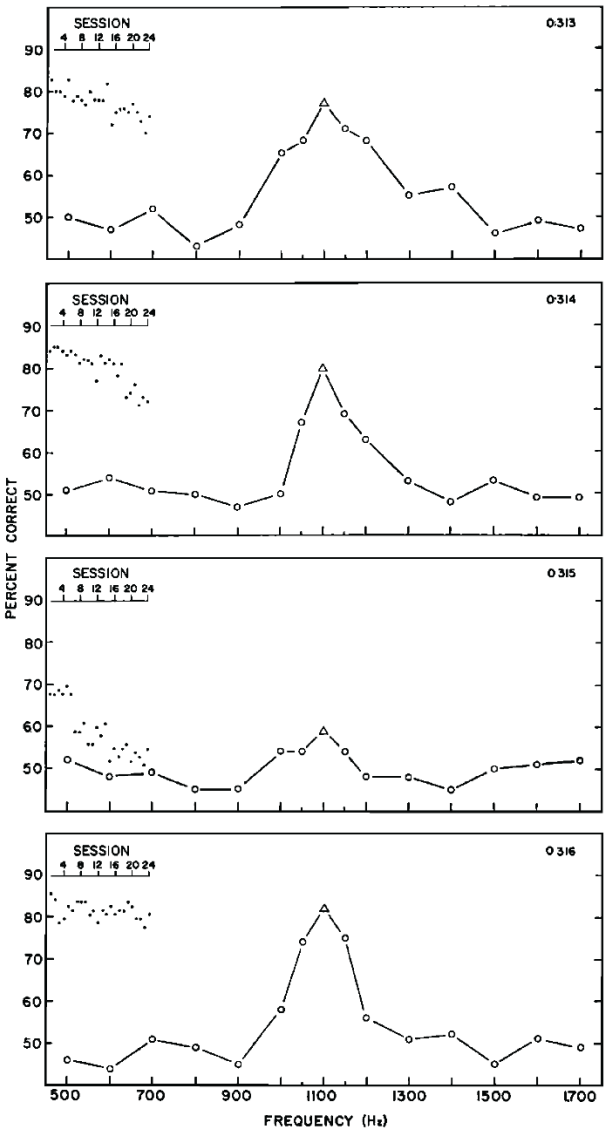

Figure 1. Frequency-Response Profile from Greenberg & Larkin (1968), Showing Detection Accuracy Across Probe Frequencies Surround A 1100 Hz Standard. Detection declined with increasing spectral distance, forming an inverted V-shaped profile consistent with a frequency-tune attentional filter. Reproduced with permission from: Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-Response Characteristic of Auditory Observers Detecting Signals of a Single Frequency in Noise: The Probe-Signal Method. The Journal of the Acoustical Society of America, 44(6), 1513–1523. Copyright © 1968 Acoustical Society of America. Licensed under License Number 6015600658854.

Scharf et al. (1987) expanded on this work by proposing two competing explanations for the reduced detectability of distant probe tones: attentional filtering versus frequency-dependent decision bias. The former posits that

attention enhances sensitivity to expected frequencies within a narrow band, reducing detection of out-of-band probes. The latter suggests that listeners apply a decisional criterion to ignore frequencies that deviate from expectation, even if they are perceptually available. To test these alternatives and reduce the extensive training previously required, Scharf et al. adapted the probe signal task to include a cue tone presented at the beginning of each trial in order to build strong expectation on each trial. While Greenberg & Larkin built expectation over repeated blocks without using a cue, Scharf et al. aimed to generate frequency-specific attention more efficiently within a single session. They also introduced probe-only blocks to serve as control conditions, allowing comparison between expected and unexpected tones when each served as the dominant stimulus. Together, these changes enabled a more direct investigation of whether frequency expectations improve perceptual sensitivity or instead bias decision-making away from unexpected tones.

Using this modified design, Scharf et al. (1987) tested two participant groups: one received probe tones that were also pure tones, while the other was presented with complex probe stimuli. In the pure-tone condition, results replicated earlier findings: detection accuracy was best for the signal, slightly lower for probes near the expected frequency and declined progressively as probes became more spectrally distant from the signal. Tones more than two critical bands from the signal were often detected at chance. Interestingly, “near” frequency probes were detected slightly better in the uncued condition (79%) than in the cued version of the task (64%), which the authors interpreted as evidence that listeners may adjust the attentional filter to include expected probe

frequencies in the absence of a reinforcing cue. The addition of probe-only blocks allowed the authors to test whether detection of previously rare probe frequencies improved once they became expected. However, even when these frequencies were overrepresented, performance did not return to the level observed for the signal in earlier blocks, indicating that the detection advantage was not solely due to frequency of presentation. These results supported the hypothesis of a frequency-specific attentional filter, while also highlighting the complexity of the filter's boundaries depending on task context.

To further test whether poor detection of distant probe frequencies reflects perceptual insensitivity or a decision-level bias, Scharf et al. introduced complex sounds as probe stimuli. These probes differed from the pure-tone signal in timbre but were designed to fall within the same spectral region, including narrowband noise centred on the signal frequency, two-tone complexes, and harmonically structured stimuli sharing the fundamental frequency (i.e. perceived pitch) of the signal tone. If detection failures were purely due to distance in frequency from the attended frequency band, then probes within the same critical band, even if complex, should be well detected. However, performance remained reduced for these complex probes, particularly those that were more perceptually distinct from the signal tone. These findings were inconsistent with both the attentional filter and the frequency-dependent criterion hypothesis. Despite falling within the same spectral band and being potentially salient due to their distinct acoustic structure, complex probes were still detected less accurately. This suggests that listeners neither ignored the probes due to expectation mismatch alone, nor failed to detect them because they fell outside the filter's passband.

To determine whether detection of unexpected probes could be improved with awareness or reinforcement, Scharf et al. introduced several manipulations. First, they provided trial-by-trial feedback. Contrary to predictions based on a decision-level account, feedback led to even poorer detection of unexpected probes. In another condition, participants were explicitly informed that multiple tones would be presented throughout the session, including sounds that deviated from the signal. This instruction led to modest improvement in probe detection but did not eliminate the performance gap. Even when the experimenters themselves completed the task with full awareness of the stimulus structure, detection of unexpected probes remained impaired. These findings argue against the hypothesis that performance differences are due solely to conscious rejection of deviant tones. Instead, they suggest that stimulus statistics shape perceptual filters automatically, prior to decision-making, and that these filters are resistant to top-down override.

To account for these persistent performance differences, Scharf et al. proposed two potential mechanisms by which attentional filters might be implemented. One possibility is peripheral filtering at the level of the cochlea, where efferent projections from the olivocochlear bundle suppress responses to frequencies outside the expected range while enhancing those within it (Lukas, 1980, 1981; McCallum et al., 1983; Oatman & Anderson, 1977; Wiederhold, 1970). Alternatively, filtering may occur at a more central level, with neural responses to expected stimuli weighted more heavily than those to unexpected probes (Buus et al., 1986; Green, 1958). The finding that complex probes with similar perceived pitch (a cortically derived feature) to the signal tone but different

spectral range (at the level of the cochlea) were still poorly detected suggests that some aspect of filtering occurs early in the auditory pathway, possibly prior to full perceptual analysis. This peripheral locus of filtering may also explain why top-down strategies, such as instruction or feedback, failed to eliminate probe detection deficits. Additionally, these effects may relate to early neural markers of expectation such as processing mismatch negativity (Näätänen et al., 2007; Näätänen & Picton, 1987), which is elicited by violations of an established regularity even in passive listening conditions.

While Scharf et al. (1987) advanced our understanding of attentional filtering by introducing complex probes, Wright & Dai (1994) provided a critical extension by manipulating tone duration. Prior to their work, it was generally assumed that detection performance in the probe signal task was unaffected by stimulus duration. However, Wright and Dai showed that the shape of the attentional filter, defined as percent correct as a function of probe frequency, was strongly modulated by tone duration. When the signal and probe tones were brief (5 ms), the filter shape was broad, indicating more uniform detection across frequencies. When tones were long (295 ms), the filter shape was sharper, with greater selectivity for the expected frequency. These effects occurred despite matched intensity thresholds across frequencies using the continuous notched-noise method, ruling out simple auditory sensitivity as an explanation.

To further test whether performance differences were due to duration or unexpected frequency, Wright & Dai (1994) included trials where the probe duration differed from the signal tone duration. Surprisingly, detection was poor even when the probe and signal shared the same frequency but differed in

duration. This suggests that detection depends not only on spectral overlap but also on whether the probe duration matches the listener's temporal expectation. These results pose a challenge to critical band-based accounts of attentional filtering and suggest that listeners monitor not just frequency but also duration-specific expectations.

A potential limitation of this study, however, lies in the regularity of probe presentation. Because probes occurred on every fourth trial, listeners may have implicitly learned this pattern (Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996)(Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996), thereby reducing the true unpredictability of the probes. This raises the possibility that the filter adaptations observed may have been partially influenced by higher-order regularities in task structure, in addition to low-level acoustic features.

Building on earlier work, Mondor & Bregman (1994) shifted the probe signal paradigm from detection to identification by introducing a duration judgement task. In their study, tones of varying frequencies were presented well above threshold, in silence, and participants were explicitly informed that they might hear sounds of different frequencies. Crucially, participants were not required to detect whether a tone was present, but instead to judge whether a tone was long or short. This design ensured that frequency, the manipulated dimension, was orthogonal to the participants' task of judging tone duration, paralleling tone-in-noise detection paradigms, where frequency is likewise varied independently of the listener's objective, which is merely to detect the presence of a tone. These designs reduce the likelihood that frequency expectations directly influence response decisions,

making them powerful tools for studying attentional filtering independent of explicit target recognition.

A notable innovation in Mondor and Bregman's study was the inclusion of four probe frequencies presented in the same block, as opposed to prior studies that used only one probe per block and estimated filter shapes across sessions. The findings showed a clear probe signal effect: accuracy declined systematically as probe frequencies deviated from the expected tone, despite frequency being task irrelevant.

In a second experiment, Mondor & Bregman (1994) aimed to dissociate the probe signal effect from simple frequency overrepresentation by cueing the expected frequency on each trial rather than allowing expectation to build gradually. Three signal frequencies (667, 1000, and 1500 Hz) were used, with a cue tone indicating the expected frequency before the target tone. One a given trial, the cue and target frequencies either matched (i.e. expected signal) or mismatched (i.e. unexpected probe). Because each frequency was presented equally often, the task avoided bias due to unequal exposure. Results showed that performance declined on mismatch trials, indicating that frequency-based expectations can form rapidly within a single trial. Conversely, when the interval between the cue tone and the target tone exceeded 1000 ms, mismatch costs were reduced, suggesting that the influence of the frequency cue on attentional filtering diminishes over time. This may reflect a decay in the working memory trace of the cue tone, such that tones deviating from the cued frequency are no longer treated as unexpected. These findings demonstrate that both detection and identification tasks can be used to study the probe signal effect and that

frequency-specific expectations can form dynamically in a temporally constrained window.

In sum, both detection and duration judgement paradigms are effective tools for studying the probe signal effect. They demonstrate that attentional filtering can arise even when the manipulated dimension is orthogonal to the task, and that expectations can influence perceptual performance under a range of stimulus conditions.

Throughout this thesis, the terms *attention* and *expectation* are used in line with their operational definitions in the auditory and behavioural literature. Attention refers to the selective enhancement of specific sensory features or dimensions, such as tone frequency or spatial location, consistent with theories of sensory prioritisation in auditory processing (Bizley & Cohen, 2013; Woods & McDermott, 2018). Expectation is used to describe predictions formed from learned statistical regularities, whether acquired implicitly or cued explicitly, as conceptualised in the predictive coding and mismatch negativity literature (Friston, 2005; Wacongne et al., 2012). These definitions align with the dominant usage in the auditory neuroscience literature and form the basis for how attention and expectation are operationalised throughout the thesis.

1.1.4 Spatial Hearing and Vertical Localisation

While much research on auditory attention has focused on the spectral and temporal features of sound, attention must be flexibly allocated across space. Auditory spatial cues allow listeners to orient toward behaviourally relevant

sounds in their environment and are crucial for parsing complex auditory environments. Spatial hearing relies on the integration of binaural differences in timing and level (interaural time differences, ITDs; interaural level differences, ILDs) as well as monaural spectral cues shaped by the external ear (Blauert, 1997; Middlebrooks, 2015). While ITDs and ILDs are the primary basis for azimuthal localisation, spectral cues are especially important for vertical and front-back discrimination.

Vertical sound localisation remains comparatively understudied despite its ecological and computational relevance. Unlike azimuthal cues, which are based on geometric interaural disparities, vertical localisation relies on frequency-dependent spectral notches shaped by the pinnae. The sound arriving at the eardrum reflects a combination of the source spectrum and direction-dependent filtering by the external ear (i.e. the HRTF), which introduces spectral notches used for vertical localisation (Figure 2). These high-frequency notches vary systematically with elevation and are altered when the pinnae are removed. To accurately infer source location, the auditory system must separate these spatial cues from the intrinsic spectral content of the sound, a process described as a form of spectral deconvolution (Middlebrooks, 2015). The necessity of learning these mappings is supported by studies showing that animals deprived of pinnae early in life fail to develop accurate vertical localisation abilities (Parsons et al., 1999; Schnupp et al., 1998). In small mammals like mice, reliable interaural time and level differences are often limited by head size and frequency sensitivity, especially for narrow azimuthal separations (Ito et al., 2020; Lauer et al., 2011). These limitations increase reliance on high-frequency spectral notches introduced

by the pinnae, which vary systematically with sound elevation (Behrens & Klump, 2016; Keating & King, 2013; Si et al., 2022). These cues are specific to the morphology of an individual's ears and must be learned through experience.

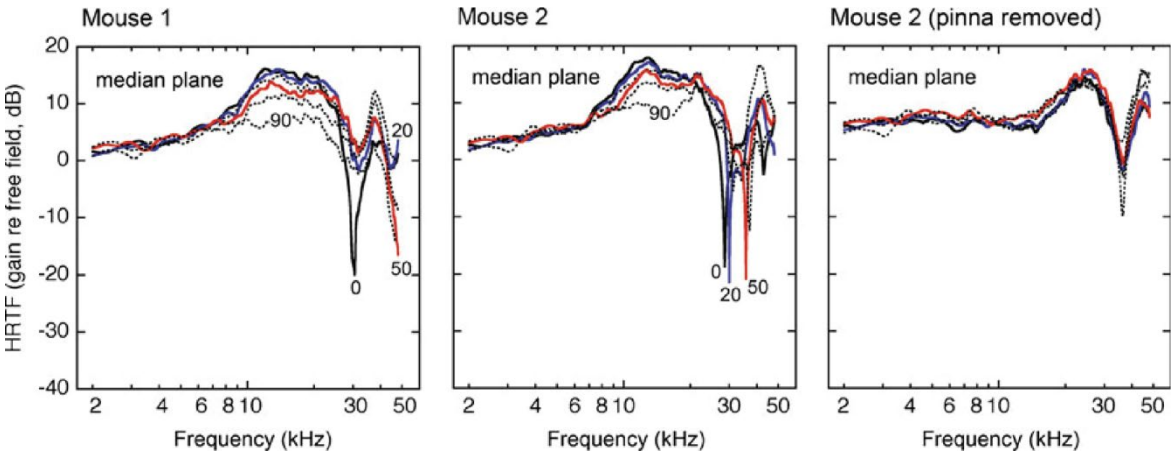

Figure 2. Spectral Notch Patterns in Mouse HRTFs Before and After Pinna Removal. Head-related transfer functions (HRTFs) recorded from six vertical-plane sound locations in two mice (left and centre) show consistent spectral notch patterns at high frequencies, which vary systematically with elevation and support local sound localisation. In the same mouse post-manipulation (right), these notches are markedly reduced or absent, indicating that removal of the pinna disrupts the spectral cues necessary for elevation encoding. The extent and laterality of pinna removal were not specific in the original publication. Figure reproduced with permission from Lauer et al. (2011), *Journal of the Association for Research in Otolaryngology*, Springer Nature. License ID: 5996600501283.

Disruption of spectral cues, such as through pinna moulds or removal, impairs vertical localisation in humans, but several studies have shown that adaptation to altered spectral input in adulthood is possible with experience. Hofman et al. (1998) altered vertical localisation ability by inserting custom-made moulds into the concha of both ears, thereby altering the spectral filtering properties of the pinnae. These moulds modified the topography of the concha, systematically shifting or attenuating the natural spectral notches used for elevation judgement. Although spectral cues remained partially informative, initial localisation performance was severely degraded. However, with several weeks of experience,

participants gradually reacquired accurate elevation perception, demonstrating adaptive remapping of spectral associations.

Building on this work, Van Wanrooij & Van Opstal (2005) introduced a concha-fitted ear mould to one ear, creating unilateral spectral disruption, and tracked localisation performance over several weeks of naturalistic wear. By testing localisation with both unilateral and bilateral moulds before and after adaptation, they demonstrated that listeners learned a new spectral-to-spatial mapping for the perturbed ear, rather than simply reweighting binaural cues. This adaptation was ear-specific and improved localisation even when both ears were disrupted, confirming that the auditory system can flexibly recalibrate spatial associations when given altered, but informative, input.

Earlier work by Shinn-Cunningham et al. (1998) used headphone-based presentation of spatialised stimuli to artificially remap the relationship between head-related transfer functions and perceived sound direction. Although the pinnae themselves were unaltered, listeners were trained with feedback to learn a new association between spectral cues and source location. Results showed that while adaptation was incomplete, response bias decreased substantially with experience, suggesting that listeners can extract and reassign meaning to novel spectral patterns when provided with training.

Together, these studies demonstrate that vertical localisation is not fixed but supported by plastic, cue-specific spatial representations that can be updated through perceptual experience. This ability to remap spatial cue associations in response to altered input underscores the adaptability of the auditory system in both real and virtual contexts.

Ferrets have been a central model for studying the behavioural and neural basis of spectral cue processing and adaptive spatial hearing. Behaviourally, they show robust vertical and front-back localisation abilities, relying on spectral cues when interaural differences are minimal or ambiguous (King et al., 2001). Developmental studies provide strong evidence that these cues must be calibrated through experience: Parsons et al. (1999) found that ferrets deprived of natural spectral input from early in life, via bilateral removal of the pinnae and conchae, exhibited lasting impairments in vertical and front-back localisation, even after extensive behavioural training. These findings parallel human studies of spectral remapping and underscore the need for early exposure to stable pinna cues to form accurate elevation representations. At the neural level, Bizley et al. (2007) showed that bilateral auditory cortex lesions impair vertical localisation in the vertical plane, confirming a cortical role in the processing of spectral spatial information. Although most studies of cortical plasticity in ferrets have focused on azimuthal cues, they provide important mechanistic insights, such as the involvement of cortical feedback pathways (Nodal et al., 2012), that may generalise to vertical localisation.

Although ferrets have been instrumental in establishing the developmental and cortical bases of spectral cue processing, mice offer a complementary model for linking behavioural performance to neural mechanisms using genetic and optical tools. While their small head size limits the reliability of binaural cues, mice possess exceptionally sensitive high-frequency hearing that extends well above 60 kHz (Lauer et al., 2011), making them theoretically well-suited for exploiting the high-frequency spectral notches required for vertical localisation. Behavioural

studies confirm that mice can use these cues under controlled conditions: Lauer et al. (2011) showed that mice could discriminate vertical speaker positions separated by 90°, although accuracy declined at smaller separations. This raises the question of whether vertical localisation acuity in mice has been underestimated, and whether more refined behavioural protocols could reveal greater spatial resolution than previously reported. At the neural level, Ito et al. (2020) demonstrated that neurons in the mouse superior colliculus form a topographic map of azimuthal space, and that this map depends critically on spectral cues, particularly for encoding frontal receptive fields. Although their study focused on azimuth, the demonstrated reliance on spectral cues highlights the broader role of these cues in spatial hearing, including their primary role in encoding elevation.

At the subcortical level, spectral localisation cues are thought to be processed by the dorsal cochlear nucleus (DCN), which integrates auditory input with somatosensory signals from the pinnae (Balmer & Trussell, 2021; Imig et al., 2000; Parsons et al., 2001; Reiss & Young, 2005). These integrated signals provide the basis for encoding learned, location-specific spectral features and may help compensate for weak interaural spatial information. The current thesis builds on this premise by investigating how vertical sound localisation is learned, represented, and adapted across changing acoustic conditions. Spatial hearing, like frequency-based expectation, relies on prior experience and learned associations, making it a natural extension of this thesis's focus on statistics-guided auditory attention.

1.1.5 Present Thesis

The work presented in this thesis aims to develop and pilot probe signal tasks that can be employed, without much adaptation, in both ferrets and to study neural correlates of selective attention across species. While building on the foundational work of Greenberg & Larkin (1968) and Mondor & Bregman (1994), the tasks developed here diverge from previous implementations in several aspects, including the removal of trial-by-trial cueing and adapting the task design to one that can be trained in non-human species. As discussed in Section *1.1.3,* most probe signal investigations have utilised cues at the start of every trial to guide attention, even though Scharf et al. (1987) demonstrated that such cues are not necessary for establishing frequency-based expectation. Instead, expectations can emerge gradually and without conscious awareness. To better capture this process, the tasks in this thesis were designed to operate without trial-by-trial cueing, allowing statistical regularities to shape behaviour over time.

The existing probe signal literature is based almost entirely on human psychophysics. Adapting such tasks for animal models, particularly ferrets, presents both conceptual and methodological challenges. Most notably, to the knowledge of the author, a duration discrimination task like that of Mondor & Bregman (1994) has not previously been attempted in non-human species. While tone-in-noise detection has been studied in ferrets (David et al., 2012; Fritz et al., 2003, 2005a, 2005b, 2007) and non-human primates (Freedman et al., 2001; Lemus et al., 2009; Russ et al., 2007), these studies have not incorporated statistical manipulation of frequency to investigate expectation. Therefore, our

specific task is novel for ferrets as well. As such, it remains an open question whether non-human animals can form frequency expectations in a manner analogous to humans, and whether such expectations influence behaviour even when they are not directly task-relevant.

The final empirical chapter shifts focus from frequency-based expectation to spatial auditory processing in mice. Rather than using a probe signal structure, this chapter employs a vertical oddball paradigm in which the spatial location of sound is manipulated across trials. Pupillometry and ear posture tracking are used to assess physiological responses to deviant sound locations, providing a non-invasive measure of orienting and arousal. To investigate the role of natural spectral cues, the experiment includes a manipulation of the pinnae, disrupting head-related transfer functions that normally support elevation perception. A second experiment introduces an active go/no-go localisation task, in which mice learn to discriminate between elevated and azimuthal sound sources. This design allows for the investigation of whether animals can learn to adjust to degraded spatial cues following partial pinnae removal. By taking advantage of the mouse's exceptionally broad hearing range, critical for resolving the spectral notches used in vertical localisation, and their compatibility with high-resolution physiological tools such as pupillometry and ear tracking in a head-fixed state, these experiments are uniquely positioned to probe the use and adaptive plasticity of spectral cues. Together, these experiments extend the thesis's broader theme of statistics-driven auditory attention into the spatial domain and highlight the potential of combining behavioural and physiological measures to study expectation and sensory learning.

2 PROBE SIGNAL: TONE IN NOISE DETECTION

Research on attention has its origin in early attempts to understand how the brain copes with the vast influx of sensory information. Broadbent's seminal filter theory (1958) proposed that attention acts as a bottleneck, which only allows select stimuli to pass through and be processed. Although formulated decades ago, Broadbent's model is not far removed from current accounts of auditory attention, which similarly describe a filter that allows listeners to attend to a specific sound or voice amidst competing auditory stimuli (Woods & McDermott, 2018).

Probe signal paradigms have been proposed as a way to study statistics-driven selective auditory attention, especially how expectation shapes behavioural responses to stimuli (Zhao et al., 2022). The corresponding probe signal effect refers to enhanced performance on a perceptual task for expected "signal" tones, and reduced performance for unexpected "probe" tones. The effect is thought to arise from the attentional filter being centred on the expected (Mondor & Bregman, 1994).

Greenberg & Larkin (1968) introduced the probe signal method, a psychophysical task aimed at examining how auditory attention is shaped by expectations. In their experiments, listeners needed to detect tones embedded in broadband noise using a two-alternative forced-choice (2AFC) paradigm. Participants were first presented with trials in which only the signal tone was

embedded in noise, allowing them to become familiar with and attuned to that frequency. Once this expectation had been established, occasional, “probe” tones at different frequencies replaced the signal tone on a small number of trials within each block. Detection rates varied systematically as a function of the probe tone’s frequency relative to the signal tone. Probes closer in frequency to the signal were detected more accurately, whereas detection declined for more distant probes, with the farthest frequencies approaching chance-level performance. This gradient in performance supports the presence of a frequency-specific attentional filter centred on the expected signal. Greenberg and Larkin’s findings demonstrated that auditory attention can selectively enhance the processing of expected stimuli while suppressing irrelevant input. Moreover, their results suggested that this attentional filter is dynamic, adapting to task demands and shaped by prior exposure and expectation.

Work by Zhao et al. (2022) and Luthra et al. (2024) has recently expanded this research to further elucidate the dynamic nature of the attentional mechanism. The former demonstrated that robust, frequency-selective auditory attention effects could be replicated in a less well-controlled online experiment. Using online versions of the probe signal paradigm, they validated tone detection tasks that efficiently established previously known perceptual thresholds and replicated classic frequency-selective probe signal effects. Participants reliably showed the expected probe signal effect, in which high-probability tones were detected with greater accuracy compared to rare, low-probability probes. The study showed that, contrary to the 24 testing sessions employed by Greenberg &

Larkin (1968), a thirty-minute online session was enough to reveal robust attentional filtering effects.

Building upon this work, Luthra et al. (2024) used an online tone-in-noise detection task to examine how statistical learning over task-irrelevant frequency distributions shapes auditory perception. In their first experiment, they replicated the basic probe signal effect. Subsequent experiments introduced more complex probability structures, revealing that detection accuracy scaled with both probability of presentation and the proximity to the signal. Tones closer in frequency to the signal showed attenuated suppression relative to more distant probes, consistent with a graded attentional filter. Importantly, this pattern held even when the high-probability tone was positioned off-centre in the distribution, suggesting that suppression followed the statistical peak rather than absolute frequency. When two high-probability tones were presented (e.g., in a bimodal distribution), detection exhibited a “dual spotlight” pattern, with strongest suppression for probe frequencies distant from both signal frequencies. As a key conceptual contribution, the authors introduced the concept of statistical deafening to describe a lasting suppression of sensitivity for low-probability frequencies, arguing that suppression, not enhancement, is the primary driver of the probe signal effect. This concept reframes the probe signal effect as a consequence of persistent, distributionally driven suppression rather than transient attentional enhancement.

This chapter presents data from both ferrets and humans collected using a shared go/no-go tone-in-noise task designed to probe selective auditory attention under comparable conditions. The primary aim was to establish a robust

behavioural task for investigating the probe signal effect in ferrets. In addition to assessing whether the behavioural effect generalises across mammalian species, the task was developed to provide an animal model suitable for future investigations of its neural basis. Unlike the 2AFC paradigm used by our collaborators (Luthra et al., 2024; Zhao et al., 2022), the go/no-go design is more readily trainable in animals (Aoki et al., 2017; Homma et al., 2016; Walker et al., 2017) and facilitates within-species and cross-species comparisons. It also offers practical advantages for animal psychophysics: ferrets typically respond by licking a water spout, and the reduced movement demands of the go/no-go format support shorter reaction times and more stable conditions for neural recordings. To evaluate whether the task structure itself influences behavioural outcomes, complementary data were collected from human participants using the same paradigm in an online setting. This cross-species, cross-task comparison enables both the validation of the ferret as a model for human auditory attention and an assessment of how task design shapes the expression of the probe signal effect.

2.1 ANIMAL METHODS

2.1.1 Subjects

Four naive adult female pigmented ferrets (aged 6-14 months) were trained in this study. Animals were housed together in a large pen with free access to food pellets and environmental enrichment. Training typically took place twice daily for five consecutive days, followed by two rest days. Each session lasted for as long as ferrets showed interest in performing the task, which was typically around 50-120 trials for the final testing stage. On training days, drinking water served as a task reward, and was supplemented as wet food in the evening to ensure that each ferret received at least 60 ml/kg of water daily. On rest days, water was provided ad lib. Body weight was measured daily during training. If more than 10% of body weight was lost, animals were taken off the water regulation schedule until their weight recovered. Tympanometry examinations were carried out to confirm the ferrets' ear health. The animal procedures were approved by the University of Oxford Committee on Animal Care and Ethical Review and were carried out under license from the UK Home Office, in accordance with the Animals (Scientific Procedures) Act 1986.

2.1.2 Apparatus

Ferrets were trained on the tone-in-noise detection task in custom-built testing chambers constructed from double glazing units with a vacuum layer for sound insulation. Each chamber measured approximately 45 x 62 x 54 cm and was fitted with a ceiling lined with sound-absorbing foam. One wall of the chamber was outfitted with three plastic nose poke tubes, each equipped with an inner waterspout and an LED beam (Figure 3). The central spout, designated as the “response spout”, was used to trigger trials. The right spout was the “reward spout”, while the left spout was not used in this paradigm. A loudspeaker (FRS 8; Visaton, Crewe, UK) with a flat frequency response (± 2 dB) from 0.2 to 20 kHz was mounted above the central spout. The animals' nose pokes were detected by disruption of the LED beam at the opening of each tube. Water delivery was controlled by solenoids, and the tasks of data collection, reward delivery, and stimulus generation were automated via custom MATLAB (The Mathworks, Natick, MA, USA) code running on a computer and communicating with TDT real-time processors (RP2 or RM1; Tucker-Davis Technologies, Alachua, FL, USA).

This version of the probe signal task used a go/no-go paradigm where ferrets needed to continuously poke the centre spout until they heard a tone, after which they released within a short reaction time window. If they held and released the centre spout at the correct time, they received a water reward from the right spout.

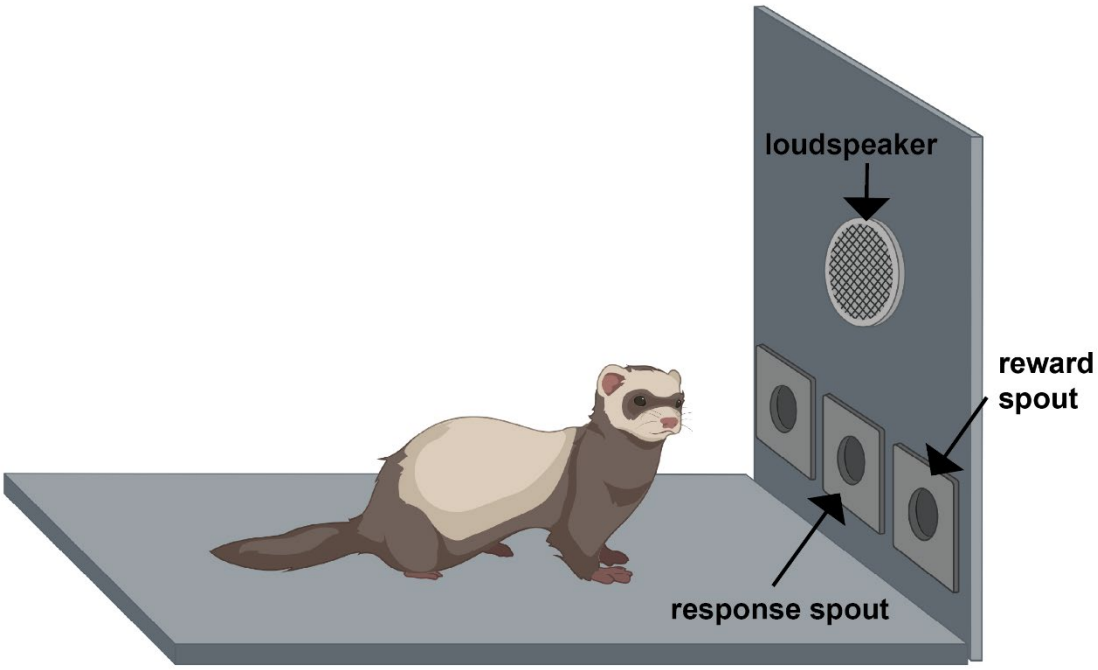

Figure 3. Schematic of the Testing Chamber Used for Ferret Behavioural Training. Three nose poke holes are located below the speaker. For the tone-in-noise task, the central and right spout are used. The ferret can move freely within the box and chooses when to initiate trials.

2.1.2.1 Training

Stage 1: Spout Training

In the initial "spout training stage", ferrets familiarised themselves with the testing chamber and nose poke holes. They were trained to keep their nose in the centre spout for increasing durations. This was achieved by rewarding them for holding at the centre spout for increasing durations of time to receive a water reward (~ 0.3 ml) at the right spout, starting from 0.1 s and increasing in steps with a mean step size of 0.3 s across trials and sessions up to a final hold time of 2 s. No sounds were played during this stage of training, and there was no time-out for unsuccessful attempts; ferrets could immediately retry a centre nose poke

until they achieved the required hold time and obtained the reward. On average, ferrets required 13 sessions to reach the final hold time and complete this stage of training.

Stage 2: Introducing Pure Tones

Following successful spout training, we introduced a 500 ms pure tone stimulus at 60 dB SPLA that was randomly presented at one of five time points during a trial. The frequency of the tone was equally likely to be 1890 Hz, 2381 Hz, 3000 Hz, 3780 Hz, or 4762 Hz across trials. Ferrets initiated a trial by poking the centre spout and were required to continue poking ("no-go") until the tone ("go") was presented. Animals were rewarded if they released within 650 ms of stimulus onset ("response window"). In the case of an early spout release before the response window ("false alarms") or lack of release during the response window ("misses"), ferrets would not receive a reward but could initiate a new trial without a time out period. We chose to introduce all five testing frequencies during training to avoid an overfamiliarity with any one stimulus. To progress to the next stage, ferrets were required to achieve a hit rate of 65% or higher. One ferret (F2206) failed to reach this criterion and was moved forward after 35 sessions. Excluding this animal, the remaining ferrets required an average of 23.67 sessions to reach criterion performance.

Stage 3: Introducing Background Noise

The final stage of training was otherwise identical to stage 2 above but introduced: (a) a continuous broadband noise ("background noise"; 20 Hz –12.5 kHz) at 55 dB SPLA throughout the task; and (b) a 5 s time out for misses and false alarms during which the background noise paused, and no new trials could

be triggered. Once a performance of 65% correct trials was reached across 2 consecutive sessions, ferrets moved to the testing stage. On average, ferrets completed 7 sessions at this stage before progressing.

Prior to determining individual thresholds as described in the following stage, we collected a control dataset using the full task structure outlined here, including continuous background noise and timeout penalties. In this control condition, all pure tones were presented with equal probability at a fixed sound intensity of 30 dB SPLA (-25 SNR). This dataset has been included in this thesis to characterise baseline performance across the stimulus frequency.

Stage 4: Determining Thresholds

Data from human participants suggested that presenting pure tones near each individual's threshold was essential to observe the probe signal effect (unpublished observations from Lori Holt's Lab). In ferrets, thresholds were measured using the tone-in-noise task design with an adaptive "three-down, one-up" staircase procedure (Levitt, 1971). After three consecutive correct responses, the tone level was reduced on the following trial. To help animals approach their threshold more efficiently, the step size for sound level decreases was set to 5 dB during the first 25 trials. A single incorrect response led to a 2 dB increase in tone level on the next trial. Unless either of these criteria were met, the sound level remained unchanged. The background noise level remained constant throughout the task. Tone frequency was fixed within a session, and different frequencies were tested across sessions.

Individual ferrets' thresholds were determined by taking the mode signal-to-noise ratio (SNR) per session, then taking the average of this value across

sessions for each of the five frequencies. The results are shown in Figure 4. These SNR values were used to present each frequency at each ferret's threshold sound intensity in the probe signal version of the task (Stage 5, below).

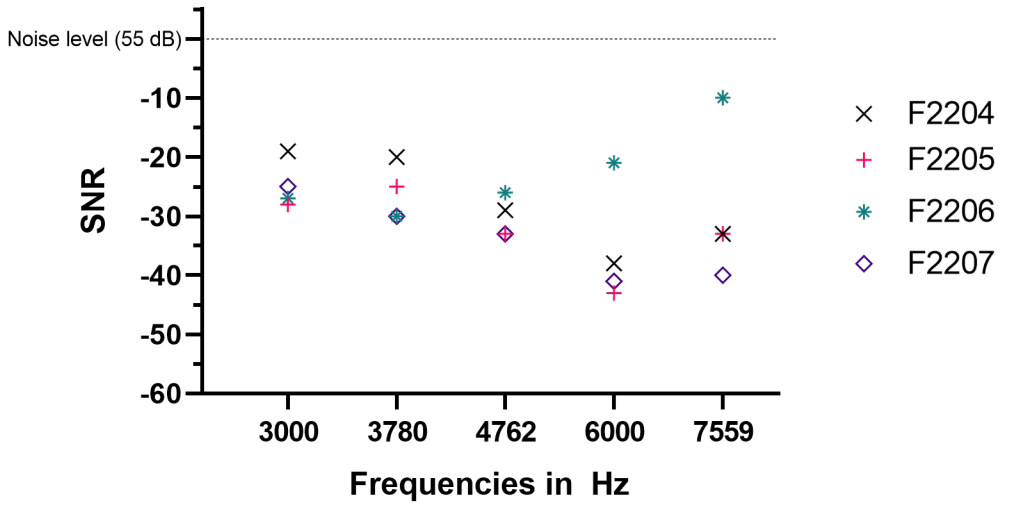

Figure 4. Signal to Noise Ratios Across Frequencies for Individual Ferrets. This plot visualises the signal to noise ratio (SNR = signal dB - noise dB) calculated for each frequency across different ferrets. Noise levels were consistently set at 55 dB. Each marker represents the threshold for a specific frequency, determined for each ferret by averaging the modal SNR values from all sessions. The SNR values were set as the respective sound intensity for each frequency and ferret during the task. Of note is that ferret F2206 had difficulties detecting higher frequencies while the other ferrets tended to have greater SNRs.

Stage 5: Testing

Ferrets completed a go/no-go probe signal task which required them to detect a pure tone in the presence of continuous background noise. This task was identical to Stage 3, but introduced variations in the probability of tone frequencies, with one tone being more probable to occur than the other four (Figure 5). On the majority (80%) of trials, the signal tone frequency was played while the four probe tone frequencies were played on a randomly interleaved 20% of trials, at a probability of 5% each. The signal frequency was randomly chosen for each behavioural session as either the central (4762 Hz), highest (7559 Hz), or

lowest (3000 Hz) tone frequency at the start of a session, to avoid ferrets developing an overall expectation of a single frequency across sessions.

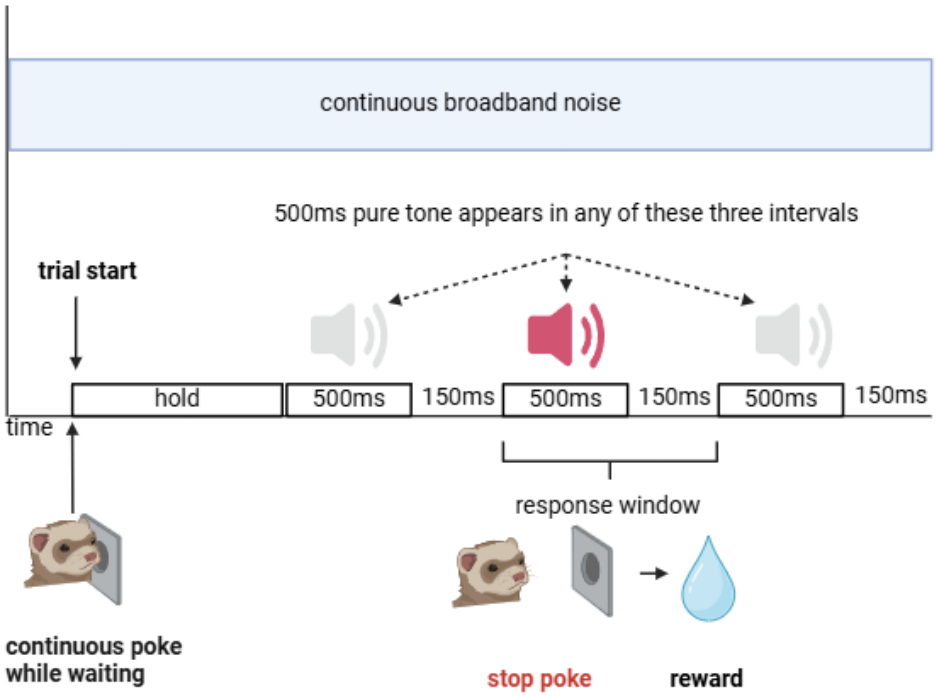

Figure 5. Schematic of the Go/No-Go Probe Signal Task with Continuous Broadband Noise. This diagram illustrates a timeline of a single trial in the behaviour task. Ferrets were required to distinguish a pure tone against a backdrop of continuous broadband noise. The ferret initiated a trial by nose poking in the centre spout, and was required to remain in the spout continuously until a tone was played in order to receive a water reward. Light blue arrows indicate potential times of the tones and a red speaker icon indicates the hypothetical time of the tone on this trial. The varied frequency tones (3000 Hz, 3780 Hz, 4762 Hz, 6000 Hz, 7559 Hz) included both signal and probe frequencies, with the signal frequency randomly assigned from the set of frequencies at the beginning of each session.

2.1.3 Data Analysis

All primary data processing and statistical analyses were conducted using custom-written scripts in MATLAB (R2024b and R2023b; MathWorks, Natick, MA, USA) and Python (version 3.12).

Classical statistical analyses, including repeated-measures ANOVAs and Welch’s t-tests, were performed using MATLAB's Statistics and Machine Learning Toolbox. Holm-Bonferroni corrections for multiple comparisons and Greenhouse-Geisser corrections for sphericity violations were applied where appropriate.

In addition, Bayesian computational modelling was implemented in Python using CmdStanPy to allow full posterior estimation of model parameters and more flexible handling of hierarchical structures. Bayesian inference was chosen to complement the Frequentist analyses by providing richer uncertainty quantification and model comparisons based on posterior predictive checks. Figures were generated in Python.

2.2 ANIMAL RESULTS

2.2.1 Classical Analyses

To establish a baseline for frequency-dependent performance, we analysed a control dataset collected using the same task structure as the main experiment, but without statistical manipulation of tone probability. All tones were presented at a fixed intensity, and no individual thresholds had yet been applied. As shown in Figure 6, performance was generally high across all frequencies, with group-level accuracy exceeding 80%. Plotted reaction times suggest a slight advantage for the two highest frequencies.

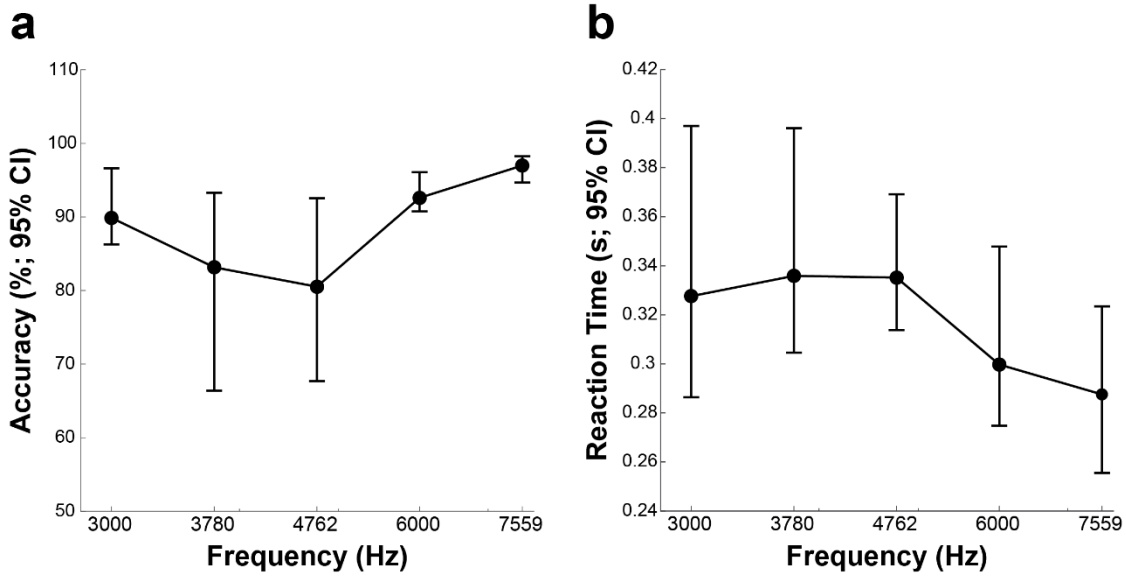

Figure 6. Group-Level Accuracy and Reaction Time for Control (Equiprobable) Sessions. (a) Mean accuracy (% correct) and **(b)** mean reaction time (seconds) across frequencies during control sessions, where all tones were equiprobable and presented without manipulation of signal probability. Data represent group-level means across ferrets, with 95% bias-corrected and accelerated (BCa) confidence intervals computed via bootstrapping. Accuracy exceeded 80% across all frequencies, and no significant pairwise differences were observed. Reaction times were generally comparable, though responses to 7559 Hz were significantly faster than to 4762 Hz. **Note.** Control sessions were conducted before threshold calibration and therefore used the same fixed sound intensity (30 dB, - 25 dB SNR) for all animals. The main experimental dataset was collected at individually determined

thresholds. This distinction is important in light of prior human studies indicating that probe signal effects emerge only when stimuli are presented at or near perceptual threshold. Ferret F2206 was excluded from these control data, consistent with all subsequent analyses.

Pairwise comparisons across tones confirmed that, for accuracy, none of the differences were significant. For reaction time, responses to 7559 Hz were significantly faster than to 4762 Hz ($p = 0.0007$); all other comparisons were non-significant. This indicates that, in the absence of statistical manipulation, behavioural performance was broadly comparable across frequencies. The effects observed in the probe signal task can therefore be attributed to the imposed statistical structure of the stimulus presentation rather than intrinsic differences in detectability between frequencies.

Figure 7 shows group-level accuracy and reaction time across frequencies, separated by signal frequency condition. Accuracy (percent correct) is plotted for all trials, and reaction times are limited to correct trials. Each dot reflects the group mean, with 95% confidence intervals. Accuracy across all animals and frequencies averaged 71% ($SD = 45.4\%$), indicating that animals successfully learned the task. Reaction times fell within a plausible range ($M = 0.29$ s, $SD = 0.106$ s; range: 0.225-0.398 s). One animal (F2206) was excluded from these plots and subsequent statistical analyses due to consistently lower performance ($M = 51.02\%$) and documented high-frequency hearing loss (see Figure 4). While prior studies predict improved performance at the signal frequency, this pattern is not consistently evident at the group level in either accuracy or reaction time.

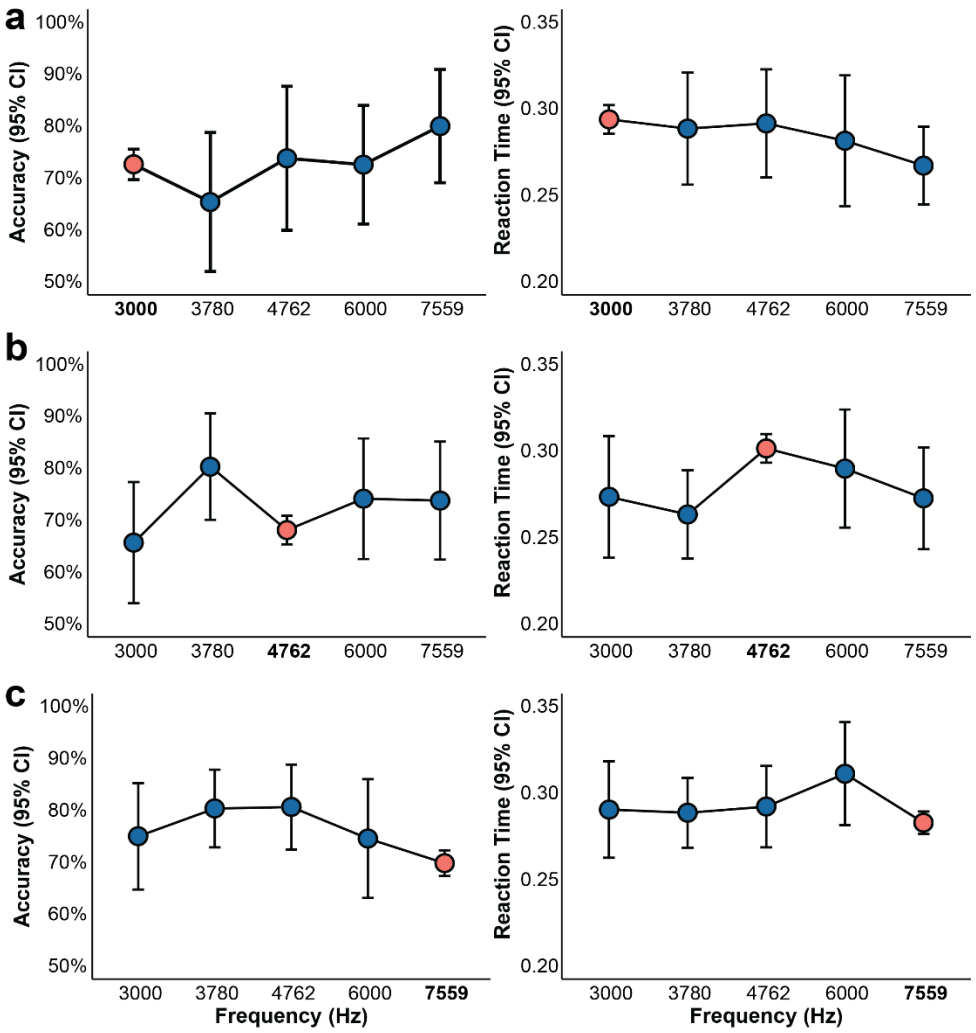

Figure 7. Group-level Accuracy and Reaction Time Across Probe Frequencies for Each Signal Condition. To determine whether accuracy and reaction time varied across frequencies within each signal condition, we conducted one-way ANOVAs treating frequency as a within-subjects factor. Based on the probe signal literature, we predicted higher accuracy and faster reaction times for signal trials relative to probe trials. However, most comparisons revealed no significant effect of frequency. Accuracy did not vary by frequency in the 3000 Hz or 4762 Hz signal conditions but showed a weak effect in the 7559 Hz condition ($\boldsymbol{F(4, 1680) = 2.68, p = .03}$) with post hoc comparisons failing to reach significance after Holm-Bonferroni correction. Reaction times showed a frequency effect only in the 4762 Hz condition ($\boldsymbol{F(4, 925) = 2.42, p = .047}$), but again, no pairwise comparisons were significant.

To more directly test the key contrast of interest, we additionally ran paired-sample t-tests comparing mean performance on signal versus probe trials, collapsed across probe frequencies within each signal condition. This complements the ANOVA by increasing statistical power and directly targeting the

signal-versus-probe hypothesis. Although the paired t-tests revealed no significant differences, the plots (Figure 8) suggest that individual ferrets showed divergent patterns. For instance, F2204 showed faster reaction times to probe trials in the 3000 Hz and 4762 Hz signal conditions, while F2205 was slower to respond to probes in the 7559 Hz condition. For accuracy, signal 4762 Hz and 7559 Hz conditions show higher hit rates for the unexpected probes than signals.

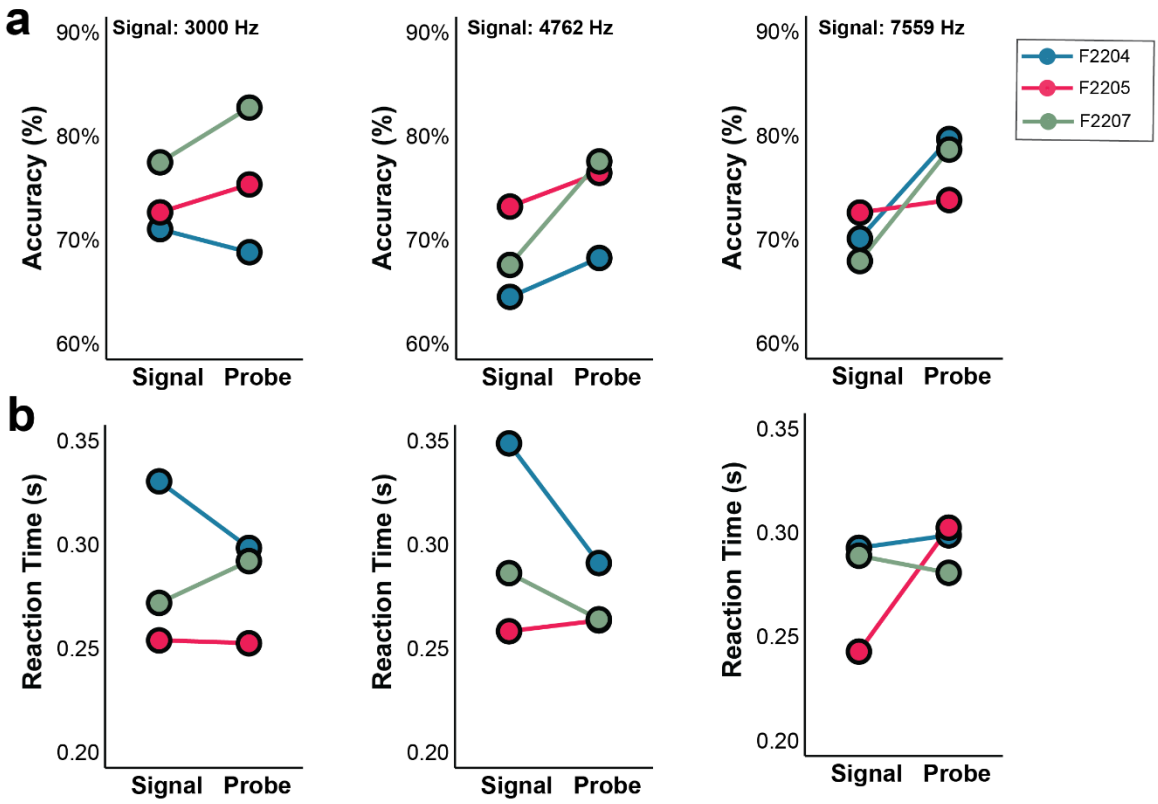

Figure 8. Performance Differences Between Signal and Probe Trials for Each Ferret Across the Three Signal Frequency Conditions. (a) Accuracy (percent correct) and **(b)** reaction times are shown separately for each signal frequency condition. Each coloured line represents an individual ferret, with markers corresponding to the mean performance per condition. Error bars were omitted, as paired t-tests were conducted on ferret-level means.

2.2.1.1 Motivating the Inclusion of *Time Bin* in the Reaction Time and Accuracy Models

Because tones occurred at one of three fixed onset times within each trial (0.65 s, 1.3 s, and 1.95 s), we tested whether response behaviour varied systematically across these time bins. This analysis served to evaluate whether latency effects from tone timing should be explicitly modelled in subsequent Bayesian analyses of both reaction time and accuracy.

Repeated-measures ANOVAs were conducted separately for accuracy and reaction time for each signal condition, with time bin as a within-subject factor and stimulus type (signal vs. probe) as a second factor. Greenhouse-Geisser corrections were applied where necessary.

Significant effects of time bin on reaction time were observed for the 3000 Hz and 4762 Hz signal conditions. In the 3000 Hz signal condition, tones presented later in the trial elicited faster responses, $F(2, 6) = 20.04, p = .039$; post hoc comparisons confirmed that responses were significantly faster at 1.95 s than at 0.65 s ($p = .008$). The 4762 Hz signal condition showed a similar pattern, $F(2, 6) = 23.44, p = .036$, with the same direction of effect (Figure 9). No such effects emerged for the 7559 Hz condition.

Similar results were seen for accuracy where all conditions had a main effect of time bin, with lower accuracy at later sound onset (e.g. 3000 Hz Signal Condition $F(2, 6) = 7.09, p = .026$). These findings supported the inclusion of sound onset time (bin) as a fixed effect in all subsequent Bayesian models of both reaction time and accuracy.

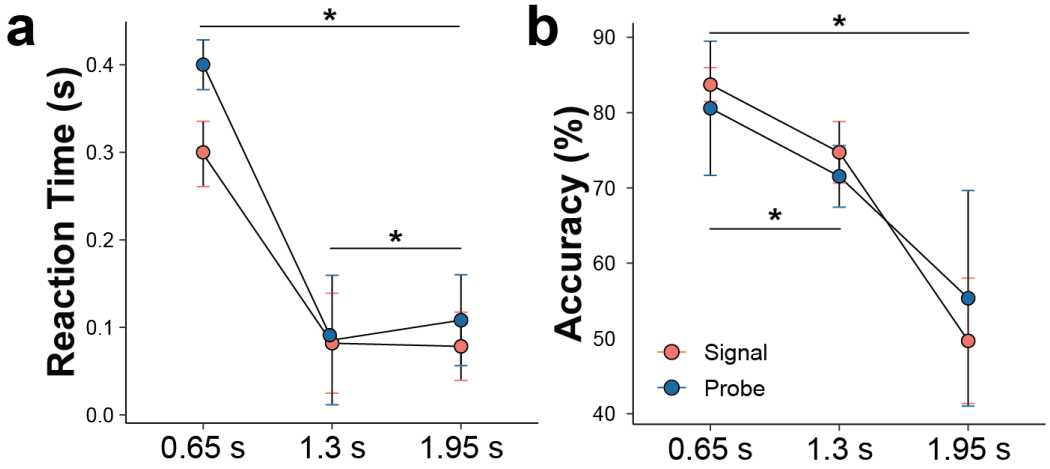

Figure 9. Reaction Times and Accuracy by Sound Onset Time in the 3000 Hz Signal Condition. Group-averaged reaction time **(a)** and accuracy (b) are shown for signal and probe tones across the three fixed sound onset times (0.65 s, 1.3 s, 1.95 s). Asterisks indicate significant main effects of time bin; pairwise comparisons confirmed faster responses and lower accuracy for later time bins. Error bars represent standard ±SEM.

In addition to subject-level ANOVAs and t-tests, we fitted generalised and linear mixed-effects models (GLME/LME) to evaluate frequency, time bin, and interaction effects on accuracy and reaction time. However, these models were often unstable or failed to converge, particularly when interaction terms were included. For example, frequency x bin interaction terms produced implausible coefficients with standard errors in the tens of millions, indicating overfitting or numerical instability. Reaction time models also could not be extended to include time bin effects due to convergence issues, limiting the scope of classical modelling.

Given the lack of consistent effects in the ANOVAs and t-tests, despite visible individual differences in performance, and the limitations of the mixed-effects models we attempted, a more flexible modelling approach was warranted. While GLME/LME models are generally well-suited for unbalanced datasets and trial-

level analyses, in this case their application was constrained by instability and convergence failures. These problems were particularly acute when modelling interaction terms or attempting to include multiple time-varying predictors.

We therefore turn to a Bayesian framework that retains the benefits of trial-level modelling while providing greater flexibility in model specification and estimation. It accommodates between-subject variability and unbalanced data without requiring large-sample approximations. Moreover, it yields posterior distributions for each parameter, allowing more interpretable statements about the probability and magnitude of effects than frequentist confidence intervals or p-values permit. This approach facilitates the detection of subtle effects that may be obscured by noise or diluted in subject-level averages.

2.2.2 Bayesian Modelling

Bayesian models were used to investigate subtle effects that might not have been captured by simpler statistical approaches. The primary aim was to determine whether a probe signal effect was present for both accuracy and reaction time across the three signal frequency conditions. Separate models were constructed for each measure and condition. This modelling strategy also allowed us to overcome the limitations encountered with classical GLME/LME approaches, including the exploration of predictors that could not be stably estimated otherwise.

Model Selection: Based on significant main effects of time bin in the ANOVAs, the final Bayesian models for both accuracy and reaction time included

fixed effects of group (signal vs. probe) and time bin (0.65 s, 1.3 s, or 1.95 s), as well as their interaction. We fitted Bayesian hierarchical models, using a generalised linear structure (logit-Bernoulli) for binary trial outcomes (correct/incorrect) and a linear Gaussian structure for continuous reaction times. Random intercepts for each ferret were included to account for individual differences. This structure captured both temporal and condition-dependent influences while maintaining model parsimony.

All models used 94% highest density intervals (HDI) for parameter estimation, with convergence assessed via effective sample sizes (ESS), and $\hat{R}$ values. Sampling was performed with 2000 posterior draws, a target acceptance rate of 0.99, and a maximum tree depth of 20. All models successfully converged.

Accuracy Model Equation

$$logit\big(P(Correct)\big) = {\gamma_0}^{(i)} + \beta_1 \cdot Group + \beta_2 \cdot Bin + \beta_3 \cdot Group\ x\ Bin$$

Where:

- $P(Correct)$: Probability of a correct response
- ${\gamma_0}^{(i)}$: Random intercept for each ferret (i)
- β_1: Fixed effect of stimulus type (Probe = 1, Signal = 0)
- β_2: Fixed effect of stimulus onset (0.65 s, 1.3 s, 1.95 s)
- β_3: Interaction effect

Reaction Time (reaction time) Model Equation

$$RT = {\gamma_0}^{(i)} + \beta_1 \cdot Group + \beta_2 \cdot \mathrm{Bin} + \beta_3 \cdot (\mathrm{Group\ x\ Bin}) + \in$$

Where:

- RT: Observed reaction time

- $\gamma_0{}^{(i)}$: Random intercept for each ferret (i)
- β_1: Fixed effect of stimulus type (probe vs signal) on reaction time.
- β_2: Fixed effect of time point (0.65 s, 1.3 s, or 1.95 s) on reaction time.
- β_3: Fixed effect of the interaction
- $\in$: normally distributed error term

2.2.2.1 Modelling Reveals Unexpected Accuracy Advantage for Probes

The Bayesian model for accuracy revealed a significant main effect of stimulus type, with higher accuracy for probe tones than for signal tones in the 4762 Hz signal condition. Accuracy also declined with increasing stimulus delay. The significant interaction at both 1.3 and 1.95 s showed that the difference between probe and signal accuracy also decreased over time (Figure 10). The 7559 Hz signal condition similarly had significant effects for stimulus type, and 1.95 s sound onset, as well as their interaction. This suggests that the probe advantage persisted across sound onset times, although the magnitude of the advantage decreased as overall accuracy declined. The 3000 Hz signal condition also showed significant main effects for sound onset, but no other effects or interactions. Full parameter estimates for the 4762 Hz signal modal are shown below; estimates for other conditions appear in the appendix.

Table 1. Parameter Estimates for Percent Correct Model (4762 Hz Signal). Asterisks denote fixed effects for which the 95% HDI excludes zero.

Parameter	Mean	SD	HDI 3%	HDI 97%	ESS Bulk	ESS Tail	$\widehat{R}$
Intercept	1.00	0.53	-0.19	2.08	1177	1293	1
Group*	1.03	0.24	0.57	1.51	1485	2279	1
Bin 1.3 s	-0.05	0.12	-0.29	0.19	2327	2548	1
Bin 1.95 s	-1.52	0.13	-1.24	-0.03	1724	2547	1
Group x Bin 1.3 s	-0.64	0.31	-1.24	-0.03	1724	2457	1
Group x Bin 1.95 s	-0.79	0.31	-1.41	-0.22	1691	2416	1

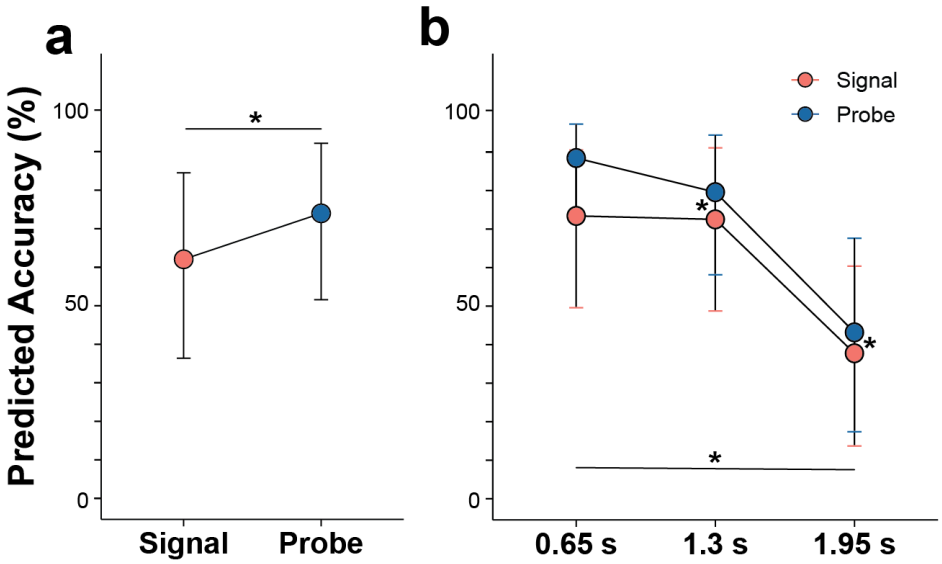

Figure 10. Modelled Accuracy for Signal Frequency 4762 Hz Condition. (a) Posterior estimates show that accuracy was significantly higher for probe tones compared to signal tones. **(b)** Predicted accuracy across time bins for signal and probe tones. Accuracy decreased as the delay between trial onset and tone presentation increased. A significant Group x Bin interaction was present at both time points, with a stronger accuracy advantage for probes than signals compared to the 0.65 s baseline. Horizontal significance bars indicate main effects; the vertical asterisks reflect the significant interaction. Error bars represent 95% highest density intervals (HDIs). Asterisks denote fixed effects for which the 95% HDI excluded zero.

2.2.2.2 Significant Reaction Time Model for CF 4762 Hz

The Bayesian model for reaction time revealed a significant negative effect for group (signal/probe) in the 4762 Hz signal condition, indicating that ferrets responded faster to probes than to signals. This result contrasts previous human studies, in which responses to probes are slower than to the expected signal frequency. To account for variability in stimulus onset timing, time bin was included as a fixed effect in all models. In the 4762 Hz signal condition, both the later 1.3 s and 1.95 s delays were associated with significantly faster responses relative to the 0.65 s reference bin, indicating that longer delays between trial onset and sound onset facilitated faster reaction times. A significant interaction at 1.3 s suggests that the probe reaction time advantage weakens at this intermediate time point relative to 0.65 s (Figure 11). The remaining models showed consistent main effects of time bin, with both 300 Hz and 7559 Hz conditions exhibiting faster responses for longer delays, which also aligns with the raw ANOVA findings (Section **Error! Reference source not found.**). The probe effect was also significant in the 7559 Hz condition but did not reach significance in the 3000 Hz condition. Full parameter estimates for the 4762 Hz condition are reported in Table 2.

Table 2. Parameter Estimates for Reaction Time Model (4762 Hz Signal). Asterisks indicate significant effects at $p < 0.05$)

Parameter	Mean	SD	HDI 3%	HDI 97%	ESS Bulk	ESS Tail	$\widehat{R}$
Intercept	0.41	0.07	0.26	0.55	1132	1098	1
Group*	-0.14	0.05	-0.24	-0.05	2035	2356	1
Bin 1.3s*	-0.25	0.03	-0.39	-0.28	2609	2694	1
Bin 1.95s*	-0.34	0.03	-0.39	-0.28	2530	2917	1
Group x Bin 1.3s*	0.16	0.07	0.04	0.30	2251	2604	1
Group x Bin 1.95s	0.11	0.07	-0.02	0.24	2073	2497	1

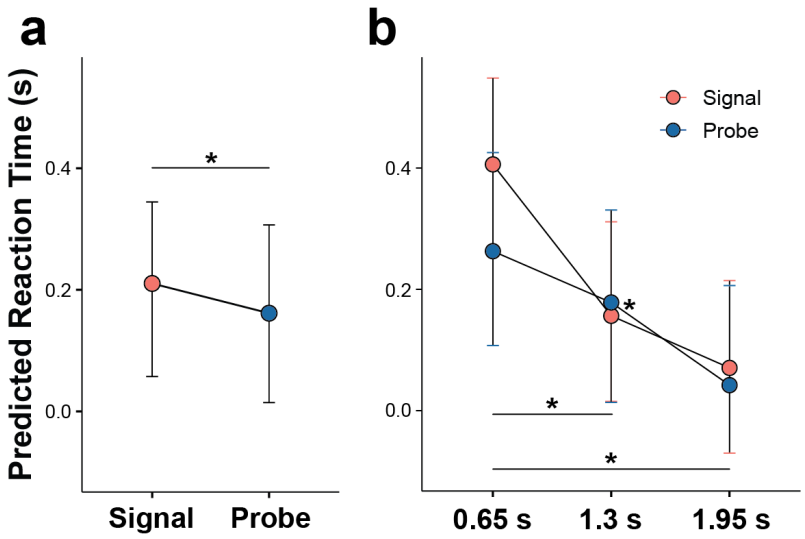

Figure 11. Modelled Predictions for Reaction Times for the 4762 Hz Condition. (a) Predicted reaction times for the main effect of Group (probe/signal) indicate faster responses to probe tones than to signal tones. **(b)** Predicted reaction times for group across time bins. A main effect of Bin was observed, with longer delays between trial onset and stimulus onset associated with faster reaction times (horizontal significance bars). A significant Group x Bin interaction emerged at 1.3 s, where the difference between signal and probe trials was most pronounced. Error bars represent 95% highest density intervals (HDIs). Asterisks indicate fixed effects with 95% HDIs that exclude zero.

The observed influence of time bin on both accuracy and reaction time indicates that ferrets may have adopted a timing-based strategy. Because the tones occurred at one of three fixed latencies (0.65 s, 1.3 s, or 1.95 s), the conditional probability of tone onset increased over time, producing a non-flat hazard rate. Reaction times were significantly faster for later tone onsets, consistent with growing temporal expectation. However, accuracy declined across time bins, suggesting a shift toward anticipatory responding rather than improved detection. This pattern implies that animals were not simply accumulating sensory evidence but instead relying on the rising probability of tone onset, at the expense of discriminative precision.

2.2.2.3 Changes within Session

To examine how reaction times evolved within sessions, we fitted linear mixed-effects (LME) models including trial number, group (signal vs. probe) and their interaction, with random intercepts for ferret and session. This model structure allowed us to directly test whether the difference in reaction times between trial type changed as a function of trial progression. That is, whether a bias developed or shifted over the course of the session. LME was chosen over a Bayesian model here because the focus was on assessing a linear relationship with trial number across individual data points, where model convergence and interpretability were more straightforward in a frequentist framework. Moreover, trial number was treated as a continuous predictor, making LME well-suited for quantifying gradual within-session changes while accounting for repeated measures at the ferret and session level.

For the 3000 Hz signal condition, a significant interaction ($p = 0.043$) indicated that responses to signal trials became faster over the course of the session while responses to probe trials became slower, suggesting that the expectation of the signal frequency may build slowly over the course of the session (Figure 12). A similar effect was observed for the 4762 Hz signal condition ($p = 0.049$), while no interaction was found for the 7559 Hz condition.

The same analysis was run for trial-level responses using generalised linear mixed-effects models (GLME) with a binomial function. While reaction time is a continuous variable suitable for standard LMEs, trial-by-trial responses are binary (correct vs. incorrect). A binomial model, classified as a “generalised” LME,

is therefore required to appropriately capture the distribution and structure of the data. While all models converged, no interactions reached significance.

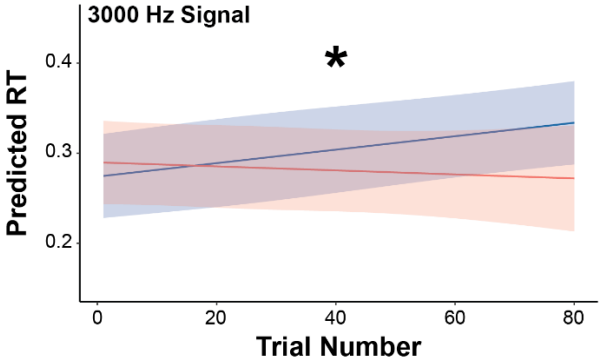

Figure 12. Linear Mixed-Effects Model Estimates of Reaction Time Across Trial Number. Estimated marginal means of reaction time are shown across trials for the 3000 Hz signal frequency condition. Lines represent predicted reaction time for signal (red) and probe (blue) tones, with shaded areas indicating 95% confidence intervals. Asterisks denote significant interaction terms ($p < .05$).

2.3 HUMAN METHODS

Our ferret findings revealed the opposite pattern to the probe signal effect, namely, improved performance on probe rather than signal trials, which contradicts the human literature (e.g. Dai & Buus, 1991; Greenberg & Larkin, 1968; Luthra et al., 2024). To clarify this discrepancy, we developed a human paradigm that more closely mirrors the ferret go/no-go task than previous human studies, which have typically employed two-alternative forced choice (2AFC) procedures. The goal was to determine whether task structure alone could account for differences in attentional filtering between species. Prior work has shown that performance thresholds in a pitch discrimination task vary significantly across task types (2AFC vs. go/no-go) in ferrets but not in humans (Walker et al., 2017). Based on this, we expected humans to exhibit the typical probe signal effect even under go/no-go conditions. This comparison allowed us to determine whether the absence of a probe cost in ferrets reflects species-specific attentional strategies or simply the influence of task design.

2.3.1 Participants

Twenty adults (7 female, 13 male) were recruited through Prolific (www.prolific.com). Participants self-reported fluency in English and normal hearing, no history of mental health problems, neurodevelopmental disorders, head injuries, or chronic health conditions. The sample was diverse in terms of nationality and residence, including individuals from South Africa (n = 3), Spain (n = 1), the Netherlands (n = 1), Portugal (n = 2), Hungary (n = 1), the United

Kingdom (n = 4) and the United States (n = 8). Ethnic representation included participants who identified as White (n = 9), Black (n = 8), Mixed (n = 1) and Asian (n = 2). Due to a known technical error on Pavlovia (Open Science Tools, Nottingham, UK), the site hosting the experiment, data from two participants could not be included in the final analysis.

2.3.2 Stimuli

The task was made in PsychoJS, a tool by Pavlovia Surveys (Open Science Tools, Nottingham, UK) using adapted code by Zhao et al. (2022), Milne et al. (2021), and custom in-house JavaScript code. Data collection was run online, hosted by Pavlovia. It consisted of three main components: a headphone check (Milne et al., 2021), an adaptive staircase (based on Zhao et al., 2022 but adapted to fit task needs), and a probe signal tone-in-noise detection task.

Using Praat 6.4.04 (Boersma & Weenink, 2024), a Gaussian white noise presented diotically (stereo) with a band-pass filter between 80–8000 Hz and an RMS amplitude of 0.000399 (26 dB SPL) was generated. This followed the same procedure as described in Zhao et al. (2022) and the noise was played throughout the experiment. Pure tone stimuli were taken from the same paper as well and exported as flac files at frequencies 800, 920, 1000, 1080, and 1200 Hz.

2.3.3 Experimental Design

The goal of this human task was to replicate the ferret tone-in-noise task as closely as possible. As such, the basic experimental go/no-go design was consistent between species. In our version, participants were asked to hold the spacebar to start a trial and to release once they heard a sound. This is similar to the ferrets who were trained to hold their snout in a nose poke hole and to release when they heard the sound.

At the beginning of the task, participants were instructed to set their volume to 50%, then hit a “play” button to start continuous noise pulses. Participants were then told to reduce their volume until the noise was no longer audible, then increase the volume slowly until the noise was barely audible. The same volume of noise was maintained throughout the experiment.

To maintain consistency across online participants, it was important that they all listened to the stimuli over headphones, rather than through free-field computer speakers that may contain more environmental noise and acoustical cues (e.g. room reverberation, head-related localisation cues). In this study, participants’ use of headphones was verified through a dichotic Huggins Pitch test, as developed by Milne et al. (2021) and later adapted for online deployment on Pavlovia and PsychoJS by Sijia Zhao (available at: https://run.pavlovia.org/sijiazhao/headphones-check/). This auditory check consisted of six trials; each trial included three intervals of one-second white noise. Within these trials, one randomly selected interval was modified to include a Huggins Pitch stimulus, as defined by Cramer & Huggins (1958). For this stimulus, standard white noise was directed to the left ear, while phase-shifted

white noise was delivered to the right ear over a narrowly defined frequency range centred at 600 Hz (±6%), creating a perceivable Huggins Pitch.

On-screen instructions informed participants that they would be presented with three distinct sound intervals, and needed to identify which interval included a subtle tone. To proceed to the main experiment, participants were required to correctly identify the tone in all six trials. If unsuccessful on their initial attempt, they were granted one additional opportunity to achieve the required accuracy before being excluded from further participation. The headphone verification process lasted around three minutes.

2.3.3.1 Adaptive Staircase Threshold Setting Procedure

The process for the adaptive staircase followed the same principles as the ferret version where participants had to get 3 consecutive trials correct to lower the dB level, while one incorrect response was enough to move up one dB level (Levitt, 1971). The code was adapted from Zhao et al. (2022) with a few notable differences: (1) instead of a three-alternative forced-choice task, we designed a go/no-go version where participants held the space bar to trigger a trial and released when they heard a tone, (2) from staircase level 21 to level 13, a faster descent of -3 steps was employed to ensure participants were able to reach threshold during each of the three staircase blocks (40 trials each), (3) feedback was given after each trial, (4) time outs of 5 s were employed whenever a trial was incorrect to mimic the design of the ferret task, and (5) the progress bar was removed.

Continuous broadband noise was set to approximately 66 dB SPLA. Each block began with a signal-to-noise ratio (SNR) of –13.75 dB. One step (whether increase or decrease) was equal to 0.75 dB and the maximum SNR that could be reached was –28.75 dB. Participants had the opportunity to rest between blocks. In contrast to the ferret task, participants only heard 1 kHz pure tones during the thresholding process as Zhao et al. (2022) showed that this was sufficient in estimating human thresholds for a tone in noise detection task.

2.3.3.2 Tone in Noise Detection Task

Upon completion of the adaptive staircase, participants moved to the main experiment, which consisted of a classic probe-signal task (Greenberg & Larkin, 1968) comprising 12 blocks of 32 trials each (384 trials total). Participants triggered each trial with a space bar press, matching the self-paced structure used in the ferret paradigm, and could therefore take breaks either between blocks or between individual trials. Tones could occur at one of three equally probable time points (0.65 s, 1.3 s, or 1.95 s after trial onset; Figure 13), identical to the timing used in the ferret paradigm. Tone duration and the response window were also matched across species to enable direct comparison of temporal attention dynamics. Each block contained an 80/20 split between signal and probe trials, with feedback after every trial and time-outs for incorrect responses. The only difference from the ferret implementation was the specific frequencies used: the signal was fixed at 1000 Hz, with probe tones drawn from 800, 920, 1080, and 1200 Hz, matching those used in prior human probe signal studies and allowing

direct comparison with the existing literature. Participants were also informed of their accuracy after each block.

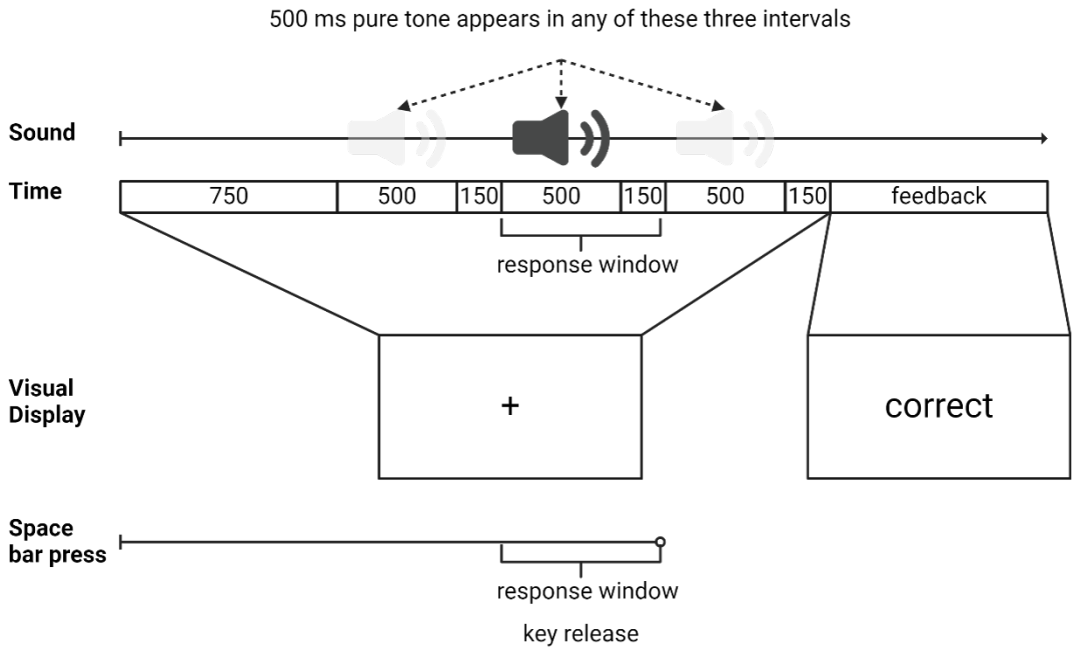

Figure 13. Trial Structure for Main Task. Participants triggered a trial by pressing the space bar, then hold until they hear a tone.

2.4 HUMAN RESULTS

2.4.1 Accuracy

Figure 14 shows the average accuracy across the five frequencies tested in the human go/no-go task. Performance was highest at the expected signal frequency (1000 Hz), as hypothesised, and declined for both closer and more distant probe tones, with the lowest accuracy observed at 1200 Hz. A one-way ANOVA confirmed a significant effect of frequency on percent correct, $F(4,432) = 25.93$, $p < .001$. Tukey's HSD revealed that, relative to the signal frequency (1kHz), performance was significantly worse for 1200 Hz ($M = 66.85\%$) than for the signal frequency ($M = 80.23\%$, $p < .001$). This suggests that while the performance was highest for the expected signal frequency, the decline in accuracy was not uniform across probe tones, with only the highest probe frequency (1200 Hz) differing significantly from the signal.

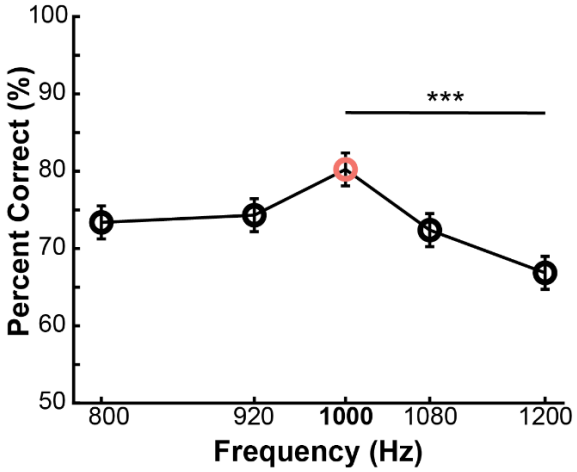

Figure 14. Accuracy Across Frequencies in the Human Probe Signal Task. Accuracy (percent correct) is plotted as a function of tone frequency. The expected signal frequency (1000 Hz) is highlighted in salmon. Performance was significantly better at 1000 Hz compared to 1200 Hz. Error bars represent the mean ±SEM. Asterisks indicate significant effects at $p < 0.05$

2.4.2 Reaction Time

Figure 15 shows average reaction times across frequencies. The fastest responses were observed for the most distant probe tones (800 and 1200 Hz), while responses were slower for frequencies closer to the signal frequency. This pattern is opposite of what was hypothesised based on the probe signal effect and contrasts with findings from previous studies using 2AFC paradigms. It does, however, align with the ferret results.

A one-way ANOVA revealed a significant effect of frequency on reaction times $F(4, 3421) = 4.75, p < .001$. Tukey's HSD indicated that responses to 1200 Hz ($M = 0.37$ s) were significantly faster than those to the signal frequency ($M = 0.4$ s, $p = 0.0123$). No other pairwise comparisons to the signal yielded significant results.

These findings indicate that the fastest responses occurred for the highest probe frequency (1200 Hz), which also elicited the largest accuracy drop, suggesting a potential speed-accuracy trade-off.

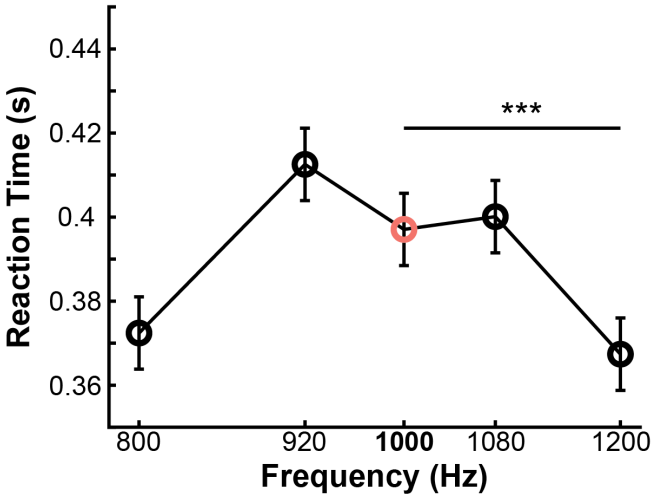

Figure 15. Mean Reaction Times as a Function of Tone Frequency (Mean ± SEM). Reaction times were significantly faster for the most distant probe (1200 Hz) compared to the expected signal tone, suggesting an inverted probe signal effect. Asterisks indicate significant effects at $p < 0.05$.

2.4.3 Time Bin Analyses

As there were significant differences in performance across time bins in the ferrets, we analysed time bins for the humans too using repeated-measures ANOVAs.

Accuracy

Figure 16 shows that performance declined across time bins for both tone probability types. However, probe trials appear to exhibit a steeper decline than signal trials. The difference between conditions was smallest at 0.65 s ($\Delta M = 0.8\%$) but widened substantially at 1.3 s ($\Delta M = 8.6\%$) and 1.95 s ($\Delta M = 7.8\%$), suggesting that the probe signal effect may be more pronounced at later time points.

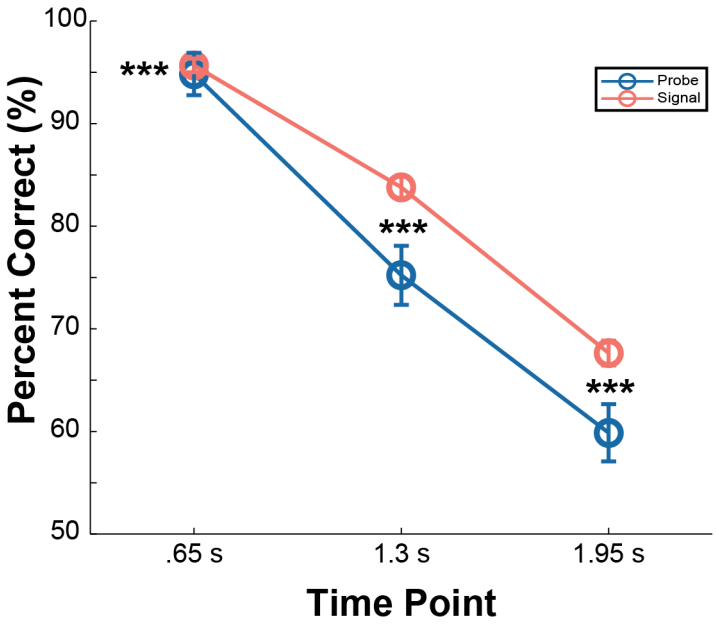

Figure 16. Percent Correct Performance Across Time Bins for Signal and Probe Trials (Mean ±SEM). Group-averaged accuracy is plotted across three time bins (0.65, 1.3, and 1.95 s) for signal (salmon) and probe (blue) trials. Asterisks denote significant differences ($p < .001$) between probe and signal at a given time point.

A repeated-measures ANOVA was conducted to evaluate the effect of time point (0.65, 1.3, 1.95s) and tone probability (probe, signal) on accuracy. The main effect of tone probability type was significant, $F(1, 14) = 81.14, p < .01$, with signal tones consistently eliciting higher accuracy than probe tones. The main effect of time point was not significant, nor was the interaction between time point and tone probability. Post hoc comparisons within each time point confirmed that accuracy for signal tones was significantly higher at each time point. While the difference was statistically significant even at the earliest time, the separation became much larger for the two later time points, suggesting that the advantage for signal trials increased with time between trial onset and stimulus onset.

Reaction Time

Figure 17 displays average reaction times for probe and signal tones across the three time bins. Consistent with previous findings in ferrets, participants responded more quickly the longer they had to wait before hearing the tone. Reaction times decreased steadily. In addition, while reaction times were similar for probe and signal tones at the first time point, they began to separate by the second and were most distinct by the third time point.

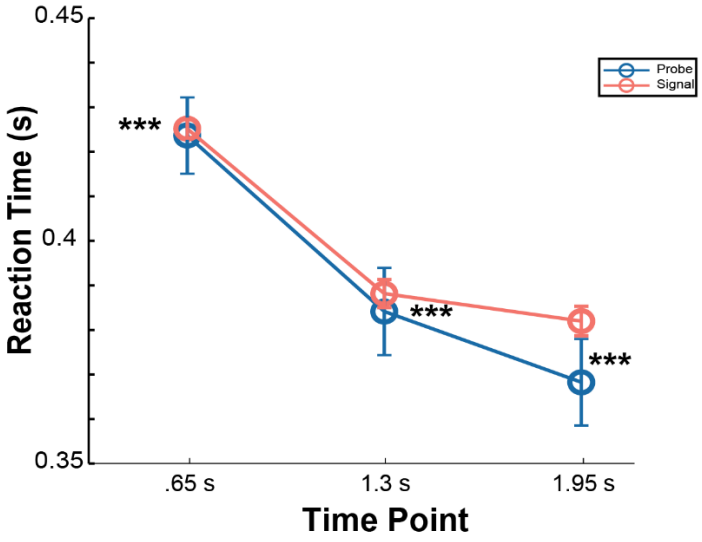

Figure 17. Reaction Time Across Three Onset Time Points for Probe and Signal Tones. A repeated-measures ANOVA showed a significant main effect of tone probability type ($\boldsymbol{F(1,12) = 116.60, p < .001}$) indicating that participants responded more quickly to probe tones than to signal tones, as previously shown. Neither the main effect of time bin, nor the interaction reached significance. Post hoc comparisons confirmed that the probe advantage was significant at all three time points. Asterisks indicate significant effects at $p < 0.05$.

Although not all main effects reached significance, the behavioural trends observed in both accuracy and reaction time suggest that accuracy declines, while reaction time improves the longer the gap between trial and sound onset was. Importantly, the advantage for signal tones in accuracy and the advantage of probe tones in reaction time remained stable or appeared to widen with increasing delay between trial onset and tone presentation. These patterns, while not always statistically significant, point to systematic modulations that may become clearer with a larger sample size or more sensitive modelling approaches.

2.4.4 Comparison of Early Ferret Behaviour with Human Probe Signal Effects

Unlike humans, who completed only a single testing session, ferrets underwent repeated exposure to the task over many sessions, performing two sessions per day across several weeks. This difference in task exposure raises the possibility that the divergent behavioural patterns observed in humans and ferrets may be partly driven by the amount of training or task familiarity. To assess whether ferret behaviour more closely resembled the human pattern in the first sessions after training, we analysed accuracy and reaction time data from the first few sessions of each ferret in each condition.

An illustrative example from Ferret F2207 in the 7559 Hz signal condition is shown in Figure 18. In the first two sessions, accuracy was higher for signal tones, consistent with the human results and the canonical probe signal effect described in the psychophysical literature. From session three onward, this trend reverses, with probe tone accuracy surpassing that of signal tones. A similar pattern is observed in reaction time: during the first two sessions, responses were faster for signal tones, consistent with the expected probe signal effect, though this contrasts with our own human dataset in which probes elicited faster reaction times than signals. From the third session onward, the ferret responded faster to probe tones. Session-wise plots for all ferrets and conditions are shown in Appendix 1 7.1.2.

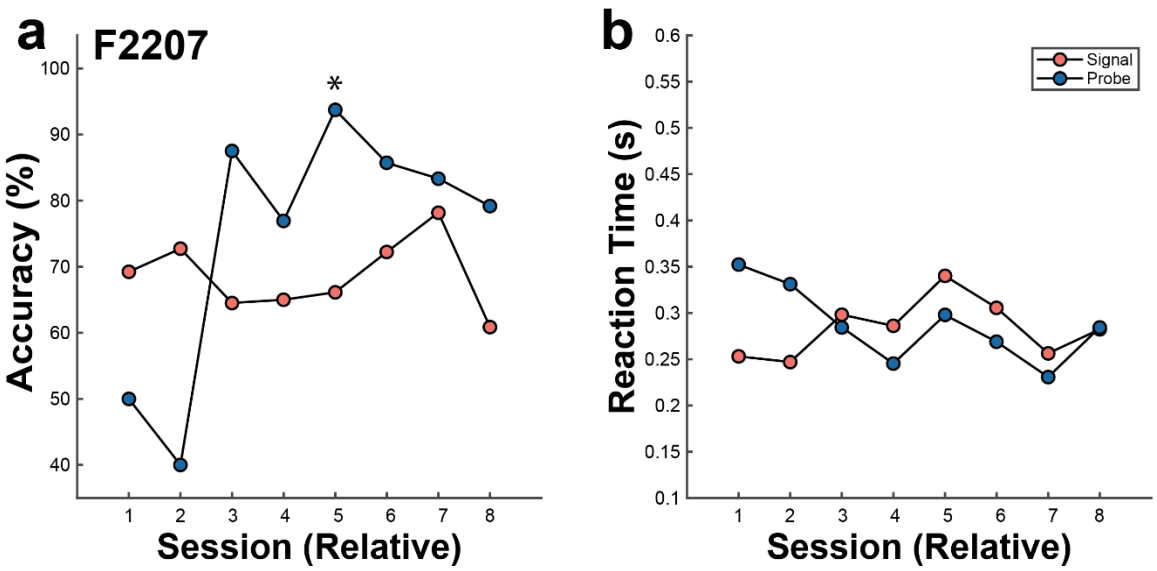

Figure 18. Early Session Performance for Ferret F2207 in the 7559 Hz Signal Condition. (a) Accuracy and **(b)** reaction time are shown across the first eight sessions (relative to the first session for this condition). Accuracy and RT patterns initially favoured signal tones, before reversing in later sessions. Session-wise comparisons were performed to evaluate early probe signal effects. Asterisks represent a significant pairwise comparison.

It is worth noting that unlike the human participants, who completed a fixed number of trials, ferrets were free to terminate a session at any point by disengaging with the task (i.e. not triggering new trials). As a result, the number of trials contributing to each session varied substantially across animals and conditions, reducing the statistical power of session-wise comparisons and potentially explaining the absence of significant effects in early sessions. Nonetheless, the patterns observed in some early sessions suggest that the probe signal effect observed in the human literature may also emerge in ferrets at the onset of task exposure, before reversing with continued training.

2.4.5 Threshold Comparison to Zhao et al. (2022)

Our experiment used a staircase procedure from Zhao et al. (2022), but customised their code to be a go/no-go task where participants needed to press and hold the space bar until they heard the tone. The original version asked participants to report at which of three time points a tone occurred (see Figure 19 for visualisation). Figure 20 shows histograms from the present study (Figure 20b) and the original staircase procedure (Figure 20a). While both versions used the same sound files, SNR performance was worse for the go/no-go than for the 3AFC, suggesting that holding the space bar and releasing immediately during sound presentation is more difficult than waiting until all three options have been presented until making a decision.

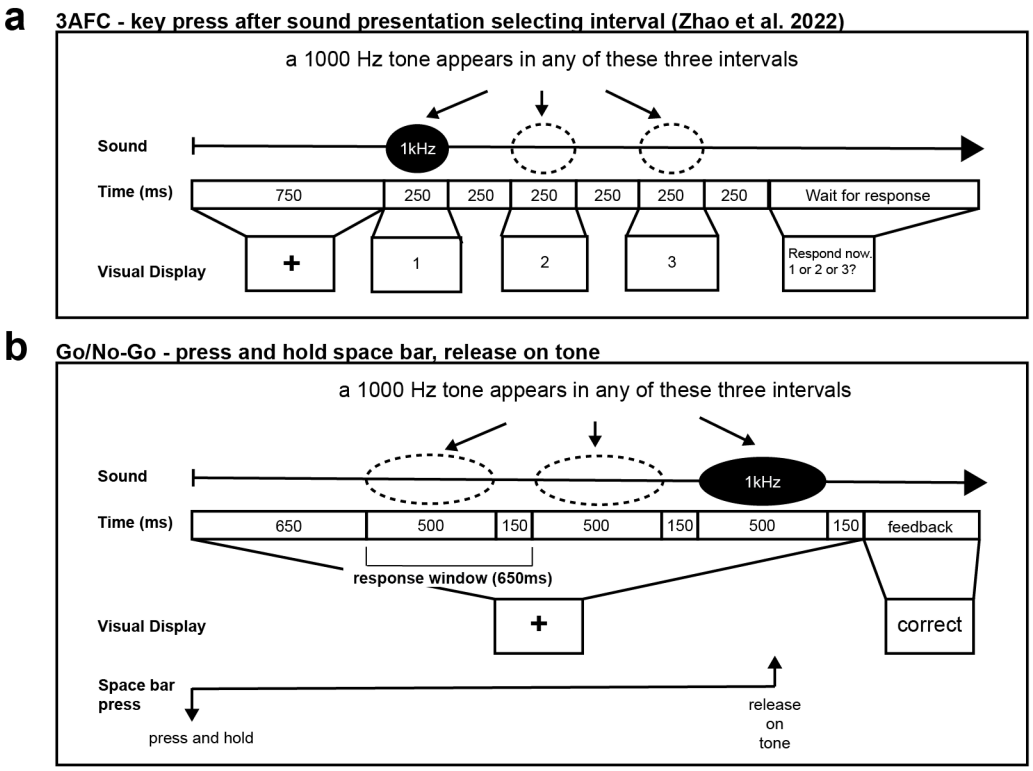

Figure 19. Comparison of Adaptive Staircase Trial Structure. (a) The original task structure from Zhao et al. (2022) where a tone could occur in 1 of 3 time intervals. After all time intervals were played, participants were asked to indicate with keys in which of the intervals they heard the sound. **(b)** Our task structure where participants needed to release the spacebar to indicate when they heard the tone, which would end the trial followed by feedback. Sound presentation was also double as long (500 ms instead of 250 ms) and ITIs were 150 ms. A response window was 650 ms (sound presentation + ITI). Both task versions played the same continuous white noise throughout. Note that contrary to the ferrets, only the common frequency (1kHz) was used to determine the threshold, which was then applied to all frequencies.

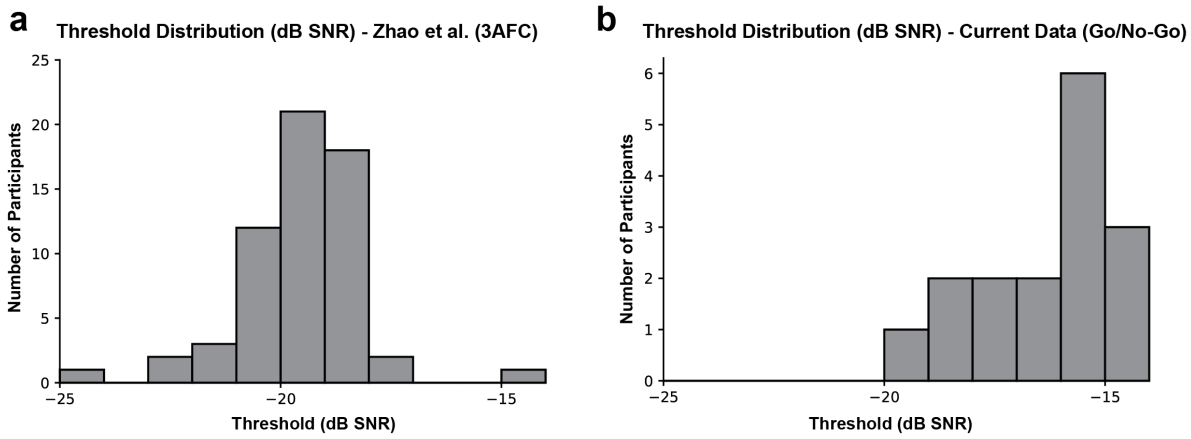

Figure 20. Comparison of Participant Thresholds Between Our Task and Zhao et al. (2022). **(a)** Participant thresholds from Zhao et al. (reproduced with permission and altered to align with the format of this chapter). The majority of participants reached a greater SNR between -19 and -20, a 5dB difference to the results of panel (b) and showed more of a normal distribution. **(b)** Participant thresholds from our go/no-go task with most participants reaching an SNR between -15 and -16. Thresholds were determined the same way in both cases, by taken the mode of each block and averaging them. Both versions used the same sound files, the only difference was the response mechanism (3AFC vs. go/no-go space bar hold and release).

2.5 DISCUSSION

Performance on the tone-in-noise detection task revealed a divergence from classical predictions of expectation-driven models of auditory attention. In humans, accuracy followed the traditional probe signal pattern, with higher correct detection rates for frequent signal tones relative to rare probes. However, reaction times were paradoxically faster for low-probability probe tones, a dissociation from the original behavioural reports (Dai & Buus, 1991; Greenberg & Larkin, 1968). Ferrets similarly violated expectation-based predictions, exhibiting superior detection accuracy and faster responses for probes compared to signals. These findings challenge the assumption that probabilistic structure alone governs attentional deployment. Instead, under certain task conditions, unexpected stimuli may preferentially capture attention due to their behavioural salience. Within the predictive coding framework, such salience effects may reflect precision-weighted prediction errors. Predictive coding posits that the brain continuously generates top-down predictions of sensory input and compares them to incoming signals, with mismatches, termed prediction errors, driving updates to perceptual representations and guiding behavioural responses (Clark, 2013; Friston, 2005; Rao & Ballard, 1999). When these errors are assigned high precision, such as for behaviourally relevant deviations, they exert stronger influence on downstream perceptual inference and motor responses (Auksztulewicz & Friston, 2015; Feldman & Friston, 2010; Heilbron & Chait, 2018).

The behavioural enhancement for rare probes in ferrets provides a basis for evaluating whether the underlying mechanism aligns more closely with stimulus-

specific adaptation (SSA) or mismatch negativity (MMN). Both have been proposed to explain neural sensitivity to deviant auditory events. SSA reflects reduced responsiveness to repeated tones, typically under passive conditions and without requiring engagement (Antunes et al., 2010; Ulanovsky et al., 2003). MMN, in contrast, is thought to reflect the detection of prediction violations within an internal sensory model and is considered a signature of predictive coding mechanisms operating in hierarchically organised auditory circuits (Friston, 2005; Garrido et al., 2009; Heilbron & Chait, 2018; Nelken & Ulanovsky, 2007; Winkler et al., 1996). In the present study, the control condition with equiprobable tones yielded uniformly high accuracy across frequencies, ruling out inherent differences in detectability. Thus, the probe advantage observed under structured conditions likely reflects the formation of sensory expectations and their violation. This pattern aligns more closely with MMN than SSA, particularly given that SSA, although also observed in auditory cortex, is typically characterised as an automatic response to stimulus regularity and can occur even in anaesthetised animals (Malmierca et al., 2009). In contrast, MMN is associated with violations of sensory predictions and is often modulated by cognitive factors such as attention, task demands, and learning context (Näätänen et al., 2007; Nelken & Ulanovsky, 2007). The sustained vigilance and motor demands of the current task provide a context in which predictive coding mechanisms are thought to be especially influential (Feldman & Friston, 2010). Although neural activity was not recorded, the behavioural profile suggests that ferrets engaged prediction-based computations consistent with MMN and precision-weighted coding.

While this interpretation favours MMN as the more plausible mechanism, it is important to note that SSA and MMN are not mutually exclusive. Several authors have argued that SSA may reflect a cellular-level process that contributes to or underlies MMN (Farley et al., 2010; Nelken & Ulanovsky, 2007). However, MMN encompasses additional cognitive and contextual influences, including abstract rule violations and attentional modulation, which SSA alone cannot explain. Although neural recordings were not obtained, the behavioural profile is consistent with prediction-based computations commonly attributed to MMN. However, given the likely overlap in underlying mechanisms, it is plausible that both SSA- and MMN-like processes contribute to deviance detection in this context.

In humans, the pattern was more complex. Accuracy remained highest for frequent signal tones, but reaction times were consistently faster for probes. This behavioural dissociation suggests that multiple mechanisms operate concurrently during auditory detection, with expectation-driven filtering influencing accuracy and saliency-driven orienting modulating response speed. Dual-process accounts have described such interactions in contexts requiring sustained vigilance under uncertainty (Adam & De Bettencourt, 2019; Esterman et al., 2013). Fluctuations in attentional state may transiently enhance sensitivity to deviant events, allowing faster responses on some probe trials despite a broader decision bias toward signal tones.

Task demands may also shape the relative contribution of these attentional mechanisms. Escera et al. (2002) demonstrated that deviant auditory stimuli speeded reaction times without impairing accuracy in a visual 2AFC task, whereas Luthra et al. (2024) observed a classical probe signal effect in accuracy but no RT

differences in a 2AFC tone-in-noise task using identical auditory stimuli to the present study. The go/no-go design employed here differs fundamentally from 2AFC paradigms by requiring continuous monitoring and immediate motor output, rather than comparative judgment. These demands may preferentially amplify saliency-driven orienting while attenuating expectation-based filtering, providing a plausible account for the reaction time divergence observed across studies.

Reaction times decreased while accuracy declined as the interval between trial onset and tone onset increased, a pattern consistent with prior work on temporal expectation (Seibold et al., 2023). Longer pre-stimulus intervals likely facilitated anticipatory motor preparation, allowing faster responses once the tone appeared (Devine et al., 2019). However, these gains in speed came at the cost of reduced detection accuracy, consistent with at least a partial reliance on internal timing cues over stimulus-evoked sensory evidence. This trade-off may reflect a shift, in part, toward a timing-based strategy that prioritised response readiness over perceptual precision.

Reaction times also changed systematically within sessions. As sessions progressed, ferrets responded increasingly quickly to signal tones and more slowly to probes, resulting in a crossover pattern. This divergence may reflect a gradual reallocation of attention, with signal tones attracting greater priority over time. Whereas humans completed the task in a fixed 1-hour session, ferrets performed fewer trials per session and self-terminated after approximately 15–25 minutes. Canonical probe signal effects might have emerged in ferrets with longer sessions, but voluntary trial initiation imposes limits on task duration. Future work

could explore whether adjusting reward volume or motivational structure extends engagement sufficiently to reveal later-emerging expectation effects.

Thresholding procedures differed across species and may have contributed to performance asymmetries. In the human task, thresholds were calibrated only for the signal frequency (1000 Hz) and then applied uniformly to all probe tones. Although consistent with prior studies (Luthra et al., 2024; Zhao et al., 2022), this approach assumes equal detectability across frequencies, an assumption challenged by the sharply sloping human audiogram around 1000 Hz (Heffner & Heffner, 2007). Frequencies above 1000 Hz may have been more easily detectable, while lower ones remained closer to threshold, potentially introducing salience differences that influenced attentional allocation. In contrast, ferret thresholds were estimated separately for each frequency, ensuring comparable difficulty across tones. These audiogram-aligned thresholds (Kelly et al., 1986), may have enabled more balanced detection performance, particularly for rare probes. As such, classical probe signal effects in human studies may partly reflect threshold calibration asymmetries rather than purely expectation-based filtering.

Further support for the role of training history in shaping attentional deployment comes from the observation that ferrets displayed a canonical probe signal effect in initial sessions, with higher accuracy and faster reaction times to frequent signal tones than to rare probes. This early human-like pattern was gradually replaced by an advantage for probe tones in both accuracy and speed, indicating a shift in strategy. One possibility is that predictive filtering governed performance early in training, but with extended exposure, this gave way to saliency-based orienting as the frequent signal lost behavioural relevance. Such a

transition may reflect a reweighting of prediction errors within a predictive coding framework, as overlearned stimuli become less informative. These dynamics suggest that the divergent performance profiles observed in humans and ferrets may stem more from differential training exposure, rather than from fundamental differences in attentional architecture. Extending training in humans or expanding the stimulus set in ferrets may therefore alter the balance between prediction-guided and saliency-driven attention.

Cross-species comparisons further highlight the need for caution when interpreting differences in attentional deployment. Walker et al. (2017) reported that ferrets performed worse on 2AFC pitch discrimination tasks compared to go/no-go formats, illustrating that task structure can systematically affect behavioural outcomes even within the same species. A similar influence of task format was evident in the present study. Human participants using a go/no-go design exhibited higher accuracy than those tested under a 2AFC format in Luthra et al. (2024), despite identical stimuli. Moreover, while Luthra et al. (2024) observed a classic probe signal effect in accuracy, our go/no-go task revealed a dissociation between accuracy and reaction time, suggesting that task demands, rather than cognitive architecture alone, shaped attentional allocation. This divergence between studies underscores that even subtle differences in task design can modulate whether predictive filtering or saliency-driven orienting dominates. In the ferrets, the go/no-go format similarly favoured enhanced detection of unexpected stimuli, consistent with heightened saliency sensitivity under continuous vigilance demands (Clayton et al., 2021; Frederick et al., 2011). These observations caution against attributing behavioural differences solely to

species. Instead, they highlight the need for carefully matched task structures and motivational contexts when drawing cross-species comparisons.

While the cross-species approach and parallel task structures represent a major strength of this study, some limitations should be acknowledged. Differences in thresholding procedures, motivational context, and training histories between species were not fully equated and may have influenced attentional deployment. Future work should systematically control these factors to distinguish between intrinsic cognitive strategies and adaptations driven by task design or context.

These findings illustrate that the expression of expectation-driven and saliency-driven attentional strategies is not intrinsic to a species or task domain but is dynamically shaped by the interaction of task demands, stimulus timing, motivational context, and training history. The contrast between go/no-go and 2AFC formats reinforces that attentional strategies are shaped not only by stimulus features but also by the temporal and decisional demands of the task.

Future research should determine the conditions under which expectation-based and saliency-driven strategies dominate, and whether these reflect discrete transitions or continuous interactions. It is also essential to establish whether these dynamics differ systematically between ferrets and humans. Clarifying such mechanisms will allow behavioural paradigms to isolate genuine species differences rather than artefacts of task design or motivation. This, in turn, will strengthen the case for the ferret as a model system for probing the neural basis of auditory selective attention at the single-cell level. Such a model is urgently needed, given that selective attention declines with age and is impaired in

neurological and neurodegenerative conditions (Dobreva et al., 2011; Glyde et al., 2011; Hasanzadeh Pashang et al., 2021; Mayer et al., 2012; Ouchi et al., 2015). Without it, progress toward understanding and treating these deficits will remain limited.

3 PROBE SIGNAL: DURATION DISCRIMINATION

Auditory selective attention is a skill we rely on daily, often without realising it. For instance, we may prioritise the podcast we are listening to over the conversations of others on public transport, or we may focus on our friend's voice amidst the busy background of a coffee shop. Both humans and nonhuman animals must identify and attend to behaviourally relevant auditory stimuli, while filtering out irrelevant noise, to effectively navigate and respond to their environments (Bizley & Cohen, 2013; Woods & McDermott, 2018).

Statistics-driven attention, specifically, describes how expectation can modulate attention and influence auditory perception (Greenberg & Larkin, 1968). Probe signal tasks were designed to test precisely this type of attention. The classical task requires participants to detect tones against background noise, which they come to expect to be a certain frequency (the "signal"). However, on a smaller percentage of trials, this expectation is violated and other, typically adjacent, frequencies are presented instead ("probes"). Due to this violation of expectation, humans are slower to respond to and worse at detecting unexpected probes compared to the expected signal (Greenberg & Larkin, 1968). This probe signal effect is assumed to be due to an attentional filter being formed around the expected frequency. Unexpected probe frequencies outside this filter are not inherently attended to leading to slower response times and reduced detection

accuracy that characterise this effect (Dai & Buus, 1991; Greenberg & Larkin, 1968; Luthra et al., 2024; Wright & Dai, 1994).

Mondor & Bregman (1994) developed a probe signal task where the statistical manipulation of stimulus probability was applied in a sound dimension (tone frequency) that is orthogonal to the task (sound duration). Rather than the traditional probe signal methods of Greenberg & Larkin (1968), they used a duration judgement task designed to be independent of frequency (Allan & Kristofferson, 1974; Woods et al., 1979). In this two-alternative forced-choice (2AFC) task, participants first heard a cue tone followed by either a short or long tone and were asked to report the sound duration as “short” or “long” with a key press. The statistical manipulation targeted frequency rather than the task-relevant dimension of duration. Specifically, 75% of trials presented the same tone frequency as the cue tone for duration judgement, while the remaining 25% consisted of one of four alternative, unexpected probe frequencies that did not match the cue frequency. This specific task design was used to study whether attention would still be allocated to the expected frequency (based simply on stimulus statistics) when frequency was not required for the task itself. Indeed, their experiments show that response time and accuracy are strongly influenced by the probability of tone frequency, where unexpected tones led to both longer reaction times and lower accuracy.

Using a duration judgement task similar to that of Mondor & Bregman (1994), we developed a novel paradigm to study this probe signal effect in ferrets. The primary experimental aim was to test the hypothesis that the probe signal effect previously reported in human listeners would also hold in this common animal

model of human hearing. In the long term, the task could then be used to study the neural basis of the probe signal effect, by carrying out this task in ferrets while using microelectrodes to measure the spiking responses of auditory cortical neurons.

As our task needed to be suitable for animals, we had to modify the human probe signal design (Greenberg & Larkin, 1968; Mondor & Bregman, 1994) in several ways, such as by using water rewards and time-outs to motivate the animals. In the human probe signal tasks, there were neither rewards nor punishments, so the necessary addition of these could also affect attentional filter shapes and the neural circuits involved. Secondly, animals require extensive training on a psychophysical task over weeks or months to learn to perform the task, while humans can complete the task in one session, guided by verbal instructions. The degree of training and stimulus exposure required for animal studies may alter the probe signal effect. Our task also differed from that of Mondor & Bregman in that we did not include frequency cue tones, which in their study were brief sounds presented prior to the target tone to direct attention to a specific frequency region. Recent work in humans showed that cues are not necessary to produce a probe signal effect (Luthra et al., 2024; Zhao et al., 2022). Instead, we anticipated that ferrets would develop an expectation of frequency over time as they gained experience with the task. Our hypothesis for this two-alternative forced choice duration discrimination task was that ferrets would show a similar probe signal effect to humans. That is, that they would show faster reaction times to the most probable tone frequencies even though the task requires reporting on an orthogonal sound dimension (namely, duration).

3.1 METHODS

3.1.1 Subjects

A further four naive female adult pigmented ferrets (aged 6-14 months) were used for this task. Two animals were culled due to unrelated illness. One animal was culled during training and another during data collection of the second task variation. Housing, water deprivation, and tympanometry followed the same principles described in Section 2.1.1. The procedures were approved by the University of Oxford Committee on Animal Care and Ethical Review and were carried out under license from the UK Home Office, in accordance with the Animals (Scientific Procedures) Act 1986.

3.1.2 Apparatus

The same testing chambers and setup used in Chapter 2 were used for this study (0) with the exception that all three spouts were used for this 2AFC task (Figure 21). The centre spout was used to trigger a trial, while the outer two spouts were used to respond (left for short duration tones and right for long duration tones). Ferrets received rewards at the corresponding response spout (left or right).

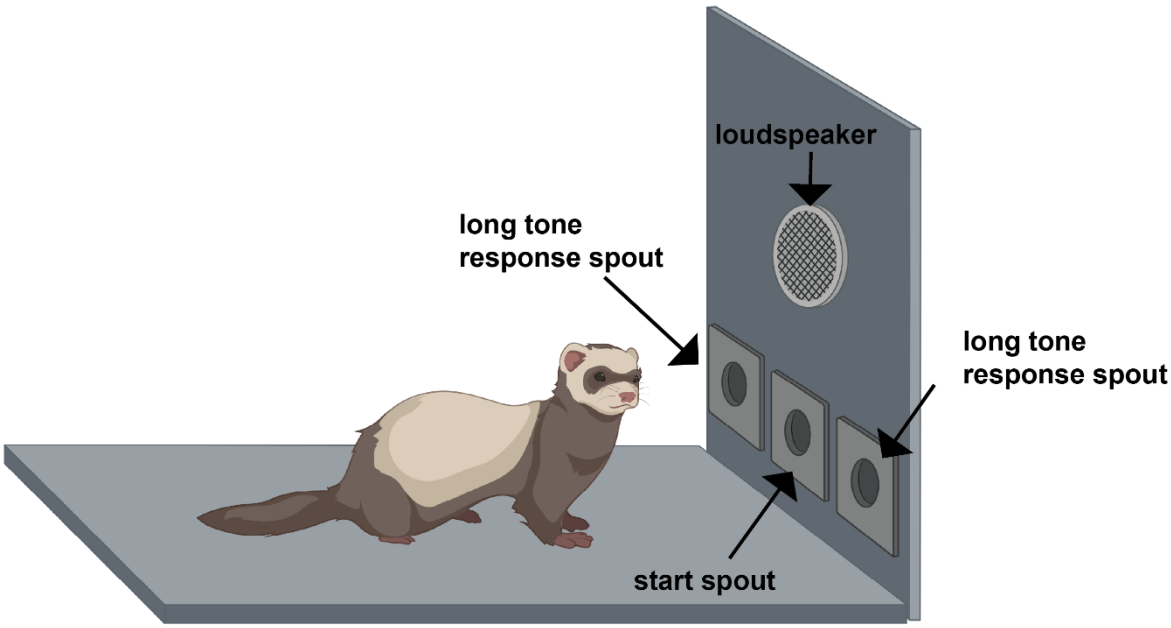

Figure 21. Schematic Drawing of the Testing Chamber. The ferrets have three nose poke holes with inner waterspouts. A poke is detected when the ferret breaks the LED beam that runs across the entrance of the tube. Ferrets must initiate a trial using the centre spout. A tone is then played through the speaker. Then a response is indicated based on the ferret's peripheral spout choice.

3.1.3 Training

3.1.3.1 Stage 1: Spout Training

Prior to being introduced to sounds, the animals were habituated to the testing chamber and the nose poke holes (Figure 21). In this first phase, the ferrets were trained to initiate a trial by activating the central 'start' spout before then activating either of the peripheral spouts. If this sequence of nose poke behaviour was carried out, a water reward (~ 0.3 ml) was delivered from the activated peripheral spout. To mitigate the development of a preference biases for one of the spouts, a peripheral spout was deactivated after being selected in three consecutive trials, thereby compelling the ferret to choose the alternative. Ferrets spent an average

of 10.33 sessions (~ 5–6 days) in this stage before they were able to confidently receive water rewards in quick succession.

3.1.3.2 Stage 2: Introducing Main Stimulus Frequency

After ferrets were able to trigger rewards in stage 1, a 3000 Hz (60dB SPLA) pure tone stimulus was introduced. The objective of these pre-training stages was to teach the animals to discern and report the duration of a stimulus, either 50 ms or 400 ms, to obtain a water reward. Upon initiating a trial via the central spout, the pure tone was emitted repeatedly (50 ms for short duration tones and 400 ms for long duration tones, with a 1000 ms interstimulus interval of silence) until the correct spout was selected: the left spout for the 50 ms tone and the right spout for the 400 ms tone. Incorrect choices did not incur penalties; however, the reward was larger if the first nose-poke accurately matched the tone duration. Following an incorrect attempt, the same tone duration was presented in the subsequent “correction trial”, which were omitted from analysis. Ferrets advanced to stage 3 upon achieving a hit rate of at least 65% for three consecutive days, a criterion applied to all subsequent training stages as well. On average, this took ferrets 19.25 sessions (~ 9 - 10 days).

3.1.3.3 Stage 3: Introducing Time Outs

Stage 3 involved a similar task as described in Stage 2, but with the modification that animals were no longer granted unlimited opportunities to activate spouts for water rewards during each trial. Instead, an incorrect spout

selection aborted the trial and resulted in a timeout. This timeout began with a 500 ms burst of broadband noise (65dB SPLA; to indicate to the ferret that it made a wrong decision), followed by a 7-second period during which the spouts were deactivated. Following the timeout, the ferret could initiate a new trial at the centre spout. On average, it took ferrets 31.67 sessions (~ 15 - 16 days) to reach criterion and progress to Stage 4.

3.1.3.4 Stage 4: Single Stimulus Presentation

Stage 4 was similar to Stage 3, but instead of the tone repeating until a choice was made, Stage 4 featured a single presentation of the stimulus on each trial prior to decision-making. Ferrets needed fewer sessions to reach the criterion: 8.67 sessions on average (~ 4 - 5 days).

3.1.3.5 Stage 5: Testing

Upon achieving a success rate of at least 65% correct trials across three consecutive sessions in Stage 4, the ferrets progressed to the final testing stage. The reward structure of the testing stage was similar to stage 4, requiring ferrets to respond on each trial at the correct peripheral spout based on the duration of tone presented to earn a water reward. The primary difference in the testing stage 5 was the introduction of four novel “probe” tone frequencies alongside the familiar “signal” tone frequency of 3000 Hz. The signal frequency was presented on 80% of trials. On the remaining 20% of randomly interspersed trials, one of the four new frequencies was presented (5% probability each). Despite these

frequency variations, the core task, duration judgement, remained the same. Accuracy was measured as the percentage of correct trials, and reaction times were recorded, as in stage 2-4.

3.1.4 Experimental Design and Task Variations

All pure tone stimuli used in this study were generated using a custom MATLAB script. Each tone consisted of a single sine wave at the desired target frequency, which was presented at a constant sound pressure level. Tones were calibrated with a GRAS Audiometer Calibration System to ensure consistency in their intensity (60 dB SPLA) across frequencies. To avoid transients at onset and offset, a 10 ms cosine ramp was applied.

3.1.4.1 Initial Frequency Range (Task Variation 1)

Five tone frequencies, evenly spaced, were selected based on their relatively stable perceived sound level across this frequency range, as indicated by ferret audiograms (Kelly et al., 1986). The 3000 Hz frequency served as the expected signal frequency, and was presented on 80% of the trials, while the frequencies 1890, 2381, 3780 and 4762 Hz were designated as probes and were each presented on 5% of trials. All stimuli were presented at 60 dB SPLA in silence, with each speaker individually calibrated.

3.1.4.2 Upwards Shifted Frequency Range (Task Variation 2)

Given the extensive training with 3000 Hz pure tones, there was an inquiry into whether the animals could generalise the probe signal effect to a new set of frequency probabilities. To determine if observed behaviours were genuinely due to the probe signal effect rather than frequency familiarity, a novel distribution was implemented.

The standard signal frequency was adjusted to 5000 Hz, and probe frequencies of 3150, 3969, 6300, and 7937 Hz were selected to maintain equivalent octave intervals to the first set of frequencies tested (variation 1). The choice of a 5000 Hz signal, with the closest probe at 3150 Hz approximating the familiar 3000 Hz used in training and variation 1, allowed for an overlap of the two frequency ranges.

3.1.4.3 Varying Expected Frequency per Session (Task Variation 3)

For this experimental stage, we chose to prevent the formation of long-term frequency expectations by varying the signal frequency randomly across testing sessions, in order to test if ferrets could adapt to stimulus statistics and show probe signal effects within a single session. The frequencies presented in this variation of the task were 3000, 3780, 4762, 6000 and 7559 Hz. The signal was randomly selected at the start of each session to be either the lowest (3000 Hz), central (4762 Hz) or highest (7559 Hz) frequency in this set. The other 4

frequencies served as probes. This variation aimed to reset the animals' expectations each session, avoiding biases formed from previous exposures.

3.1.4.4 Complex Inharmonic Sounds (Task Variation 4)

To further counteract familiarity with frequencies, we introduced a complex inharmonic sound as the probe instead of merely varying the pure tone frequencies. The experimental setup was adjusted to present only two stimuli: the standard 4762 Hz pure tone (comprising 80% of trials) and the complex inharmonic sound (comprising 20% of trials). The complex inharmonic sound was generated by summing sine waves at the following frequencies: 3779 (50% of the reference frequency), 5669 (75%), 7559 (reference frequency), 9827 Hz (130%), and 13,229 Hz (179%). These frequencies were deliberately chosen to avoid harmonic relationships, resulting in an inharmonic sound with a broad spectral range from 3779 Hz to 13,229 Hz. The inharmonic relationships between these components produced a “unpleasant” quality to human listeners, chosen to be obviously different in spectrum and timbre to the previously used pure tones. Short sound durations were set to 100 ms (previous levels commonly used 50 ms) and long durations were set to 200 ms (previous levels used 200–400 ms). This adjustment not only introduced novelty but also accelerated data collection, with the probe appearing more frequently in trials compared to the previous setup.

3.1.4.5 Determining Ideal Tone Durations

This novel behavioural paradigm for ferrets used tone durations of 50 ms (short) and 200 - 400 ms (long). However, these may not have matched the difficulty of the human psychophysics tasks (Luthra et al., 2024; Mondor & Bregman, 1994), which used durations of 50 ms and 90 ms, possibly making the duration discrimination too easy for the ferrets to attend carefully to the sounds. To determine an appropriate difficulty level more accurately, a thresholding task was developed using the same 2AFC framework. This task presented one of eight potential durations (ranging from 50 ms to 250 ms in increments of 25 ms), randomised across trials. The durations were categorised as “short” (50 ms to 125 ms) and “long” (175 ms to 250 ms), with water rewards and timeouts provided based on appropriate spout selection, as described above.

3.1.5 Data Analysis

Accuracy and reaction time data were analysed using generalised linear mixed-effects models (GLMEs) and linear mixed-effects models (LMEs), respectively. These model types were selected to accommodate the repeated-measures and hierarchical structure of the dataset, with trials nested within sessions and sessions within individual animals (Baayen et al., 2008). GLMEs were used to model binary accuracy outcomes using logistic regression, while LMEs were applied to continuous reaction time data (Gelman & Hill, 2007).

For both model types, fixed effects included StimulusType (signal vs. probe) and Duration (short vs. long), with interaction terms included to test whether the

effect of frequency probability varied by tone duration. Ferret ID, Session ID, and Time of Day (morning vs. evening) were included as random intercepts to account for between-subject and session-level variability. For reaction time models, Trial Number was also included as an additional fixed effect in exploratory analyses to assess learning or fatigue effects within sessions. Where higher-order interactions (e.g. StimulusType × Duration × TrialNumber) were not supported by the data, models were simplified to avoid overfitting.

To evaluate both global (group-level) and stimulus-specific effects, separate models were run with either the full set of five frequencies (frequency-level models) or a binary coding for StimulusType (signal vs. probe). The former allowed identification of individual probes driving effects, while the latter tested the central hypothesis regarding the attentional impact of signal probability.

All models were implemented in MATLAB (MathWorks, Natick, MA, USA) and R (R Core Team, 2024) using custom-written scripts in RStudio (Posit Team, 2025). Where significant interactions were observed, post hoc comparisons were conducted using estimated marginal means (EMMs), and Holm–Bonferroni corrections were applied for multiple comparisons. Additional repeated-measures ANOVAs and t-tests were used to complement mixed-effects modelling where appropriate.

3.2 RESULTS

The aim of this chapter was to determine whether a probe signal effect could be observed in ferrets performing a 2AFC duration discrimination task when the

manipulated variable (frequency) was orthogonal to the task. A single “signal” frequency was presented on 80% of trials in each session, while four other “probe” frequencies were presented equally for the remaining 20%. Correction trials and trials with reaction times larger than two standard deviations from the mean of a given session were excluded from all data analyses.

To test for differences in accuracy and response times between signal and probe frequencies, we first fitted generalised linear mixed effects models (GLMEs) and linear mixed effects models (LMEs), respectively. These models were selected to account for the hierarchical structure of the data (sessions nested within animals) and allowed us to model both fixed effects of stimulus conditions and random effects of subject and session. While initial models included frequency as a fixed factor, the primary hypothesis concerns the group-level contrast between signal and probe frequencies. Therefore, only group-level results are presented in the main summary figures and tables; frequency-level models and full results tables are included in the Appendix (Appendix 7.2.2 to7.2.3). *StimulusType* (signal vs. probe frequency) and *Duration* (short vs. long tone) were included as fixed effects in all models. *Ferret ID, Session ID,* and *Time of Day* (morning or evening session) were included as random effects in all models. The variable *Duration* was not part of the original hypothesis but was included to account for within-task variation. While trial number was not included in the initial LME formulation, it was added in follow-up models to explore within-session dynamics, given the capacity of mixed-effects models to flexibly incorporate trial-level covariates.

Where significant interactions were present, follow-up comparisons were conducted using estimated marginal means. In such cases, plots focus on the interactions, as main effects should not be interpreted in isolation (Nelder, 1977). For reaction times, model structure was adjusted across variations depending on the significance of higher-order terms, with interaction terms included only when justified by the data to avoid overfitting.

To provide an overview of the key findings, we present a set of summary plots and tables capturing group-level effects across task variations as well as raw data plotted by frequency and ferret. Group-level plots are primarily modelled estimates (mean ±95% CI) but where only one main effect was significant or none at all, raw data (mean ±SD) was also plotted for comparison. Raw and modelled data are labelled by their axes titles. Detailed tables for all models are provided in the Appendix.

3.2.1 Accuracy

As shown in Figure 22, raw accuracy data plotted across individual frequencies did not reveal a consistent probe signal effect. In each task variation, performance varied across frequencies and animals, with no systematic advantage for the signal frequency (bolded on the x-axis). In most cases (e.g. Variation 1, Figure 22b), signal frequency performance was similar to neighbouring probes, except for Figure 22f, where performance for the neighbouring probe frequency was lower than for the signal frequency. Individual ferret traces further highlight this variability, often crossing one another and showing divergent profiles across frequency. Together, these patterns suggest an absence of stable probe signal effects at the raw data level, contrasting with prior human work (e.g. Mondor & Bregman, 1994) where the effect typically emerges as a pronounced dip in accuracy for infrequent probes.

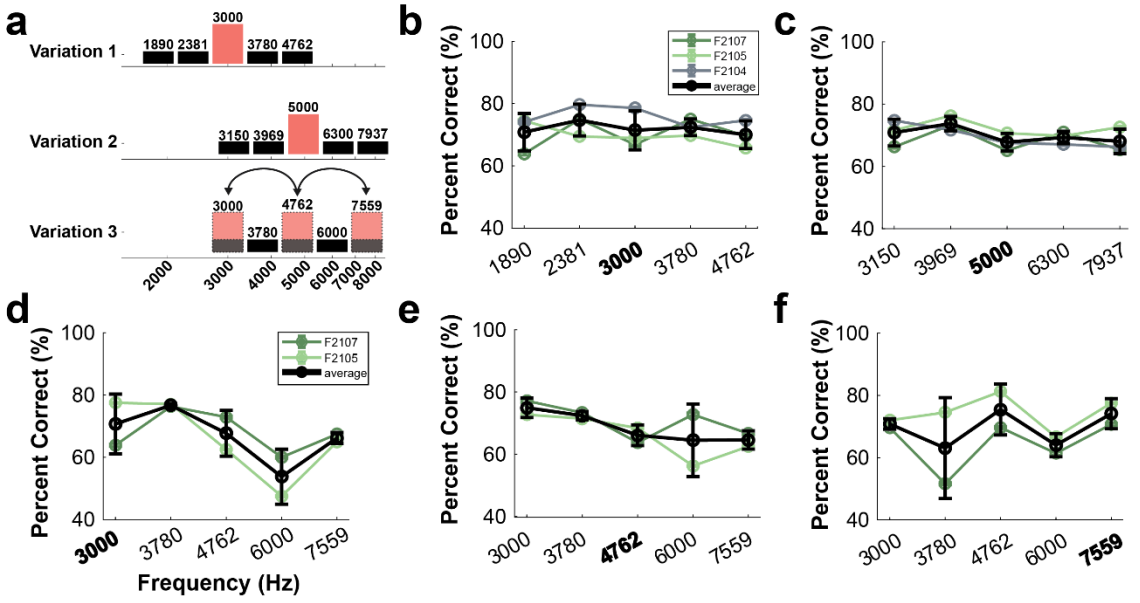

Figure 22. Frequency Structure and Raw Percent Correct Performance Across Task Variations. (a) illustrates the frequency configuration used in each task variation. The signal frequency is marked in salmon. For variation 3, the signal frequency was randomly assigned on a per-

session basis from one of three possible tones (3000, 4762, or 7559 Hz). **(b-f)** show raw percent correct values for each frequency, plotted separately for each variation. Thin coloured lines indicate performance of individual ferrets; thick black lines represent across-animal average (mean ±SD). Frequency labels on the x-axis are bolded to indicate signal tones. Variation 1 is plotted in b, Variation 2 in c, and Variation 3 in the bottom row.

To assess the effects of stimulus type and tone duration on accuracy, we employed mixed-effects modelling. A significant interaction between stimulus type (signal vs. probe) and tone duration (short vs. long) emerged, indicating that the accuracy advantage shifted depending on tone duration. Specifically, accuracy was higher for signal tones on short-duration trials and for probe tones on long-duration trials. This pattern suggests that the probe-signal effect is more pronounced with brief tones. A summary of the results for each variation is presented in Table 3, with detailed model results available in Appendix Section 0.

In Variation 1, where the task and frequency manipulations were novel to the ferrets, accuracy was generally higher for long tones regardless of stimulus type. This likely reflects the early difficulty of the task, with ferrets either struggling to correctly discriminate tone duration or exhibiting a right spout bias, favouring long-duration responses irrespective of the actual stimulus.

The significant interaction across all variations indicates that tone duration plays a substantial role in modulating attentional mechanisms. Specifically, shorter tones may facilitate the use of attentional filters based on expectation (probe-signal effect), while longer tones might engage saliency-driven attention, enhancing the detection of unexpected events. This finding extends previous human studies by directly examining the interplay between tone duration and attentional modulation in duration discrimination tasks.

Table 3. Summary of Significant Effects in Generalised Linear Mixed-Effects (GLME) Models for Percent Correct (PC). Each row represents one task variation or condition. Check

marks indicate significant fixed effects or their interactions (✓). Crosses indicate non-significant main effects (✗). The final column summarises post hoc comparisons for the significant *StimulusType x Duration* interaction. In all cases, accuracy tended to be higher for signal tones at short durations and higher for probe tones at long durations, although this was not always statistically significant for both duration levels. **Appendix Table references are provided for full model outputs and detailed results.**

Task/Condition	StimType (Signal/Probe)	Duration	G x Dur	Appendix Table	Interaction Notes
Var 1	✗	✓	✓	Table 36	Significant difference between short and long tones within both signal and probe groups; no sign difference between signal and probe within each tone duration level
Var 2	✓	✓	✓	Table 39	Short duration: signal > probe; Long duration: probe > signal
Var 3 – 3000 Hz Signal	✗	✓	✓	Table 42	Trend for short: signal > probe (but n.s.); Long duration: probe > signal
Var 3 – 4762 Hz Signal	✗	✓	✓	Table 45	Long duration: probe > signal
Var 3 – 7559 Hz Signal	✗	✓	✓	Table 48	Same as above
Var 4 – complex inharmonic probe	✓	✓	✓	Table 51	Short duration: signal > probe; Long duration: probe > signal

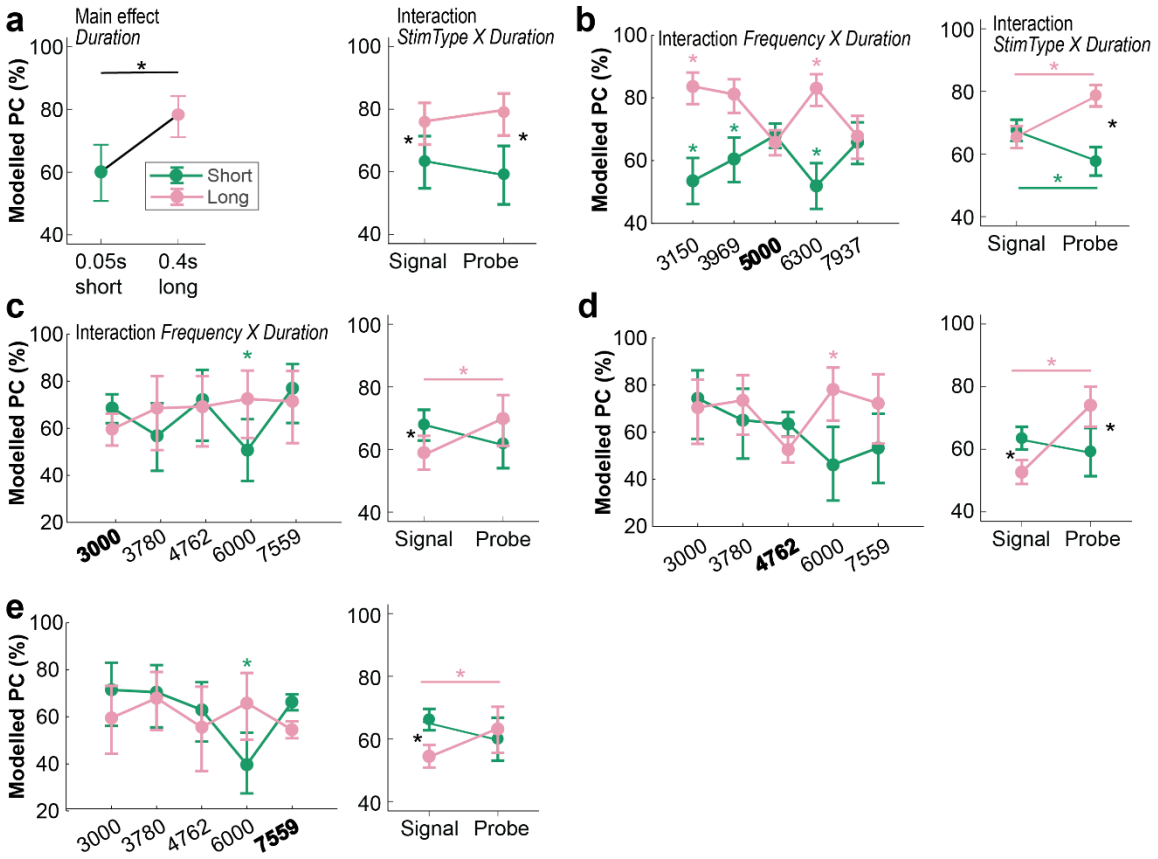

Figure 23. Modelled Accuracy Across Task Variations. Each panel shows significant effects from mixed-effects models assessing the interaction between stimulus duration (short vs. long) and either frequency (left plots) or stimulus type (probe vs. signal; right plots). **(a)** Task Variation 1 showed a main effect of duration (left), and a stimulus type x duration interaction (right). **(b-e)** Task Variation 2 and 3 showed consistent frequency x duration interactions at the frequency-level (left), and stimulus type x duration at the group-level (right). Asterisks with coloured bars denote significant differences between signal and probe for the corresponding tone duration; black asterisks indicate significant differences between durations within a stimulus type; coloured asterisks in the left plots show significant difference from the signal frequency. Error bars show 95% confidence intervals for the modelled estimates.

Finally, performance was examined in a fourth task variation that differed from the main design by introducing a broadband inharmonic probe sound, perceptually distinct from the 4762 Hz pure tone signal. The aim was to assess whether a more acoustically salient probe would amplify the probe signal effect. A highly significant *StimType x Duration* interaction emerged ($\beta = 2.81$, $p < .001$), reflecting the same crossover pattern observed in earlier variations: higher accuracy for signal tones on short-duration trials and for probe sounds on long-duration trials (Figure 24). This effect was especially pronounced in Variation 4, supporting the hypothesis that increased perceptual saliency can enhance the probe signal effect.

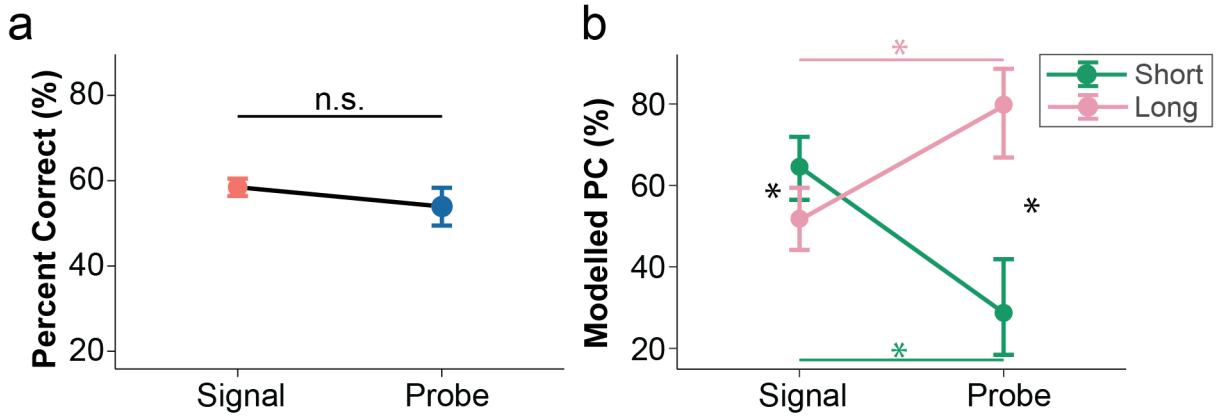

Figure 24. Accuracy Results for Task Variation 4 (inharmonic complex as probe). (a) Raw accuracy (mean ±SEM) for signal and probes. No significant difference was observed. **(b)** Modelled accuracy (% correct) as a function of *StimulusType* (Signal vs. Probe) and *Duration* (Short = 0.1 s, Long

= 0.2 s). Modelled values are plotted as point estimates with ±95% confidence intervals. A significant *StimulusType x Duration* interaction indicated reduced performance for short probes and enhanced performance for long probes, relative to the signal tone.

3.2.2 Reaction Time

As shown in Figure 25, raw reaction time data plotted across individual frequencies revealed little evidence of a probe signal effect. Reaction times were relatively stable across frequencies within each variation and showed minimal separation between signal (bolded) and probe frequencies. Individual animal traces overlapped substantially, with no systematic advantage for either signal or probe tones.

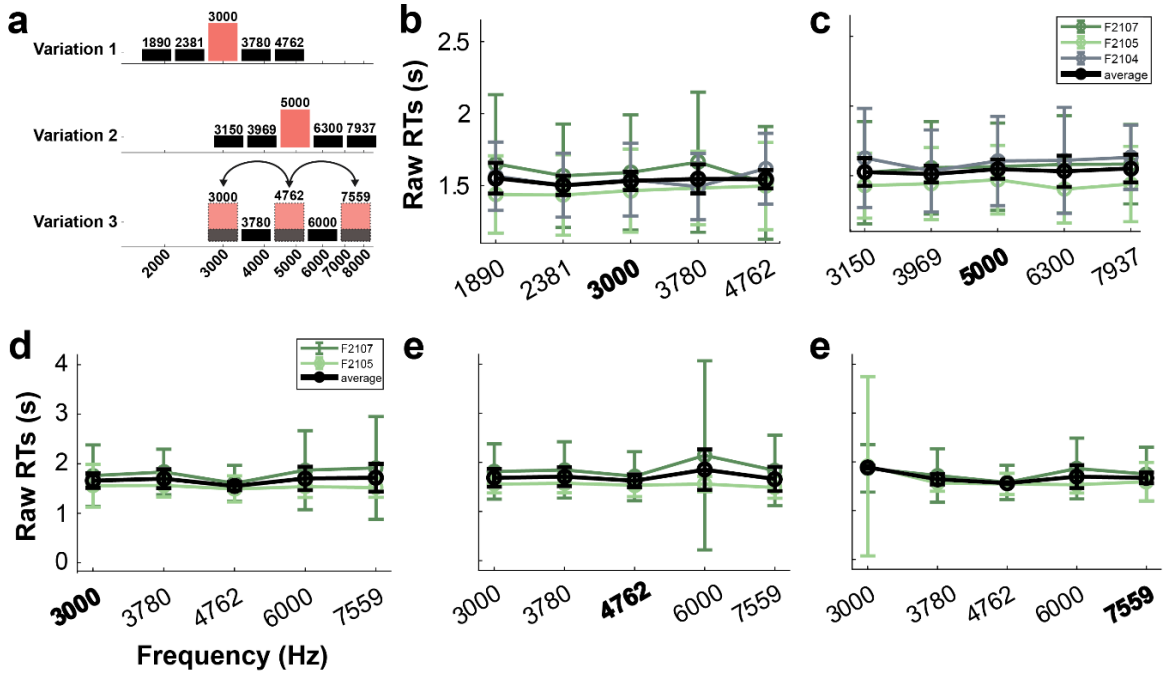

Figure 25. Frequency Structure and Raw Reaction Time Across Task Variations. (a) illustrates the frequency configuration used in each task variation. The signal frequency is marked in salmon. For variation 3, the signal frequency was randomly assigned on a per-session basis from one of three possible tones (3000, 4762, or 7559 Hz). **(b-f)** show raw reaction times for each frequency, plotted separately for each variation. Thin coloured lines indicate performance of individual ferrets; thick black lines represent across-animal average (mean ±SD). Frequency labels on the x-axis are bolded to indicate signal tones. Variation 1 is plotted in b, Variation 2 in c, and Variation 3 in the bottom row.

To assess whether any frequency-based differences might emerge at a distributional level, we conducted Kolmogorov-Smirnov (KS) tests comparing signal and probe reaction time distributions for Task Variation 3, which had the largest dataset. These tests revealed statistically significant differences across all signal frequencies and both ferrets (all $p < 0.001$). Full histograms and cumulative distribution functions (CDF) are provided in the Appendix (7.2.1, Figure 53, Figure 58, and Table 15).

While these findings confirmed overall distributional differences between signal and probe trials, these effects were not apparent in the raw reaction time data. Moreover, KS tests do not indicate the direction or magnitude of reaction time differences. To more precisely evaluate how reaction times were shaped by experimental factors, we applied linear mixed-effects (LME) models, which allowed us to estimate fixed effects of duration and stimulus type (probe vs. signal) while controlling for subject- and session-level variability.

Mixed-effects modelling revealed a consistent effect of tone duration on reaction times: longer tones elicited significantly faster responses, indicating a general performance advantage that parallels, but does not fully mirror, the duration-related interaction seen in accuracy (Figure 26, summarised in Table 5, detailed results in Appendix Section 0). Reaction time also decreased as sessions progressed, reflected in a significant main effect of *Trial Number* across several conditions, suggesting learning-related or motivation influences (e.g. Figure 26d).

No consistent differences were observed between signal and probe tones across variations. However, in the 4762 Hz signal condition of Variation 3, a significant *StimType x Duration x Trial Number* interaction revealed diverging

response patterns: signal tone reaction times decreased over the session, while probe tone responses remained stable, an effect restricted to long-duration trials (Figure 26g–h). This may reflect the gradual accumulation of frequency-based expectations, suggesting that ferrets form attentional filters more slowly than humans. Unlike human participants, who show rapid onset of the probe signal effect (Armstrong et al., 2017), ferrets may require more extensive exposure to high-probability tones to bias response speed. The relatively short task duration (approx. 20 min per session), determined by the animals' voluntary engagement, may have limited the opportunity for stable expectations to emerge, potentially contributing to the absence of group-level reaction time effects in other condition.

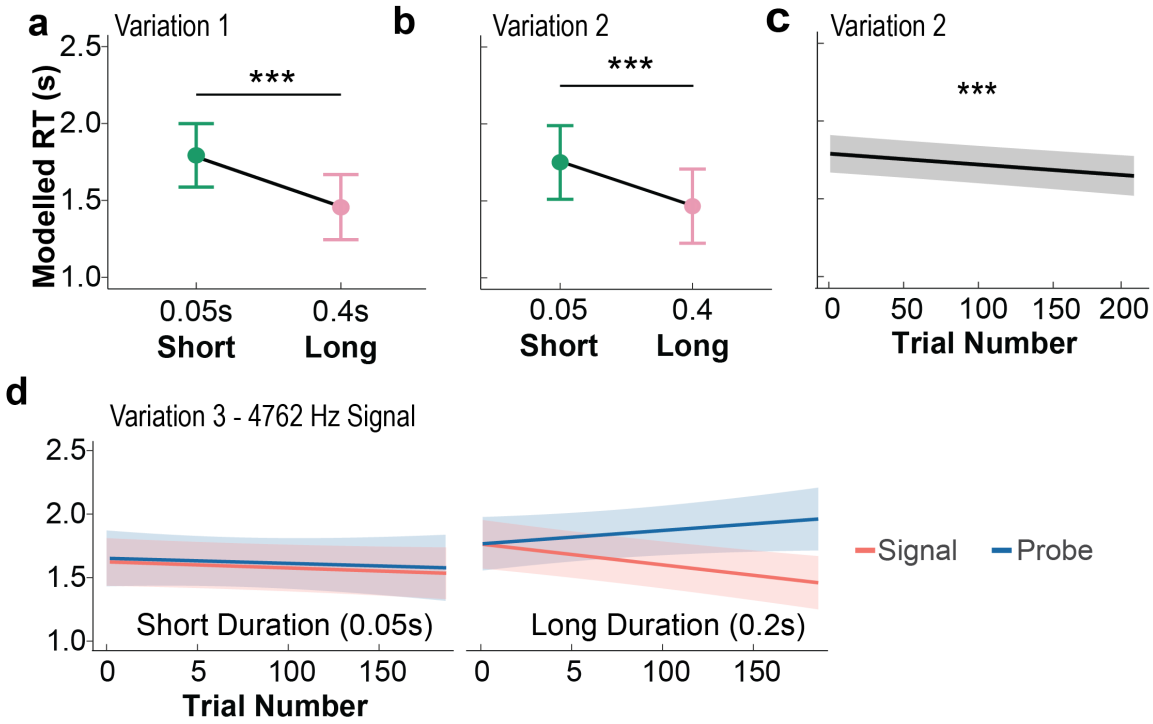

Figure 26. Significant Modelled Reaction Time Effects Across Task Variations. (a-b) Modelled reaction times (RTs) in Variations 1 and 2 showed a significant main effect of tone duration, with faster responses to long tones compared to short. **(c)** In Variation 2, RTs also decreased across the session. **(d)** Three-way interaction between stimulus type, tone duration, and trial number for 4762 Hz signal condition of Variation 3. Signal RTs became faster over time on long-duration trials, while probe RTs remained stable. Error bars and shaded regions show 95% confidence intervals. Asterisks denote statistical significance (***p <.001).

Table 4. Summary of Significant Effects from Linear Mixed-Effects Models of Reaction Time (reaction time). Each row represents a task variation or condition from Task Variation 3, with signal frequency indicated for the Variation 3 conditions. Columns denote whether main effects or interactions were significant. ✓ indicates a significant effect or where the direction of the effect is noted (e.g. "long < short"), ✗ indicates a non-significant effect, — indicates that the effect or interaction was not included in the model due to lack of significant main effects. For the Duration and Trial # columns, directionality is provided where applicable. The Notes summarises directionality and relevant post hoc interpretations. **Appendix Table references are provided for full model outputs and detailed results.**

Task/ Condition	StimType (Probe/ Signal)	Duration	Trial #	G x Dur	G x Trial	Dur x Trial	G x Dur x Trial	Appendix Table	Notes
Var 1	✗	long < short	✗	✗	✗		✗	Table 54	Faster reaction times for long durations
Var 2	✗	long < short	reaction times decreased over session	✗	✗	—	—	Table 56	Faster reaction times for long durations; reaction times become faster as session progresses
Var 3 – 3000 Hz Signal	✗	✗	✗	—	—	—	—	Table 58	No evidence of probe signal effect in reaction time
Var 3 – 4762 Hz Signal	✗	✓	✗	✗	✗	✓	✓	Table 60	For long durations, probe trials become slower over trials while signal trials may become faster.
Var 3 – 7559 Hz Signal	✗	✗	✗	—	—	—	—	Table 62	No evidence of probe signal effect in reaction time
Var 4 – complex inharmonic probe	✗	✓	✗	✗	✗	✓	✗	Table 64	reaction times to long duration become slower throughout a session; reaction times to short duration stay stable.

In Variation 4, reaction times showed a significant *Trial Number x Duration* interaction (β = .0042, p = 0.0003), with responses slowing over time for long-duration trials (Figure 27b). This contrasts with the 4762 Hz signal condition in Variation 3, where reaction times became faster over time for long-duration signal trials. Signal-probe differences were not observed in this or other variations, suggesting an absence of such an effect in reaction time.

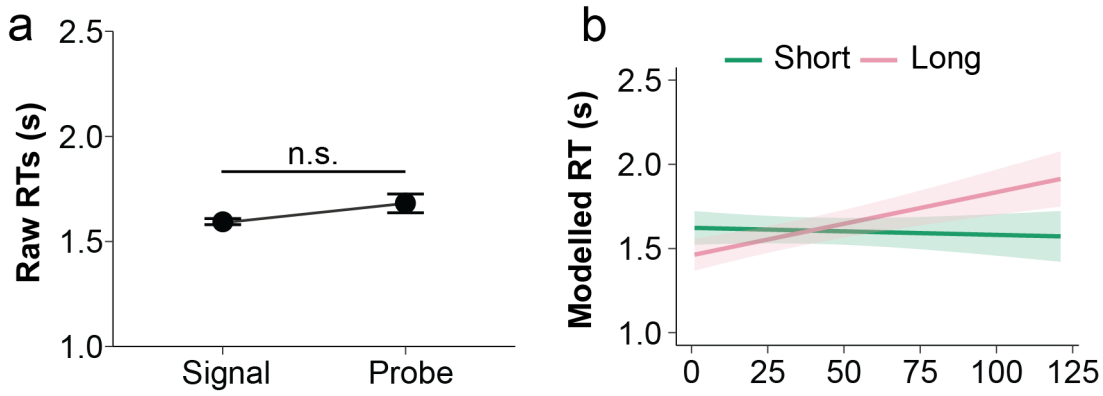

Figure 27. Reaction Times in Task Variation 4 (inharmonic probe tone). (a) Raw reaction times (mean ± SEM) showed no significant difference between signal and probe trials. **(b)** Modelled reaction time by *Trial Number* and *Duration*. A significant *Duration* × *Trial Number* interaction was observed: reaction times increased over time for long durations but remained stable for short durations. Error bars represent 95% CIs.

In sum, tone duration exerted consistent effects on reaction times, whereas frequency-based effects were more limited and often interacted with trial progression. These findings suggest that attentional filtering based on frequency may emerge more slowly or less robustly in ferrets than in humans.

3.2.3 Determining Ideal Tone Durations

To assess whether ferrets were performing near perceptual threshold, we ran two animals (F2105 and F2107) on a version of the duration discrimination task using eight tone durations ranging from 50 to 250 ms. Ferrets reported whether each tone was “short” or “long” via a left or right spout response. Psychometric curves were constructed based on the percentage of right output right choices (interpreted as “long” responses) at each duration. Both ferrets exhibited the expected increase in “long” responses with increase duration (Figure 28).

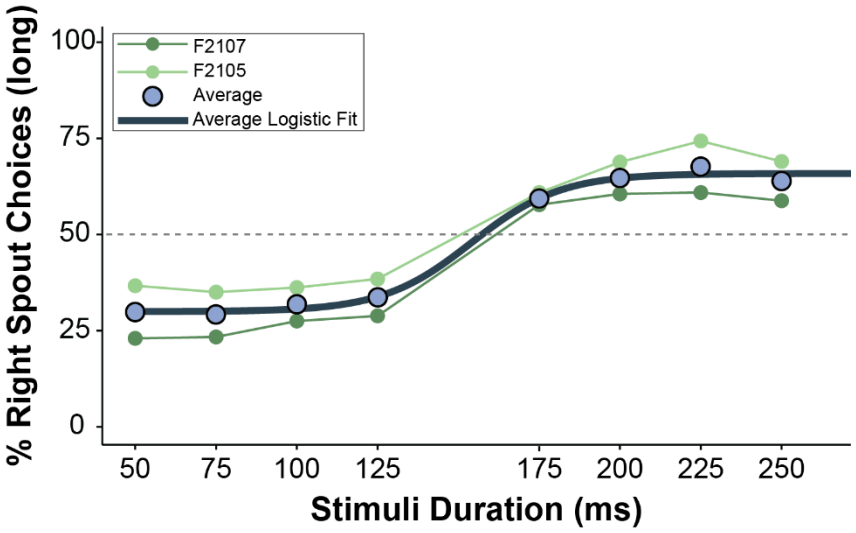

Figure 28. Psychometric Performance in the Duration Discrimination Task. Raw percentage "long" responses are plotted for each duration tested in the duration categorisation task (50–250 ms), shown separately of Ferrets F2105 (light green) and F2107 (dark green). Each point reflects the proportion of right spout choices, which corresponded to "long" responses. Average values across both ferrets are plotted in blue (circle markers), and a three-parameter logistic function was fit to the averaged data (dark blue line). The fitted curve provides a close approximation of the sigmoidal shape expected for perceptual categorisation, with a lower asymptote near 30%, upper asymptote near 66%, and an inflection point around 154 ms. The horizontal dashed line marks the 50% point, indicating the perceptual threshold.

Averaged psychometric data were fit with a four-parameter logistic function to quantify perceptual sensitivity. The fitted curve captured the sigmoidal shape of the data well (pseudo-$R^2 = 0.995$), with an estimated lower asymptote of 30%, upper asymptote of 66%, and point of subjective equality (PSE) at 154 ms (Figure 28). These estimates indicate that durations between 125 and 200 ms were most perceptually ambiguous, while shorter and longer durations elicited more categorical responses.

To quantify perceptual difficulty across task variations, we used the fitted logistic function to estimate the absolute distance of each fixed tone duration from threshold (defined as 50% "long" responses).

$$f(x) = b + \frac{d - b}{1 + exp(-\frac{x - c}{a})}$$

Where:

- a = slope parameter
- b = lower asymptote,
- c = inflection point or PSE (point of subjective equality),
- d = upper asymptote.

As shown in Figure 29, short durations used in Variations 1–3 (50 ms) and 4 (100 ms) had absolute distances from threshold of 20.0% and 19.3%, respectively. Long durations in Variations 3–4 (200 ms) were slightly closer to threshold (14.6%), while the 400 ms tones used in Variations 1–2 were more distant (15.9%), suggesting they were more readily categorised as “long”. Although durations fell well outside the most ambiguous range, these rather modest differences in perceptual distance may nonetheless contribute to variation-specific effects on performance. Notably, Variation 4 included both the shortest perceptual distance for its long duration (200 ms), one of the lowest distances for its short duration (100 ms) and also showed the strongest deviance detection effect in accuracy. While no direct causal link can be established, this alignment raises the possibility that even small differences in perceptual ambiguity may shape attentional dynamics across variations.

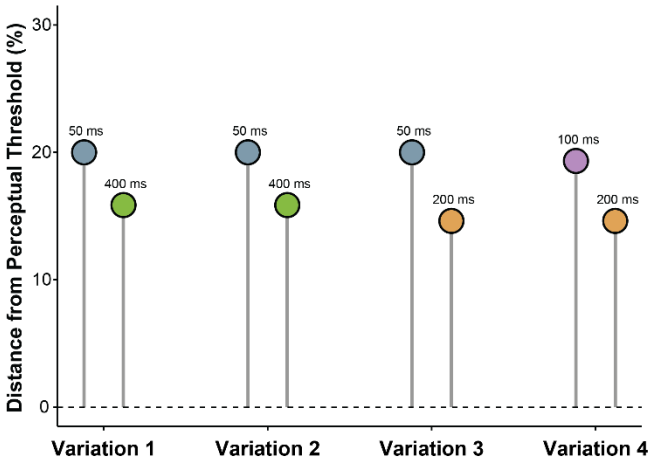

Figure 29. Perceptual Distance from Threshold Across Task Variations. Lollipop plots depict the absolute deviation from 50% "long" responses for each duration used in Task Variations 1-4, estimated from the averaged logistic function fit. Greater distances indicate durations that were more easily categorised as "short" or "long", whereas smaller distances reflect greater ambiguity near threshold. Vertical lines connect each point to threshold (0%, dashed line). Dot colour denotes tone duration.

3.3 DISCUSSION

This chapter investigated whether manipulations of frequency, when orthogonal to the task, would produce a probe signal effect in ferrets, as has been reported in humans (Luthra et al., 2024[1]; Mondor & Bregman, 1994). Although the probe signal paradigm is well established in human auditory research, this study represents the first adaptation of the duration discrimination version for non-human animals. Based on prior findings, we hypothesised that ferrets would exhibit faster reaction times for expected (signal) frequencies and slower responses for unexpected (probe) frequencies, with increasing latency as a function of distance from the signal frequency (Luthra et al., 2024; Mondor & Bregman, 1994). We also predicted higher accuracy for signal frequencies, consistent with earlier reports of expectation-based attentional filtering in detection and discrimination tasks (Mondor & Bregman, 1994; Schröger & Wolff, 1998).

Before presenting the behavioural findings, it is important to compare the design of this study to prior human work, particularly Luthra et al. (2024), who employed a similar two-alternative forced-choice 2AFC duration discrimination paradigm. Both studies, along with that of Mondor & Bregman (1994), manipulated frequency as a task-irrelevant dimension while varying tone duration. However, key methodological differences may account for the variation in observed effects. Luthra et al. and Mondor & Bregman used single-session

[1] This recent publication was made by our collaborators with whom we hold a mutual grant on cross-species investigations of statistics-driven attention.

designs with minimal feedback, whereas our ferrets underwent extensive multi-session training with continuous feedback.

In Luthra et al. (2024), participants encountered five frequencies under either unimodal or bimodal statistical distributions and received feedback only during initial familiarisation blocks, completing the entire task in a single session. In contrast, ferrets in the present study underwent extensive multi-session training under a consistent unimodal distribution, with feedback provided on every trial. Additional differences included session-wise reassignment of the signal frequency and the introduction of an inharmonic, spectrally complex probe stimulus in later variations. Mondor & Bregman (1994) also used a single-session design with minimal feedback and further directed attention by presenting a cue tone before each trial. These differences in training structure, feedback, and cueing likely shaped attentional strategies and may explain why the probe signal effect in ferrets emerged primarily in accuracy rather than reaction time.

Ferrets showed higher accuracy for signal tones during short-duration trials, consistent with the classic probe signal effect. However, at longer durations, this pattern reversed, with accuracy for probes often surpassing that for signals. This reversal implies that attentional deployment was initially guided by prior frequency statistics but shifted toward saliency-based processing as more sensory information became available during extended tone presentations.

One interpretation of this reversal is that infrequent probe tones may have been perceived as lasting longer than equally timed signal tones. Human studies have shown that deviant or unexpected sounds are often judged to have longer durations than expected ones (Tse et al., 2004; Ulrich et al., 2006). In our task,

probe tones were not only rare but, in one condition, also spectrally distinct, which may have enhanced their perceptual salience and strengthened this “time’s subjective expansion” (TSE). This condition also saw the largest difference between accuracy on long vs short duration probe trials. Such elongation could improve categorisation accuracy during long-duration trials, particularly when more processing time is available. At the same time, it can explain the low accuracy for short-duration probe trials.

This perceptual expansion may relate to the neural mechanisms underlying mismatch negativity (MMN), a component of the auditory evoked potential thought to index prediction error signals elicited when sensory input violates learned regularities (Näätänen et al., 2007). Both MMN and duration expansion have been associated with expectation violation and are likely to reflect complementary manifestations of predictive processing. Within predictive coding frameworks, MMN arises when bottom-up input diverges from top-down predictions (Friston, 2005; Garrido et al., 2009; Wacongne et al., 2012). The tendency to judge unexpected events as longer supports the idea that prediction error may influence not only neural response dynamics but also the subjective experience of time.

Reaction time patterns showed a consistent temporal advantage for long-duration tones across task variations, with faster responses compared to short tones. In some conditions, reaction times decreased further over the course of a session, suggesting contributions from learning or changes in motivational state. A selective effect emerged in the 4762 Hz signal condition of Variation 3, where signal reaction times progressively declined across trials while probe reaction times remained stable, an effect limited to long-duration tones. This may indicate

a gradual transition from deviance-driven orienting toward an expectation-based attentional filter as the session progressed.

Compared to humans, who typically form statistics-driven expectations within a single session, ferrets appeared to require more extended exposure before selectively enhancing responses to signal frequencies. The relatively short sessions (~20 minutes), driven by voluntary disengagement, likely constrained the consolidation of frequency-based attention. Future studies could extend engagement time, for example by reducing reward amounts per trial, to determine whether expectation-based filtering strengthens with additional exposure. Differences in motivational contingencies likely contributed to the observed attentional dynamics. In our task, ferrets were water-restricted and received water rewards only for correct responses, with time-outs following errors, a structure that incentivised accuracy and discouraged impulsive responding. By contrast, human participants typically performed probe signal tasks without tangible rewards or penalties, encouraging faster but potentially less cautious strategies. This difference may explain why ferrets exhibited frequency-related effects primarily in accuracy rather than reaction time, adopting a conservative response policy when faced with uncertainty.

Neural and behavioural studies support the strong influence of motivational demands on attentional strategies. David et al. (2012) showed that reward contingencies modulate spectrotemporal receptive fields in ferret auditory cortex, suggesting that task structure can shape sensory coding. Behavioural evidence from rats similarly indicates that water restriction encourages more deliberate, accuracy-focused responding (Masís et al., 2023).

Having established the broader dynamics of accuracy and response strategy, we next turn to evaluating how our data inform theoretical models of frequency-specific attentional filtering. In particular, we examine performance patterns in the complex inharmonic probe condition, which provides a novel test of the attentional band hypothesis proposed by Greenberg & Larkin (1968).

The complex probe was constructed by presenting five non-harmonically related frequencies, creating a broadband, noisy percept distinct from the pure tones used during training. Its spectral range (3779 Hz to 13,229 Hz) encompassed the expected signal frequency, but its inharmonic structure extended beyond the equivalent rectangular bandwidth (ERB) typically associated with effective frequency-specific attention in ferrets (Alves-Pinto et al., 2016).

This design invites comparison to earlier human work using complex or broadband stimuli in probe signal paradigms. Scharf et al. (1987) demonstrated that probes perceptually dissimilar to the signal tone were detected less reliably, even when falling within the same critical bandwidth. Their findings challenged a strict interpretation of the attentional band hypothesis, suggesting that attentional benefits are not determined solely by spectral overlap, but are modulated by stimulus familiarity and harmonic structure. In our study, despite partial spectral overlap with the expected frequency, the novelty and inharmonicity of the complex probe likely impaired attentional filtering, increasing perceptual demands. Consistent with observations by Escera et al. (1998), novel, spectrally rich sounds may override selective attention and elicit behavioural costs even when task-irrelevant. Signal–probe differences in accuracy were largest in the complex probe

condition, supporting the view that attentional filters are less effective when the probe is perceptually distinct from prior expectations.

The pattern observed in the complex probe condition complements recent human findings. Luthra et al. (2024) reported that in a bimodal frequency distribution, probe frequencies did not elicit the expected impairments in reaction time, despite evidence from other paradigms supporting dual-frequency attentional enhancement. They proposed that perceptual dissimilarity between probes and signals, combined with task-specific biases, may disrupt the efficacy of attentional filters. Similarly, in our study, the complex probe's lack of harmonic structure and broad spectral spread likely placed it outside the effective range of attentional filtering, even though its frequency range encompassed the signal tone. These converging observations point to a more nuanced model of selective auditory attention, where both statistical expectations and perceptual similarity shape the success of attentional deployment.

A further complication arises from the possibility of frequency-duration biases, as proposed by Luthra et al. (2024). In their study, participants exhibited a tendency to categorise lower-frequency tones as longer and higher-frequency tones as shorter, potentially confounding measures of frequency-specific attention. We assessed this possibility in our ferret data by examining the proportion of "long" responses across frequencies. Response patterns remained stable across signal assignments and frequency groups, indicating that ferrets did not exhibit a comparable bias (Table 68 through Table 71 in Appendix 1). This divergence may reflect differences in species-specific auditory environments: humans experience natural acoustic environments where low-frequency sounds

persist longer due to reverberation (Traer & McDermott, 2016), whereas laboratory-reared ferrets are unlikely to develop such statistical priors.

Our results revealed a consistent advantage for long-duration tones, most evident in the reaction time data. Across task variations, long tones generally elicited faster responses, raising the possibility of a response-side bias, as long tones always required a rightward spout response. If ferrets exhibited a lateralised motor preference, comparable to handedness in humans, this could account for faster responses to long tones, independent of tone properties. However, several lines of evidence argue against this interpretation. First, the overall proportion of long and short responses remained balanced across task variations, suggesting that animals did not default to the right spout. Second, the accuracy models revealed a consistent *StimulusType* × *Duration* interaction: signal tones produced higher accuracy at short durations, while probe tones elicited better performance at long durations. A static motor bias would not predict such systematic reversals aligned with stimulus statistics and temporal structure.

Further evidence against a pure motor bias account comes from the reaction time models themselves. In Variations 1 and 2, long tones elicited significantly faster responses without stimulus-specific interactions, consistent with the idea that early in training, response-side biases may have played a greater role. However, later task variations either showed significant interactions involving stimulus type and duration or no main effects of duration at all. For example, in Variation 4, reaction times for long tones slowed across the session while short-tone responses remained stable. A persistent rightward bias would predict the opposite pattern.

Rather than reflecting a static motor preference, the faster responses for long tones are better explained by greater sensory evidence accumulation prior to stimulus offset. This interpretation is most consistent with Variations 1 and 2, which featured the largest duration contrast (50 ms versus 400 ms) and showed main effects of duration in reaction time without stimulus-specific interactions. In these cases, the longer stimulus likely allowed animals to accumulate more auditory evidence, leading to greater decisional certainty and faster responses. In contrast, Variations 3 and 4 involved smaller duration differences and more complex probe conditions. Here, duration effects on reaction time co-occurred with higher-order interactions, indicating that the dynamics of evidence accumulation became more intricately modulated by task difficulty and attentional deployment.

Accuracy results, however, complicate a straightforward evidence accumulation account. Across task variations, we consistently observed significant *StimulusType* × *Duration* interactions: accuracy was higher for signal tones at short durations but shifted in favour of probe tones at long durations. This crossover suggests a time-dependent change in attentional strategy. When sensory evidence was limited, animals relied more heavily on expectations, enhancing detection of high-probability signals. As stimulus duration increased, attention appeared to reallocate toward salient or unexpected stimuli, improving probe performance.

The observed reversal, from signal prioritisation during short tones to improved probe detection during long tones, provides further evidence for dynamic adjustment of the attentional filter. Under temporal constraints, attention

appears to focus narrowly on expected frequencies, aligning with classical models of statistics-driven selective attention (Greenberg & Larkin, 1968; Mondor & Bregman, 1994). When time allows, this filter becomes more permissive, enabling the detection of salient deviations from expectation. Future work could test whether this flexibility reflects true reallocation of attentional resources or arises from shifts in decision boundaries under varying cognitive load, for instance by parametrically manipulating tone durations around perceptual threshold.

Much of the discussion thus far has focused on group-level effects averaged across probe frequencies. However, frequency-specific analyses revealed a more nuanced picture. While group-level models consistently showed signal–probe differences modulated by duration, frequency-level models indicated that not all probes contributed equally to the observed patterns. Some probes exhibited stronger impairments relative to signals, while others did not differ significantly. Reaction time data, by contrast, remained largely unaffected by frequency-specific manipulations across all variations.

In Variation 1, which reused the frequency range from training centred on 3000 Hz, none of the probes differed significantly from the signal in accuracy or interacted with tone duration. This likely reflects high familiarity with all tested frequencies, blunting any potential attentional asymmetries. Nonetheless, the group-level model still revealed a significant *StimulusType* × *Duration* interaction, suggesting that subtle effects of statistical structure could be detected when collapsing across probes, even if not apparent at the individual frequency level.

In Variation 2, several probes elicited performance impairments and duration-dependent interactions, although these effects did not follow a systematic gradient

across frequency distance from the signal, unlike human studies. By Variation 3, a particularly salient probe emerged: the 6000 Hz tone. Across all three signal contexts (3000, 4762, and 7559 Hz), the 6000 Hz probe consistently impaired accuracy and interacted with tone duration, suggesting that it occupied a unique perceptual status relative to the signal tones. A similar pattern was observed for the complex inharmonic probe in Variation 4, reinforcing the idea that certain probes can exert disproportionate influence depending on their spectral distinctiveness.

The disproportionate impact of the 6000 Hz probe likely reflects spectral sensitivity in ferrets. Audiometric data show peak ferret sensitivity between 8 and 12 kHz, with improved sensitivity emerging around 6 kHz (Kelly et al., 1986). Thus, although 6000 Hz lies slightly below the audiometric peak, it occupies a region where ferrets are increasingly sensitive, potentially rendering it more perceptually salient than neighbouring tones. Under the attentional filter framework, attention is biased toward the expected frequency, but salient, unexpected stimuli can capture attention and interfere with filtering processes. The need to suppress such attention-capturing stimuli could heighten uncertainty, impair decision-making, and selectively degrade performance on probe trials.

Theoretical accounts of perceptual decision-making reinforce this interpretation. Liu & Watanabe (2012) showed that when stimuli are ambiguous or conflict with expectations, decision boundaries widen, slowing evidence accumulation and promoting more cautious responses. Mulder et al. (2013, 2014) similarly demonstrated that under high uncertainty, drift rates decrease and decision thresholds increase, prolonging response times and reducing accuracy.

In the present task, short tones (e.g. 50 ms) offered minimal opportunity for sensory sampling. For salient but unexpected probes, this temporal constraint likely compounded uncertainty, leading to reduced accuracy. In contrast, longer durations permitted sufficient evidence accumulation to overcome initial conflict, enabling improved probe detection relative to signals.

Tone duration played a central role in shaping behavioural performance in this task. For short tones, ferrets exhibited higher accuracy for signal frequencies compared to probes, consistent with classical accounts of expectation-driven attentional filtering. This likely reflects the limited time available for sensory evidence accumulation, promoting reliance on top-down expectations to guide decisions. In contrast, longer tones often yielded superior performance for probes, suggesting that extended stimulus duration enabled a shift toward bottom-up saliency-driven attention. Rather than enhancing detection of expected tones, the additional processing time allowed unexpected, salient features to attract attention, improving detection of probe tones.

This reversal, from signal prioritisation under temporal constraint to probe enhancement with greater sensory evidence, highlights the dynamic interplay between expectation and stimulus-driven processes. Although findings are based on two ferrets, the consistency of the effects across animals and task variations strengthens the interpretation. Future work could test whether this crossover depends systematically on the perceptual difficulty of the duration discrimination, for instance by manipulating tone durations around the point of subjective equality identified in the psychophysical thresholding task. Such manipulations would help

clarify whether the observed shift arises from flexible attentional deployment or from difficulty-driven changes in decision boundaries.

Electrophysiological studies provide a strong basis for interpreting these behavioural dynamics in terms of attentional plasticity. In ferrets, spectrotemporal receptive fields (STRFs) in auditory cortex can undergo rapid, reversible reshaping depending on attentional focus and task demands (Fritz et al., 2003, 2005a; Yin et al., 2014).

Such plasticity selectively enhances neural gain for expected stimuli while suppressing distractors and can occur within seconds. In the context of the present task, these mechanisms could explain why attentional benefits for signal tones diminish on longer trials, as cortical representations shift toward unexpected stimuli when more evidence is available.

Although neural recordings were not collected in this study, the behavioural results are consistent with models of flexible STRF reorganisation. Under temporal constraints, cortical tuning may remain focused around the expected signal frequency, supporting faster and more accurate responses to signals. As tone duration increases, however, tuning may gradually shift toward salient, unexpected features, facilitating improved detection of probe tones.

Our psychophysical thresholding experiment aligns with the idea that perceptual difficulty modulates attentional dynamics. The estimated point of subjective equality (PSE) between "short" and "long" durations was approximately 154 ms. Stimuli near threshold are likely to increase task difficulty and cognitive load, conditions that have been associated with stronger attentional modulation of STRFs (Atiani et al., 2009).

Using durations around this perceptual boundary in future studies could amplify the neural signatures of attentional reorientation, offering a sensitive assay for dynamic changes in frequency-specific tuning during auditory discrimination tasks.

At the same time, it remains unclear whether the shift from signal-focused to probe-enhanced performance would persist under heightened perceptual strain. Increased task difficulty may reinforce top-down expectations, maintaining a narrow attentional filter around the signal frequency even at longer durations. Systematically varying tone duration around the PSE could clarify how perceptual ambiguity interacts with temporal constraints to shape attentional flexibility.

Finally, while STRF plasticity has been well documented over multi-second timescales, the durations used in this study ranged from only 50 to 400 ms. This raises an open question: can frequency-specific receptive field tuning reorganise within the sub-second timeframes of individual trials? If so, this would provide strong evidence for rapid and flexible attentional mechanisms capable of reorienting in real time to salient or unexpected stimuli. Testing this possibility through neural recordings during behaviour would offer valuable insights into the temporal granularity of attentional filtering in the auditory system.

A key limitation of the present study is the reduced sample size. Although four ferrets were initially trained, only two completed all task variations due to unrelated health issues. While mixed-effects modelling helped account for repeated measures and individual variability, a larger cohort would enhance the generalisability of the findings and allow firmer conclusions about species-typical attentional strategies. Future work should also incorporate a repeated-measures

design in humans aligned more closely with the ferret training structure, to directly test whether and how attentional allocation shifts across sessions.

Another methodological constraint concerns the use of multiple low-probability probe frequencies in the first three task variations. Each probe was presented on only 5% of trials, limiting trial numbers per condition and introducing variability in perceptual distance from the signal. This design choice was modelled after earlier human studies, such as Mondor and Bregman (1994) and Luthra et al. (2024), to maintain comparability across species. While group-level analyses revealed clearer and more consistent probe signal effects than frequency-level models, future studies might consider using a single, well-characterised probe at a higher probability (e.g., 20%) to maximise statistical power. However, this would come at the cost of losing the ability to examine differences across probes, and should therefore be weighed against the specific goals of the experimental design.

The training structure also introduced potential biases. In Variation 1, ferrets were tested using the same frequency set employed during training, with 3000 Hz consistently reinforced as the signal frequency. This may have encouraged a familiarity-based bias. By contrast, later variations introduced novel signal frequencies (Variations 2 and 4) or session-wise counterbalancing (Variation 3), which likely reduced reliance on memorised tones. Variation 1 did not exhibit the crossover interaction between stimulus type and duration seen in other variations, instead showing better performance for both signals and probes at long durations. This outcome could reflect either an early difficulty effect or an overlearned attentional bias towards familiar tones. Future applications of this paradigm should avoid reusing training frequencies in the test phase and counterbalance signal

assignments to disentangle the effects of familiarity from those of statistical expectation.

Despite these limitations, this study provides the first demonstration of a probe signal effect in a non-human species using a duration discrimination task. By showing that short trials favour accuracy for expected tones while long trials facilitate detection of unexpected probes, the results highlight a dynamic and temporally flexible deployment of auditory attention. These findings lay the groundwork for future studies integrating behavioural and neural recordings to trace the real-time evolution of attentional filters, and to test how statistical expectations interact with stimulus properties under changing temporal constraints.

4 VERTICAL AUDITORY SPATIAL DISCRIMINATION IN MICE IN BOTH PASSIVE AND ACTIVE LISTENING DESIGNS

4.1 INTRODUCTION

Sound localisation enables mammals to detect prey and predators and plays an important role in social communication (Blauert, 1997). It also contributes to auditory attention, as exemplified in the “cocktail party effect”, where listeners must monitor relevant voices against competing speakers and background noise (Grothe et al., 2010; Woods & McDermott, 2018). The present study focuses on vertical sound localisation, a dimension less commonly explored.

To localise sounds, mammals rely on two types of cues: binaural and monaural cues (Schnupp et al., 2011). Binaural cues are made up of interaural time differences (ITD), and interaural level differences (ILD). Sounds that are off centre, i.e. not perfectly located between both ears, will arrive first at the ear closer to the sound origin, and later at the far ear. This time difference in sound information arriving at each ear makes up the ITD. Additionally, soundwaves, particularly those of high frequencies, are more intense at the ear closer to the sound location, because the head and torso cast an acoustic shadow creating differences in sound intensity at the ears (ILD). These binaural cues are thought

to be primarily used for horizontal sound localisation, though to an extent, also play a role in vertical localisation (Carlile, 2014; Parsons et al., 1999).

Monaural, pinna-based cues play a primary role in vertical sound localisation and can also become important when binaural cues are ambiguous. These cues arise from the direction-dependent filtering by the outer ear, which attenuates specific frequency bands and creates spectral notches that vary systematically with the elevation of the sound source (see Figure 90 for examples; Blauert, 1969; Gardner & Gardner, 1973; Grothe et al., 2010; Musicant et al., 1990).

Both binaural and monaural spatial cues are captured in Head-Related Transfer Functions (HRTFs), which characterise how incoming sound is filtered by the listener's body, head, and outer ears before reaching the tympanic membrane. Each HRTF corresponds to a specific source location and ear, reflecting the direction-dependent filtering unique to that acoustic path. While individual HRTFs describe the monaural filtering for each ear, ILD and ITD can be derived by comparing the pair of HRTFs for the same source location. In this way, HRTFs collectively encompass both binaural and monaural spatial information (Grothe et al., 2010; Stitt et al., 2019). Although this framework applies broadly across mammalian species, the relative contribution of different cues can vary. In small rodents such as mice, ITDs are negligible due to their small interaural distance, and ILDs are limited by their lack of low-frequency hearing range. Consequently, mice are thought to rely more heavily on direction-dependent spectral filtering by the pinnae for sound localisation (Allen & Ison, 2010; Behrens & Klump, 2016; Ehret & Dreyer, 1984; Ito et al., 2020).

Monaural cues, and by extension HRTFs, can be substantially altered by manipulations of the outer ear, often rendering vertical localisation markedly impaired (Parsons et al., 1999). While prior research in humans has shown that localisation performance can recover over time, presumably via cue reweighting or relearning (Hofman et al., 1998), it remains unclear whether similar adaptive processes can occur in other species. The present chapter investigates the impact of bilateral pinnae manipulation on vertical sound localisation in mice, assessing both the immediate consequences of altered HRTFs and the extent to which animals can compensate or relearn by employing a passive oddball paradigm and an active go/no-go discrimination task.

Oddball tasks present sequences of repeated “standard” stimuli interspersed with infrequent “deviant” stimuli. These paradigms are commonly used to probe auditory novelty detection, selective attention, and predictive processing (Alanazi et al., 2023; Voola et al., 2022). A hallmark of the oddball paradigm is the *oddball effect*, whereby deviant stimuli elicit larger neural or other physiological responses than standard stimuli (Kiani et al., 2022; Tabas & von Kriegstein, 2021). This stands in contrast to the *probe-signal effect* discussed in other chapters, where behavioural performance is typically *better* for expected stimuli and *worse* for unexpected ones. Here, we use a vertical spatial oddball paradigm to assess sound discrimination in passively listening mice, using pupillometry to track stimulus-evoked changes in pupil diameter.

Pupil diameter is a sensitive indicator of arousal and cognitive processes such as attention and surprise and is influenced by neuromodulators including noradrenaline and acetylcholine (Aston-Jones & Cohen, 2005; Larsen & Waters,

2018). In mice, phasic pupil dilations, so called microdilations, occur even in the absence of overt movement and are thought to reflect task engagement and stimulus processing (McGinley, Vinck, et al., 2015; Reimer et al., 2016). Importantly, pupil responses have been linked to violations of expectation and decision uncertainty in both humans and animals (Alnæs et al., 2014; Montes-Lourido et al., 2021; Preuschoff et al., 2011), and have recently been validated as reliable markers of auditory novelty and discrimination across stimulus types in animal models (Montes-Lourido et al., 2021). In this context, pupillometry offers a non-invasive measure of deviance detection, making it well-suited for use in our passive oddball paradigm.

Unlike pupil diameter, which is a well-established measure of arousal state in mice (de Gee et al., 2020; McGinley, David, et al., 2015; McGinley, Vinck, et al., 2015; Reimer et al., 2016), ear movements have received comparatively little attention in the neuroscience literature. While some studies have assessed facial movements, including ears, this work has largely centred on nociception, most notably through the development of the Mouse Grimace Scale (Langford et al., 2010). According to this scale, a "happy" mouse holds its ears in a neutral, upright position, while increasing pain causes the ears to move back and towards the fur. Beyond this literature, only a small number of studies have explored facial expressions in response to other emotional states: one in mice (Dolensek et al., 2020) and one in rats (Finlayson et al., 2016). These studies demonstrate that facial musculature, including ear position, is modulated by affective states such as pleasure, malaise, and fear. Additionally, anecdotal reports in pet-owner guides

suggest that ear posture may indicate internal states such as vigilance or contentment (BeChewy, 2014; J. Walker, n.d.), hinting at a broader role of ear movements in response to auditory stimuli. To our knowledge, the present experiments represent the first systematic attempt to quantify ear movements in response to auditory stimuli in mice. The contribution of ear posture and kinematics to auditory processing, particularly in mice, has remained largely unexplored. Whether ear movements reflect internal states related to arousal and attention remains an open question. Given their accessibility and dynamic nature, ear movements may offer a promising but underexplored behavioural index of sensory processing. The current study begins to address this possibility by examining ear position in the context of an auditory oddball paradigm, alongside established pupillometry, and by quantifying ear movement trajectories during a go/no-go discrimination task, assessing their relationship with stimulus timing and behavioural responses.

Together, these experiments address whether gross alterations of the pinnae impair vertical sound localisation in mice, and whether animals can relearn localisation following such disruption. By combining pupillometry and ear tracking, this work contributes to our understanding of spatial hearing and its underlying mechanisms thus addressing a largely understudied area. More broadly, we aim to establish mice as a viable model for investigating vertical sound discrimination, an ability often assumed to be limited in this species (Lauer et al., 2011), thereby opening new avenues for mechanistic and circuit-level studies of spatial hearing.

4.2 VERTICAL ODDBALL PARADIGM: ACROSS INTACT AND PINNAE-MANIPULATED COHORTS

To investigate the impact of pinnae manipulation on vertical sound processing, we first employed a passive oddball paradigm using pupillometry to measure deviance detection. By comparing control mice with littermates who had undergone bilateral pinnae removal, we assessed whether disrupting monaural spectral cues modulates physiological responses to unexpected changes in sound elevation. A major strength of this paradigm is that it does not require training: mice can be tested shortly after habituation to the experimental setup, allowing for efficient and minimally invasive assessment of auditory processing. Furthermore, because deviance detection reflects sensitivity to statistical regularities, this task links spatial hearing with broader themes of auditory attention and predictive processing that are addressed throughout this thesis.

4.2.1 Methods

4.2.1.1 Animals

Surgical and animal handling procedures were carried out in accordance with the ethical guidelines of the National Institutes of Health (NIH; USA) and were approved by the Institutional Animal Care and Use Committee (IACUC) of Baylor College of Medicine (Houston, Texas, USA). Eight male C57bl/6j mice aged 8-

weeks old (JAX: 000664; The Jackson Laboratory, Bar Harbor, ME, USA) were used for this experiment. Mice received ad libitum food and water and were individually housed after headpost implantation. Mice were kept on a regular light-dark cycle, and all experiments were done during the light phase.

4.2.1.2 Surgical Procedures and Habituation

Custom titanium head posts were implanted under isoflurane anaesthesia following preoperative administration of buprenorphine extended release and fluids. In a subset of animals (n=4), the external pinnae were bilaterally removed, leaving ~ 2 mm of the base intact. These mice received dexamethasone during surgery and topical antibiotic treatment post-operatively.

Animals were given two days to recover before habituation commenced. On postoperative day 3, mice were introduced to the head-fixation setup in the experimental booth without auditory stimulation. Habituation sessions were conducted daily, followed by forage rewards in their home cage. These procedures were carried out in line with recent recommendations to refine rodent head-fixation to improve welfare and engagement (Barkus et al., 2022). During head fixation, mice were positioned on a freely rotating cylindrical treadmill to permit spontaneous locomotion.

4.2.1.3 Apparatus

The experiment was conducted in a walk-in sound-attenuated chamber using a setup mounted on a vibration-isolated optical table to minimise mechanical vibrations. Two loudspeakers were mounted on a shared post positioned 30° to the right in the horizontal plane, at a distance of 50 cm. The bottom speaker was aligned with the animal's azimuth, and the top speaker was elevated by 60° and angled downward toward the animal.

Pupil and ear movements were recorded using two infrared-sensitive cameras positioned to the right of the mouse. The pupil camera recorded at 15 Hz and the ear camera recorded at 50 Hz. Illumination was provided by infrared LED lights and a ring light to ensure image quality. Pupil size was further modulated by a dim near-ultraviolet LED (405–410 nm) placed above the mouse, which was adjusted to produce mid-range pupil diameters under steady-state lighting conditions and by passive wakefulness, indicated by upright posture and lack of locomotion.

Locomotor activity was monitored with a rotary optical encoder mounted to the treadmill axis. While the pupil data were recorded without issue, an in-progress hardware-software optimisation led to intermittent frame loss in the ear camera. Although this reduced the temporal resolution of the ear tracking, the data retained sufficient quality for exploratory analysis.

4.2.1.4 Design

The experiment was designed as a passive oddball task using custom LabVIEW code, in which head-fixed mice were exposed to sequences of pink noise bursts (50 dB SPL) presented from either the top or bottom speaker. Each stimulus was 150 ms in duration with an interstimulus interval (ISI) of 2850 ms. The experiment consisted of three sessions conducted on consecutive days, with the orders of sessions randomised across animals.

Each session consisted of 300 trials and lasted approximately 15 minutes. 90% of trials were played from the “standard” speaker location, while the remaining 10% were played from the unexpected, “deviant” speaker location. Mice were exposed to three session types: (1) a *top speaker deviant* condition, in which the top speaker (60° elevation) served as the unexpected deviant location, while the bottom speaker served as the standard location; (2) a *bottom speaker deviant* condition, with the reverse proportions; and (3) a *control* condition, in which both speakers presented stimuli with equal probability (50% each). The primary analysis focused on within-condition comparisons of pupil responses to standard and deviant trials to assess the presence of an oddball effect and its modulation by pinnae removal.

4.2.1.5 Pupillometry Analysis

Pupil size and exposed eye area were measured from video footage taken during the task using an established model from the McGinley lab (Mridha et al., 2021) using DeepLabCut (Mathis et al., 2018). The network was trained on approximately 1000 manually labelled frames, with eight markers placed around the pupil and eight around the eyelids, ensuring even spacing for accurate tracking. Figure 30 shows an example set of image frames from the preprocessing output, illustrating the detected markers used for analysis.

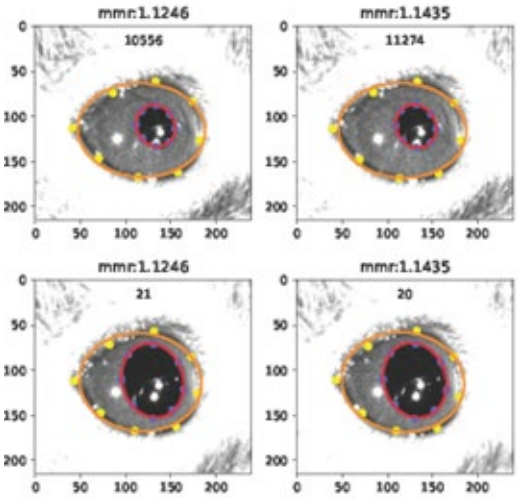

Figure 30. Example frames showing 8-point ellipses around eye and pupil. Each point represents a marker as described. Frames originate from oddball data collected for this thesis, not training data.

To correct for missing data due to blinks, frames were excluded from ellipse fitting if two or more markers had a DeepLabCut-assigned likelihood (confidence score) below 0.1. Blink detection was further defined using a custom algorithm that identified outliers in the z-scored temporal derivative of the pupil size time series. Missing data were linearly interpolated within a 150 ms window. The time

series was then resampled to 50 Hz, low-pass filtered at 3 Hz, and normalised to a percentage of the 99.9th percentile of the time series. For analysis, pupil dilation was computed in 20 ms bins across trials, ensuring temporal alignment with other data variables. This binning allowed for within-trial comparisons of dilation timing and magnitude relative to sound presentation while maintaining consistency across conditions. Pupil diameter values were baseline-subtracted, and all data presented in this chapter reflect the percentage change in dilation from baseline.

4.2.1.6 Exploratory Eye Tracking and State Analysis

In addition to pupil tracking, a secondary camera recorded ear movements for exploratory analysis. Initial pilot data from the oddball task were used to train the ear tracking algorithm in Lightning Pose (Biderman et al., 2024). Each frame was labelled with five markers: two at the base of the ear where the pinna meets the head, one at the apex of the pinna, one on the posterior surface, and one at the eye (Figure 31).

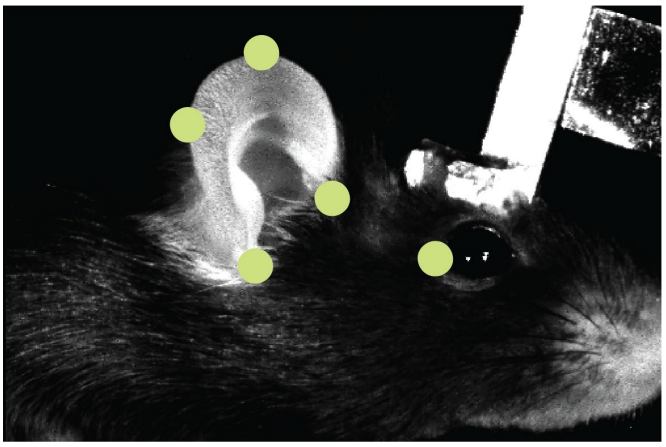

Figure 31. Example Frame with Position of Five Markers Used for Analysis.

To characterise ear kinematics, both ear position (X, Y) and velocity (X, Y) were tracked in a two-dimensional Euclidean space for each marker. The 20-dimensional feature vectors (X and Y positions and velocities for five tracked markers), were processed separately for each mouse and session using t-distributed Stochastic Neighbour Embedding (t-SNE). This projection mapped the high-dimensional kinematics into a three-dimensional latent, preserving local structure while facilitating clustering and visualization. To identify distinct ear positions, k-means clustering was applied within the t-SNE space.

4.2.2 Results

4.2.2.1 Pupillometry Results: Deviance Detection and Spatial Saliency Effects

4.2.2.1.1 Control Condition: Spatial Saliency Bias in Pupil Responses

Independent-samples Welch's t-tests (

Table 4) revealed significantly greater peak pupil dilation and delayed response latency for sounds originating from the top speaker location compared to the bottom speaker, in both cohorts. This occurred despite equal presentation probabilities across locations in the control condition, indicating an intrinsic spatial salience bias whereby sounds originating from above elicited stronger pupil responses than those presented at azimuth (Figure 32).

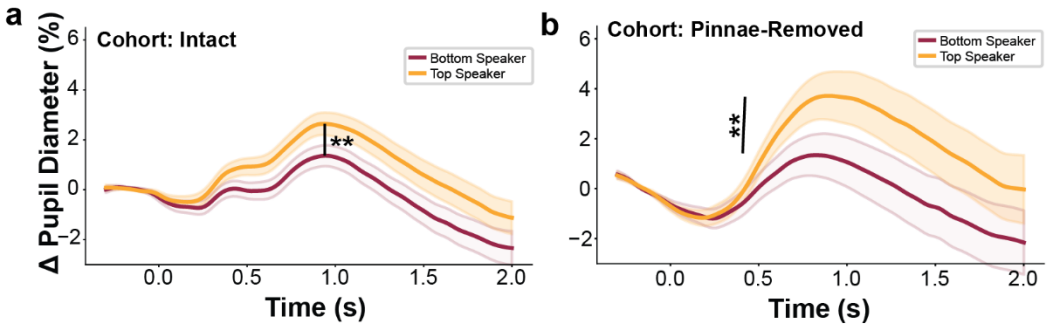

Figure 32. Pupil Responses to Equally Probable Sounds from the Top and Bottom Speakers in the Control Condition. (a) Pinnae-intact cohort, **(b)** Pinnae-removed cohort. Coloured lines indicate speaker location. Shaded areas indicate ±95% confidence intervals. X = 0 indicates sound onset.

This baseline asymmetry has important implications for interpreting the oddball conditions. Specifically, it introduces a confound where deviant responses may appear enhanced simply because they originate from the more salient top speaker, while standard responses may be inflated when the deviant comes from the less salient bottom speaker. Consequently, observed effects in the oddball task must be interpreted in light of this elevation-driven response bias, especially since it was still present in mice whose pinnae-based cues were significantly disrupted.

Table 5. Pupil Dilation and Peak Timing Across the Control Condition and Cohorts. Abbreviation "btm" stands for trials where sounds originated from the bottom speaker and "top" for trials where sounds originated from the top speaker. Speaker location was equiprobable in the control condition. Significant results are indicated with *.

Condition	Cohort	M_{pupil} (btm)	M_{pupil} (top)	t_{pupil}	p_{pupil}	$M_{peakTime}$ (btm)	$M_{peakTime}$ (top)	$t_{peakTime}$	$p_{peakTime}$
Control	Intact	4.608	5.586	-2.896	0.004*	0.846	0.937	-2.72	0.007*
Control	Removed	6.578	8.698	-3.026	0.003*	0.734	0.859	-2.87	0.004*

4.2.2.1.2. Deviance Detection in Intact and Pinnae-Removed Mice

In the pinnae-intact cohort, deviant trials originating from the top speaker elicited significantly greater pupil dilation and delayed peak responses relative to standard trials (Figure 33). However, given the spatial salience bias observed in the control condition, this apparent oddball effect likely reflects enhanced responses to elevated sounds rather than genuine sensitivity to statistical deviance. When the deviant was presented from the bottom speaker, no amplitude difference was observed, although deviant trials showed significantly delayed peak dilation relative to standards (Table 6).

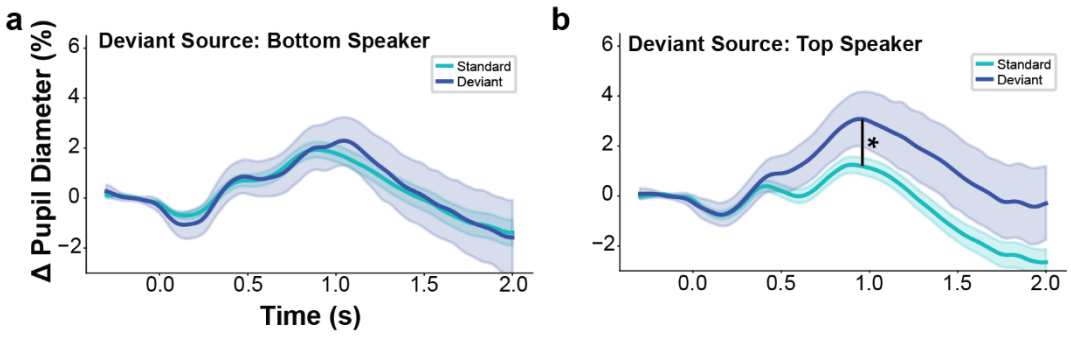

Figure 33. Pupil Responses in the Intact Cohort for (a) bottom speaker deviant and **(b**) top speaker deviant conditions. Shaded areas reflect ±95% confidence intervals. Asterisk indicates significant difference. Coloured lines correspond to stimulus type. X=0 represents time onset. Pupil diameter traces are baseline-subtracted.

Table 6. Pupil Dilation and Peak Timing Across the Oddball Conditions for each Cohort. Abbreviation btm stands for trials where sounds originated from the bottom speaker and top for trials where sounds originated from the top speaker. The deviant speaker mean is bolded. Significant results are indicated with *.

Condition	Cohort	M_{pupil} (btm)	M_{pupil} (top)	t_{pupil}	p_{pupil}	$M_{peakTime}$ (btm)	$M_{peakTime}$ (top)	$t_{peakTime}$	$p_{peakTime}$
Bottom deviant	Intact	**4.99**	4.95	-0.08	0.93	**0.99**	0.87	- 2.46	0.015*
Top deviant	Intact	4.68	**6.10**	-2.17	0.031*	0.81	**0.92**	-2.01	0.046*
Bottom deviant	Removed	**5.35**	9.94	4.67	0.00001*	**0.73**	0.92	2.65	0.009*
Top deviant	Removed	6.86	**7.92**	-0.998	0.32	0.88	**0.91**	-0.44	0.658

The pinnae-removed cohort exhibited no significant difference in pupil dilation amplitude or peak latency in the top speaker condition (Table 6, Figure 34). In the bottom deviant condition, however, top-originating standards evoked greater and later pupil responses than bottom-originating deviants. The absence of a deviance effect, but the persistence of an elevation-driven bias in mice with significantly altered pinnae, suggests that residual localisation cues remained accessible. The small portion of pinnae tissue left intact at the base, approximately 2 mm in height, may have been sufficient to receive direction-dependent spectral filtering, allowing for partial preservation of elevation-related information.

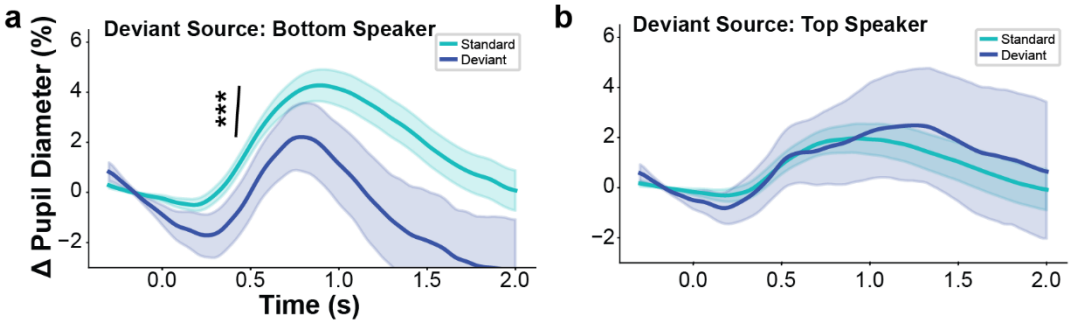

Figure 34. Pupil Responses in the Pinnae-Removed Cohort for (a) bottom speaker deviant and **(b)** top speaker deviant conditions. Shaded areas reflect ±95% confidence intervals. Asterisk indicates significant difference. Coloured lines correspond to stimulus type. X=0 represents time onset. Pupil diameter traces are baseline-subtracted.

4.2.2.1.2.1 Direct Comparison of Deviant and Standard Trials Within Speaker Location

We conducted an additional analysis to disentangle the effects of spatial location from deviance probability. Trials were selected such that standard and deviant tones originated from the same physical speaker, using data from separate oddball blocks in which the speaker served either as the deviant or the standard source. This approach allowed us to isolate the contribution of deviance while controlling for spatial position. As shown in Figure 35, pupil dilations in the intact cohort were generally larger on deviant trials compared to standard trials from the same speaker location. In contrast, the pinnae-removed cohort showed the opposite pattern at the top speaker location, with greater dilation to standard tones. However, none of these comparisons reached statistical significance based on peak pupil amplitude (intact top speaker: $t(143) = –1.88$, $p = 0.062$; intact bottom speaker: $t(152) = –0.37$, $p = 0.709$; pinnae-removed top: $t(119) = 1.77$, $p = 0.079$; pinnae-removed bottom: $t(71) = 1.51$, $p = 0.135$), suggesting that deviance-related pupil effects cannot be explained by spatial location alone. Although no reliable within-speaker deviance effects were observed in either cohort, the pupil traces from pinnae-removed mice differed in shape and amplitude from those of intact animals. These differences may reflect altered spectral processing following

the loss or alteration of pinna-related elevation cues.

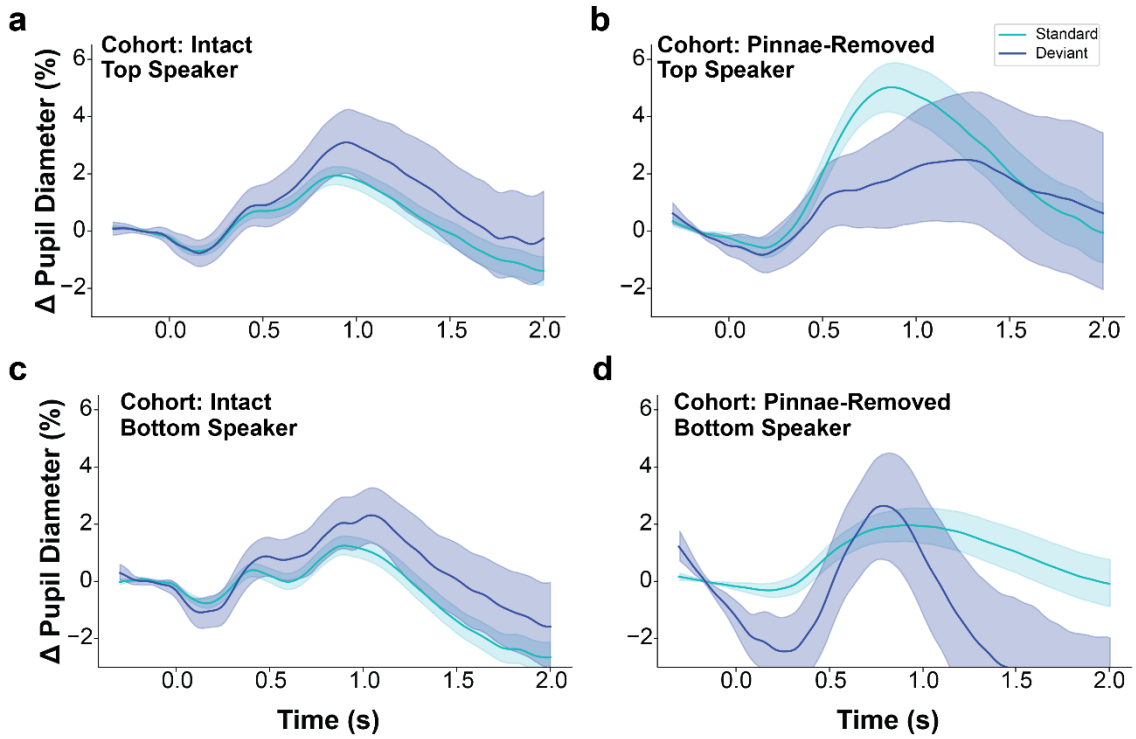

Figure 35. Pupil responses to deviant and standard tones originating from the same speaker location, drawn from separate oddball conditions. Mean pupil dilation time courses (±95% CI) are shown for deviant (blue) and standard (cyan) trials delivered from the same physical speaker location. Trials were pooled across two separate oddball conditions, one in which the speaker acted as the deviant source and one in which it acted as the standard source, allowing for direct comparison of deviant and standard responses while controlling for spatial location. **(a, b)** Top speaker responses for intact and pinnae-removed cohorts, respectively. **(c, d)** Bottom speaker responses for intact and pinnae-removed cohorts. No comparisons reached significance based on peak pupil dilation.

4.2.2.1.3 Controlling for Spatial Saliency Bias in the Oddball Effects

The absence of a consistent oddball effect in both the top and bottom speaker deviant conditions in the intact cohort, combined with the elevation-driven pupil bias observed in the control condition and the pinnae-removed cohort, suggests that spatial salience may have masked any underlying sensitivity to statistical deviance. To account for this confound, we adjusted pupil responses in the oddball conditions by subtracting a cohort-specific saliency index: the

difference in peak dilation between top and bottom speaker trials in the 50/50 control condition. This correction was applied only to top-originating sounds, specifically, to deviant trials in the top deviant condition and standard trials in the bottom deviant condition, thereby isolating the contribution of statistical deviance from vertical salience (Table 7).

In the pinnae-intact cohort, corrected data revealed a trend toward significance in the bottom deviant condition, suggesting that an underlying deviance effect may have been masked by the competing influence of spatial salience. By contrast, the initially significant effect in the top deviant condition was no longer statistically reliable following correction, indicating that this difference was predominantly driven by elevation rather than by deviance.

In the pinnae-removed cohort, the reversed pattern in the bottom deviant condition persisted after correction, with top-originating standards continuing to evoke significantly greater pupil responses than bottom-originating deviants. No difference was observed in the top deviant condition, reinforcing the conclusion that deviance detection was impaired by the loss of pinnae cues, while the salience bias for elevated sounds remained.

Table 7. Adjusted Peak Pupil Dilation and Statistical Comparisons for Deviant and Standard Trials Across Oddball Conditions. Mean peak pupil dilation responses are shown for standard (M_{std}) and deviant (M_{dev}) trials. The spatial saliency effect was subtracted from trials involving the top speaker (for bottom condition, M_{std} ; for top condition, M_{dev}). This yielded the corresponding M_{adj} for the top speaker trials. Significant t-tests comparing M_{adj} to the other trial type are indicated with an asterisk ($p<.05$).

Cohort	Condition	t	p	Spatial Saliency Effect (%)	M_{std} (%)	M_{dev} (%)	M_{adj} (%)
Intact	Bottom	-1.872	.063	0.978	**4.95**	4.99	**3.97**
Intact	Top	0.672	.503		4.68	**6.10**	**5.12**
Removed	Bottom	2.517	.014*	2.120	**9.94**	5.35	**7.82**
Removed	Top	-1.015	.313		6.87	**7.92**	**5.80**

4.2.2.1.4. Variability and Consistency of Pupil Responses

Given the impaired deviance detection observed in the pinnae-removed cohort, we next examined whether pupil responses differed in their trial-to-trial consistency. Initial visual inspection suggested greater variability in this group, with noticeably larger standard deviations and wider confidence intervals (Figure 34). To assess this systematically, we quantified trial-to-trial variability in peak pupil dilation using a two-factor analysis with cohort (intact vs. removed) and session condition (bottom deviant, top deviant, control) as factors. Indeed, the results showed significantly greater standard deviations in the pinnae-removed cohort across all conditions ($F_{(1,30)} = 203.12, p < .0001$).

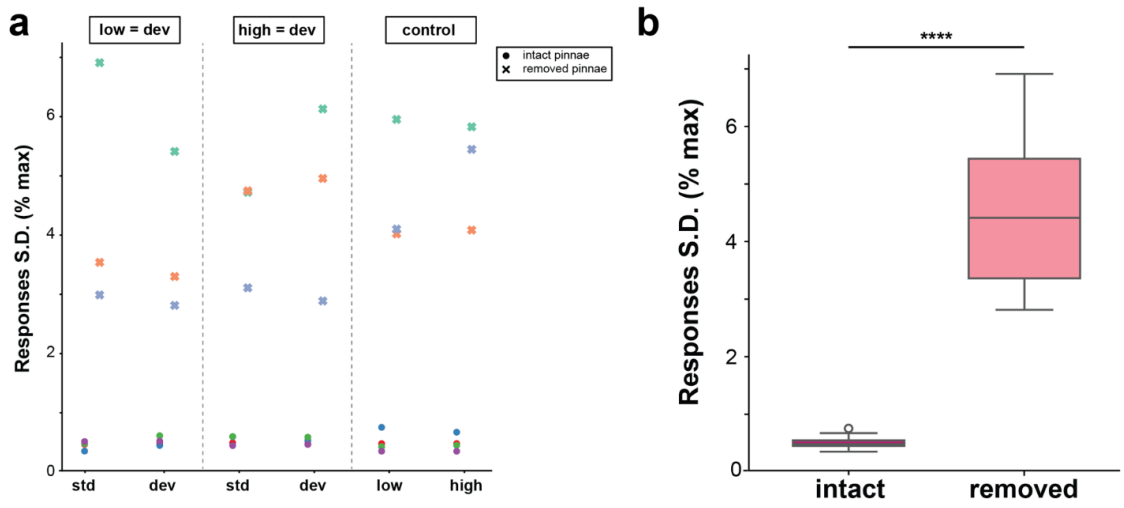

Figure 36. Standard Deviation (SD) of Pupil Responses Across Cohorts. (a) Standard deviation of pupil responses for individual mice, with circles representing the intact pinnae cohort and crosses representing the pinnae-removed cohort. SD is plotted separately for standard (std) and deviant (dev) trials within the experimental conditions (low speaker is deviant, high speaker is deviant) and between speaker location (low/high) for the control condition. The pinnae-removed cohort exhibits substantially higher trial-by-trial variability in pupil responses compared to the intact cohort. **(b)** Aggregated standard deviation data between conditions, plotted for each cohort. A two-way ANOVA revealed a significant difference in variability between the two cohorts ($p<0.0001$), as indicated by the significance bar.

These findings suggest that altered pinna-based spectral cues increased trial-to-trial variability in pupil responses. Although the pinnae-removed cohort continued to exhibit an elevation-related salience bias, the heightened trial-to-trial variability suggests that vertical spatial localisation was nonetheless impaired, likely due to increased perceptual uncertainty. This instability likely contributed to impaired deviance detection, where perceptual uncertainty may have diminished the salience of statistical irregularities in sound source location.

4.2.2.1.5. Modelling Pupil Size with Linear Mixed-Effects Models: Contributions of Stimulus Type and Behavioural Covariates

To complement the group-level analyses above, we fitted linear mixed-effects models at the single-trial level to examine whether observed pupil dilation differences could be attributed specifically to stimulus type, or whether they were better explained by behavioural and physiological covariates. Models were run separately for each condition and included fixed effects for stimulus type, baseline pupil size, peak pupil latency, wheel speed, and trial history (number of trials since last deviant), with a random intercept for mouse. This approach accounted for the hierarchical structure of the data (trials nested within mice) and addressed a key interpretive challenge: in the raw data, sounds from the top speaker often evoked greater pupil dilation regardless of deviance status. The models thus helped isolate deviance-related effects from other sources of variability. All continuous predictors were mean-centred to facilitate interpretation of model estimates and reduce multicollinearity (Figure 37; see also Figure 59 and Figure 60 in the Appendix for model diagnostics).

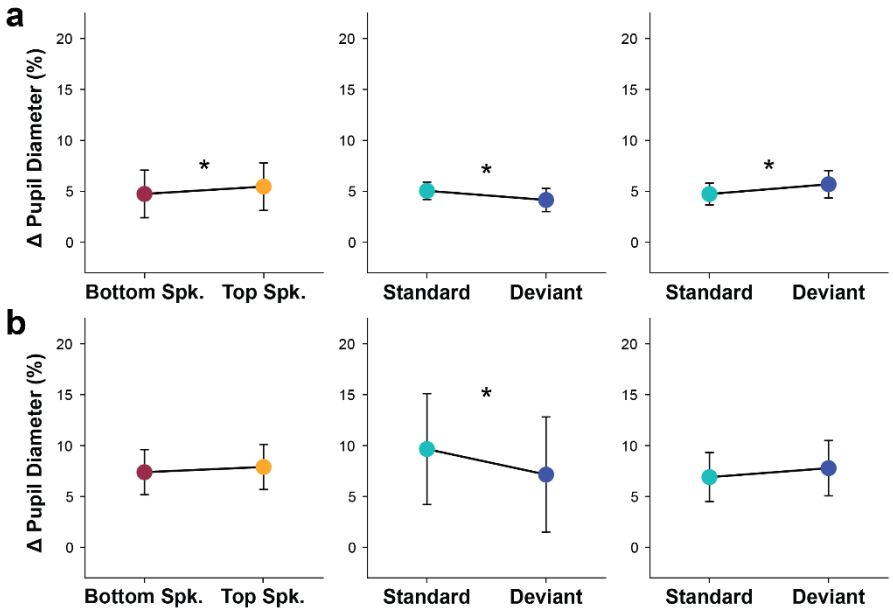

Figure 37. Model-Predicted Pupil Dilation Responses by Stimulus Type Across Spatial Deviance Conditions and Cohorts. Each subplot shows estimated marginal means (±95% confidence intervals) for modelled pupil dilation responses in the control (left), bottom speaker deviant (middle), and top speaker deviant (right) conditions. Data are shown separately for the **pinnae-intact (a)** and **pinnae-removed (b)** cohorts. Stimulus type reflects either spatial origin in control sessions (bottom vs. top speaker location) or deviance status in oddball sessions (standard vs. deviant). Asterisks indicate a significant main effect of stimulus type (*$p < 0.05$); "n.s" indicates non-significant effects. Marginal estimates are derived from linear mixed-effects models run separately for each condition.

Across both cohorts, delayed pupil peaks, faster wheel movement, and smaller baseline pupil diameters were consistently associated with larger stimulus-evoked pupil responses.

Stimulus type (standard vs. deviant) effects in the models largely paralleled the t-test results, with one notable exception. In the pinnae-intact cohort, the bottom speaker condition, initially non-significant in the raw t-test comparison, approached significance following adjustment for saliency bias and reached significance in the LME model. This convergence across methods suggests that the intrinsic bias, together with behavioural variability across trials, may have

obscured a true oddball effect, which became detectable only when behavioural covariates were statistically controlled. In the pinnae-removed cohort, only the bottom speaker condition showed a significant stimulus effect, reflecting a persistent elevation bias that remained evident across unadjusted comparisons, bias-corrected analyses, and mixed-effects modelling. The consistency of this effect across all analytic approaches underscores its robustness and further supports the conclusion that deviance detection was absent in this cohort. Notably, model fit was poorer in this group, with higher leave-one-out (LOO) values and wider confidence intervals, consistent with increased variability and greater susceptibility to internal state fluctuations following disruption of pinnae-based cues.

Table 8. Summary of Fixed Effects from Linear Mixed Effects (LME) Models Predicting Pupil Dilation. Models were fit separately for each cohort (pinnae-intact, pinnae-removed) and condition (Top Speaker Deviant, Bottom Speaker Deviant, Control). A checkmark (✓) indicates a significant effect ($p < .05$); a tilde (~) indicates a trend-level effect ($p \approx .059$); and a cross (✗) indicates a non-significant effect. All models included random intercepts for mouse ID. Trial history reflects whether the previous trial was a deviant. For full coefficients and confidence intervals, see Table 72–Table 77 in the Appendix.

Condition	Cohort	Stimulus Type	Peak Pupil Time	Wheel Speed	Baseline Pupil	Trial History
Control (50/50)	Intact	Dev > Std	✓	✓	✓	✗
	Pinnae-Removed	✗	✓	✓	✓	~ (p = .059)
Bottom Speaker Deviant	Intact	Std > Dev	✓	✓	✓	✗
	Pinnae-Removed	Std > Dev	✓	✓	✓	✗
Top Speaker Deviant	Intact	Dev > Std	✓	✓	✓	✗
	Pinnae-Removed	✗	✓	✓	✓	✗

4.2.2.2 Exploratory Analysis: Joint Dynamics of Ear Kinematics and Pupil States

Ear tracking data from two 30-minute sessions from a separate cohort of oddball mice with intact pinnae, were used to explore the relationship between ear kinematic states and pupil size, as well as how these features vary in relation to sound presentation during passive listening.

K-means clustering applied to ear and velocity data identified five distinct ear states (Figure 38a). Clusters 1 and 2 were characterised by high and moderate ear velocity, respectively, and did not correspond to distinct ear tip positions (Figure 38b–d), suggesting that they reflect dynamic ear movements rather than postural states. Clusters 3–5 exhibited stable ear positions, forming distinct spatial groupings and corresponding to identifiable ear postures (Figure 38c–d, f). These included retracted, perked, neutral, and protracted states. Movement-related clusters (1–2) were more likely to occur during walking (Figure 38e), confirming that they reflect active motor states. Still-frame examples (Figure 38f) illustrate the spatial configuration of the ears in each cluster, supporting their classification into either dynamic or postural categories.

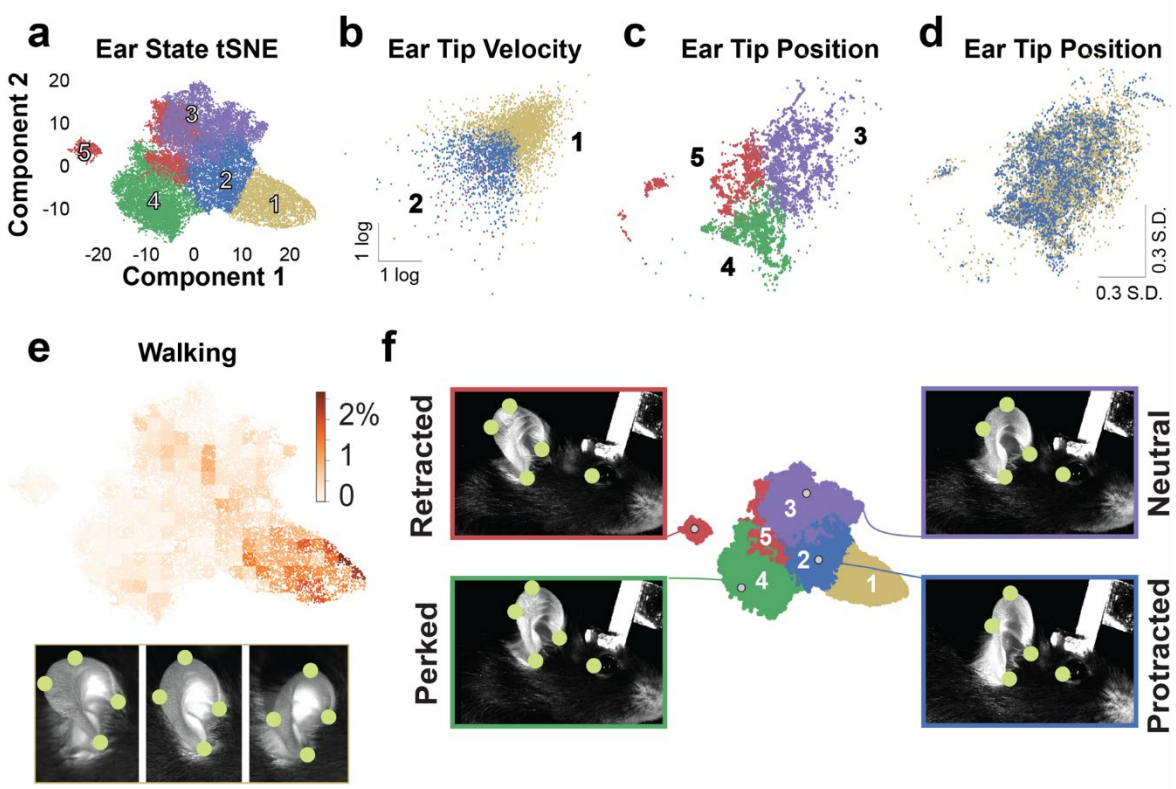

Figure 38. Identification of Ear States Using t-SNE and k-Means Clustering. (a) t-SNE representation of ear kinematics, showing five distinct clusters. **(b)** Scatter plot of x vs. y velocity for the ear tip, showing that clusters 1 and 2 correspond to ear movement velocity. **(c)** Scatter plot of x vs. y ear tip positions for clusters 3-5, demonstrating that these clusters correspond to stable ear positions. **(d)** Scatter plot of ear tip positions for clusters 1-2, showing that movement-related states exhibit more widely distributed ear positions. **(e)** Probability of being in movement-related clusters 1-2 during walking, example images of cluster 1 below. **(f)** Example frames corresponding to k-means clusters associated with stable ear positions.

To examine the relationship between arousal and ear state, we analysed the distribution of small, medium, and large pupil diameters across the t-SNE space (Figure 39, top row). Large pupil diameters were predominantly associated with cluster 1, indicating that this high-velocity ear state is linked to heightened arousal. Small pupils were most frequently found in cluster 3, corresponding to a neutral posture and likely reflecting a resting state. Medium pupil diameters, the most common category, were distributed across clusters 2–5, including perked and retracted states, suggesting intermediate arousal levels and stable posture.

The dynamics of ear movement relative to sound onset are illustrated in Figure 39 (bottom row). Before and after sound onset, ear velocity was most associated with clusters reflecting stable positions (e.g. perked or retracted). At sound onset, however, ear velocity increased within cluster 2, suggesting that this cluster may reflect a slow, directed movement toward the sound source, potentially an orienting response.

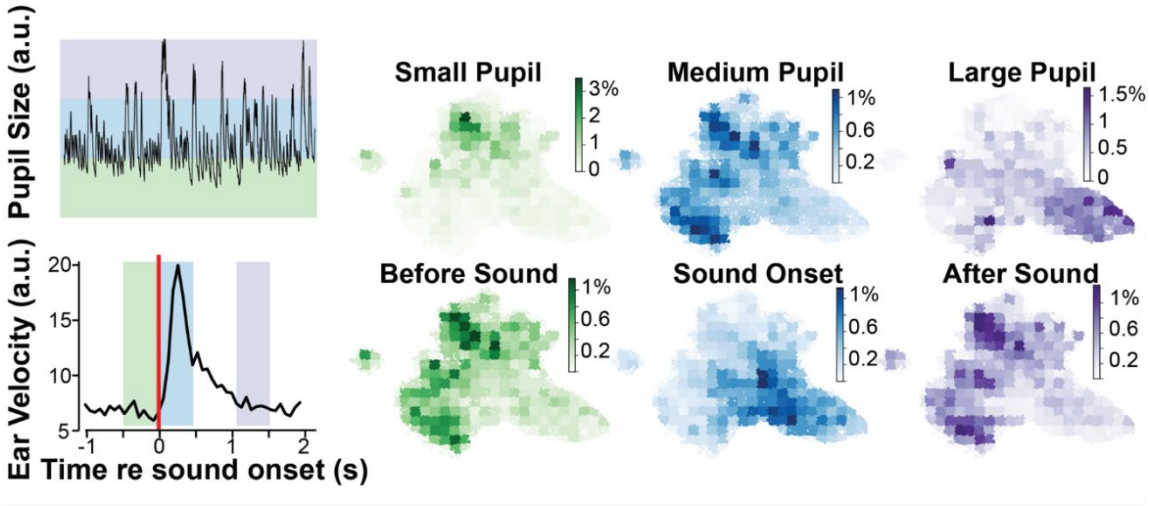

Figure 39. Mapping Pupil Size and Ear Velocity onto Ear State Clusters. *(***Top row***)* Distribution of small, medium, and large pupil sizes within the t-SNE space. *(***Bottom row***)* Ear velocity mapped onto the t-SNE space relative to sound presentation, showing distinct patterns before, during, and after sound onset.

In summary, cluster 1 reflects a state of high arousal and locomotion, defined by large pupil size, high ear velocity and walking behaviour. Cluster 2 encompasses slower ear movements that are especially pronounced at sound onset, suggesting that it may reflect orienting responses. However, the ear may also be in this cluster outside of sound-evoked periods, but such instances have not yet been systematically examined. Clusters 3-5 represent stable ear postures associated with mid-range pupil diameters and periods of relative resting state. These findings support the interpretation that ear kinematics index behavioural

and sensory responsiveness in mice. See Table 9 for a summary of the ear states.

Table 9. Summary of Ear State Clusters Derived from K-Means Clustering of Ear Tip Velocity and Position.

Cluster	Movement Type	Posture Characteristics	Associated Arousal (Pupil Size)	Behavioural Context	Sound Responsiveness
1	High velocity	—	Large	Locomotion	Briefly increased at sound onset; may reflect locomotor response to sound onset
2	Moderate velocity	—	Medium	Listening	Increases at sound onset (possible orienting response)
3	—	Neutral	Small	Resting	—
4	—	Perked	Medium	Stable posture	—
5	—	Retracted	Medium	Stable posture	—

4.3 BEHAVIOURAL VERTICAL SOUND DISCRIMINATION

This experiment involved the development and validation of a novel vertical auditory discrimination task for mice, using a custom training protocol designed to assess adaptive sound localisation following pinnae manipulation. The primary objective was to determine whether mice could relearn vertical sound localisation after their monaural spectral cues were altered by modifying the external ear. Vertical localisation critically depends on these spectral cues, which are shaped by the filtering properties of the pinnae (Grothe et al., 2010). By altering the pinnae, we aimed to disrupt these cues and assess whether mice could adapt to the modified spectral information and successfully perform the task using a head-fixed go/no-go paradigm. Research indicates that individuals can adapt to spectral modifications over time, relearning to accurately localise sounds despite altered cue processing (Van Wanrooij & Van Opstal, 2005).

To ensure that the chosen speaker locations would support reliable task performance, spatial separation was based on prior behavioural data showing that mice reach or slightly exceed a *d′* of 1 when discriminating sounds separated by 90° in elevation (Lauer et al., 2011). This angle was therefore used in the current task to provide sufficient baseline performance for detecting the effects of pinnae manipulation and subsequent relearning. The task was initially piloted as part of a grant application, during which quinine punishment was introduced once animals had reached high hit rates. Observations from ongoing training cohorts confirmed

that this timing was optimal, as early quinine exposure led to poorer performance and reduced task engagement.

4.3.1 Methods

4.3.1.1 Animals

Ten adult mice were used in this experiment: eight male CBA/CaJ (JAX:000654) and two female B6.CAST-Cdh23^{Ahl+}/Kjn mice (JAX:002756), all obtained from The Jackson Laboratory. CBA/CaJ mice were selected because they exhibit age-related hearing loss later in life compared to C57BL/6J mice, typically beginning around 16 to 18 months (Park et al., 2024), whereas B6.CAST-Cdh23^{Ahl+}/Kjn mice carry a protective allele that prevents early-onset hearing loss (Johnson et al., 1997).

The inclusion of the second strain was motivated by handling challenges presented by the CBA mice, which frequently exhibited hyperactivity and escape-prone behaviour. These traits were described in the JAX Handbook (2009) but are not explicitly listed on the Jackson Laboratory website. Despite handling challenges, CBA mice ultimately performed well in the task. Due to limited availability of B6.CAST mice, only two females could be obtained, resulting in an imbalance in both strain and sex. However, the primary analyses focused on the CBA cohort, as they constituted the majority of the dataset and successfully progressed to the pinnae-manipulation stage.

Mice were food-restricted for five consecutive days each week, with ad libitum water access. Food was available ad libitum in home cages from Friday afternoon to Sunday morning. During food restriction, body weight was monitored daily, and post-session feeding was adjusted to maintain animals at appropriate weights.

4.3.1.2 Apparatus and Stimuli

Mice were trained in a head-fixed go/no-go auditory discrimination task while freely running on a wheel which reduced stress associated with head restraint and allowed for more naturalistic behaviour. The apparatus was identical to that described in 4.2.1.3, except for the use of a custom-built speaker dome containing 32 equally spaced speakers (Figure 40), that replaced the previous two-speaker pole configuration. A custom-designed three-spout lick sensor was used to monitor behavioural responses and administer fluids. One spout delivered 10% sucrose solution as a reward (6 µL); the second dispensed 100 µM quinine solution as punishment for false alarms (3 µL), and the third removed residual liquids via suction.

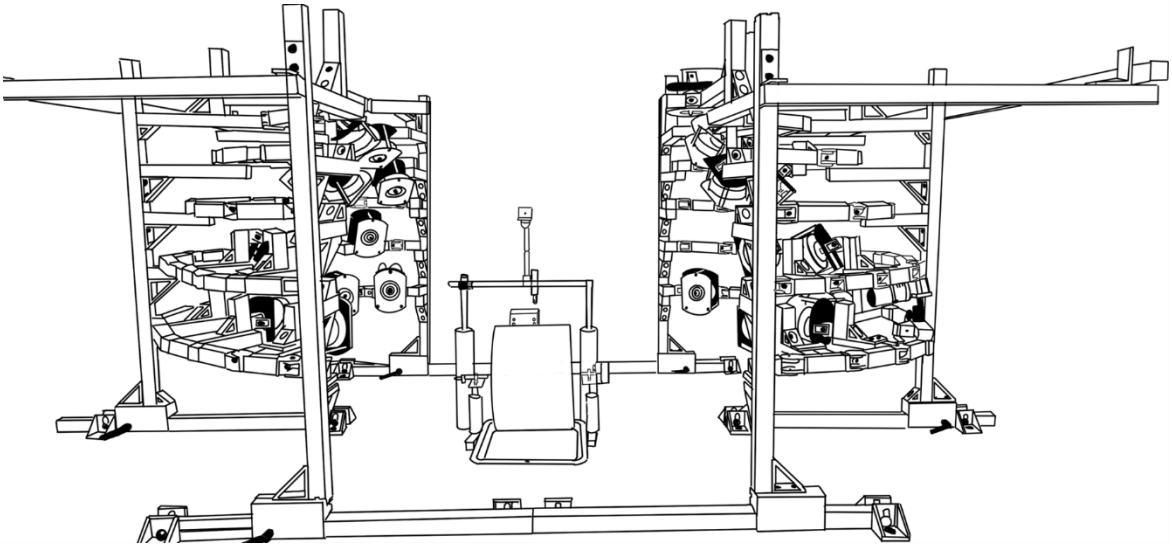

Figure 40. Schematic of Custom Apparatus for Sound Delivery, Behaviour, and Recording. A custom speaker dome array consisting of 32 speakers was designed for sound localisation investigations and was piloted for this behavioural task.

Stimuli were 4 s pink noise bursts (65 dB SPL) presented from one of two spatially distinct speakers. The bottom speaker was positioned 22.5° to the right of the mouse in the horizontal plane. The top speaker was located directly above the head (0° horizontal, 90° vertical). The CBA mice were trained to associate the bottom speaker with reward (go trials), whereas the B6.CAST mice were trained to respond to the top speaker as the go stimulus. The interstimulus interval (ISI) ranged from 3 to 5 s of silence, sampled from a truncated Gaussian distribution.

4.3.1.3 Behavioural Training

Stage 1: Lick Training (Habituation)

Mice were head-fixed and allowed to explore the task environment while freely running on a wheel for 15 minutes per session. Mice could lick the spout ad libitum to receive 3 µL of 10% sucrose solution per lick. This stage helped mice

habituate to the task environment and served to establish a basic reward association while promoting stable licking behaviour.

Stage 2: Association Training

Only go trials were presented during this stage. At 2 s into the 4 s stimulus, mice received a free sucrose reward (3 µL) if they had not licked prior to this point. If the mouse licked before the automatic reward, it received an increased reward (6 µL). Sound presentation terminated upon the first lick. This structure encouraged early licking within the stimulus window and reinforced the association between sound presentation, spatial location, and reward. Mice advanced to the next stage once they consistently licked prior to the automatic reward. This took 3 days on average.

Stage 3: No-Go Introduction

The no-go stimulus was introduced gradually. Initially, 10% of trials were presented from the non-target speaker (no-go). The proportion of no-go trials was incrementally increased across sessions until a 50/50 go/no-go ratio was reached. During this phase, incorrect licking on no-go trials was not punished. Instead, no-go trials resulted in continued stimulus presentation for the full 4 s without reward. Hit trials, on the other hand, led to immediate sound termination and sucrose delivery upon the first lick.

Stage 4: Quinine Punishment Introduction

Mice initially completed training sessions with a balanced 50/50 go/no-go trial structure, as described in the previous stage. Quinine punishment was introduced once mice achieved high hit rates but maintained elevated false alarm

rates, with the specific aim of reducing inappropriate licking on no-go trials without reducing correct responses on go trials. In this phase, licking on a no-go trial triggered immediate stimulus termination and delivery of 2 µL of 100 µM quinine solution. Training continued until mice achieved stable discrimination performance.

4.3.1.4 Pinnae Manipulation

Once mice reached high performance (*mean d′* = 2.89, *SD* = 1.18), they underwent bilateral surgical removal of the upper third of the pinna under isoflurane anaesthesia. Note that this was a more conservative manipulation compared to the near-complete removal used in Experiment 4.2. Post-surgical recovery lasted 48 hours with free access to food and water. Following recovery, mice resumed training under identical conditions to assess relearning after spatial cue disruption.

4.3.1.5 Behavioural Measures and Analysis

Mice performed approximately 150 trials per 30-min session. Trials were categorised as hits (lick on go trial), misses (no lick on go trial), false alarms (lick on no-go trial), and correct rejections (no lick on no-go trial). Perceptual sensitivity (*d′*) was calculated using a signal detection theory formula. To prevent infinite *d′* values in cases of perfect performance, adjusted hit and false alarm rates were applied to all sessions.

Adjusted hit rate (HR) and false alarm rate (FAR) were calculated as:

$$HR = \frac{(Hits+0.5)}{(Total\ Go\ Trials+1)} \qquad FAR = \frac{(False\ Alarms+0.5)}{(Total\ No{-}Go\ Trials+1)}$$

Perceptual sensitivity (*d'*) was computed using:

$$d' = Z(HR) - Z(FAR)$$

Where Z represents the z-score.

4.3.1.6 Ear Tracking

The algorithm used for ear tracking was previously described in Section 4.2.1.6. In addition, an exploratory analysis was conducted to examine ear trajectories, defined as the movement of the ear tip marker in Euclidean space over time. The goal was to determine whether ear movement patterns differed depending on sound location, specifically comparing trajectories evoked by stimuli presented at the mouse's azimuth (horizontal plane) versus 90° (vertical plane). By tracking the spatiotemporal evolution of ear position relative to stimulus onset, this approach allowed us to assess whether ear movements exhibit systematic differences based on the spatial origin of the sound, potentially reflecting adaptive orienting behaviours.

Figure 41 provides a schematic overview of the trajectory tracking approach. A rectangular region of interest (ROI) was defined around the ear tip marker in the original tracking frame (left panel). Within this region, the ear's position was sampled at 41 time points, spanning -1.5 s to +1.5 s relative to sound onset, corresponding to a bin width of 75 ms per sample. The trajectory was

reconstructed by linking the marker positions across frames. The grey rectangle in the left panel corresponds to the plotting area shown in the right panel.

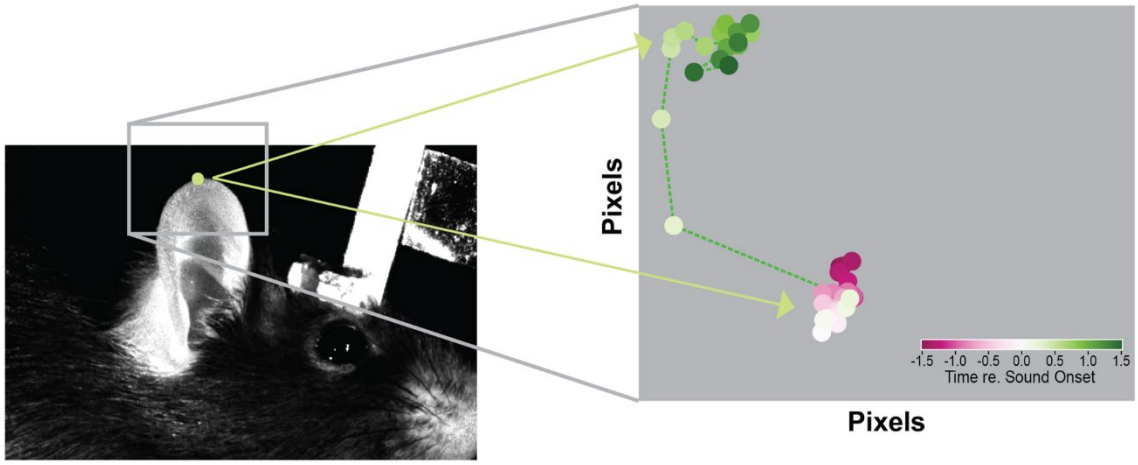

Figure 41. Schematic Representation of Ear Trajectory Analysis. Ear tip movement was tracked over time in relation to sound onset to assess whether trajectory differed depending on sound location and response type (correct/incorrect). The left panel depicts the tracked ear tip marker (highlighted within the grey rectangle), which corresponds to the grey rectangular plotting space in the right panel. In the right panel, individual markers represent ear tip location at discrete time points from 1.5 s before to 1.5 s after sound onset (41 total time bins). Marker colour indicates time relative to sound onset, with earlier time points in this case shown in pink and later points in green. The dashed line indicates the directionality of movement and arrows highlight the overall trajectory pattern in this example.

To visualise trajectories, individual data points were colour-coded based on their temporal distance from stimulus onset. The resulting paths illustrated the direction of ear movement across the trial window. This approach allowed us to explore whether ear trajectories differed systematically depending on sound location. Trajectories were plotted separately for different trial types (e.g. go vs. no-go, hit and miss, false alarm/correct rejection).

4.3.2 Results

4.3.2.1 Behavioural

Sensitivity (*d'*) was assessed for each mouse's best-performing session (i.e. the session with the highest *d'*). During peak performance, mice exhibited high perceptual sensitivity, with a mean *d'* of 3.13 (*SD* = 0.86, *N* = 6; Figure 42a). Response rates revealed a strong differentiation between hit rate (HR) and false alarm rate (FAR). HR was high ($M = 0.95, SD = 0.04$), whereas FAR remained low ($M = 0.35, SD = 0.08$), with a significant difference between the two ($t(5) = 18.37, p < .0001$; Figure 42b).

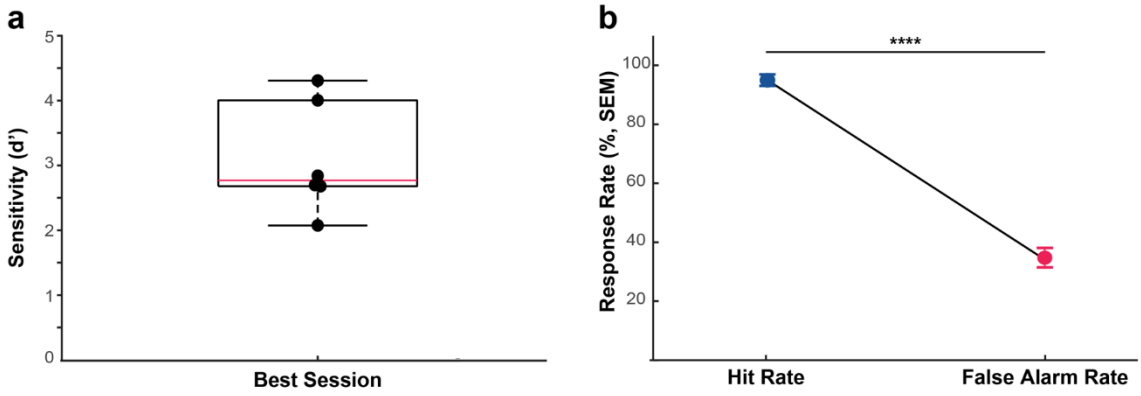

Figure 42. Behavioural Performance Before Pinnae-Manipulation (a) Best session performance. Box plot showing the distribution of *d'* values for the highest-performing session of each mouse (n=5). **(b)** Hit rate vs. false alarm rate. Mean response rates (±SEM) for hit and false alarm averaged across all pre-manipulation sessions post-quinine introduction. A paired t-test confirmed that the hit rate was significantly higher than the false alarm rate.

To assess the effect of quinine, *d'* was compared between the best session prior to quinine introduction and the best session following quinine administration but before pinnae manipulation. Sensitivity significantly improved following the

introduction of quinine as a negative reinforcement, increasing from $M = 0.88$ $(SD = 0.59)$ pre-quinine to $M = 3.13$ $(SD = 0.86)$ post-quinine $(t(5) = -5.23, p = .0034$; Figure 43).

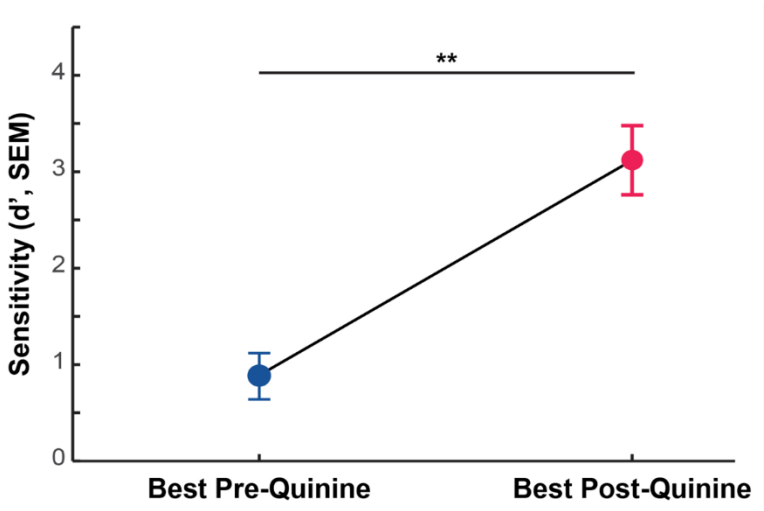

Figure 43. Effect of quinine introduction. Mean ***d′*** (±SEM) of the best session before and the best session after quinine introduction as a negative reinforcement. A paired t-test confirmed a significant improvement in ***d′*** following quinine.

4.3.2.1.1 Learning Curve Analysis

To quantify task learning, performance (sensitivity, *d′*) was modelled as a function of training session using a sigmoid function:

$$d'(x) = \frac{A}{1 + \exp\left(-B(x - x_0)\right)}$$

Where:

- A represents the asymptote, reflecting the estimated maximum *d′* level,
- B is the learning rate (slope), indicating the steepness of the transition from early acquisition phase to the plateau,
- x_0 is the midpoint, the session at which performance reaches 50% of the asymptote.

This model was fit separately for each mouse to capture subject-specific variability in learning.

Figure 44a shows individual learning trajectories and corresponding sigmoid fits. Mice W5321, W5323, and W5326 demonstrated steep performance increases, reaching *d′* > 4, whereas others showed more gradual improvement and plateaued at lower levels. Figure 44b displays individual slope parameters (B), which ranged from 0.20 to 0.82. Mouse W5321 exhibited the steepest learning curve ($B = 0.82$). Parameter estimates for all mice are reported in Table 10.

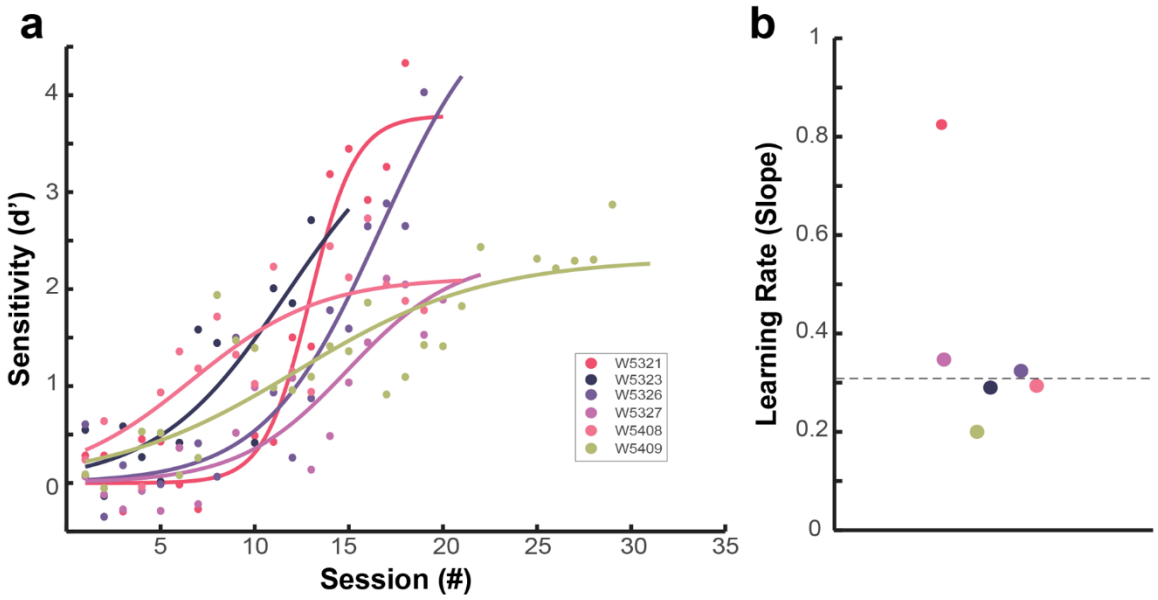

Figure 44. Learning Curves and Individual Learning Rates Across Mice. (a) Sigmoid learning curves fitted to ***d′*** (sensitivity) across sessions for individual mice (n=6). Each mouse is plotted in a unique colour, with raw data points overlaid. All pre-pinnae manipulation sessions are plotted. Sigmoid fits were constrained within biologically plausible limits. **(b) Learning rates (slopes) for individual mice**, extracted from the fitted sigmoid curves. Each point represents the slope (B parameter) for a single mouse, coloured to match the learning curves in (a). The dashed line indicates the median learning rate across all mice.

To complement sigmoid analysis, we computed time to criterion, defined as the first session in which each mouse first reached $d' \geq 2$. This threshold is commonly used to indicate successful task acquisition in similar paradigms. Time to criterion ranged from 9 to 17 sessions, reflecting individual differences in learning speed (Table 10).

To assess whether learning dynamics were predictive of final task performance, we examined correlations between asymptote (A) and both midpoint (x_0) and learning rate (B). Neither correlation was significant ($r = 0.59, p = 0.21$ for A and x_0; $r = 0.26, p = 0.66$ for d' and B), suggesting that faster acquisition did not necessarily correspond to higher ultimate performance.

A group-level sigmoid fit was also conducted across all sessions and mice (Figure 61 in Appendix 3), resulting in estimated parameters of $B = 0.50,\ A = 2.08,\ x_0 = 31.17$. However, due to the pronounced heterogeneity in individual learning curves, the group-level model was not used in the primary analysis.

Table 10. Sigmoid Parameter Estimates and Time to Criterion for Individual Mice. Sigmoid function parameters estimated for each mouse, capturing individual learning trajectories. A represents the estimated performance (*d'*) plateau reached by each mouse. B reflects the learning rate, with higher values indicating steeper learning curves. x0 denotes the session at which performance reaches half of the asymptote. Time to criterion represents the session number at which a *d'* of 2 was first reached.

Mouse	Asymptote (A)	Slope (B)	Midpoint (x0)	Time to Criterion
W5321	3.7897	0.8247	12.927	16
W5323	3.9113	0.2900	11.704	13
W5326	5.2298	0.3237	16.672	17
W5327	2.3349	0.3471	14.947	17
W5408	2.1193	0.2934	6.646	19
W5409	2.3157	0.2000	12.252	30

4.3.2.1.2 Impact of Pinnae Manipulation on Performance (d')

To evaluate the effect of pinnae manipulation on vertical sound localisation, *d'* values were compared between each mouse's best-performing pre-manipulation session and their first post-surgical session. As expected, performance declined significantly following pinnae removal, with *d'* decreasing from $M = 3.69$ to $M = 1.64$, $t(2) = 6.96, p = .02$ (**Figure 45**).

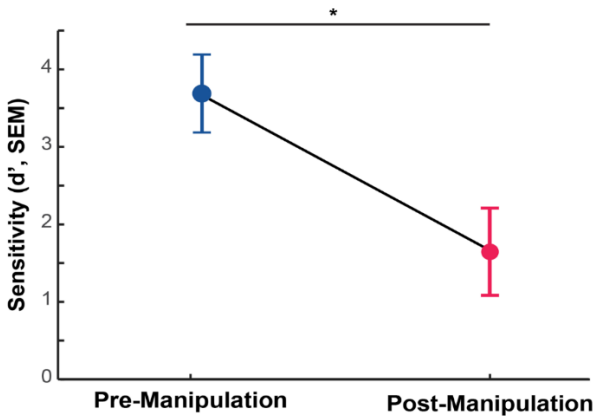

Figure 45. Effect of pinnae manipulation. Mean *d′* (±SEM) in the final session before pinnae-manipulation compared to the first session post-manipulation. A paired t-test confirmed that vertical localisation was significantly impacted by this manipulation.

4.3.2.1.3 Relearning Following Pinnae Manipulation

To determine whether mice could reacquire task performance following disruption of spectral cues, *d′* values from each animal's best pre-manipulation session were compared to their best post-manipulation session using a paired t-test. Sensitivity ultimately returned to near pre-manipulation levels ($M = 3.65, SD = 0.69$), with no significant difference between peak and relearned performance ($t(4) = -1.50, p = .2$; Figure 46).

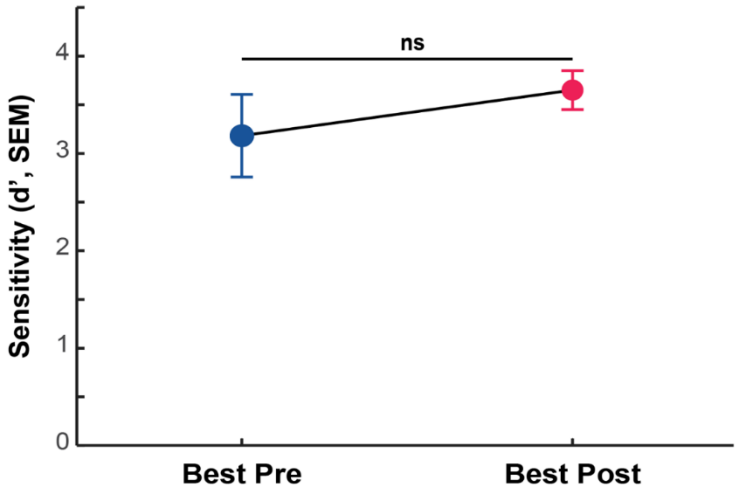

Figure 46. Relearning post-pinnae manipulation. Mean ***d′*** (±SEM) for pre-manipulation peak performance compared to post-manipulation peak performance. A paired t-test showed no significant difference, suggesting that mice successfully adapted to altered acoustic cues.

To formally assess changes in performance over time, we fitted a linear mixed-effects model predicting *d′* as a function of session, with random intercepts for each mouse. The model revealed a significant main effect of *Session* ($\beta = 0.081,\ SE = 0.016, t(70) = 5.03, p < .0001$), indicating that *d′* significantly improved with additional training (Figure 47). The random intercept variance for *Mouse ID* was estimated at *0.46* ($95\%\ CI: 0.22 - 0.96$), suggesting moderate variability in baseline *d′* across subjects. The observed upward trajectory suggests that, despite the disruption of spectral cues, mice gradually adapted to rely on altered or residual localisation cues to guide task performance. A summary of individual trajectories is visualised in Figure 48, which also includes non-

manipulated animals for reference.

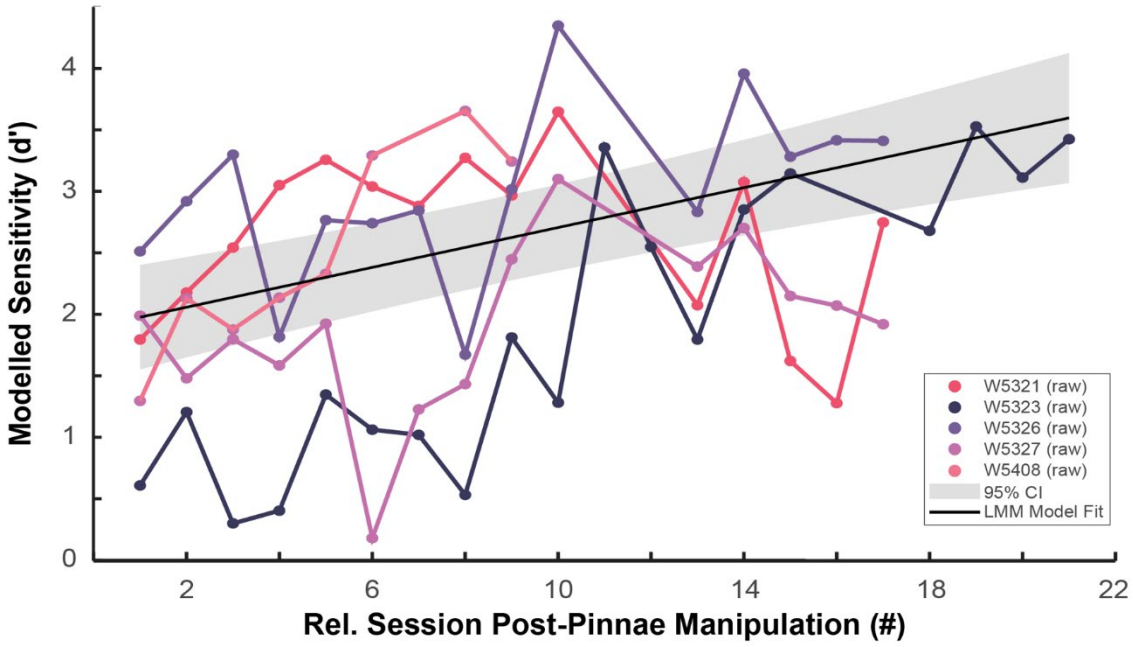

Figure 47. Relearning Trajectory Following Pinnae Removal. A linear mixed-effects model predicting ***d'*** recovery over sessions following pinnae removal. Each point represents an individual session for a given mouse, colour-coded by mouse ID. The black line represents the model fit, with the shaded region representing the 95% confidence interval.

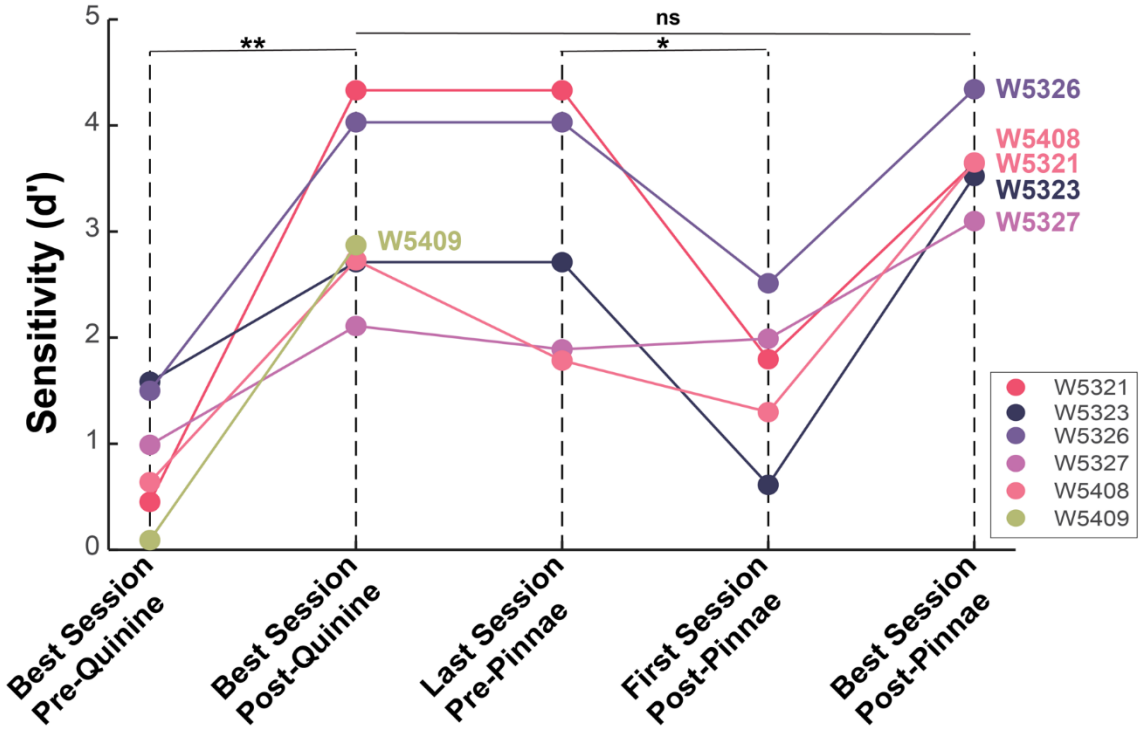

Figure 48. Summary of Behavioural Performance Across Quinine Exposure and Pinnae Removal. The plot illustrates individual mouse performance (d') at key times points throughout training, quinine exposure, and pinnae manipulation. Each coloured line represents an individual mouse. **Best Session Pre-Quinine:** The highest d' value recorded before quinine introduction. **Best Session Post-Quinine:** The highest d' value achieved after introducing quinine as a negative reinforcement, indicating peak task proficiency. **Last Session Pre-Pinnae:** The final session immediately before pinnae manipulation, serving as a baseline for the effects of manipulation. In many cases, this was equivalent to the highest d' achieved. **First Session Post-Pinnae:** The session immediately after pinnae manipulation, showing the initial impact of manipulation on vertical sound localisation. **Best Session Post-Pinnae:** The highest performance recorded after pinnae manipulation, reflecting adaptation to altered spatial cues, comparable to pre-manipulation

performance. Vertical dashed lines separate distinct training phases. Asterisks indicate statistically significant differences and "ns" denotes a non-significant comparison.

4.3.2.2 Ear Trajectory

To expand upon the behavioural findings, we conducted an exploratory ear trajectory analysis to determine whether ear movements were influenced by sound location. Figure 49 and Figure 50 illustrate example ear tip trajectory plots from a single session in which the bottom speaker served as the go. These analyses provide preliminary evidence that mice adjust ear position in response to spatial sound cues, and that such adjustments may relate to response accuracy.

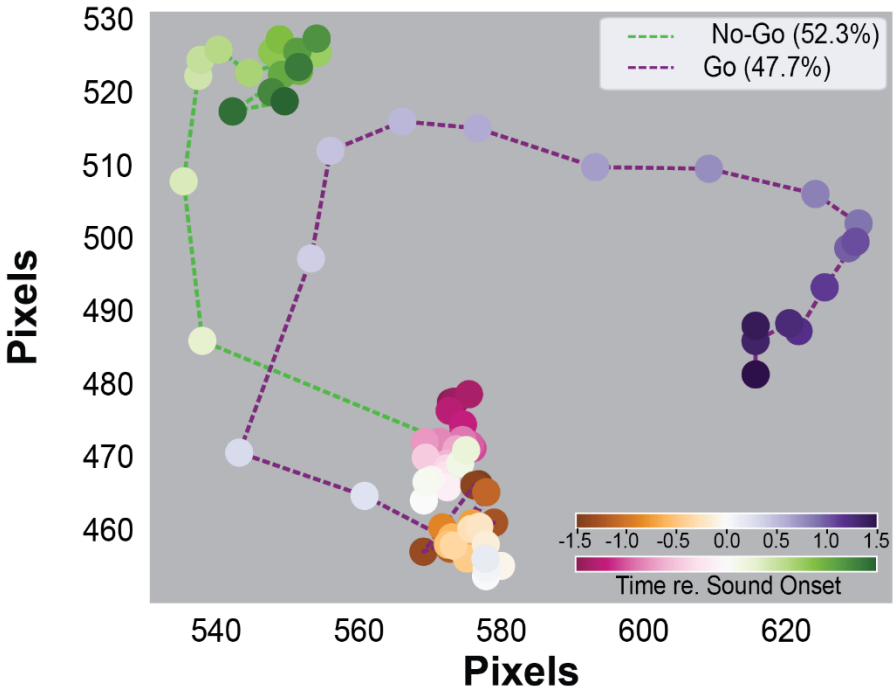

Figure 49. Ear Trajectory During Go and No-Go Trials. Example ear trajectories from a single session, spanning 1.5s before to 1.5s after sound onset. Markers indicate time points, with colour representing time relative to sound onset (legend). Dashed lines connect consecutive time points. The go stimulus was presented from the bottom speaker. The ear moves in distinct directions for go (purple) and no-go (green) trials.

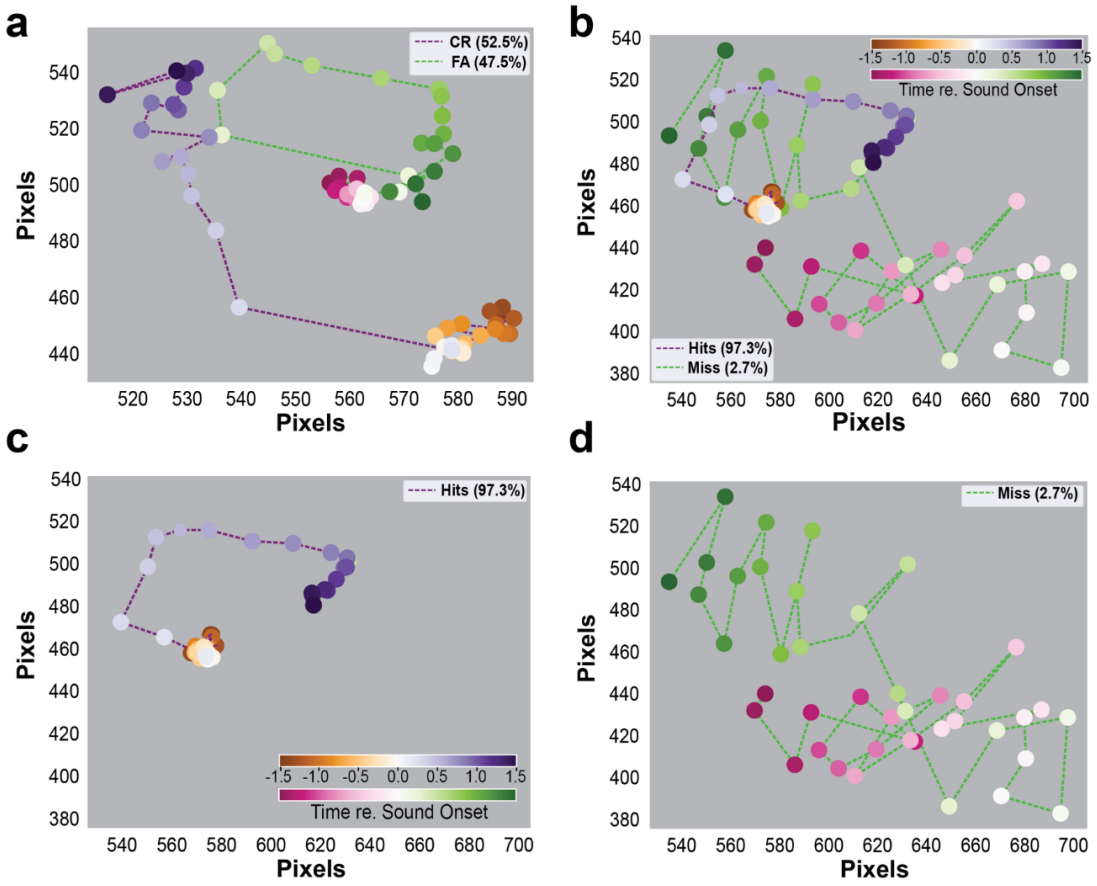

Figure 50. Ear Trajectories Separated by Trial Type. Trajectories from the same session as Figure 49 subdivided by response type. **(a)** No-go trials separated into correct rejections (CR, dashed purple) and false alarms (FA, dashed green). **(b)** Go trials (hits) and miss trials overlaid for comparison. **(c)** Isolated hit trials. **(d)** Isolated miss trials. Colour coding represents time relative to sound onset.

Across both go and no-go trials, ear position was initially similar, possibly reflecting a shared preparatory state. Trajectories began to diverge shortly after sound onset. In go trials, the ear tip moved forward and downward, suggesting orientation towards the bottom speaker. In no-go trials, the ear shifted slightly backward and upward, consistent with an orientation toward the elevated (90°) speaker.

Figure 50a separates no-go trials by response type. In false alarms (FAs), the ear initially followed the upward trajectory typical of correct rejections (CRs), but then veered forward, resembling the trajectory of a go response. This suggests

that misclassification of sound location may be reflected in, or driven by, misoriented ear movements. CRs maintained an upward trajectory throughout, consistent with correct localisation of the top speaker.

Go trials were almost entirely composed of hits (97.3%), but we nonetheless visualised the few miss trials Figure 50b and Figure 50d. While hit trajectories followed the expected forward movement toward the bottom speaker, miss trials did not exhibit a consistent pattern, potentially reflecting either uncertainty about the sound source or attentional lapses. Further work would be required to interpret these rare miss responses.

4.4 DISCUSSION

This study investigated the mechanisms underlying vertical auditory spatial processing and the contribution of monaural spectral cues to spatial attention and behaviour. Using both passive (oddball) and active (go/no-go) paradigms, we examined how mice process sound elevation and how pinnae manipulation alters performance. Our findings revealed distinct signatures of spatial discrimination and behavioural flexibility, offering insight into the role of statistical learning and sensory plasticity in auditory perception. We also examined ear movements as a novel physiological measure of auditory orienting, identifying ear states that co-occur with pupil dilation and behavioural responses. The discussion is organised by paradigm, beginning with the passive oddball task, followed by behavioural performance in the discrimination task, and concluding with the effects of pinnae removal and the implications of ear movements for auditory attention.

4.4.1 Spatial Salience and Disrupted Deviance Sensitivity in Passive Sound Presentation

Across conditions, pupil-linked responses were modulated more strongly by sound elevation than by statistical deviance. In both cohorts, top-originating sounds evoked larger and later pupil dilations than sounds at azimuth, even when probability was held constant. This intrinsic salience of elevated sounds likely reflects an ecologically grounded attentional bias toward the overhead field, where aerial threats are more likely to appear. Rodents possess binocular visual

coverage of the upper visual field (Wallace et al., 2013), and both behavioural and neurophysiological studies confirm that overhead stimuli elicit heightened defensive responses (Blanchard et al., 2011; Salay et al., 2018; Wei et al., 2015). Although the stimuli used here were static and probabilistic rather than looming-like, the direction-specific arousal responses observed suggest that elevation alone may function as a salient sensory dimension, capable of triggering vigilance-related pupil changes.

This spatial salience bias not only persisted but remained dominant even in the deviance conditions, where stimulus probability was manipulated. In the pinnae-intact cohort, deviant responses appeared enhanced when the deviant was presented from the top speaker, but this effect was not distinct from the general salience effect observed in the control condition. After accounting for elevation bias, the only deviance-related response emerged in the bottom speaker condition: deviant trials from the bottom speaker now evoked significantly greater pupil dilation than top-originating standards. This suggests that spatial salience had previously masked the effect, inflating standard responses due to the inherent salience of elevated sounds. Once this bias was corrected, a clearer signature of deviance-related processing became evident.

In the pinnae-removed cohort, the salience effect persisted, but no reliable evidence of deviance sensitivity was observed. Despite near-complete removal of the external ears, mice continued to show enhanced pupil responses to top-originating sounds. This finding indicates that some elevation cues, potentially mediated by residual pinna bases or head shadowing, were sufficient to support spatial salience tracking. In the absence of intact spectral filtering, stable spatial

encoding may have degraded, increasing uncertainty and disrupting the formation of reliable statistical prediction.

These findings are consistent with frameworks that distinguish between salience-driven responses and predictive coding mechanisms. While deviance detection depends on establishing and updating sensory expectations based on stimulus regularities (Carbajal & Malmierca, 2018; Näätänen et al., 2007), salience responses may reflect more reflexive or low-level attentional capture (Parmentier, 2014). From this perspective, the persistence of an elevation-driven pupil response under cue degradation may reflect the relative robustness of arousal-linked responses to salient spatial features compared to the more fragile computations required for detecting violations of expectation.

In the laboratory, rodents are frequently approached from above, particularly during handling procedures. Such interactions may sensitise animals to the overhead field, further amplifying innate vigilance. Although direct evidence for rapid recalibration of vertical spatial threat sensitivity in the auditory system is lacking, related findings in the visual domain suggest that subcortical circuits can maintain directional selectivity even under degraded sensory input. For example, looming-sensitive neurons in the superior colliculus continue to drive robust fear responses despite variation in visual input (Shang et al., 2015). Whether a similar preservation of elevation sensitivity occurs in the auditory system remains to be tested. Nevertheless, the present results indicate that, while deviance detection fails without precise spatial filtering from intact pinnae, general arousal to spatial cues is maintained and may reflect a conserved survival mechanism. Future

studies could quantify the spectral and spatial information preserved after this level of pinnae removal by recording from implanted ear microphones to estimate head-related transfer functions (HRTFs). Identifying which cues remain intact may clarify how coarse threat detection is retained in the absence of statistical deviance sensitivity.

4.4.2 Vertical Sound Discrimination in an Active Go/No-go Task

In the active vertical sound discrimination task, mice reliably distinguished between sound sources originating from different elevations with excellent accuracy (high *d′*). This finding challenges the prevailing assumption that mice possess poor vertical localisation abilities. Lauer et al. (2011), for example, reported that CBA/129 mice barely exceeded a *d′* of 1 when discriminating sounds separated by 90°, using a similar strain and the same spatial configuration. Based on these results, the authors concluded that mice may lack the capacity for precise vertical localisation, particularly compared to horizontal discrimination. However, mice in the present study reached *d′* values exceeding 4, indicating highly robust and stable vertical sound discrimination ability. These findings suggest that, under appropriate training conditions, mice are capable of more precise vertical auditory discrimination than previously recognised under standard behavioural testing protocols.

A likely explanation for this discrepancy lies in the differences in task structure and reinforcement history. In our study, mice underwent a gradual shaping

procedure: training began with free reward delivery, followed by progressive introduction of no-go trials. Quinine-based punishment was introduced only after mice consistently achieved hit rates above 80%. This structure enabled animals to first establish a strong association between the go stimulus and reward, before being required to suppress responses to the non-target speaker.

Lauer et al. (2011) employed a two-alternative forced-choice (2AFC) design in which incorrect responses to the spatial deviant (“warning” stimulus) resulted in a mild tongue shock. Although their mice also began with reward-only trials, the shock was introduced as soon as spatial discrimination trials were added. Because of its aversiveness, warning trials were presented infrequently (15%) to minimise disruption of task engagement. In our paradigm, go and no-go trials were presented with equal frequency, and quinine was introduced only after animals demonstrated reliable identification of the go stimulus.

Importantly, prior to quinine, mice were not punished for licking, allowing them to explore the task structure freely. Many continued to respond during no-go trials, which could suggest incomplete learning. However, the rapid drop in false alarm rates following quinine introduction, without a corresponding decrease in hit rate, indicates that the discrimination had been acquired but was not consistently expressed in the absence of behavioural cost. This suggests that when punishment is decoupled from initial task acquisition, animals can learn the stimulus-reward contingencies without the interference of stress or suppression effects. Thus, quinine acted not to *establish* learning but to *reveal* and stabilise an already-learned distinction.

Other contextual differences between studies may have contributed to the observed performance disparity. While Lauer et al. used 60-minute testing sessions, our sessions were limited to 30 minutes, potentially helping to maintain engagement and reduce fatigue (de Gee et al., 2024). Our mice were food-restricted and received a sucrose solution as a reward, whereas Lauer et al.'s were water-restricted and received juice. Although water deprivation is often assumed to enhance motivation (Barkus et al., 2022), our mice outperformed theirs on the same vertical separation (0° vs. 90°). This suggests that reward type and motivational context might influence task performance. However, Goltstein et al. (2018) found that both food and water restriction support comparable behavioural engagement depending on the paradigm. Thus, the deprivation regime alone is unlikely to account for the observed behavioural advantage.

4.4.2.1 Recovery of Vertical Discrimination Following Pinnae Manipulation

Lauer et al. (2011) estimated HRTFs in mice and demonstrated that pinna removal substantially altered spectral features associated with elevation, including high-frequency notches that vary systematically with vertical angle (Figure 2). Although the extent and laterality of their manipulation were not specified, limiting direct comparison to our protocol, their physiological results underscore the critical role of the external ear in shaping elevation-specific auditory input. In the horizontal plane, they also showed that unilateral pinna removal or folding impaired spatial discrimination, consistent with the sensitivity of auditory spatial coding to changes in ear morphology. Importantly, mice have excellent high-

frequency hearing (Heffner & Heffner, 2007), and the spectral cues that encode elevation occur in these higher frequency ranges. This suggests that mice should, in principle, be well-equipped to perform vertical localisation, provided these cues are available.

In our study, the upper third of the pinnae was bilaterally removed to minimise post-surgical complications while still producing substantial alterations to spectral cues. That mice fully reacquired vertical discrimination following this manipulation, suggests that partial spectral cues remained intact or that compensatory plasticity occurred via recruitment of residual localisation mechanisms. While interaural level differences (ILDs) and interaural time differences (ITDs) are generally thought to contribute minimally to vertical localisation, particularly in small mammals, some limited vertical tuning may persist through these channels when pinnae-based cues are compromised (Parsons et al., 1999). However, it is more likely that the preserved ability to discriminate elevation reflected plastic reweighting of degraded spectral input or the development of novel spectral-to-spatial mappings.

Evidence from ferrets supports this interpretation. Parsons et al. (1999) found that complete bilateral removal of the pinnae and conchae in infancy produced long-lasting deficits in vertical and front-back localisation that persisted into adulthood, despite behavioural training. This suggests that spectral input is essential during development for acquiring the capacity for vertical localisation. In contrast, the mice in our study underwent partial pinnae removal in adulthood, after they had already learned the localisation task. This likely enabled recruitment

of compensatory mechanisms such as plastic reweighting of residual cues, highlighting a distinction between irreversible developmental disruption and adaptive reorganisation in adulthood.

Related findings in humans further support the idea that vertical localisation can be reacquired in adulthood following distortion of spectral input. Hofman et al. (1998) showed that adults fitted with bilateral ear moulds, which altered pinna shape, exhibited initial deficits in vertical localisation but gradually regained accuracy over several weeks. Crucially, removal of the moulds immediately restored baseline performance, indicating that listeners retained the original mapping and developed a new one in parallel. Behavioural adjustment to altered input has also been observed in the context of unilateral spectral disruption, with improvements specific to the affected ear (Van Wanrooij & Van Opstal, 2005). These studies suggest that the adult auditory system can flexibly maintain multiple spectral-to-spatial transformations depending on context.

The recovery observed in our mice may reflect a comparable capacity for plasticity. Although our manipulation was permanent, the ability to regain high levels of performance suggests that new mappings can form in adulthood without requiring complete developmental plasticity. Whether the original spectral-spatial transformations were reserved remains unknown. Future studies, using reversible manipulations such as temporary occlusion of the pinna or fitted ear moulds, could determine whether multiple spatial mappings are maintained in parallel and how they are accessed during adaptive task performance.

4.4.3 Ear Movements and Auditory Attention

Although pupil dilation has become a widely used index of internal state in cognitive and systems neuroscience, the behavioural relevance of ear movements, particularly in rodents, remains comparatively underexplored. While the pinnae are known to support acoustic filtering, recent evidence suggests that their movements may also provide a temporally precise readout of sensory processing. In cats, Populin et al. (2004) demonstrated that ear movements can function as rapid, directionally specific “acoustic saccades,” initiated toward auditory targets and modulated by eye position, suggesting an integration of sensory and motor signals within the superior colliculus. Similarly, guinea pigs exhibit a fast pinna twitch reflex, known as the Preyer reflex, in response to loud sounds, which has been used as a reliable behavioural index in startle paradigms and tinnitus assessment (e.g. Berger et al., 2013; Wallace et al., 2024). In mice, Clayton et al. (2024) reported that brief broadband sounds evoke highly stereotyped facial twitches, including movement of the pinnae, that precede pupil dilation and track auditory features such as stimulus envelope and salience. Although their primary metric was motion energy from the vibrissal cheek region, pinna movement was consistently observed and modulated across conditions. These findings collectively indicate that ear movements can serve not only an acoustic function but also reflect orienting behaviour and sensory prioritisation. The present study builds on this work by quantifying ear position and motion in head-fixed mice during sound presentation, assessing whether these movements index stimulus identity or deviance in a structured and interpretable manner.

As a first step, we analysed video recordings from head-fixed mice in the passive oddball task. While the data were collected in a stimulus-driven context, our focus was not on stimulus-specific responses at that time but rather on the classification of spontaneous ear movement. Using t-SNE on framewise ear velocity and position, followed by k-means clustering, we identified five clusters reflecting separable ear states. Two clusters captured dynamic movements likely related to arousal and orienting, while the remaining three corresponded to static postures such as perked, neutral, and retracted positions. These postures may indicate tonic differences in arousal or attentional engagement.

The emergence of interpretable clusters, some of which aligned with pupil dilation, locomotion, or stimulus onset, suggests that ear movements encode behaviourally relevant information. Although preliminary, this classification highlights ear kinematics as a continuous, quantifiable readout of internal state, with the added potential for spatial specificity due to the acoustic role of the pinnae. These features make ear posture a promising behavioural signal not only for indexing arousal but also for investigating affective state and sensory decision-making in rodents.

To further investigate the relationship between ear kinematics and spatial context, we analysed ear trajectories in the go/no-go discrimination task. Here, movements differed by trial type: on go trials (bottom speaker), the ears moved forward and downward; on no-go trials (top speaker), they shifted backward and upward. This directional dissociation suggests that ear movement functions as an orienting response, with trajectories reflecting the spatial configuration of upcoming stimuli. Strikingly, false alarm trials began with no-go-like trajectories

but transitioned toward go-like profiles, raising the possibility that shifts in ear movement may reflect, or even precede, shifts in perceptual decision-making. Future work will expand this analysis to a larger sample of mice to determine whether ear trajectory patterns generalise across animals and whether they correlate with overall task performance.

Together, these findings demonstrate that ear movements provide a structured, temporally precise behavioural signal, even in head-fixed contexts where traditional orienting behaviours are constrained. Ear posture may thus provide a complementary index of sensory engagement and attentional state in rodent models of auditory processing. Building on this work, we aim to investigate the neural mechanisms linking ear movements to auditory processing. In particular, we plan to manipulate specific neuronal subtypes within the dorsal cochlear nucleus (DCN), which integrates auditory and multisensory inputs, including proprioceptive signals from head and ear position (Apostolides & Trussell, 2014). Granule cells within the DCN receive convergent input from diverse sensory modalities and are implicated in processing head-related cues (Kanold & Young, 2001). Given the DCN's role in sensory-motor transformation, modulating its activity may reveal how central circuits govern ear movements that support spatial hearing. The DCN is also a highly plastic structure (Fujino & Oertel, 2003; Tzounopoulos et al., 2004), shaped by neuromodulatory influences and Hebbian learning mechanisms (Beebe et al., 2023; Caporale & Dan, 2008; Mellott et al., 2011; Roberts et al., 2006; Stefanescu & Shore, 2017), yet the conditions under which this plasticity is engaged remain poorly understood. Future

work will examine whether these mechanisms are recruited during specific learning states, such as those induced by spatial cue disruption, to support behavioural refinement of vertical localisation through adaptive circuit-level organisation.

4.4.4 Broader Implications and Future Directions

Collectively, these findings offer new insight into the mechanisms supporting vertical sound localisation and the broader principles of sensory-motor flexibility and behavioural state monitoring. By integrating controlled behavioural paradigms with physiological readouts such as pupil size and ear movement, this work delineates how mice interpret vertical spatial cues and how these processes recalibrate following disruption. The identification of an innate salience bias for elevated sounds, its partial preservation after pinnae removal, and the recovery of localisation performance through training, all underscore the plasticity inherent to vertical auditory processing. The inclusion of ear posture as a behavioural readout introduces a novel, spatially specific index of orienting behaviour, with potential to uncover how sensory input, internal state, and motor planning interact. Together, these results not only deepen our understanding of the behavioural and computational basis of elevation perception but also establish a foundation for future investigations into the neural circuits that mediate spatial hearing.

5 GENERAL DISCUSSION

Auditory perception relies on the extraction of structure from a dynamically changing environment. Expectations formed from regularities in sensory input can guide attention towards likely events, enabling more efficient detection and discrimination. Equally important, however, is the capacity to rapidly detect deviations from these regularities, particularly in contexts of uncertainty or threat. Over longer timescales, persistent changes in sensory input, such as alterations to spatial cues, demand more fundamental plasticity, requiring the auditory system to remap internal models of the environment. Together, short- and long-term adaptability allow organisms to maintain perceptual stability while remaining sensitive to meaningful change.

This thesis investigated how auditory behaviour is shaped by these two types of structure: short-term statistical regularities and long-term spatial experience. Using a series of behavioural paradigms across ferrets, humans, and mice, the work examined how expectations based on frequency and spatial cues influence detection, discrimination, and orienting responses. Although the experimental designs were grounded in extensive prior literature on auditory attention and plasticity, results across all tasks diverged from initial hypotheses. These divergences highlight the critical influence of task structure, sensory timing, reinforcement contingencies, and species-specific ecological pressures in shaping

auditory attention. Even small methodological differences proved sufficient to alter attentional dynamics, suggesting that attentional models based on human psychophysics may not fully capture the diversity of strategies deployed across contexts and species. The findings contribute to existing theoretical frameworks by illustrating how auditory attention flexibly integrates stimulus statistics, task demands, and ecological relevance across different contexts and species.

The first two empirical chapters of this thesis explored how short-term statistical regularities influence auditory detection and discrimination. Using frequency probability as the manipulated dimension, the tone-in-noise detection and duration discrimination tasks probed whether animals would exhibit classic expectation-driven advantages or, alternatively, sensitivity to deviant stimuli. Despite close alignment with previous probe signal task designs in humans (Greenberg & Larkin, 1968; Mondor & Bregman, 1994; Scharf et al., 1987), results differed from the initial hypotheses, revealing that attentional allocation depended critically on task timing and stimulus duration. They generally suggest that attentional allocation dynamically balances between expectation-driven and deviance-driven orienting, with the relative dominance of each process shifting depending on temporal and behavioural context.

In the ferret duration discrimination task, expectation-based benefits were evident for short tones, where high-probability signal tones were detected more accurately. However, this pattern reversed at longer durations, with low-probability probe tones being favoured, consistent with a shift towards saliency-driven orienting. In the tone-in-noise task, ferrets consistently performed better for probe tones in both accuracy and reaction time while humans showed higher accuracy

for signals, but paradoxically faster reaction times for probes. These differences suggest that the balance between expectation-driven and saliency-driven attention is sensitive to both temporal characteristics and broader task demands, a theme explored further in the following section.

Across the empirical chapters of this thesis, a consistent principle emerged: whether an effect of stimulus probability was observed, and whether it favoured expected or unexpected stimuli, seemed to depend on task structure, reinforcement contingencies, and motivational context. In ferrets, water deprivation placed a high behavioural cost on errors, likely encouraging a general prioritisation of accuracy over speed to maximise the likelihood of obtaining reward. In humans, the self-paced design and low-stakes monetary reward likely incentivised rapid responding, with participants favouring speed over careful discrimination in order to complete the task sooner. Differences between our human results and those of Luthra et al. (2024), who used a 2AFC tone-in-noise task with the same stimuli and observed a probe signal effect, further emphasise the influence of task format and motivational structure. Designing comparable behavioural tasks across species will require matching not only stimulus features, but also motivational structures, such as introducing intermittent trial-by-trial rewards for humans to more closely parallel animal paradigms. Without such considerations, cross-species comparisons of attention risk misinterpreting differences in cognitive strategies that arise from task-induced motivational divergences rather than true species differences.

In contrast, the vertical sound localisation task in mice did not probe attentional filtering, but instead demonstrated how careful reinforcement shaping

could reveal perceptual abilities that had previously been underestimated. By delaying the introduction of negative reinforcement until high hit rates were achieved, and by replacing time-outs with quinine to suppress false alarms, mice reached levels of spatial discrimination that far exceeded prior reports and demonstrated a capacity for vertical localisation previously thought unlikely in this species (Lauer et al., 2011). This striking improvement compared to earlier studies provides strong evidence that behavioural performance in mice is highly sensitive to task design, and that careful optimisation of training protocols is essential for accurately assessing species-specific perceptual and cognitive abilities.

Our success also highlights the broader risks associated with suboptimal task design. Early assumptions that mice were fundamentally poor at vertical localisation may have discouraged the use of this genetically accessible model for studying spatial hearing. Yet, mice offer substantial experimental advantages, including access to transgenic lines, compatibility with precise circuit manipulation techniques such as optogenetics and chemogenetics, and suitability for high-resolution behavioural and physiological recording. Careful optimisation of task design not only reveals latent perceptual capacities but also opens new opportunities for linking behaviour to neural mechanisms in species that are otherwise ideally positioned for mechanistic investigation.

Beyond demonstrating high localisation performance, the vertical sound localisation task also shows that mice can adapt flexibly to altered spectral cues following partial pinna removal. After an initial decline in performance, animals rapidly regained their ability to discriminate vertical sound locations, reaching accuracy levels comparable to their pre-manipulation performance within only a

few sessions. This behavioural recovery indicates that mice are capable of remapping altered spectral information onto spatial representations in adulthood, rather than relying solely on fixed mappings established during development.

Previous work in other species, including ferrets and humans, has shown that adult spatial hearing retains a degree of plasticity following changes in peripheral input (Hofman et al., 1998; Kacelnik et al., 2006). However, evidence for such adaptive remapping in mice has been limited. Early behavioural studies suggested that mice might be poor candidates for investigating vertical localisation due to limited accuracy in spatial discrimination tasks. The present findings challenge this assumption, demonstrating that mice are capable of both excellent vertical sound localisation performance and adaptive remapping following peripheral disruption. Further experiments will be necessary to fully characterise the limits of this ability, including assessments of minimal audible angle (e.g. as measured by Parsons et al., 1999 in ferrets).

Critically, this behavioural evidence lays a new foundation for investigating the cellular and circuit mechanisms underlying spatial learning in the dorsal cochlear nucleus (DCN). The DCN is uniquely positioned to process these cues through its integration of auditory and multisensory inputs (Reiss & Young, 2005). The rapid behavioural adaptation observed following ear manipulation suggests that plasticity mechanisms within the DCN, such as synaptic changes at granule cell inputs, neuromodulatory influences, and instructive top-down signals, can support realignment of spatial representations in response to altered sensory input (Caporale & Dan, 2008; Fujino & Oertel, 2003; Roberts et al., 2006; Tzounopoulos

et al., 2004; Wu et al., 2015). This creates a unique opportunity to directly link perceptual adaptation to underlying changes in DCN circuitry.

Beyond basic science, understanding the mechanisms that enable flexible remapping of spatial cues has important translational implications. Dysfunction in DCN plasticity and neuromodulation has been implicated in disorders such as tinnitus and hyperacusis (Shore et al., 2016), where maladaptive changes following peripheral damage are thought to drive pathological hyperactivity and altered multisensory integration (Middleton et al., 2011; Shore et al., 2008). By establishing a behavioural paradigm in which spatial remapping can be robustly induced and tracked in mice, this work provides a new platform for investigating how healthy plasticity is maintained, and how its disruption may contribute to sensory processing disorders (Glyde et al., 2011). More broadly, these findings underscore the importance of flexible, context-sensitive plasticity mechanisms in maintaining accurate spatial perception across the lifespan.

Most prior studies examining multimodal integration in the dorsal cochlear nucleus have relied on static or artificially manipulated ear positions (e.g. Ryugo et al., 2003; Shore, 2005; Young & Davis, 2002), limiting our understanding of how natural ear kinematics influence auditory processing. Although these approaches have provided valuable mechanistic insights, they do not capture the dynamic ear movements and postures that animals exhibit across different behavioural and arousal states. This thesis introduces a new methodological advancement by systematically tracking natural ear kinematics in mice during auditory tasks, classifying distinct ear states associated with different behavioural contexts, and linking these states to internal arousal indices measured through pupil dynamics.

Exploratory ear tracking data revealed that mice adopt distinct ear postures depending on their engagement with the task, spontaneous movements, and stimulus presentation. Shifts in ear position were not random but structured, suggesting that ear kinematics may actively modulate the spectral filtering cues critical for spatial hearing. This work complements existing evidence that pupil diameter tracks arousal-related brain states (de Gee et al., 2020; McGinley, David, et al., 2015; Reimer et al., 2016), extending these insights into the subcortical auditory system where such relationships have remained largely unexplored. By simultaneously monitoring pupil size and ear posture, this thesis provides a temporally resolved, non-invasive window into internal state fluctuations during auditory perception.

Understanding these dynamics has important implications for interpreting neural coding in auditory brainstem structures. In species like mice, flexible ear positioning may help compensate for limitations in interaural cues by dynamically adjusting spectral filtering, enhancing vertical and front-back localisation (Allen & Ison, 2010; Ito et al., 2020). Although humans lack prominent ear mobility, small vestigial movements related to attention have been observed (Strauss et al., 2020), and broader adaptations such as head and shoulder adjustments maintain spatial acuity across changing acoustic environments (Grothe et al., 2010). Thus, the study of natural ear kinematics is not limited to understanding rodent models but has broader relevance for understanding how flexible filtering of sound cues supports spatial hearing across species.

Looking forward, these methodological advances open new possibilities for investigating how ear posture and brain state interact to shape spatial hearing and

auditory plasticity. Even in the present work, exploratory analysis revealed that incorrect ear orienting was sometimes associated with false alarm errors in a vertical sound localisation task, suggesting that ear posture can provide a real-time behavioural window into spatial decision-making. By combining ear tracking, pupillometry, and neural recordings, future studies can map how naturalistic behaviour influences DCN activity in real time, linking external kinematic states to internal sensory processing dynamics. Such approaches could provide new insights into adaptive recalibration following peripheral disruptions, as well as into maladaptive processes underlying conditions like tinnitus, where abnormal integration of multimodal inputs may play a role. By establishing these tools, this work lays the foundation for an integrated behavioural-neural framework to probe auditory spatial processing in dynamic, naturalistic conditions.

While the preceding sections focused on the mechanisms underlying spatial hearing and auditory plasticity, the experiments in this thesis collectively provide a broader perspective on how attentional strategies and sensory adaptation are shaped across species. By directly comparing ferrets and humans on structurally similar auditory tasks, and by situating mice within an ecological framework of spatial hearing and deviance detection, this work highlights how evolutionary pressures, ecological roles, and motivational contexts influence the balance between expectation-driven and deviance-driven attention. The following section synthesises these cross-species findings, exploring how ecological pressures and behavioural contexts may have shaped the evolution of attentional strategies in auditory perception.

Understanding how attention and sensory adaptation operate across species offers critical insights into the flexibility and constraints of perceptual systems. However, as the experiments in this thesis illustrate, apparent similarities in task structure do not guarantee identical underlying mechanisms. Species differ not only in their sensory and motor capabilities but also in their ecological roles and evolutionary pressures, which shape the priorities and strategies of their attentional systems. Cross-species comparisons must therefore be interpreted carefully, with attention to how task design, motivational context, and ecological relevance interact.

Direct comparisons between ferrets and humans in the tone-in-noise and duration discrimination tasks revealed differences in attentional deployment. In humans, accuracy remained highest for expected (signal) tones, consistent with classical expectation-driven filtering models, while reaction times paradoxically favoured low-probability probe tones. In ferrets, behavioural patterns suggested a greater prioritisation of deviance. In the tone-in-noise task, probe tones were detected more accurately and more rapidly than signal tones, and in the duration discrimination task, expectation effects were observed only for short tone durations, giving way to probe advantages at longer tone durations. Even when external task structure is matched, internal attentional strategies may differ, potentially reflecting differences in motivational priorities, cognitive architecture, or ecological demands.

The mouse experiments, although not involving direct task comparisons to ferrets and humans, similarly highlight the role of ecological pressures in shaping attentional deployment. In the passive spatial oddball paradigm, pupil dilation

responses revealed a strong elevation bias, with sounds from the upper speaker eliciting greater responses regardless of deviance status. This pattern is consistent with the idea that prey species such as mice prioritise potential aerial threats (Wallace et al., 2013; Yilmaz & Meister, 2013), leading to a stronger weighting of overhead spatial locations and heightened sensitivity to deviant or unexpected auditory events (Rogalla et al., 2020). Moreover, the rapid recovery of vertical sound localisation ability following partial pinna removal further illustrates the plasticity and adaptability of spatial hearing in a species facing constant environmental threats.

Together, these findings suggest that attentional strategies reflect an ecological trade-off between efficiency and vigilance. Ferrets, as mesopredators (Prugh et al., 2009), may balance the need for predictive filtering when hunting with heightened sensitivity to unexpected stimuli to avoid predation. Humans, as apex predators, may favour expectation-driven models of perception, consistent with predictive coding frameworks optimised for stable environments (De-Wit et al., 2010). Mice, as obligate prey, appear to prioritise salience and deviance detection over probabilistic filtering, consistent with a broader strategy of adapting dynamically to potential threats in their environment. Although some aspects remain speculative, these interpretations are grounded in observed cross-species differences and suggest that attentional architectures may be shaped not only by immediate task demands but also by broader evolutionary pressures influencing sensory priorities.

At the same time, these cross-species comparisons underscore the need for caution. Future work combining comparative behavioural assays with neural

recording and manipulation across species will be critical for determining how deeply these attentional differences are rooted in neural architecture versus task-induced strategies. Nevertheless, the patterns observed across ferrets, mice, and humans suggest that ecological roles offer a powerful framework for interpreting divergences in auditory attention and sensory adaptation.

Across species and sensory tasks, this thesis provides evidence that auditory attention and spatial perception are shaped by a dynamic interplay between environmental structure, internal state, and possibly ecological priorities. By combining behavioural paradigms across ferrets, humans, and mice, and by implementing pupillometry and developing ear tracking methods in mice, this work offers new insights into how expectation-driven and deviance-driven processes coexist and compete, and how flexible spatial hearing is maintained even after peripheral disruption. While further comparative and mechanistic studies will be important to deepen and extend these interpretations, the patterns observed here highlight the value of integrating behavioural paradigms with targeted neural investigations. Understanding the mechanisms that support selective attention and spatial hearing is not only critical for basic neuroscience but also has translational relevance, as both processes are known to decline with age (Dobreva et al., 2011; Glyde et al., 2011). The approaches developed here provide a foundation for investigating how sensory expectations, stimulus salience, and adaptive mechanisms interact to shape auditory perception across species.

6 REFERENCES

11 Common Mouse Behaviors Explained | BeChewy. (2014). Chewy. https://be.chewy.com/11-common-mouse-behaviors-explained/

Adam, K. C. S., & De Bettencourt, M. T. (2019). Fluctuations of attention and working memory. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/JOC.70

Alain, C., & Arnott, S. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience : A Journal and Virtual Library*, *5*(1), d202. https://doi.org/10.2741/ALAIN

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(21), 12301–12306. https://doi.org/10.1073/PNAS.211209098/ASSET/CEDB0B44-F803-4238-8151-49853C56DBF3/ASSETS/GRAPHIC/PQ2112090003.JPEG

Alain, C., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *Neuroreport*, *6*(1), 140–144. https://doi.org/10.1097/00001756-199412300-00036

Alanazi, F. I., Kalia, S. K., Hodaie, M., Lopez Rios, A. L., Lozano, A. M., Milosevic, L., & Hutchison, W. D. (2023). Top-down control of human motor thalamic neuronal activity during the auditory oddball task. *Npj Parkinson's Disease*, *9*(1). https://doi.org/10.1038/s41531-023-00493-1

Alho, K., Rinne, T., Herron, T. J., & Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. *Hearing Research*, *307*, 29–41. https://doi.org/10.1016/J.HEARES.2013.08.001

Allan, L. G., & Kristofferson, A. B. (1974). Psychophysical theories of duration discrimination. *Perception & Psychophysics*, *16*(1), 26–34. https://doi.org/10.3758/BF03203244/METRICS

Allen, P. D., & Ison, J. R. (2010). Sensitivity of the mouse to changes in azimuthal sound location: Angular separation, spectral composition, and sound level. *Behavioral Neuroscience*, *124*(2), 265–277. https://doi.org/10.1037/A0018913

Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, *14*(4), 1–1. https://doi.org/10.1167/14.4.1

Alves-Pinto, A., Sollini, J., Wells, T., & Sumner, C. J. (2016). Behavioural estimates of auditory filter widths in ferrets using notched-noise maskers. *The Journal of the Acoustical Society of America*, *139*(2), EL19. https://doi.org/10.1121/1.4941772

Antunes, F. M., Nelken, I., Covey, E., & Malmierca, M. S. (2010). Stimulus-specific adaptation in the auditory thalamus of the anesthetized rat. *PloS One*, *5*(11). https://doi.org/10.1371/JOURNAL.PONE.0014071

Aoki, R., Tsubota, T., Goya, Y., & Benucci, A. (2017). An automated platform for high-throughput mouse behavior and physiology with voluntary head-fixation. *Nature Communications 2017 8:1*, *8*(1), 1–9. https://doi.org/10.1038/s41467-017-01371-0

Apostolides, P. F., & Trussell, L. O. (2014). Superficial stellate cells of the dorsal cochlear nucleus. *Frontiers in Neural Circuits*, *8*(JUNE), 96941. https://doi.org/10.3389/FNCIR.2014.00063/BIBTEX

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). https://doi.org/10.1098/RSTB.2016.0047

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage*, *22*(1), 401–408. https://doi.org/10.1016/J.NEUROIMAGE.2004.01.014

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(Volume 28, 2005), 403–450. https://doi.org/10.1146/ANNUREV.NEURO.28.061604.135709/CITE/REFWORKS

Atiani, S., Elhilali, M., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Task Difficulty and Performance Induce Diverse Adaptive Patterns in Gain and Shape of Primary Auditory Cortical Receptive Fields. *Neuron*, *61*(3), 467–480. https://doi.org/10.1016/j.neuron.2008.12.027

Auksztulewicz, R., & Friston, K. (2015). Attentional Enhancement of Auditory Mismatch Responses: a DCM/MEG Study. *Cerebral Cortex*, *25*(11), 4273–4283. https://doi.org/10.1093/CERCOR/BHU323

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/J.JML.2007.12.005

Balmer, T. S., & Trussell, L. O. (2021). Trigeminal Contributions to the Dorsal Cochlear Nucleus in Mouse. *Frontiers in Neuroscience*, *15*, 715954. https://doi.org/10.3389/FNINS.2021.715954/BIBTEX

Barkus, C., Bergmann, C., Branco, T., Carandini, M., Chadderton, P. T., Galiñanes, G. L., Gilmour, G., Huber, D., Huxter, J. R., Khan, A. G., King, A. J., Maravall, M., O'Mahony, T., Ragan, C. I., Robinson, E. S. J., Schaefer, A. T., Schultz, S. R., Sengpiel, F., & Prescott, M. J. (2022). Refinements to rodent head fixation and fluid/food control for neuroscience. *Journal of Neuroscience Methods*, *381*, 109705. https://doi.org/10.1016/J.JNEUMETH.2022.109705

Beebe, N. L., Herrera, Y. N., Noftz, W. A., Roberts, M. T., & Schofield, B. R. (2023). Characterization of three cholinergic inputs to the cochlear nucleus. *Journal of Chemical Neuroanatomy*, *131*. https://doi.org/10.1016/J.JCHEMNEU.2023.102284

Behrens, D., & Klump, G. M. (2016). Comparison of mouse minimum audible angle determined in prepulse inhibition and operant conditioning procedures. *Hearing Research*, *333*, 167–178. https://doi.org/10.1016/J.HEARES.2016.01.011

Berger, J. I., Coomber, B., Shackleton, T. M., Palmer, A. R., & Wallace, M. N. (2013). A novel behavioural approach to detecting tinnitus in the guinea pig. *Journal of Neuroscience Methods*, *213*(2), 188. https://doi.org/10.1016/J.JNEUMETH.2012.12.023

Biderman, D., Whiteway, M. R., Hurwitz, C., Greenspan, N., Lee, R. S., Vishnubhotla, A., Warren, R., Pedraja, F., Noone, D., Schartner, M. M., Huntenburg, J. M., Khanal, A., Meijer, G. T., Noel, J. P., Pan-Vazquez, A., Socha, K. Z., Urai, A. E., Cunningham, J. P., Sawtell, N. B., & Paninski, L. (2024). Lightning Pose: improved animal pose estimation via semi-supervised learning, Bayesian ensembling and cloud-native open-source tools. *Nature Methods*, *21*(7), 1316–1328. https://doi.org/10.1038/S41592-024-02319-1

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience 2013 14:10*, *14*(10), 693–707. https://doi.org/10.1038/nrn3565

Bizley, J. K., Nodal, F. R., Parsons, C. H., & King, A. J. (2007). Role of auditory cortex in sound localization in the midsagittal plane. *Journal of Neurophysiology*, *98*(3), 1763–1774. https://doi.org/10.1152/JN.00444.2007/ASSET/IMAGES/LARGE/Z9K0090783770010.JPEG

Bizley, J. K., Walker, K. M. M., Silverman, B. W., King, A. J., & Schnupp, J. W. H. (2009). Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. *Journal of Neuroscience*, *29*(7), 2064–2075. https://doi.org/10.1523/JNEUROSCI.4755-08.2009

Blanchard, D. C., Griebel, G., Pobbe, R., & Blanchard, R. J. (2011). Risk assessment as an evolved threat detection and analysis process. *Neuroscience & Biobehavioral Reviews*, *35*(4), 991–998. https://doi.org/10.1016/J.NEUBIOREV.2010.10.016

Blauert, J. (1969). Sound localization in the median plane. *Acta Acustica United with Acustica*, *22*(4), 205–213.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.

Boersma, P., & Weenink, D. (2024). *Praat: doing phonetics by computer* (6.4.12).

Bregman, A. (1990). Chapter 1: The Auditory Scene. In *Auditory Scene Analysis: The Perceptual Organization of Sounds*.

Broadbent, D. E. (1958). *Perception and communication* [Book]. Pergamon Press, c1958.

Brodbeck, C., Jiao, A., Hong, L. E., & Simon, J. Z. (2020). Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. *PLOS Biology*, *18*(10), e3000883. https://doi.org/10.1371/JOURNAL.PBIO.3000883

Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(1), 361–365. https://doi.org/10.1073/PNAS.0907658106/SUPPL_FILE/PNAS.200907658SI.PDF

Buus, S., Schorer, E., Florentine, M., & Zwicker, E. (1986). Decision rules in detection of simple and complex tones. *The Journal of the Acoustical Society of America*, *80*(6), 1646–1657. https://doi.org/10.1121/1.394329

Caporale, N., & Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience*, *31*, 25–46. https://doi.org/10.1146/ANNUREV.NEURO.31.060407.125639

Carbajal, G. V., & Malmierca, M. S. (2018). The Neuronal Basis of Predictive Coding Along the Auditory Pathway: From the Subcortical Roots to Cortical Deviance Detection. *Trends in Hearing*, *22*, 2331216518784822. https://doi.org/10.1177/2331216518784822

Carlile, S. (2014). The plastic ear and perceptual relearning in auditory spatial perception. *Frontiers in Neuroscience*, *8*(8 JUL), 98671. https://doi.org/10.3389/FNINS.2014.00237/BIBTEX

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 115–127. https://doi.org/10.1037/0096-1523.27.1.115

Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/COGP.1998.0681

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clayton, K. K., Asokan, M. M., Watanabe, Y., Hancock, K. E., & Polley, D. B. (2021). Behavioral Approaches to Study Top-Down Influences on Active Listening. *Frontiers in Neuroscience*, *15*, 666627. https://doi.org/10.3389/FNINS.2021.666627

Cramer, E. M., & Huggins, W. H. (1958). Creation of Pitch through Binaural Interaction. *The Journal of the Acoustical Society of America*, *30*(5), 413–417. https://doi.org/10.1121/1.1909628

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(4), 643–656. https://doi.org/10.1037/0096-1523.30.4.643

Dai, Best, & Shinn-Cunningham. (2018). Sensorineural hearing loss degrades behavioral and physiological measures of human spatial selective auditory attention. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), E3286–E3295. https://doi.org/10.1073/PNAS.1721226115/-/DCSUPPLEMENTAL

Dai, H., & Buus, S. (1991). Effect of gating the masker on frequency-selective listening. *The Journal of the Acoustical Society of America*, *89*(4), 1816–1818. https://doi.org/10.1121/1.401015

David, S. V., Fritz, J. B., & Shamma, S. A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences*, *109*(6), 2144–2149. https://doi.org/10.1073/PNAS.1117717109

de Gee, J. W., Mridha, Z., Hudson, M., Shi, Y., Ramsaywak, H., Smith, S., Karediya, N., Thompson, M., Jaspe, K., Jiang, H., Zhang, W., & McGinley, M. J. (2024). Strategic stabilization of arousal boosts sustained attention. *Current Biology*, *34*(18), 4114-4128.e6. https://doi.org/10.1016/J.CUB.2024.07.070

de Gee, J. W., Tsetsos, K., Schwabe, L., Urai, A. E., McCormick, D., McGinley, M. J., & Donner, T. H. (2020). Pupil-linked phasic arousal predicts a reduction of choice bias across species and decision domains. *ELife*, *9*, 1–25. https://doi.org/10.7554/ELIFE.54014

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? In *Trends in Cognitive Sciences* (Vol. 22, Issue 9). https://doi.org/10.1016/j.tics.2018.06.002

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(Volume 18, 1995), 193–222. https://doi.org/10.1146/ANNUREV.NE.18.030195.001205/CITE/REFWORKS

Devine, C. A., Gaffney, C., Loughnane, G., Kelly, S. P., & O'Connell, R. G. (2019). The role of premature evidence accumulation in making difficult perceptual decisions under temporal uncertainty. *ELife*, *8*. https://doi.org/10.7554/ELIFE.48526

De-Wit, L., Machilsen, B., & Putzeys, T. (2010). Predictive Coding and the Neural Response to Predictable Stimuli. *The Journal of Neuroscience*, *30*(26), 8702. https://doi.org/10.1523/JNEUROSCI.2248-10.2010

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. https://doi.org/10.1152/JN.00297.2011/ASSET/IMAGES/LARGE/Z9K0121110960009.JPEG

Dobreva, M. S., O'Neill, W. E., & Paige, G. D. (2011). Influence of aging on human sound localization. *Journal of Neurophysiology*, *105*(5), 2471. https://doi.org/10.1152/JN.00951.2010

Dolensek, N., Gehrlach, D. A., Klein, A. S., & Gogolla, N. (2020). Facial expressions of emotion states and their neuronal correlates in mice. *Science*, *368*(6486). https://doi.org/10.1126/SCIENCE.AAZ9468/SUPPL_FILE/AAZ9468S3.MP4

Ehret, G., & Dreyer, A. (1984). Localization of tones and noise in the horizontal plane by unrestrained house mice (Mus musculus). *The Journal of Experimental Biology*, *109*, 163–174. https://doi.org/10.1242/JEB.109.1.163

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural Mechanisms of Involuntary Attention to Acoustic Novelty and Change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. https://doi.org/10.1162/089892998562997

Escera, C., Corral, M. J., & Yago, E. (2002). An electrophysiological and behavioral investigation of involuntary attention towards auditory frequency, duration and intensity changes. *Cognitive Brain Research*, *14*(3), 325–332. https://doi.org/10.1016/S0926-6410(02)00135-0

Esterman, M., Noonan, S. K., Rosenberg, M., & Degutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(11), 2712–2723. https://doi.org/10.1093/CERCOR/BHS261

Farley, B. J., Quirk, M. C., Doherty, J. J., & Christian, E. P. (2010). Stimulus-specific adaptation in auditory cortex is an NMDA-independent process distinct from the sensory novelty encoded by the mismatch negativity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(49), 16475–16484. https://doi.org/10.1523/JNEUROSCI.2793-10.2010

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 7028. https://doi.org/10.3389/FNHUM.2010.00215/XML

Finlayson, K., Lampe, J. F., Hintze, S., Würbel, H., & Melotti, L. (2016). Facial Indicators of Positive Emotions in Rats. *PLOS ONE*, *11*(11), e0166446. https://doi.org/10.1371/JOURNAL.PONE.0166446

Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Https://Doi.Org/10.1111/1467-9280.00392*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Frederick, D. E., Rojas-Líbano, D., Scott, M., & Kay, L. M. (2011). Rat behavior in go/no-go and two-alternative choice odor discrimination: differences and similarities. *Behavioral Neuroscience*, *125*(4), 588–603. https://doi.org/10.1037/A0024371

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science*, *291*(5502), 312–316. https://doi.org/10.1126/SCIENCE.291.5502.312

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/RSTB.2005.1622;WEBSITE:WEBSITE:RSJ-SITE;WGROUP:STRING:PUBLICATION

Fritz, J. B., Elhilali, M., & Shamma, S. (2005a). Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hearing Research*, *206*(1–2), 159–176. https://doi.org/10.1016/J.HEARES.2005.01.015

Fritz, J. B., Elhilali, M., & Shamma, S. A. (2005b). Differential Dynamic Plasticity of A1 Receptive Fields during Multiple Spectral Tasks. *Journal of Neuroscience*, *25*(33), 7623–7635. https://doi.org/10.1523/JNEUROSCI.1318-05.2005

Fritz, J. B., Elhilali, M., & Shamma, S. A. (2007). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *Journal of Neurophysiology*, *98*(4), 2337–2346. https://doi.org/10.1152/JN.00552.2007/ASSET/IMAGES/LARGE/Z9K0100784710007.JPEG

Fritz, J. B., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience 2003 6:11*, *6*(11), 1216–1223. https://doi.org/10.1038/nn1141

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/BUL0000210

Fujino, K., & Oertel, D. (2003). Bidirectional synaptic plasticity in the cerebellum-like mammalian dorsal cochlear nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(1), 265–270. https://doi.org/10.1073/PNAS.0135345100/ASSET/8EA63426-7A2D-4F33-BB44-0BFBBD6F0E43/ASSETS/GRAPHIC/PQ0135345005.JPEG

Gardner, M. B., & Gardner, R. S. (1973). Problem of localization in the median plane: effect of pinnae cavity occlusion. *The Journal of the Acoustical Society of America*, *53*(2), 400–408.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/J.CLINPH.2008.11.029

Glyde, H., Hickson, L., Cameron, S., & Dillon, H. (2011). Problems Hearing in Noise in Older Adults: A Review of Spatial Processing Disorder. *Trends in Amplification*, *15*(3), 116–126.

https://doi.org/10.1177/1084713811424885/ASSET/C05BC25C-ED21-4883-84D8-D939B36E26AE/ASSETS/IMAGES/LARGE/10.1177_1084713811424885-FIG1.JPG

Goltstein, P. M., Reinert, S., Glas, A., Bonhoeffer, T., & Hübener, M. (2018). Food and water restriction lead to differential learning behaviors in a head-fixed two-choice visual discrimination task for mice. *PLoS ONE*, *13*(9), e0204066. https://doi.org/10.1371/JOURNAL.PONE.0204066

Goujon, A., Didierjean, A., & Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences*, *19*(9), 524–533. https://doi.org/10.1016/J.TICS.2015.07.009

Green, D. M. (1958). Detection of Multiple Component Signals in Noise. *The Journal of the Acoustical Society of America*, *30*(10), 904–911. https://doi.org/10.1121/1.1909400

Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-Response Characteristic of Auditory Observers Detecting Signals of a Single Frequency in Noise: The Probe-Signal Method. *The Journal of the Acoustical Society of America*, *44*(6), 1513. https://doi.org/10.1121/1.1911290

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

Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiological Reviews*, *90*(3), 983–1012. https://doi.org/10.1152/PHYSREV.00026.2009/ASSET/IMAGES/LARGE/Z9J0031025480014.JPEG

Hasanzadeh Pashang, S., Zare, H., Alipour, A., & Sharif-Alhoseini, M. (2021). The effectiveness of cognitive rehabilitation in improving visual and auditory attention in ischemic stroke patients. *Acta Neurologica Belgica*, *121*(4), 915–920. https://doi.org/10.1007/S13760-020-01288-4/TABLES/5

Heffner, H. E., & Heffner, R. S. (2007). Hearing ranges of laboratory animals. *Journal of the American Association for Laboratory Animal Science : JAALAS*, *46*(1), 20–22. https://europepmc.org/article/med/17203911

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73. https://doi.org/10.1016/J.NEUROSCIENCE.2017.07.061

Hill, K. T., & Miller, L. M. (2010). Auditory Attentional Control and Selection during Cocktail Party Listening. *Cerebral Cortex*, *20*(3), 583–590. https://doi.org/10.1093/CERCOR/BHP124

Hofman, P. M., Van Riswick, J. G. A., & Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience 1998 1:5*, *1*(5), 417–421. https://doi.org/10.1038/1633

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, *16*(4). https://doi.org/10.1111/j.0956-7976.2005.01532.x

Homma, N. Y., Bajo, V. M., Happel, M. F. K., Nodal, F. R., & King, A. J. (2016). Mistuning detection performance of ferrets in a go/no-go task. *The Journal of the Acoustical Society of America*, *139*(6), EL246–EL251. https://doi.org/10.1121/1.4954378

Imig, T. J., Bibikov, N. G., Poirier, P., & Samson, F. K. (2000). Directionality derived from pinna-cue spectral notches in cat dorsal cochlear nucleus. *Journal of Neurophysiology*, *83*(2), 907–925. https://doi.org/10.1152/JN.2000.83.2.907

Ito, S., Si, Y., Feldheim, D. A., & Litke, A. M. (2020). Spectral cues are necessary to encode azimuthal auditory space in the mouse superior colliculus. *Nature Communications 2020 11:1*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-14897-7

JAX Handbook: Genetically Standardized Mice (6th ed.). (2009). The Jackson Laboratory. https://resources.jax.org/therapeutic-area-essentials/jax-handbook-genetically-standardized-mice

Johnson, K. R., Erway, L. C., Cook, S. A., Willott, J. F., & Zheng, Q. Y. (1997). A major gene affecting age-related hearing loss in C57BL/6J mice. *Hearing Research*, *114*(1–2), 83–92. https://doi.org/10.1016/S0378-5955(97)00155-X

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11793–11799. https://doi.org/10.1073/PNAS.97.22.11793/ASSET/EAC7F62B-3082-4E16-BDC3-AC70B43672A7/ASSETS/GRAPHIC/PQ2003419009.JPEG

Kacelnik, O., Nodal, F. R., Parsons, C. H., & King, A. J. (2006). Training-Induced Plasticity of Auditory Localization in Adult Mammals. *PLOS Biology*, *4*(4), e71. https://doi.org/10.1371/JOURNAL.PBIO.0040071

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object Vision in a Structured World. In *Trends in Cognitive Sciences* (Vol. 23, Issue 8). https://doi.org/10.1016/j.tics.2019.04.013

Kanold, P. O., & Young, E. D. (2001). Proprioceptive Information from the Pinna Provides Somatosensory Input to Cat Dorsal Cochlear Nucleus. *Journal of Neuroscience*, *21*(19), 7848–7858. https://doi.org/10.1523/JNEUROSCI.21-19-07848.2001

Keating, P., & King, A. J. (2013). Developmental plasticity of spatial hearing following asymmetric hearing loss: Context-dependent cue integration and its clinical implications. *Frontiers in Systems Neuroscience*, *7*(DEC), 73248. https://doi.org/10.3389/FNSYS.2013.00123/PDF

Kelly, J. B., Kavanagh, G. L., & Dalton, J. C. H. (1986). Hearing in the ferret (Mustela putorius): Thresholds for pure tone detection. *Hearing Research*, *24*(3), 269–275. https://doi.org/10.1016/0378-5955(86)90025-0

Kiani, M. M., Mousavi, Z., & Aghajan, H. (2022). The most descriptive surprise definition for brain's EEG response to visual and auditory oddball tasks. *2022 30th International Conference on Electrical Engineering, ICEE 2022*. https://doi.org/10.1109/ICEE55646.2022.9827347

King, A. J., Kacelnik, O., Mrsic-Flogel, T. D., Schnupp, J. W. H., Parsons, C. H., & Moore, D. R. (2001). How Plastic Is Spatial Hearing? *Audiology and Neurotology*, *6*(4), 182–186. https://doi.org/10.1159/000046829

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience*, *30*(Volume 30, 2007), 57–78. https://doi.org/10.1146/ANNUREV.NEURO.30.051606.094256/CITE/REFWORKS

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2). https://doi.org/10.1016/j.neuron.2012.04.034

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*(1–2), 97–126. https://doi.org/10.1016/S0010-0277(00)00155-4

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biological Psychology*, *74*(1), 1–19. https://doi.org/10.1016/J.BIOPSYCHO.2006.06.001

Kurmanavičiūtė, D., Kataja, H., Jas, M., Välilä, A., & Parkkonen, L. (2023). Target of selective auditory attention can be robustly followed with MEG. *Scientific Reports 2023 13:1*, *13*(1), 1–10. https://doi.org/10.1038/S41598-023-37959-4

Langford, D. J., Bailey, A. L., Chanda, M. L., Clarke, S. E., Drummond, T. E., Echols, S., Glick, S., Ingrao, J., Klassen-Ross, T., Lacroix-Fralish, M. L., Matsumiya, L., Sorge, R. E., Sotocinal, S. G., Tabaka, J. M., Wong, D., Van Den Maagdenberg, A. M. J. M., Ferrari, M. D., Craig, K. D., & Mogil, J. S. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nature Methods 2010 7:6*, *7*(6), 447–449. https://doi.org/10.1038/nmeth.1455

Larsen, R. S., & Waters, J. (2018). Neuromodulatory correlates of pupil dilation. *Frontiers in Neural Circuits*, *12*, 309004. https://doi.org/10.3389/FNCIR.2018.00021/BIBTEX

Lauer, A. M., Slee, S. J., & May, B. J. (2011). Acoustic Basis of Directional Acuity in Laboratory Mice. *JARO: Journal of the Association for Research in Otolaryngology*, *12*(5), 633. https://doi.org/10.1007/S10162-011-0279-Y

Lee, A. K. C., Rajaram, S., Xia, J., Bharadwaj, H., Larson, E., Hämäläinen, M. S., & Shinn-Cunningham, B. G. (2013). Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. *Frontiers in Neuroscience*, *6*(JAN). https://doi.org/10.3389/FNINS.2012.00190

Lemus, L., Hernández, A., & Romo, R. (2009). Neural encoding of auditory discrimination in ventral premotor cortex. *Proceedings of the National Academy of Sciences*, *106*(34), 14640–14645. https://doi.org/10.1073/PNAS.0907505106

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477. https://doi.org/10.1121/1.1912375

Liu, C. C., & Watanabe, T. (2012). Accounting for speed–accuracy tradeoff in perceptual learning. *Vision Research*, *61*, 107–114. https://doi.org/10.1016/J.VISRES.2011.09.007

Lukas, J. H. (1980). Human Auditory Attention: The Olivocochlear Bundle May Function as a Peripheral Filter. *Psychophysiology*, *17*(5), 444–452. https://doi.org/10.1111/J.1469-8986.1980.TB00181.X

Lukas, J. H. (1981). The Role of Efferent Inhibition in Human Auditory Attention: An Examination of the Auditory Brainstem Potentials. *International Journal of Neuroscience*, *12*(2), 137–145. https://doi.org/10.3109/00207458108985796

Luthra, S., Luor, A., Tierney, A. T., Dick, F., & Holt, L. L. (2024). Distributional learning drives statistical suppression and deafening. *BioRxiv*, 2024.09.09.612146. https://doi.org/10.1101/2024.09.09.612146

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., Pittet, A., & Clarke, S. (2001). Distinct Pathways Involved in Sound Recognition and Localization: A Human fMRI Study. *NeuroImage*, *14*(4), 802–816. https://doi.org/10.1006/NIMG.2001.0888

Malmierca, M. S., Cristaudo, S., Pérez-González, D., & Covey, E. (2009). Stimulus-Specific Adaptation in the Inferior Colliculus of the Anesthetized Rat. *Journal of Neuroscience*, *29*(17), 5483–5493. https://doi.org/10.1523/JNEUROSCI.4153-08.2009

Masís, J., Chapman, T., Rhee, J. Y., Cox, D. D., & Saxe, A. M. (2023). Strategically managing learning during perceptual decision making. *ELife*, *12*. https://doi.org/10.7554/ELIFE.64978

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience 2018 21:9*, *21*(9), 1281–1289. https://doi.org/10.1038/S41593-018-0209-Y

Mayer, A. R., Yang, Z., Yeo, R. A., Pena, A., Ling, J. M., Mannell, M. V., Stippler, M., & Mojtahed, K. (2012). A functional MRI study of multimodal selective attention following mild traumatic brain injury. *Brain Imaging and Behavior*, *6*(2), 343–354. https://doi.org/10.1007/S11682-012-9178-Z/FIGURES/5

McCallum, W. C., Curry, S. H., Pocock, P. V., & Papakostopoulos, D. (1983). Brain Event-Related Potentials as Indicators of Early Selective Processes in Auditory Target Localization. *Psychophysiology*, *20*(1), 1–17. https://doi.org/10.1111/J.1469-8986.1983.TB00891.X

McGinley, M. J., David, S. V., & McCormick, D. A. (2015). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*, *87*(1), 179–192. https://doi.org/10.1016/j.neuron.2015.05.038

McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., Tolias, A. S., Cardin, J. A., & McCormick, D. A. (2015). Waking State: Rapid Variations Modulate Neural and Behavioral Responses. *Neuron*, *87*(6), 1143–1161. https://doi.org/10.1016/J.NEURON.2015.09.012

Medalia, A., Aluma, M., Tryon, W., & Merriam, A. E. (1998). Effectiveness of Attention Training in Schizophrenia. *Schizophrenia Bulletin*, *24*(1), 147–152. https://doi.org/10.1093/OXFORDJOURNALS.SCHBUL.A033306

Mellott, J. G., Motts, S. D., & Schofield, B. R. (2011). Multiple Origins of Cholinergic Innervation of the Cochlear Nucleus. *Neuroscience*, *180*, 138. https://doi.org/10.1016/J.NEUROSCIENCE.2011.02.010

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature 2012 485:7397*, *485*(7397), 233–236. https://doi.org/10.1038/nature11020

Middlebrooks, J. C. (2015). Sound localization. *Handbook of Clinical Neurology*, *129*, 99–116. https://doi.org/10.1016/B978-0-444-62630-1.00006-8

Middlebrooks, J. C., & Onsan, Z. A. (2012). Stream segregation with high spatial acuity. *The Journal of the Acoustical Society of America*, *132*(6), 3896–3911. https://doi.org/10.1121/1.4764879

Middleton, J. W., Kiritani, T., Pedersen, C., Turner, J. G., Shepherd, G. M. G., & Tzounopoulos, T. (2011). Mice with behavioral evidence of tinnitus exhibit dorsal cochlear nucleus hyperactivity

because of decreased GABAergic inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(18), 7601–7606. https://doi.org/10.1073/PNAS.1100223108

Miller, L. M., & Recanzone, G. H. (2009). Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 5931–5935. https://doi.org/10.1073/PNAS.0901023106/SUPPL_FILE/0901023106SI.PDF

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

Mondor, T. A., & Bregman, A. S. (1994). Allocating attention to frequency regions. *Perception & Psychophysics*, *56*(3), 268–276. https://doi.org/10.3758/BF03209761/METRICS

Montes-Lourido, P., Kar, M., Kumbam, I., & Sadagopan, S. (2021). Pupillometry as a reliable metric of auditory detection and discrimination across diverse stimulus paradigms in animal models. *Scientific Reports 2021 11:1*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-82340-y

Mridha, Z., de Gee, J. W., Shi, Y., Alkashgari, R., Williams, J., Suminski, A., Ward, M. P., Zhang, W., & McGinley, M. J. (2021). Graded recruitment of pupil-linked neuromodulation by parametric stimulation of the vagus nerve. *Nature Communications 2021 12:1*, *12*(1), 1–14. https://doi.org/10.1038/S41467-021-21730-2

Mulder, M. J., Keuken, M. C., van Maanen, L., Boekel, W., Forstmann, B. U., & Wagenmakers, E. J. (2013). The speed and accuracy of perceptual decisions in a random-tone pitch task. *Attention, Perception, and Psychophysics*, *75*(5), 1048–1058. https://doi.org/10.3758/S13414-013-0447-8/FIGURES/6

Mulder, M. J., van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences - a model-based review. *Neuroscience*, *277*, 872–884. https://doi.org/10.1016/J.NEUROSCIENCE.2014.07.031/ASSET/CDB38D40-488C-40C9-82A6-A2FC9E4DE18D/MAIN.ASSETS/GR7.SML

Musicant, A. D., Chan, J. C. K., & Hind, J. E. (1990). Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons. *The Journal of the Acoustical Society of America*, *87*(2), 757–781. https://doi.org/10.1121/1.399545

Musz, E., Weber, M. J., & Thompson-Schill, S. L. (2015). Visual statistical learning is not reliably modulated by selective attention to isolated events. *Attention, Perception & Psychophysics*, *77*(1), 78–96. https://doi.org/10.3758/S13414-014-0757-5

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

Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/J.1469-8986.1987.TB00311.X

Nelder, J. A. (1977). A Reformulation of Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, *140*(1), 48. https://doi.org/10.2307/2344517

Nelken, I., & Ulanovsky, N. (2007). Mismatch negativity and stimulus-specific adaptation in animal models. *Journal of Psychophysiology*, *21*(3–4), 214–223. https://doi.org/10.1027/0269-8803.21.34.214

Niwa, M., Johnson, J. S., O'Connor, K. N., & Sutter, M. L. (2012). Active Engagement Improves Primary Auditory Cortical Neurons' Ability to Discriminate Temporal Modulation. *Journal of Neuroscience*, *32*(27), 9323–9334. https://doi.org/10.1523/JNEUROSCI.5832-11.2012

Nodal, F. R., Bajo, V. M., & King, A. J. (2012). Plasticity of spatial hearing: behavioural effects of cortical inactivation. *The Journal of Physiology*, *590*(Pt 16), 3965. https://doi.org/10.1113/JPHYSIOL.2011.222828

Oatman, L. C., & Anderson, B. W. (1977). Effects of visual attention on tone burst evoked auditory potentials. *Experimental Neurology*, *57*(1), 200–211. https://doi.org/10.1016/0014-4886(77)90057-7

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, *13*(1), 14–19. https://doi.org/10.1016/j.tics.2008.09.005

Otazu, G. H., Tai, L. H., Yang, Y., & Zador, A. M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nature Neuroscience 2009 12:5*, *12*(5), 646–654. https://doi.org/10.1038/nn.2306

Ouchi, Y., Meguro, K., Akanuma, K., Kato, Y., & Yamaguchi, S. (2015). Normal Hearing Ability but Impaired Auditory Selective Attention Associated with Prediction of Response to Donepezil in Patients with Alzheimer's Disease. *Behavioural Neurology*, *2015*(1), 540348. https://doi.org/10.1155/2015/540348

Park, C. R., Willott, J. F., & Walton, J. P. (2024). Age-related changes of auditory sensitivity across the life span of CBA/CaJ mice. *Hearing Research*, *441*. https://doi.org/10.1016/J.HEARES.2023.108921

Parmentier, F. B. R. (2014). The cognitive determinants of behavioral distraction by deviant auditory stimuli: A review. *Psychological Research*, *78*(3), 321–338. https://doi.org/10.1007/S00426-013-0534-4

Parsons, J. E., Lim, E., & Voigt, H. F. (2001). Type III units in the gerbil dorsal cochlear nucleus may be spectral notch detectors. *Annals of Biomedical Engineering*, *29*(10), 887–896. https://doi.org/10.1114/1.1408924

Parsons, Lanyon, Schnupp, & King. (1999). Effects of altering spectral cues in infancy on horizontal and vertical sound localization by adult ferrets. *Journal of Neurophysiology*, *82*(5), 2294–2309. https://doi.org/10.1152/JN.1999.82.5.2294/ASSET/IMAGES/LARGE/9K1190583012.JPEG

Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch Negativity: Different Water in the Same River. *Audiology and Neurotology*, *5*(3–4), 111–139. https://doi.org/10.1159/000013875

Populin, L. C., Tollin, D. J., & Yin, T. C. T. (2004). Effect of eye position on saccades and neuronal responses to acoustic stimuli in the superior colliculus of the behaving cat. *Journal of Neurophysiology*, *92*(4), 2151–2167. https://doi.org/10.1152/JN.00453.2004/ASSET/IMAGES/LARGE/Z9K0100441690016.JPEG

Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. In *Trends in Cognitive Sciences* (Vol. 24, Issue 1). https://doi.org/10.1016/j.tics.2019.11.003

Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, *5*(SEP), 1–12. https://doi.org/10.3389/FNINS.2011.00115/BIBTEX

Prugh, L. R., Stoner, C. J., Epps, C. W., Bean, W. T., Ripple, W. J., Laliberte, A. S., & Brashares, J. S. (2009). The Rise of the Mesopredator. *BioScience*, *59*(9), 779–791. https://doi.org/10.1525/BIO.2009.59.9.9

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580;KWRD=BIOMEDICINE

Rauschecker, J. P. (1997). Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Oto-Laryngologica. Supplementum*, *532*(532), 34–38. https://doi.org/10.3109/00016489709126142

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, *271*(1–2), 16–25. https://doi.org/10.1016/J.HEARES.2010.09.001

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Frontiers in Evolutionary Neuroscience*, *4*(MAY), 18154. https://doi.org/10.3389/FNEVO.2012.00007/BIBTEX

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6). https://doi.org/10.1016/S0022-5371(67)80149-X

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications 2016 7:1*, *7*(1), 1–7. https://doi.org/10.1038/ncomms13289

Reiss, L. A. J., & Young, E. D. (2005). Spectral Edge Sensitivity in Neural Circuits of the Dorsal Cochlear Nucleus. *Journal of Neuroscience*, *25*(14), 3680–3691. https://doi.org/10.1523/JNEUROSCI.4963-04.2005

Roberts, P. D., Portfors, C. V., Sawtell, N., & Felix, R. (2006). Model of auditory prediction in the dorsal cochlear nucleus via spike-timing dependent plasticity. *Neurocomputing*, *69*(10–12), 1191–1194. https://doi.org/10.1016/J.NEUCOM.2005.12.073

Robertson, I. H. (2001). Do We Need the "Lateral" in Unilateral Neglect? Spatially Nonselective Attention Deficits in Unilateral Neglect and Their Implications for Rehabilitation. *NeuroImage*, *14*(1), S85–S90. https://doi.org/10.1006/NIMG.2001.0838

Rogalla, M. M., Rauser, I., Schulze, K., Osterhagen, L., & Hildebrandt, K. J. (2020). Mice tune out not in: violation of prediction drives auditory saliency. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1919), 20192001. https://doi.org/10.1098/rspb.2019.2001

Romanski, L. M., Tian, B., Fritz, J. B., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience 1999 2:12*, *2*(12), 1131–1136. https://doi.org/10.1038/16056

Russ, B. E., Lee, Y. S., & Cohen, Y. E. (2007). Neural and behavioral correlates of auditory categorization. *Hearing Research*, *229*(1–2), 204–212. https://doi.org/10.1016/J.HEARES.2006.10.010

Ryugo, D. K., Haenggeli, C. A., & Doucet, J. R. (2003). Multimodal inputs to the granule cell domain of the cochlear nucleus. *Experimental Brain Research*, *153*(4), 477–485. https://doi.org/10.1007/S00221-003-1605-3

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294). https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/JMLA.1996.0032

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental Language Learning: Listening (and Learning) Out of the Corner of Your Ear. *Https://Doi-Org.Ezproxy-Prd.Bodleian.Ox.Ac.Uk/10.1111/j.1467-9280.1997.Tb00690.x*, *8*(2), 101–105. https://doi.org/10.1111/J.1467-9280.1997.TB00690.X

Salay, L. D., Ishiko, N., & Huberman, A. D. (2018). A midline thalamic circuit determines reactions to visual threat. *Nature 2018 557:7704*, *557*(7704), 183–189. https://doi.org/10.1038/s41586-018-0078-2

Santolin, C., & Saffran, J. R. (2018). Constraints on Statistical Learning Across Species. *Trends in Cognitive Sciences*, *22*(1), 52–63. https://doi.org/10.1016/J.TICS.2017.10.003

Scharf, B., Quigley, S., Aoki, C., Peachey, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Perception & Psychophysics*, *42*(3), 215–223. https://doi.org/10.3758/BF03203073/METRICS

Schnupp, J. W. H., King, A. J., & Carlile, S. (1998). Altered spectral localization cues disrupt the development of the auditory space map in the superior colliculus of the ferret. *Journal of Neurophysiology*, *79*(2), 1053–1069. https://doi.org/10.1152/JN.1998.79.2.1053/ASSET/IMAGES/LARGE/JNP.JA30F9.JPEG

Schnupp, Nelken, & King. (2011). Auditory neuroscience : making sense of sound. In *Auditory Neuroscience: Making Sense of Sound*.

Seibold, V. C., Balke, J., & Rolke, B. (2023). Temporal attention. *Frontiers in Cognition*, *2*. https://doi.org/10.3389/FCOGN.2023.1168320/FULL

Shamma, S. (2008). On the Emergence and Awareness of Auditory Objects. *PLOS Biology*, *6*(6), e155. https://doi.org/10.1371/JOURNAL.PBIO.0060155

Shang, C., Liu, Z., Chen, Z., Shi, Y., Wang, Q., Liu, S., Li, D., & Cao, P. (2015). A parvalbumin-positive excitatory visual pathway to trigger fear responses in mice. *Science*, *348*(6242), 1472–1477. https://doi.org/10.1126/SCIENCE.AAA8694/SUPPL_FILE/SHANG.SM.PDF

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Shinn-Cunningham, B. G., & Best, V. (2008). Selective Attention in Normal and Impaired Hearing. *Trends in Amplification*, *12*(4), 283–299.

https://doi.org/10.1177/1084713808325306/ASSET/IMAGES/LARGE/10.1177_1084713808325306-FIG3.JPEG

Shinn-Cunningham, B. G., Durlach, N. I., & Held, R. M. (1998). Adapting to supernormal auditory localization cues. I. Bias and resolution. *The Journal of the Acoustical Society of America*, *103*(6), 3656–3666. https://doi.org/10.1121/1.423088

Shore, S. E. (2005). Multisensory integration in the dorsal cochlear nucleus: unit responses to acoustic and trigeminal ganglion stimulation. *The European Journal of Neuroscience*, *21*(12), 3334–3348. https://doi.org/10.1111/J.1460-9568.2005.04142.X

Shore, S. E., Koehler, S., Oldakowski, M., Hughes, L. F., & Syed, S. (2008). Dorsal cochlear nucleus responses to somatosensory stimulation are enhanced after noise-induced hearing loss. *European Journal of Neuroscience*, *27*(1), 155–168. https://doi.org/10.1111/J.1460-9568.2007.05983.X

Shore, S. E., Roberts, L. E., & Langguth, B. (2016). Maladaptive plasticity in tinnitus--triggers, mechanisms and treatment. *Nature Reviews. Neurology*, *12*(3), 150–160. https://doi.org/10.1038/NRNEUROL.2016.12

Si, Y., Ito, S., Litke, A. M., & Feldheim, D. A. (2022). High-Frequency Hearing Is Required to Compute a Topographic Map of Auditory Space in the Mouse Superior Colliculus. *ENeuro*, *9*(3). https://doi.org/10.1523/ENEURO.0513-21.2022

Sohlberg, M. M., & Mateer, C. A. (2001). Improving Attention and Managing Attentional Problems. *Annals of the New York Academy of Sciences*, *931*(1), 359–375. https://doi.org/10.1111/J.1749-6632.2001.TB05790.X

Southwell, R., & Chait, M. (2018). Enhanced deviant responses in patterned relative to random sound sequences. *Cortex*, *109*. https://doi.org/10.1016/j.cortex.2018.08.032

Stecker, G. C., & Middlebrooks, J. C. (2003). Distributed coding of sound locations in the auditory cortex. *Biological Cybernetics*, *89*(5), 341–349. https://doi.org/10.1007/S00422-003-0439-1/METRICS

Stefanescu, R. A., & Shore, S. E. (2017). Muscarinic acetylcholine receptors control baseline activity and Hebbian stimulus timing-dependent plasticity in fusiform cells of the dorsal cochlear nucleus. *Journal of Neurophysiology*, *117*(3), 1229–1238. https://doi.org/10.1152/JN.00270.2016

Stevens, C., Sanders, L., & Neville, H. (2006). Neurophysiological evidence for selective auditory attention deficits in children with specific language impairment. *Brain Research*, *1111*(1), 143–152. https://doi.org/10.1016/J.BRAINRES.2006.06.114

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 of the United States of America*, *107*(50). https://doi.org/10.1073/pnas.1009020107

Stitt, P., Picinali, L., & Katz, B. F. G. (2019). Auditory Accommodation to Poorly Matched Non-Individual Spectral Localization Cues Through Active Learning. *Scientific Reports 2019 9:1*, *9*(1), 1–14. https://doi.org/10.1038/s41598-018-37873-0

Strauss, D. J., Corona-Strauss, F. I., Schroeer, A., Flotho, P., Hannemann, R., & Hackley, S. A. (2020). Vestigial auriculomotor activity indicates the direction of auditory attention in humans. *ELife*, *9*, 1–18. https://doi.org/10.7554/ELIFE.54536

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. In *Nature Reviews Neuroscience* (Vol. 15, Issue 11). https://doi.org/10.1038/nrn3838

Tabas, A., & von Kriegstein, K. (2021). Adjudicating Between Local and Global Architectures of Predictive Processing in the Subcortical Auditory Pathway. In *Frontiers in Neural Circuits* (Vol. 15). https://doi.org/10.3389/fncir.2021.644743

Teki, S., Chait, M., Kumar, S., Von Kriegstein, K., & Griffiths, T. D. (2011). Brain Bases for Auditory Stimulus-Driven Figure–Ground Segregation. *Journal of Neuroscience*, *31*(1), 164–171. https://doi.org/10.1523/JNEUROSCI.3788-10.2011

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, *292*(5515), 290–293. https://doi.org/10.1126/SCIENCE.1058911/SUPPL_FILE/1058911S2_THUMB.GIF

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–B34. https://doi.org/10.1016/J.COGNITION.2005.01.006

Traer, J., & McDermott, J. H. (2016). Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(48), E7856–E7865. https://doi.org/10.1073/PNAS.1612524113/SUPPL_FILE/PNAS.201612524SI.PDF

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception and Psychophysics*, *66*(7), 1171–1189. https://doi.org/10.3758/BF03196844/METRICS

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Tzounopoulos, T., Kim, Y., Oertel, D., & Trussell, L. O. (2004). Cell-specific, spike timing-dependent plasticities in the dorsal cochlear nucleus. *Nature Neuroscience*, *7*(7), 719–725. https://doi.org/10.1038/NN1272

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience 2003 6:4*, *6*(4), 391–398. https://doi.org/10.1038/nn1032

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Perceived duration of expected and unexpected stimuli. *Psychological Research*, *70*(2), 77–87. https://doi.org/10.1007/S00426-004-0195-4/FIGURES/3

Van Wanrooij, M. M., & Van Opstal, A. J. (2005). Relearning Sound Localization with a New Ear. *The Journal of Neuroscience*, *25*(22), 5413. https://doi.org/10.1523/JNEUROSCI.0850-05.2005

Voola, M., Nguyen, A. T., Marinovic, W., Rajan, G., & Tavora-Vieira, D. (2022). Odd-even oddball task: Evaluating event-related potentials during word discrimination compared to speech-token and tone discrimination. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.983498

Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Walker, Davies, A., Bizley, J. K., Schnupp, J. W. H., & King, A. J. (2017). Pitch discrimination performance of ferrets and humans on a go/no-go task. *BioRxiv*, 165852. https://doi.org/10.1101/165852

Walker, J. (n.d.). *8 Common Behaviours Of Pet Mice, Explained | Coops And Cages*. Retrieved June 2, 2024, from https://www.coopsandcages.com.au/blog/8-common-behaviors-pet-mice-explained/

Wallace, Greenberg, Sawinski, Rulla, Notaro, & Kerr. (2013). Rats maintain an overhead binocular field at the expense of constant fusion. *Nature 2013 498:7452*, *498*(7452), 65–69. https://doi.org/10.1038/nature12153

Wallace, M. N., Berger, J. I., Hockley, A., Sumner, C. J., Akeroyd, M. A., Palmer, A. R., & McNaughton, P. A. (2024). Identifying tinnitus in mice by tracking the motion of body markers in response to an acoustic startle. *Frontiers in Neuroscience*, *18*. https://doi.org/10.3389/FNINS.2024.1452450

Wang, B., & Theeuwes, J. (2018a). How to inhibit a distractor location? Statistical learning versus active, top-down suppression. *Attention, Perception & Psychophysics*, *80*(4), 860–870. https://doi.org/10.3758/S13414-018-1493-Z

Wang, B., & Theeuwes, J. (2018b). Statistical Regularities Modulate Attentional Capture. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 13–17. https://doi.org/10.1037/XHP0000472

Wang, B., & Theeuwes, J. (2018c). Statistical regularities modulate attentional capture independent of search strategy. *Attention, Perception, and Psychophysics*, *80*(7), 1763–1774. https://doi.org/10.3758/S13414-018-1562-3/FIGURES/6

Wei, P., Liu, N., Zhang, Z., Liu, X., Tang, Y., He, X., Wu, B., Zhou, Z., Liu, Y., Li, J., Zhang, Y., Zhou, X., Xu, L., Chen, L., Bi, G., Hu, X., Xu, F., & Wang, L. (2015). Processing of visually evoked innate fear by a non-canonical thalamic pathway. *Nature Communications 2015 6:1*, *6*(1), 1–13. https://doi.org/10.1038/ncomms7756

Wiederhold, M. L. (1970). Variations in the Effects of Electric Stimulation of the Crossed Olivocochlear Bundle on Cat Single Auditory-Nerve-Fiber Responses to Tone Bursts. *The Journal of the Acoustical Society of America*, *48*(4B), 966–977. https://doi.org/10.1121/1.1912235

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. https://doi.org/10.1016/j.tics.2009.09.003

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, *742*(1–2), 239–252. https://doi.org/10.1016/S0006-8993(96)01008-6,

Woods, D. D., Sorkin, R. D., & Boggs, G. J. (1979). Stimulus context and duration discrimination. *Perception & Psychophysics*, *26*(2), 127–132. https://doi.org/10.3758/BF03208306/METRICS

Woods, & McDermott. (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/-/DCSUPPLEMENTAL

Wright, B. A., & Dai, H. (1994). Detection of unexpected tones with short and long durations. *The Journal of the Acoustical Society of America*, *95*(2), 931–938. https://doi.org/10.1121/1.410010

Wu, C., Martel, D. T., & Shore, S. E. (2015). Transcutaneous induction of stimulus-timing-dependent plasticity in dorsal cochlear nucleus. *Frontiers in Systems Neuroscience*, *9*(AUGUST), 158651. https://doi.org/10.3389/FNSYS.2015.00116/BIBTEX

Yao, P. T., Shen, J., Chen, L., Ge, S., & Xiong, Q. (2019). Cortical ensemble activity discriminates auditory attentional states. *Molecular Brain*, *12*(1), 1–9. https://doi.org/10.1186/S13041-019-0502-Z/FIGURES/5

Yilmaz, M., & Meister, M. (2013). Rapid innate defensive responses of mice to looming visual stimuli. *Current Biology*, *23*(20), 2011–2015. https://doi.org/10.1016/j.cub.2013.08.015

Yin, P., Fritz, J. B., & Shamma, S. A. (2014). Rapid Spectrotemporal Plasticity in Primary Auditory Cortex during Behavior. *Journal of Neuroscience*, *34*(12), 4396–4408. https://doi.org/10.1523/JNEUROSCI.2799-13.2014

Young, E. D., & Davis, K. A. (2002). *Circuitry and Function of the Dorsal Cochlear Nucleus*. 160–206. https://doi.org/10.1007/978-1-4757-3654-0_5

Zatorre, R. J., Mondor, T. A., & Evans, A. C. (1999). Auditory Attention to Space and Frequency Activates Similar Cerebral Systems. *NeuroImage*, *10*(5), 544–554. https://doi.org/10.1006/NIMG.1999.0491

Zhao, S., Brown, C. A., Holt, L. L., & Dick, F. (2022). Robust and Efficient Online Auditory Psychophysics. *Trends in Hearing*, *26*. https://doi.org/10.1177/23312165221118792

7 APPENDIX

Supplementary analyses and full model outputs are provided in the following sections. These include additional distributional analyses of reaction times, detailed fixed and random effects tables for behavioural models, and learning curve fits. All statistical models are described in the main text, with corresponding full results reported here for transparency and reproducibility.

7.1 APPENDIX 1 – TONE IN NOISE

7.1.1.1 Differences in Probability Densities

To investigate if there were distinct characteristics associated with common vs. probe frequencies other than direct comparison of reaction times, distributions between frequency types were examined for each condition. Skewness and kurtosis were calculated to assess the symmetry and tail heaviness of reaction times across stimuli. Figure 51 shows density plots for visualization of distributions differences along with skewness and kurtosis.

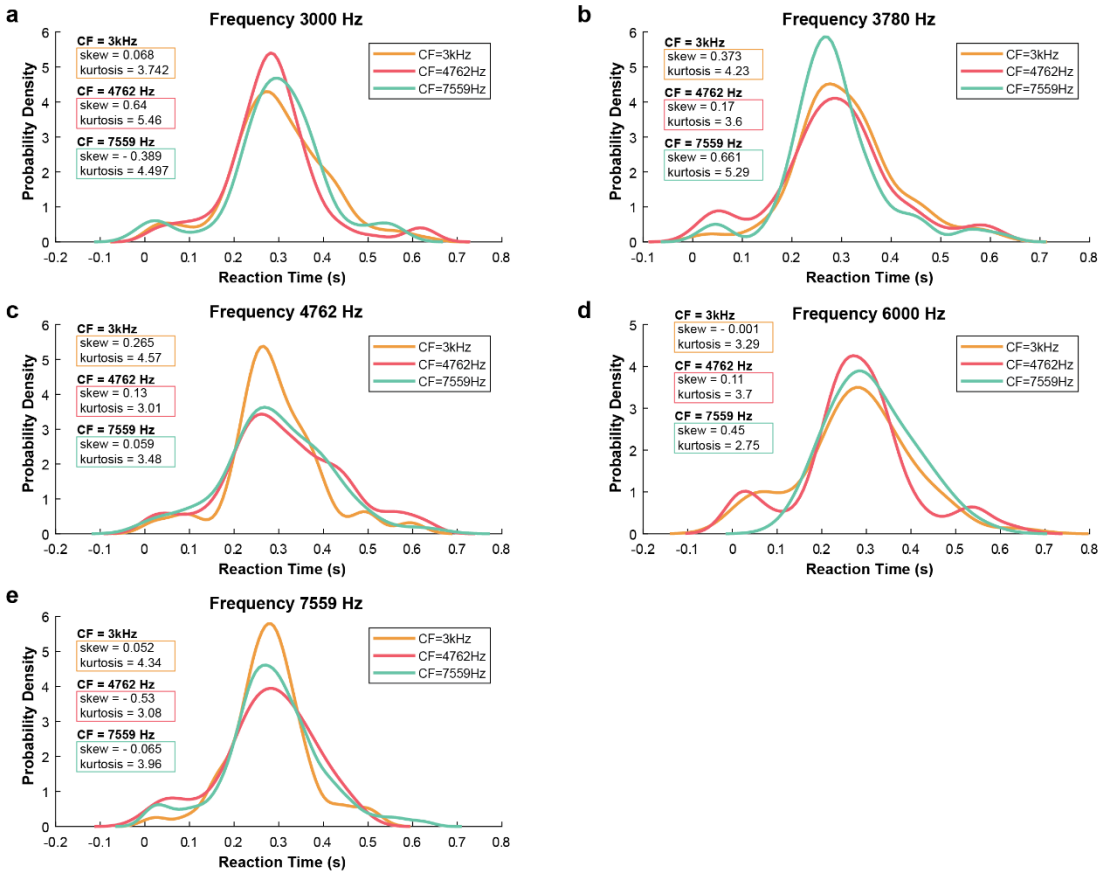

Figure 51 Kernel Density Plots, Skewness, and Kurtosis for each Signal Frequency (CF). Each panel **(a–e)** shows the kernel density estimate of reaction time distributions for a given frequency and condition (low CF, centred CF, high CF). Signal frequencies are shown in distinct colours relative to probe frequencies to visualise distributional shifts. Density plots provide an intuitive view of differences in symmetry and dispersion across stimuli. Colours indicate different conditions (low CF, centred CF, high CF). Panels a-e show distribution of reaction times for a given frequency by condition.

Skewness

Skewness values revealed differences in reaction time distributions between signal and probe frequencies, with distinct patterns based on their proximity to the signal frequency (Figure 52). Near probe frequencies generally exhibited higher skewness than the signal frequency in two of the three conditions (3000 Hz Signal and 7559 Hz Signal), indicating tails towards slower reaction times and greater

variability. This suggests that near probes, while unexpected, elicit a tendency for slower reaction times compared to the signal. In contrast, in the 4762 Hz Signal condition, near probes had skewness values similar to the signal, suggesting less disruption in reaction times and more consistent performance.

Far probe frequencies generally showed more extreme skewness values, with either stronger tails towards slower responses (positive skew) or faster responses (negative skew). For example, far probes showed high positive skew in 4762 Hz Signal (3000 Hz, 0.64) and 7559 Hz Signal (3780 Hz, 0.661), indicating a greater likelihood of slower responses and more variability. Conversely, far probes with negative skewness, such as 7559 Hz in 4762 Hz Signal (-0.53) and 3kHz in 7559 Hz Signal (-0.389), reflect faster reaction times overall, with a tendence towards very fast responses. An exception was for 3000 Hz Signal, where far probes had skewness values close to zero, reflecting symmetric reaction time distributions and low variability.

These results indicate that while the reaction time data do not align with typical predictions of the probe signal effect (i.e. faster reaction times for expected signals), they do reveal notable differences between signal and probe frequencies. Specifically, the data highlight the distinct impacts of near versus far probes on reaction time distributions. Near probes appear to evoke slower and more variable responses compared to signals, while far probes show a wider range of effects, with some eliciting slower responses and others faster responses. These findings underscore the importance of stimulus predictability and frequency proximity in shaping reaction time patterns.

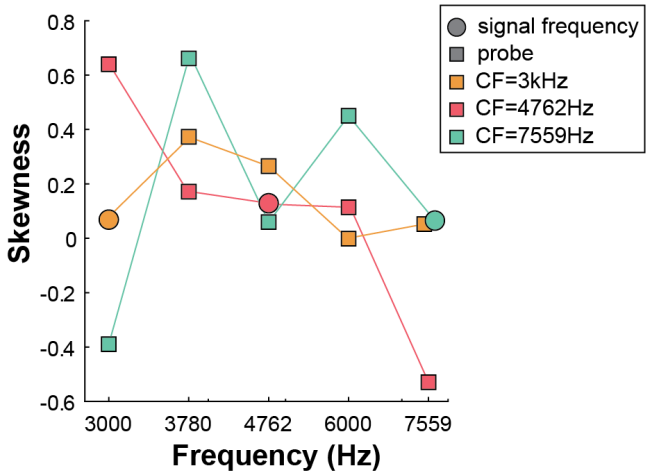

Figure 52. Summary of Skewness Values per Signal Condition. Summary of skewness values for signal and probe frequencies across conditions (low signal, centred signal, high signal). Circular markers indicate signal frequencies, while square markers represent probe frequencies. Higher skewness values reflect greater tail asymmetry towards slower responses. Note: Abbreviation CF stands for signal frequency in this case.

Kurtosis

Kurtosis values for the signal frequencies were consistently lower than those for probe frequencies across all conditions, reflecting more stable reaction times for expected stimuli (Figure 53). In the 4762 Hz Signal condition, for example, the signal frequency (4762 Hz) had the lowest kurtosis value (3.01) compared to the probes in that condition.

For the 3000 Hz Signal condition, kurtosis was higher for the probe frequencies at 3780 Hz and 4762 Hz, indicating increased variability in reaction times to these unexpected stimuli near the CF. However, at 6kHz, kurtosis decreased to its lowest value (3.29), reflecting more consistent responses, before increasing again at 7559 (4.34). A similar trend was observed in the 4762 Hz Signal condition, where the signal frequency (4762 Hz) exhibited the lowest kurtosis (3.01). The probe at 3kHz had the highest kurtosis (5.46), indicating substantial variability in reaction times, whereas the adjacent probe frequencies,

3780 Hz and 6000 Hz, had intermediate values (3.6 and 3.7 respectively). Interestingly, the probe at 7559 Hz exhibited a kurtosis similar to the signal frequency (3.08), suggesting relatively stable performance for this frequency despite it being a probe.

In the 7559 Hz Signal condition, kurtosis was lowest for the adjacent probe at 6kHz (2.75), indicating highly consistent reaction times for this frequency. In contrast, kurtosis was highest for the probes at 3780 Hz (5.29) and 3kHz (4.497), reflecting greater variability in responses. The signal frequency (7559 Hz) exhibited a moderately low kurtosis (3.96), consistent with stable reaction times for expected stimuli.

Across conditions, probe frequencies closer to the signal frequency tended to exhibit lower kurtosis values, reflecting greater consistency in reaction times. In contrast, probe frequencies further from the signal, such as 3kHz in the 4762 Hz Signal condition, and 3780 Hz in the 7559 Hz Signal condition, showed higher kurtosis values, indicating greater variability in reaction times for these unexpected stimuli.

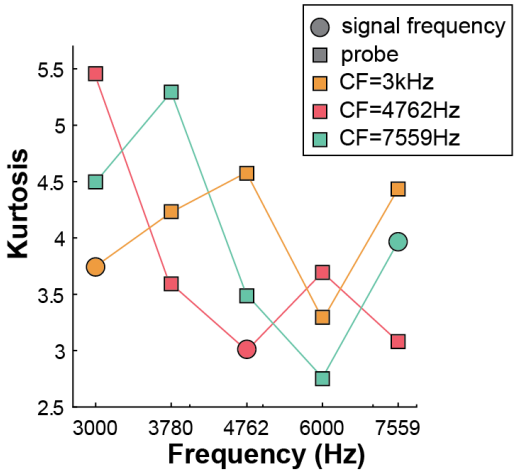

Figure 53. Summary of Kurtosis Values for Reaction Time Across Conditions. Summary of kurtosis values for reaction time distributions across signal and probe frequencies in each condition. Lower kurtosis values indicate more stable and consistent reaction times; higher values reflect greater dispersion and variability.

7.1.2 By Session Ferret Plots

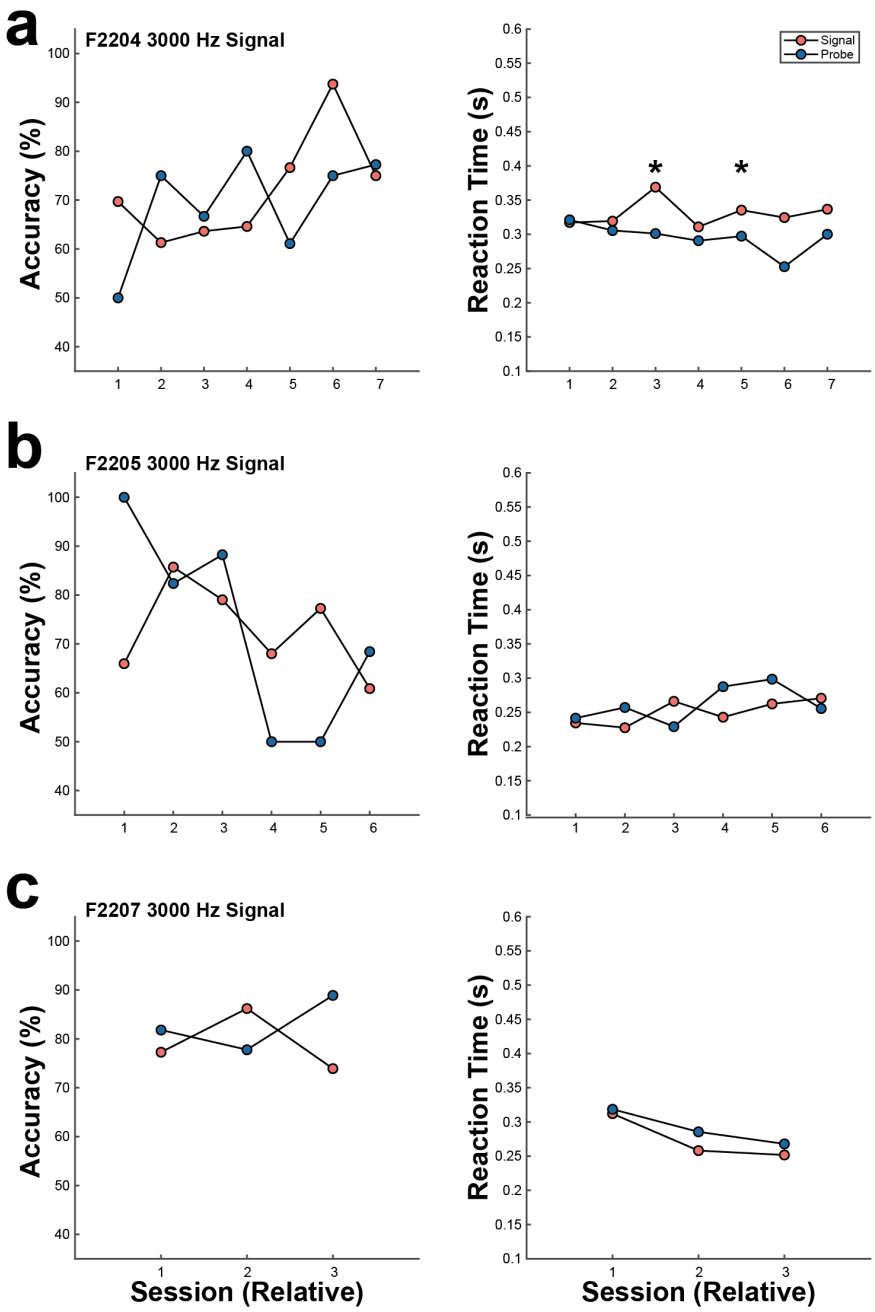

Figure 54. Accuracy and reaction time across early sessions in the 3000 Hz signal condition. Accuracy (left) and reaction time (right) across early sessions for the 3000 Hz signal condition. Each subplot shows data from one ferret (F2204, F2205, F2207). Sessions are plotted relative to the first included session for that animal and condition. Accuracy is shown as the percentage of correct responses per session for signal (red) and probe (blue) tones. Reaction times are shown for correct trials only. Asterisks indicate sessions where signal and probe responses differed significantly (Welch's *t*-test, $p < 0.05$). Data are based on up to 9 valid sessions per animal (defined as sessions with ≥ 20 correct trials). Note. The dataset contains only three valid sessions for Ferret F2207 in this condition; this reflects the available data and is not a processing error.

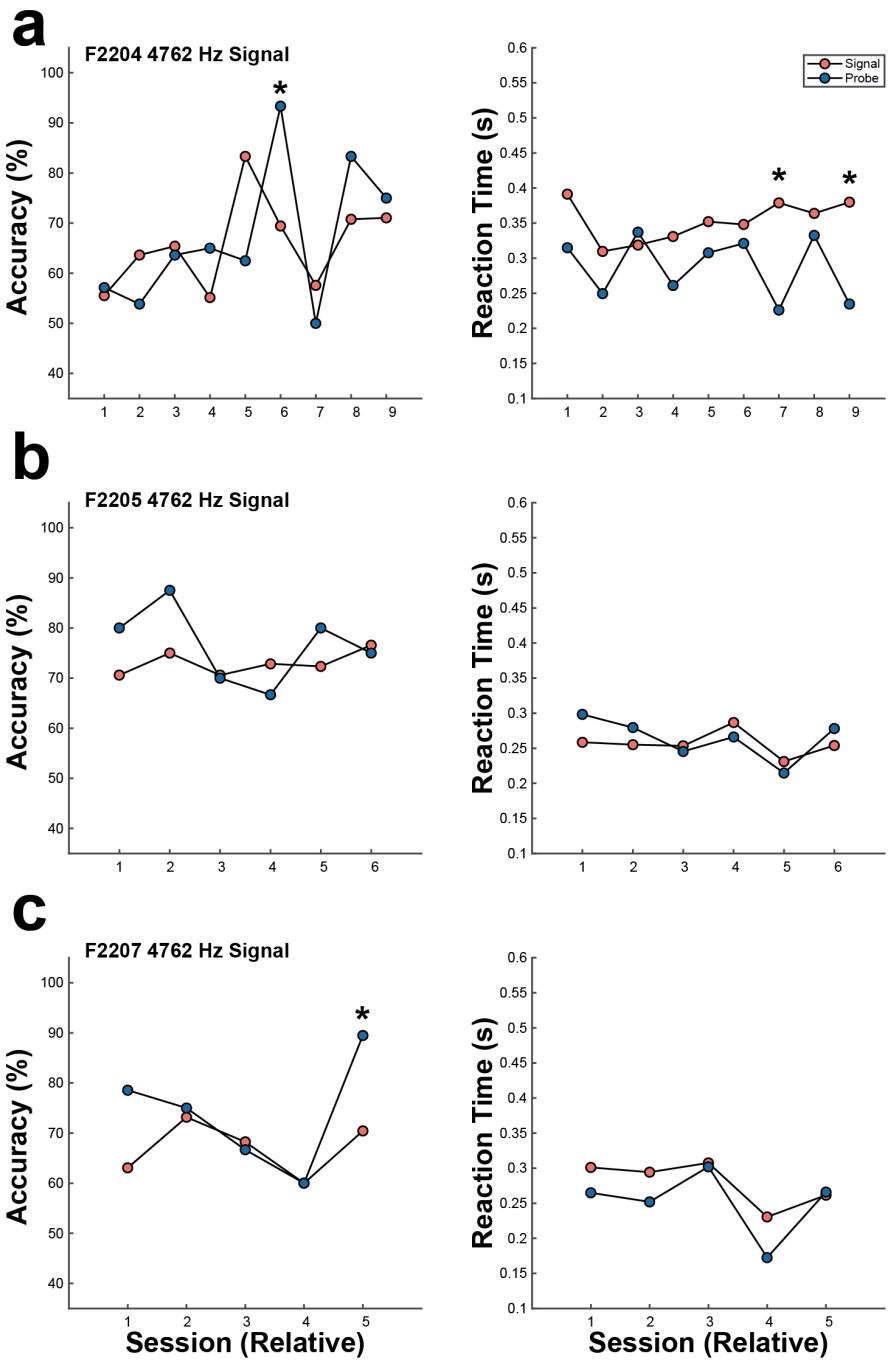

Figure 55. Accuracy and reaction time across early sessions in the 4762 Hz signal condition. Accuracy (left) and reaction time (right) across early sessions for the 4762 Hz signal condition. Each subplot shows data from one ferret (F2204, F2205, F2207). Sessions are plotted relative to the first included session for that animal and condition. Accuracy is shown as the percentage of correct responses per session for signal (red) and probe (blue) tones. Reaction times are shown for correct trials only. Asterisks indicate sessions where signal and probe responses differed significantly (Welch's *t*-test, $p < 0.05$). Data are based on up to 9 valid sessions per animal (defined as sessions with ≥ 20 correct trials).

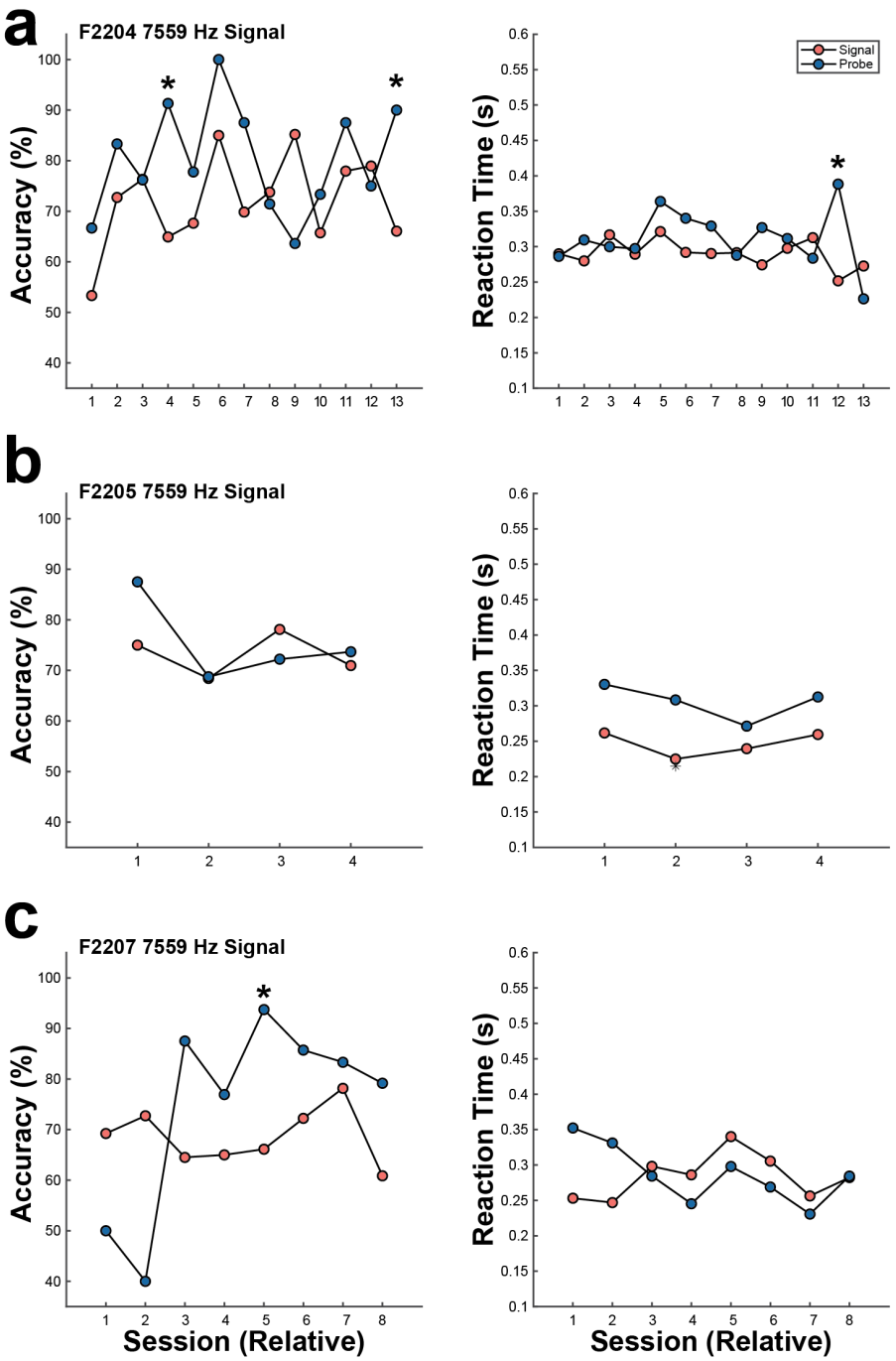

Figure 56. Accuracy and reaction time across early sessions in the 4762 Hz signal condition. Accuracy (left) and reaction time (right) across early sessions for the 7559 Hz signal condition. Each subplot shows data from one ferret (F2204, F2205, F2207). Sessions are plotted relative to the first included session for that animal and condition. Accuracy is shown as the percentage of correct responses per session for signal (red) and probe (blue) tones. Reaction times are shown for correct trials only. Asterisks indicate sessions where signal and probe responses differed significantly (Welch's *t*-test, $p < 0.05$). Data are based on up to 9 valid sessions per animal (defined as sessions with ≥ 20 correct trials).

7.1.3 Posterior Parameter Estimates for Accuracy and Reaction Time Models

Table 11. Posterior Parameter Estimates for 3000 Hz Signal Condition Accuracy Model.

Parameter	Mean	SD	HDI 3%	HDI 97%	ESS Bulk	ESS Tail	R̂
Intercept	1.74	0.49	0.70	2.86	353	125	1.01
Probe Effect	−0.38	0.28	−0.91	0.16	1515	2053	1.00
Bin 1.3 s*	−0.55	0.17	−0.89	−0.22	1911	2607	1.00
Bin 1.95 s*	−1.88	0.16	−2.21	−1.56	2201	2505	1.00
Group × Bin 1.3 s	0.27	0.37	−0.47	1.00	1794	2062	1.00
Group × Bin 1.95 s	0.45	0.36	−0.27	1.16	1761	2214	1.00

Table 12. Posterior Parameter Estimates for 7559 Hz Signal Condition Accuracy Model.

Parameter	Mean	SD	HDI 3%	HDI 97%	ESS Bulk	ESS Tail	R̂
Intercept	1.02	0.47	0.01	2.00	972	1348	1.00
Probe Effect*	0.78	0.24	0.31	1.26	1854	2278	1.00
Bin 1.3 s	−0.06	0.13	−0.32	0.20	2789	2731	1.00
Bin 1.95 s*	−1.10	0.13	−1.35	−0.84	2558	2159	1.00
Group × Bin 1.3 s*	−0.65	0.33	−1.30	−0.01	2082	2620	1.00
Group × Bin 1.95 s	−0.58	0.31	−1.17	0.04	1942	2273	1.00

*Asterisks indicate parameters for which the 95% highest density interval (HDI) excludes zero.

Table 13. Posterior Parameter Estimates for 3000 Hz Signal Condition RT Model.

Parameter	Mean	SD	HDI 3%	HDI 97%	ESS Bulk	ESS Tail	R̂
Intercept	0.30	0.07	0.16	0.45	1334	1289	1.00
Probe Effect*	0.11	0.05	0.00	0.22	1811	2292	1.00
Bin 1.3 s*	-0.22	0.03	-0.28	-0.15	2459	2708	1.00
Bin 1.95 s*	-0.23	0.03	-0.30	-0.17	2315	2516	1.00
Group × Bin 1.3 s	-0.07	0.07	-0.21	0.08	2318	2384	1.00
Group × Bin 1.95 s	-0.10	0.08	-0.25	0.05	2027	2631	1.00

Table 14. Posterior Parameter Estimates for 7559 Hz Signal Condition RT Model.

Parameter	Mean	SD	HDI 3%	HDI 97%	ESS Bulk	ESS Tail	R̂
Intercept	0.33	0.12	0.11	0.58	638	275	1.01
Probe Effect*	-0.10	0.04	-0.18	-0.01	1946	2356	1.00
Bin 1.3 s*	-0.15	0.03	-0.21	-0.10	2380	2378	1.00
Bin 1.95 s*	-0.21	0.03	-0.26	-0.15	2437	2347	1.00
Group × Bin 1.3 s	0.07	0.06	-0.05	0.20	1737	2491	1.00
Group × Bin 1.95 s	0.12	0.06	-0.00	0.25	2047	2409	1.00

*Asterisks indicate parameters for which the 95% highest density interval (HDI) excludes zero.

7.2 APPENDIX 2 – DURATION DISCRIMINATION

7.2.1 Distributional reaction time comparisons using KS tests

To assess whether reaction time distributions differed between signal and probe tones beyond mean effects, Kolmogorov-Smirnov (KS) tests were conducted separately for each signal frequency, animal, and tone duration (short, long) in Task Variation 3. Table 11 summarises the results, reporting the D-statistic (indicating the maximum distance between cumulative distributions) and associated p-values. All comparisons were significant at $p < 0.0001$, indicating reliable distributional differences between signal and probe trials across conditions.

Table 15. Kolmogorov-Smirnov (KS) Test Results for Reaction Time Distributions. Summary of KS test results comparing reaction time distributions between signal and probe tones across signal frequencies (3000 Hz, 4762 Hz, 7559 Hz), ferrets (F2105 and F2107), and tone durations (short, long). The D-statistic reflects the maximum difference between cumulative distribution functions. All comparisons were significant at $p < 0.0001$.

Signal Frequency (Hz)	Ferret	Duration	D-statistic	p-value
3000	F2105	Short	0.652	< 0.0001
3000	F2107	Short	0.381	< 0.0001
4762	F2105	Short	0.667	< 0.0001
4762	F2107	Short	0.347	< 0.0001
7559	F2105	Short	0.681	< 0.0001
7559	F2107	Short	0.39	< 0.0001
3000	F2105	Long	0.652	< 0.0001
3000	F2107	Long	0.381	< 0.0001
4762	F2105	Long	0.667	< 0.0001
4762	F2107	Long	0.347	< 0.0001
7559	F2105	Long	0.681	< 0.0001
7559	F2107	Long	0.39	< 0.0001

Histograms and cumulative distribution functions (CDFs) were generated to visualise these distributional differences (Figure 58Figure 59).

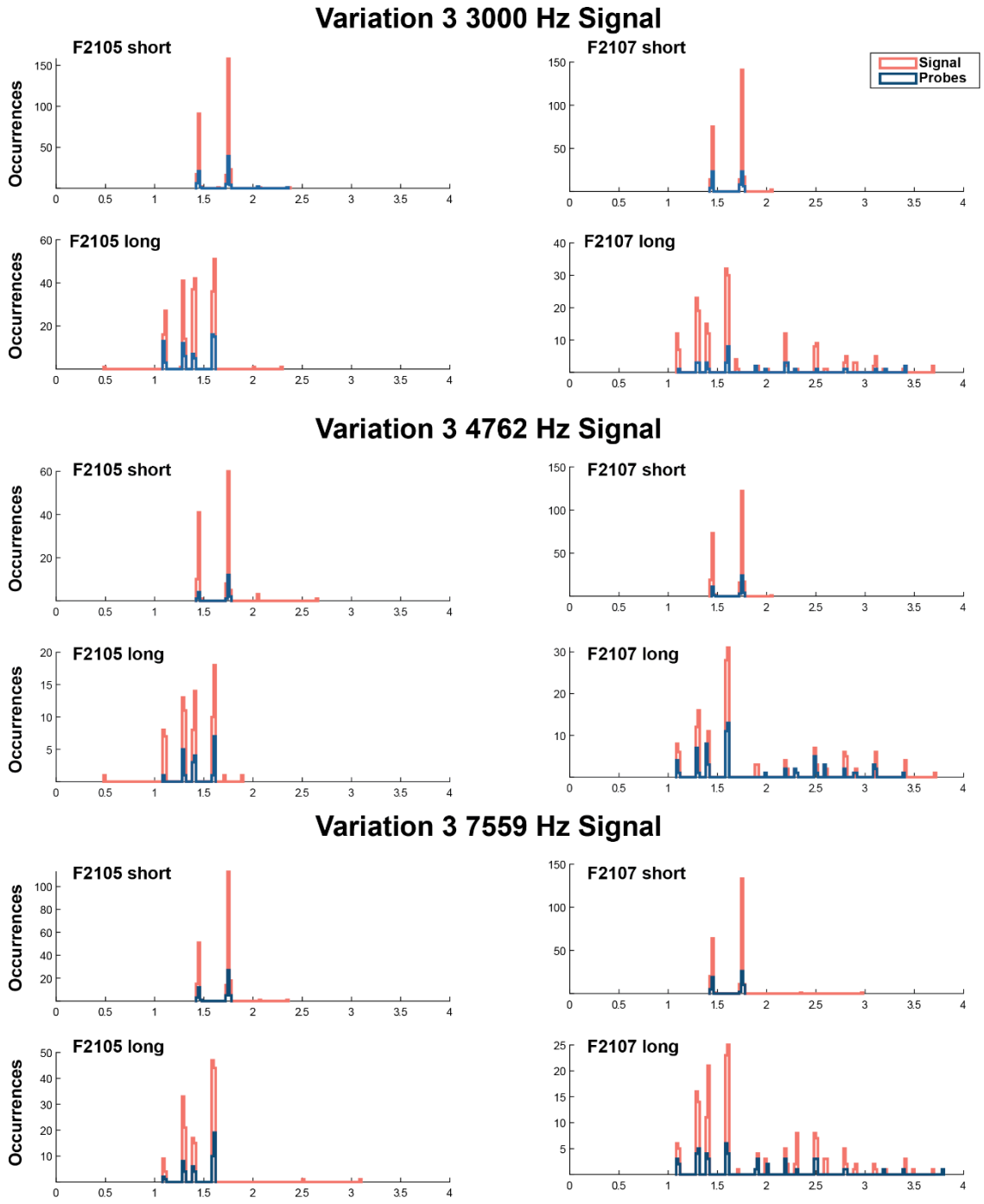

Figure 57. Reaction Time Histograms Across Variation 3 Conditions. Distribution of reaction times for signal (salmon) and probe (blue) trials, plotted separately by tone duration (short vs. long) and animal (F2105 and F2107). Rows correspond to signal frequency conditions (3000 Hz, 4762 Hz, 7559 Hz), and columns separate duration and animal. Each histogram shows the number of occurrences (y-axis) at a given reaction time bin (x-axis, in seconds). Across all conditions, signal tones generally elicited faster responses than probe tones, particularly during short-duration trials.

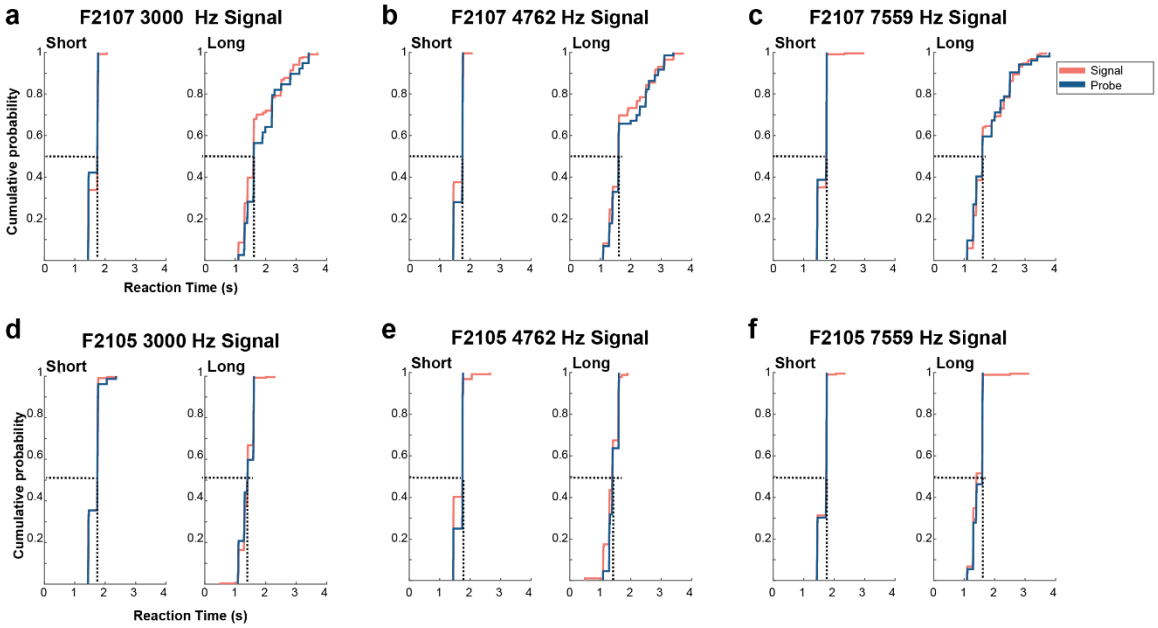

Figure 58. Cumulative Distribution Functions of Reaction Times for Task Variation 3. Cumulative distribution functions (CDFs) comparing reaction times for probe (blue) and signal (salmon) trials across common frequencies and stimulus durations. Panels represent individual ferrets (F2105 and F2107). Vertical dashed lines indicate median reaction times.

7.2.2 Frequency Models

7.2.2.1 Accuracy

Mixed-effects logistic regression models were used to predict correct responses based on tone frequency and duration.

7.2.2.1.1 Task Variation 1

Table 16. Fixed Effects Coefficients for the GLME Task: Variation 1. Summary of modelled fixed effects predicting performance accuracy. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.466	0.186	2.504	0.012*	0.101	0.832
1890 Hz	-0.288	0.217	-1.324	0.186	-0.714	0.138
2381 Hz	-0.191	0.216	-0.884	0.377	-0.613	0.232
3780 Hz	-0.331	0.207	-1.598	0.110	-0.737	0.075
4762 Hz	-0.104	0.203	-0.512	0.609	-0.502	0.294
Duration	1.710	0.193	8.842	<0.0001*	1.331	2.089
1890 * Duration	1.483	0.825	1.798	0.072	-0.134	3.100
2381 * Duration	1.207	0.800	1.509	0.131	-0.361	2.775
3780 * Duration	1.235	0.792	1.560	0.119	-0.317	2.787
4762 * Duration	0.056	0.797	0.071	0.944	-1.507	1.619

Table 17. Random Effects Coefficients for GLME Task: Variation 1. Variance estimates for random intercepts grouped by *SessionID, FerretID,* and *TimeOfDay*.

Random effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
SessionID	0.0124985	0.11180	0.0000000	0.2077249
FerretID	0.0931179	0.30515	0.1532352	0.9177002
TimeOfDay	0.0009629	0.03103	0.0000000	0.3638636

7.2.2.1.2 Task Variation 2

Fixed and random effects for Task Variation 2 are summarised below. Confidence intervals for random effects could not be estimated reliably due to low variance estimates.

Table 18. Fixed Effects Coefficients: Task Variation 2. Asterisks indicate significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.750	0.082	9.164	< 0.001*	0.590	0.911
3150 Hz	-0.834	0.151	-5.515	< 0.001*	-1.130	-0.537
3969 Hz	-0.495	0.153	-3.243	0.001*	-0.794	-0.196
6300 Hz	-0.899	0.149	-6.020	< 0.001*	-1.191	-0.606
7937 Hz	-0.125	0.150	-0.830	0.407	-0.420	0.170
Duration	-0.279	0.134	-2.088	0.037*	-0.541	-0.017
3150 * Duration	4.532	0.613	7.394	< 0.001*	3.331	5.733
3969 * Duration	3.243	0.594	5.460	< 0.001*	2.079	4.408
6300 * Duration	4.565	0.599	7.621	< 0.001*	3.391	5.739
7937 * Duration	0.497	0.549	0.905	0.366	-0.579	1.573

Table 19. Random Effects for Task: Variation 2. Note: Confidence intervals could not be computed for the random effects, resulting in NaN values, likely due to near-zero variance estimates or boundary constraints in the estimation process.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	Intercept	0.00921	0.096	NaN	NaN
SessionID	Intercept	0.01506	0.123	NaN	NaN
TimeOfDay	Intercept	6.97E-11	8.35E-06	NaN	NaN

7.2.2.1.3 Task Variation 3

7.2.2.1.3.1 Signal Frequency 3000 Hz

Table 20. Fixed Effects Coefficients for the GLME: 3000 Hz Signal. Summary of modelled fixed effects predicting performance accuracy in the 3000 Hz signal condition. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.753	0.109	6.892	<0.001*	0.539	0.967
3780 Hz	-0.481	0.296	-1.629	0.104	-1.061	0.098
4762 Hz	0.188	0.382	0.492	0.623	-0.561	0.937
6000 Hz	-0.742	0.262	-2.831	0.0047*	-1.256	-0.228
7559 Hz	0.415	0.354	1.172	0.241	-0.279	1.108
Duration	-0.386	0.116	-3.337	0.0009*	-0.613	-0.159
3780 Hz * Duration	0.863	0.477	1.81	0.07	-0.072	1.798
4762 Hz * Duration	0.241	0.523	0.461	0.645	-0.785	1.267
6000 Hz * Duration	1.307	0.449	2.914	0.0036*	0.427	2.187
7559 Hz * Duration	0.107	0.522	0.205	0.838	-0.917	1.131

Table 21. Random Effects Coefficients Task Variation 3, 3000 Hz Signal. Summary of variance estimates for random intercepts grouped by *FerretID, SessionID,* and *TimeOfDay* for the 3000 Hz signal condition. Note: Confidence intervals could not be computed for the random effects, resulting in NaN values, likely due to near-zero variance estimates or boundary constraints in the estimation process.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	Intercept	1.607×10^{-10}	0.000013	NaN	NaN
SessionID	Intercept	0.02106	0.14509	NaN	NaN
TimeOfDay	Intercept	0.00726	0.08525	NaN	NaN

7.2.2.1.3.2 *Signal Frequency 4762 Hz*

Table 22. Fixed Effects Coefficients Task Variation 3, 4762 Hz Signal. Summary of modelled fixed effects predicting performance accuracy in the 4762 Hz signal condition. Asterisks denote significant effects at $p < 0.05$. Asterisks indicate significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.558	0.079	7.087	<0.001*	0.404	0.713
3000 Hz	0.514	0.394	1.303	0.193	-0.26	1.287
3780 Hz	0.063	0.34	0.185	0.853	-0.604	0.73
6000 Hz	-0.7	0.33	-2.12	0.0341*	-1.348	-0.052
7559 Hz	-0.416	0.309	-1.35	0.177	-1.021	0.189
Duration	-0.449	0.105	-4.261	<0.001*	-0.655	-0.242
3000 Hz * Duration	0.256	0.52	0.492	0.623	-0.764	1.275
3780 Hz * Duration	0.842	0.475	1.772	0.077	-0.09	1.775
6000 Hz * Duration	1.879	0.47	3.996	<0.001*	0.957	2.802
7559 Hz * Duration	1.276	0.489	2.608	0.0092*	0.316	2.235

Table 23. Random Effects Coefficients Task Variation 3, 4762 Hz Signal. Summary of variance estimates for random intercepts grouped by *FerretID, SessionID,* and *TimeOfDay* for the 4762 Hz signal condition. Note: Confidence intervals could not be computed for the random effects, resulting in NaN values, likely due to near-zero variance estimates or boundary constraints in the estimation process.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	Intercept	3.45×10^{-11}	5.8715×10^{-6}	NaN	NaN
SessionID	Intercept	0.009429	0.09708	NaN	NaN
TimeOfDay	Intercept	0.000086	0.009292	NaN	NaN

7.2.2.1.3.3 Signal Frequency 7559 Hz

Table 24. **Fixed Effects Coefficients GLME: Variation 3, 7559 Hz Signal.** Summary of modelled fixed effects predicting performance accuracy in the 7559 Hz signal condition. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.675	0.078	8.687	<0.001*	0.523	0.827
3000 Hz	0.241	0.35	0.689	0.491	-0.446	0.928
3780 Hz	0.194	0.339	0.572	0.568	-0.472	0.86
4762 Hz	-0.144	0.292	-0.494	0.621	-0.718	0.429
6000 Hz	-1.096	0.291	-3.762	<0.001*	-1.668	-0.525
Duration	-0.495	0.107	-4.614	<0.001*	-0.706	-0.285
3000 Hz * Duration	-0.035	0.476	-0.074	0.941	-0.97	0.899
3780 Hz * Duration	0.376	0.455	0.827	0.408	-0.517	1.269
4762 Hz * Duration	0.188	0.491	0.382	0.702	-0.775	1.15
6000 Hz * Duration	1.573	0.446	3.528	<0.001*	0.699	2.448

Table 25. Random Effects Coefficients GLME: Variation 3, 7559 Hz Signal. Summary of variance estimates for random intercepts grouped by FerretID, SessionID, and TimeOfDay for the 7559 Hz signal condition. Note: Confidence intervals could not be computed for the random effects, resulting in NaN values, likely due to near-zero variance estimates or boundary constraints in the estimation process.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	Intercept	3.2×10^{-8}	0.0001789	NaN	NaN
SessionID	Intercept	7.82×10^{-10}	0.0002796	NaN	NaN
TimeOfDay	Intercept	1.57×10^{-8}	0.0003969	NaN	NaN
Residual	-	1	1	-	-

7.2.2.2 Reaction Times

Linear mixed-effects (LME) models were fitted to predict reaction times across signal and probe frequencies for each task variation. In all models, fixed effects included frequency, tone duration (short or long), and trial number, with random intercepts for *FerretID, SessionID,* and *TimeOfDay*. Where applicable, interaction terms were included. Tables summarise the modelled coefficients and variance estimates.

7.2.2.2.1 Task Variation 1

Table 26. Fixed Effects Coefficients LME: Variation 1 (3000 Hz Signal). Summary of fixed effects predicting reaction time for the 3000 Hz signal condition in Task Variation 1. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
(Intercept)	1.78000	0.05600	31.7920	0.00020*	1.67060	1.89010
1890 Hz	0.03773	0.09339	0.40400	0.68619	-0.14530	0.22080
2381 Hz	0.01739	0.08435	0.20600	0.83664	-0.14790	0.18270
3780 Hz	0.08204	0.08526	0.96200	0.33598	-0.08510	0.24910
4762 Hz	0.05253	0.08387	0.62600	0.53116	-0.11180	0.21690
Duration	-0.36360	0.02004	-18.1420	<0.0001*	-0.40290	-0.32430
TrialNumber	0.00023	0.00026	0.91300	0.36113	-0.00027	0.00073
1890 Hz x Duration	-0.02089	0.08424	-0.24800	0.80412	-0.18600	0.14420
2381 Hz x Duration	0.10940	0.08094	1.35100	0.17678	-0.04930	0.26800
3780 Hz x Duration	0.03170	0.08324	0.38100	0.70336	-0.13140	0.19480
4762 Hz x Duration	0.09515	0.08334	1.14200	0.25367	-0.06820	0.25850
1890 Hz x TrialNumber	0.00047	0.00107	0.43800	0.66123	-0.00163	0.00257
2381 Hz x TrialNumber	-0.00153	0.00103	-1.48500	0.13753	-0.00356	0.00049
3780 Hz x TrialNumber	-0.00083	0.00108	-0.77100	0.44070	-0.00294	0.00128
4762 Hz x TrialNumber	-0.00141	0.00097	-1.45600	0.14541	-0.00331	0.00049

Table 27. Random Effects Coefficients LME Variation 1. Variance components for random intercepts grouped by *FerretID, SessionID,* and *TimeOfDay* in Task Variation 1.

Random Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	0.00804	0.08965	0.03658	0.21902
SessionID	0.00049	0.02223	0.00000	0.05184
TimeOfDay	0.00001	0.00332	0.00000	0.08660
Residual	0.28530	0.53413	0.52106	0.54572

7.2.2.2.2 Task Variation 2

Table 28. Fixed Effects Coefficients LME: Variation 2. Summary of fixed effects predicting reaction time across frequencies in Task Variation 2. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
(Intercept)	1.79000	0.06091	29.39500	0.00055*	1.67107	1.90983
3150 Hz	-0.02081	0.05422	-0.38400	0.70120	-0.12708	0.08547
3969 Hz	-0.03325	0.05253	-0.63300	0.52675	-0.13620	0.06970
6300 Hz	-0.00746	0.05445	-0.13700	0.89099	-0.11418	0.09925
7937 Hz	-0.03177	0.05140	-0.61800	0.53654	-0.13250	0.06897
Duration	-0.29480	0.01245	-23.67900	<0.0001*	-0.31920	-0.27040
TrialNumber	-0.00071	0.00016	-4.46100	0.00001*	-0.00101	-0.00040
3150 Hz x Duration	0.04091	0.05204	0.78600	0.43181	-0.06108	0.14291
3969 Hz x Duration	0.01444	0.05022	0.28800	0.77365	-0.08399	0.11287
6300 Hz x Duration	0.07058	0.05187	1.36100	0.17365	-0.03108	0.17223
7937 Hz x Duration	-0.05538	0.05105	-1.08500	0.27801	-0.15542	0.04467
3150 Hz x TrialNumber	0.00045	0.00059	0.76900	0.44207	-0.00070	0.00161
3969 Hz x TrialNumber	0.00095	0.00062	1.54600	0.12225	-0.00026	0.00216
6300 Hz x TrialNumber	-0.00050	0.00063	-0.79700	0.42559	-0.00174	0.00073
7937 Hz x TrialNumber	0.00061	0.00064	0.96500	0.33475	-0.00063	0.00186

Table 29. Random Effects Coefficients LME: Variation 2. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* in Task Variation 2.

Random Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	0.01032	0.10158	0.04331	0.24781
SessionID	0.00253	0.05030	0.03486	0.06894
TimeOfDay	0.00014	0.01184	0.00000	0.10109
Residual	0.21323	0.46177	0.45374	0.46909

7.2.2.2.3 Task Variation 3

7.2.2.2.3.1 Signal Frequency 3000 Hz

Table 30. Fixed Effects Coefficients LME: Variation 3, 3000 Hz Signal. Summary of fixed effects predicting reaction time for the 3000 Hz signal condition in Task Variation 3. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
(Intercept)	3.004	0.853	3.522	0.0120*	1.33	4.68
3780 Hz	2.695	3.183	0.847	0.3973	-3.54	8.93
4762 Hz	-1.332	3.083	-0.432	0.6658	-7.37	4.71
6000 Hz	-1.531	3.135	-0.488	0.6255	-7.68	4.61
7559 Hz	-1.363	3.239	-0.421	0.6740	-7.71	4.99
Duration	1.368	0.772	1.773	0.0766	-0.14	2.88
TrialNumber	-0.0207	0.0087	-2.391	0.0176*	-0.0377	-0.0037
3780 Hz x Duration	2.691	3.192	0.843	0.3994	-3.57	8.95
4762 Hz x Duration	-1.444	3.170	-0.455	0.6489	-7.66	4.77
6000 Hz x Duration	-1.117	2.958	-0.378	0.7059	-6.91	4.68
7559 Hz x Duration	-1.337	3.018	-0.443	0.6579	-7.25	4.58
3780 Hz x TrialNumber	-0.0348	0.0327	-1.064	0.2875	-0.099	0.0293
4762 Hz x TrialNumber	0.0193	0.0312	0.617	0.5373	-0.0419	0.0805
6000 Hz x TrialNumber	0.0217	0.0329	0.659	0.5103	-0.0428	0.0861
7559 Hz x TrialNumber	0.0196	0.0335	0.585	0.5584	-0.0460	0.0852

Table 31. Random Effects Coefficients LME: Variation 3, 3000 Hz Signal. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for the 3000 Hz signal condition.

Random Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	0.0990	0.3147	0.0000	1.3327
TimeOfDay	0.1481	0.3848	0.0000	1.7743
SessionID	0.0000	0.0000	0.0000	1.6725
Residual	125.0870	11.1842	10.6587	11.5960

7.2.2.2.3.2 Signal Frequency 4762 Hz

Table 32. Fixed Effects Coefficients LME: Variation 3, 4762 Hz Signal. Summary of fixed effects predicting reaction time for the 4762 Hz signal condition in Task Variation 3. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
(Intercept)	3.097	0.529	5.854	p<.00001***	2.06	4.13
3000 Hz	-1.528	2.226	-0.687	0.4925	-5.89	2.83
3780 Hz	9.298	2.194	4.238	p<.00001***	5.00	13.60
6000 Hz	2.085	2.291	0.910	0.3629	-2.41	6.58
7559 Hz	0.179	2.417	0.074	0.9411	-4.56	4.92
Duration	0.189	0.540	0.351	0.7259	-0.87	1.25
TrialNumber	-0.0163	0.0061	-2.656	0.0080**	-0.0283	-0.0043
3000 Hz x Duration	0.139	2.202	0.063	0.9496	-4.18	4.45
3780 Hz x Duration	-2.215	2.112	-1.049	0.2945	-6.35	1.92
6000 Hz x Duration	4.584	2.343	1.957	0.0506.	-0.0071	9.18
7559 Hz x Duration	1.348	2.319	0.581	0.5613	-3.20	5.89
3000 Hz x TrialNumber	0.0176	0.0208	0.848	0.3969	-0.0232	0.0584
3780 Hz x TrialNumber	-0.0902	0.0252	-3.584	0.0004***	-0.1396	-0.0409
6000 Hz x TrialNumber	-0.0491	0.0232	-2.117	0.0345*	-0.0945	-0.0036
7559 Hz x TrialNumber	-0.0039	0.0223	-0.175	0.8614	-0.0475	0.0397

Table 33. Random Effects Coefficients LME: Variation 3, 4762 Hz Signal. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for the 4762 Hz signal condition.

Random Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	0.0000	0.0000	0.0000	0.8897
TimeOfDay	0.0000	0.0000	0.0000	1.0420
SessionID	0.0000	0.0000	0.0000	0.9837
Residual	62.9200	7.9320	7.5605	8.2184

7.2.2.2.3.3 Signal Frequency 7559 Hz

Table 34. Fixed Effects Coefficients LME: Variation 3, 7559 Hz Signal. Summary of fixed effects predicting reaction time for the 7559 Hz signal condition in Task Variation 3. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
(Intercept)	3.865	0.658	5.875	0.0002***	2.58	5.15
3000 Hz	-2.177	2.849	-0.764	0.4451	-7.76	3.41
3780 Hz	-2.187	2.522	-0.867	0.3861	-7.13	2.76
4762 Hz	-2.259	2.682	-0.842	0.3998	-7.52	3.00
6000 Hz	-0.373	3.290	-0.113	0.9098	-6.82	6.08
Duration	1.572	0.606	2.596	0.0096**	0.38	2.76
TrialNumber	-0.0324	0.0072	-4.470	p<.00001***	-0.0466	-0.0182
3000 Hz x Duration	-1.512	2.495	-0.606	0.5447	-6.40	3.38
3780 Hz x Duration	-1.629	2.273	-0.717	0.4737	-6.08	2.83
4762 Hz x Duration	-1.736	2.912	-0.596	0.5512	-7.44	3.97
6000 Hz x Duration	-0.823	2.770	-0.297	0.7664	-6.25	4.61
3000 Hz x TrialNumber	0.0318	0.0307	1.036	0.3005	-0.0283	0.0918
3780 Hz x TrialNumber	0.0310	0.0314	0.988	0.3234	-0.0305	0.0925
4762 Hz x TrialNumber	0.0317	0.0315	1.006	0.3147	-0.0301	0.0934
6000 Hz x TrialNumber	0.0063	0.0356	0.178	0.8586	-0.0634	0.0761

Table 35. Random Effects Coefficients LME: Variation 3, 7559 Hz Signal. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for the 7559 Hz signal condition.

Random Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	0.0000	0.0000	0.0000	0.7576
TimeOfDay	0.0779	0.2792	0.0000	1.4248
SessionID	0.0000	0.0000	0.0000	0.9577
Residual	79.8958	8.9384	8.5228	9.2620

7.2.3 Group-Level Models

Generalised linear mixed-effects (GLME) models and linear mixed-effects (LME) models were used to predict performance and reaction times across all animals at the group level. For all analyses, tone duration (short = 0.05 s) and stimulus type (signal vs. probe) were treated as fixed effects, along with trial number where applicable. Random intercepts were included for *FerretID, SessionID,* and *TimeOfDay*. Post hoc comparisons were conducted to interpret significant interactions.

7.2.3.1 Percent Correct

7.2.3.1.1 Task Variation 1

Table 36. Fixed Effects Coefficients: GLME Variation 1. Summary of fixed effects predicting percent correct in Task Variation 1. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.466	0.186	2.507	0.012*	0.102	0.831
Duration	1.710	0.193	8.842	<0.0001*	1.331	2.089
StimType	-0.228	0.114	-1.994	0.046*	-0.453	-0.004
Duration * StimType	1.104	0.434	2.335	0.020*	0.163	1.866

Table 37. Random Effects Coefficients GLME: Variation 1. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for percent correct in Task Variation 1.

Random effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
SessionID	0.0123551	0.11115	0.0000000	0.2070182
FerretID	0.0928492	0.30471	0.1530207	0.9164239
TimeOfDay	0.0009598	0.03098	0.0000000	0.3635483

Table 38. Post Hoc Comparisons for Task Variation 1. Pairwise comparisons of stimulus type and tone duration from the GLME model for percent correct in Task Variation 1. Asterisks indicate significant effects.

Condition Type	Variable 1	Variable 2	Estimate	SE	z/t-ratio	p-value
Short Duration	Signal	Probe	0.178	0.0993	1.793	0.073
Long Duration	Signal	Probe	-0.178	0.115	-1.543	0.1229
Signal	Short Duration	Long Duration	-0.6	0.0678	-8.847	<.0001*
Probe	Short Duration	Long Duration	-0.956	0.136	-7.01	<.0001*

7.2.3.1.2 Task Variation 2

Table 39. Fixed Effects Coefficients GLME: Variation 1. Summary of fixed effects predicting percent correct in Task Variation 2. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.750	0.082	9.203	<0.001*	0.590	0.910
Duration	-0.279	0.134	-2.087	0.037*	-0.541	-0.017
StimType	-0.582	0.082	-7.093	<0.001*	-0.742	-0.421
Duration * StimType	3.137	0.312	10.049	<0.001*	2.525	3.748

Table 40. Random Effects Coefficients GLME: Variation 2. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* in Task Variation 2. Note. NaN values for confidence intervals indicate insufficient data or variability constraints preventing reliable estimation.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	Intercept	0.01487	0.12194	NaN	NaN
SessionID	Intercept	0.00978	0.09886	NaN	NaN
TimeOfDay	Intercept	5.42×10^{-11}	7.37×10^{-6}	NaN	NaN
Residual	-	1	1	-	-

Table 41. Post Hoc Comparisons for Task Variation 2. Pairwise comparisons for stimulus type and tone duration from the GLME model for percent correct in Task Variation 2. Asterisks indicate significant effects.

Condition Type	Variable 1	Variable 2	Estimate	SE	z/t-ratio	p-value
Short Duration	Signal	Probe	0.426	0.0711	5.991	<.0001*
Long Duration	Signal	Probe	-0.674	0.083	-8.115	<.0001*
Signal	Short Duration	Long Duration	0.0978	0.0468	2.089	0.0367*
Probe	Short Duration	Long Duration	-1.0	0.0988	-10.143	<.0001*

7.2.3.1.3 Task Variation 3

7.2.3.1.3.1 Signal Frequency 3000 Hz

Table 42. Fixed Effects Coefficients GLME: Variation 3, 3000 Hz Signal. Summary of fixed effects predicting percent correct for the 3000 Hz signal condition in Task Variation 3. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.878	0.137	6.404	<0.001*	0.609	1.146
Duration	-2.565	0.771	-3.327	0.001*	-4.077	-1.053
StimType	-0.508	0.234	-2.174	0.03*	-0.966	-0.05
Duration * StimType	4.912	1.729	2.841	0.005*	1.521	8.304

Table 43. Random Effects Coefficients GLME: Variation 3, 3000 Hz Signal. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for percent correct in the 3000 Hz signal condition. Note. NaN values for confidence intervals indicate insufficient data or variability constraints preventing reliable estimation.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	Intercept	2.8643×10^{-11}	5.35×10−6	NaN	NaN
SessionID	Intercept	0.02317	0.15223	NaN	NaN
TimeOfDay	Intercept	0.00863	0.0929	NaN	NaN
Residual	-	1	1	-	-

Table 44. Post Hoc Comparisons for Task Variation 3 3000 Hz Signal. Pairwise comparisons of stimulus type and tone duration for percent correct in the 3000 Hz signal condition. Asterisks denote significant effects.

Condition Type	StimType 1	Variable 2	Estimate	SE	z/t-ratio	p-value
Short Duration	Signal	Probe	0.263	0.169	1.562	0.1183
Long Duration	Signal	Probe	-0.476	0.198	-2.409	0.016*
Signal	Short Duration	Long Duration	0.386	0.116	3.326	0.0009*
Probe	Short Duration	Long Duration	-0.353	0.233	-1.519	0.1288

7.2.3.1.3.2 Signal Frequency 4762 Hz

Table 45. Fixed Effects Coefficients GLME: Variation 3, 4762 Hz Signal. Summary of fixed effects predicting percent correct for the 4762 Hz signal condition in Task Variation 3. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.708	0.106	6.677	<0.001*	0.5	0.916
Duration	-2.991	0.702	-4.26	<0.001*	-4.368	-1.614
StimType	-0.555	0.245	-2.26	0.024*	-1.036	-0.073
Duration * StimType	7.473	1.705	4.384	<0.001*	4.129	10.816

Table 46. Random Effects Coefficients GLME: Variation 3, 4762 Hz Signal. Variance components for random intercepts grouped by *FerretID, SessionID,* and *TimeOfDay* in the 4762 Hz signal condition. Note. NaN values for confidence intervals indicate insufficient data or variability constraints preventing reliable estimation.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	Intercept	2.0898×10^{-10}	1.4453×10^{-5}	NaN	NaN
SessionID	Intercept	0.00964	0.09817	NaN	NaN
TimeOfDay	Intercept	1.89×10^{-10}	1.3748×10^{-5}	NaN	NaN
Residual	-	1	1	-	-

Table 47. Post Hoc Comparisons for Task Variation 3: 4762 Hz Signal. Pairwise comparisons of stimulus type and tone duration for percent correct in the 4762 Hz signal condition. Asterisks denote significant effects.

Condition Type	StimulusType 1	StimulusType 2	Estimate	SE	z/t-ratio	p-value
Short Duration	Signal	Probe	0.181	0.178	1.016	0.3097
Long Duration	Signal	Probe	-0.942	0.184	-5.128	<.0001*
Signal	Short Duration	Long Duration	0.449	0.105	4.264	<.0001*
Probe	Short Duration	Long Duration	-0.673	0.233	-2.888	0.0039*

7.2.3.1.3.3 Signal Frequency 7559 Hz

Table 48. Fixed Effects Coefficients GLME Variation 3: 7559 Hz Signal. Summary of fixed effects predicting percent correct for the 7559 Hz signal condition in Task Variation 3. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.84	0.106	7.888	<0.001*	0.631	1.049
Duration	-3.301	0.715	-4.614	<0.001*	-4.704	-1.898
StimType	-0.474	0.23	-2.064	0.039*	-0.924	-0.024
Duration * StimType	4.172	1.626	2.565	0.01*	0.982	7.361

Table 49. Random Effects Coefficients GLME Variation 3: 7559 Hz Signal. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* in the 7559 Hz signal condition. Note. NaN values for confidence intervals indicate insufficient data or variability constraints preventing reliable estimation.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
FerretID	Intercept	4.0983×10^{-9}	6.403×10^{-5}	NaN	NaN
SessionID	Intercept	2.2308×10^{-9}	4.723×10^{-5}	NaN	NaN
TimeOfDay	Intercept	4.735×10^{-9}	6.881×10^{-5}	NaN	NaN
Residual	-	1	1	-	-

Table 50. Post Hoc Comparisons for Task Variation 3: 7559 Hz Signal. Asterisks indicate significant effects at $p < 0.05$. Pairwise comparisons of stimulus type and tone duration for percent correct in the 7559 Hz signal condition. Asterisks denote significant effects.

Condition Type	Variable 1	Variable 2	Estimate	SE	z/t-ratio	p-value
Short Duration	Signal	Probe	0.265	0.166	1.595	0.1107
Long Duration	Signal	Probe	-0.361	0.178	-2.02	0.0434*
Signal	Short Duration	Long Duration	0.495	0.107	4.614	<.0001*
Probe	Short Duration	Long Duration	-0.131	0.219	-0.596	0.551

7.2.3.1.4 Task Variation 4

Table 51. Fixed Effects Table for GLME for Percent Correct Task Variation 4, Complex Inharmonic. Summary of fixed effects predicting percent correct for Task Variation 4, using a broadband, inharmonic probe. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	Estimate (β)	SE	z	p-value	95% CI (lower)	95% CI (upper)
Intercept	0.600	0.173	3.47	0.0005*	0.261	0.940
Duration	-0.527	0.170	-3.10	0.0019*	-8.5589	-1.9416
StimType	-1.508	0.301	-5.02	<0.0001*	-2.097	-0.919
Duration * StimType	2.811	0.456	6.17	<0.0001*	1.918	3.703

Table 52. Random Effects Table for Variation 4: Complex Inharmonic. Variance components for random intercepts grouped by *FerretID, SessionID, and TimeOfDay* for Task Variation 4.

Variable	Effect	Variance (σ^2)	95% CI (Lower)	95% CI (Upper)
FerretID	Intercept	1.768e-14	0	0.3911
SessionID	Intercept	0.01533	0	0.4377
TimeOfDay	Intercept	0.00506	0	0.6937

Table 53. Post Hoc Comparisons for GLME: Task Variation 4 (Complex Inharmonic Stimuli). Pairwise comparisons of stimulus type and tone duration from the GLME model for Task Variation 4. Holm-corrected p-values are reported. Asterisks indicate significant effects.

Condition Type	StimType 1	Variable 2	Estimate (log odds)	SE	z-ratio	p-value (Holm)
Short Duration	Signal	Probe	1.51	0.3	5.018	<.0001*
Long Duration	Signal	Probe	-1.3	0.342	-3.81	0.0001*
Signal	Short Duration	Long Duration	0.527	0.17	3.104	0.0019*
Probe	Short Duration	Long Duration	-2.28	0.421	-5.421	<.0001*

7.2.3.2 Reaction Time Models

7.2.3.2.1 Task Variation 1

Table 54. Fixed Effects Coefficients for LME: Task Variation 1. Summary of fixed effects predicting reaction times at the group level for Task Variation 1. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	β	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	1.810	0.060	30.10	3.08e-05 **	1.692	1.928
Duration	-0.944	0.102	-9.29	<0.0001 **	-1.144	-0.745
Trial Number	0.00061	0.00042	1.45	0.148	-0.00022	0.00144
StimType	0.079	0.068	1.16	0.248	-0.055	0.213
Duration x StimType	-0.0093	0.227	-0.041	0.967	-0.455	0.436
Trial Number x StimType	-0.00152	0.00097	-1.56	0.119	-0.00342	0.00039
Duration x Trial Number	-0.00159	0.00141	-1.13	0.260	-0.00436	0.00118
Duration x Trial Number x StimType	0.00279	0.00322	0.87	0.386	-0.00352	0.00909

Table 55. Random Effects Table of LME for Task Variation 1. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for reaction times in Task Variation 1.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
Session	Intercept	0.00055	0.023	0.000	0.0529
Ferret	Intercept	0.00810	0.090	0.0368	0.2200
Time of Day	Intercept	0.00002	0.004	0.000	0.0878
Residual	-	0.2851	0.534	0.5213	0.5450

7.2.3.2.2 Task Variation 2

Table 56. Fixed Effects Coefficients for LME: Task Variation 2. Summary of fixed effects predicting reaction times at the group level for Task Variation 2. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	β	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	1.790	0.061	29.23	0.00056*	1.670	1.910
Duration	-0.312	0.021	-14.63	<0.0001**	-0.354	-0.270
Trial Number	-0.00085	0.00022	-3.93	<0.0001**	-0.00127	-0.00042
StimType	-0.0117	0.036	-0.323	0.747	-0.083	0.059
Duration x StimType	0.00363	0.0048	0.075	0.940	-0.091	0.098
Trial Number x StimType	0.00021	0.0005	0.414	0.679	-0.00078	0.0012
Duration x Trial Number	0.00030	0.00029	0.997	0.319	-0.00029	0.00088
StimType x Duration x Trial Number	0.00029	0.00067	0.437	0.662	-0.00103	0.00162

Table 57. Random Effects Coefficients for LME: Task Variation 2. Variance components for random intercepts grouped by *FerretID, SessionID,* and *TimeOfDay* for reaction times in Task Variation 2.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
Session	Intercept	0.00255	0.050	0.0350	0.0691
Ferret	Intercept	0.01045	0.102	0.0436	0.2494
Time of Day	Intercept	0.00014	0.014	0.0000	0.1007
Residual	-	0.21324	0.462	0.454	0.4694

7.2.3.2.3 Task Variation 3

7.2.3.2.3.1 Signal Frequency 3000 Hz

Table 58. Fixed Effects Coefficients for LME: Task Variation 3, 3000 Hz Signal. Summary of fixed effects predicting reaction times for the 3000 Hz signal condition in Task Variation 3. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	β	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	1.640	0.108	15.16	0.0236 *	1.428	1.851
StimType	-0.004	0.029	-0.14	0.8918	-0.062	0.054
Trial Number	0.00037	0.00028	1.30	0.1931	-0.00019	0.00092
Duration	0.0473	0.0243	1.94	0.0523	-0.0004	0.0950

Table 59. Random Effects Coefficients for LME: Task Variation 3, 3000 Hz Signal. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for reaction times in the 3000 Hz signal condition.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
Session	Intercept	0.00048	0.0219	0.0000	0.0648
Time of Day	Intercept	0.00143	0.0379	0.0000	0.2260
Ferret	Intercept	0.0206	0.1436	0.0463	0.4405
Residual	–	0.1554	0.3942	0.3775	0.4111

7.2.3.2.3.2 Signal Frequency 4762 Hz

Table 60. Fixed Effects Coefficients for LME: Task Variation 3, 4762 Hz Signal. Summary of fixed effects predicting reaction times for the 4762 Hz signal condition in Task Variation 3. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	β	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	1.624	0.095	17.01	0.0231 *	1.437	1.811
StimType	0.0282	0.073	0.39	0.697	-0.114	0.170
Trial Number	-0.00049	0.00039	-1.25	0.210	-0.00126	0.00028
Duration	0.140	0.046	3.03	0.0025 **	0.0496	0.231
StimType × Trial Number	0.000081	0.00087	0.09	0.926	-0.00162	0.00178
StimType × Duration	-0.0265	0.099	-0.27	0.789	-0.221	0.168
Trial Number × Duration	-0.00118	0.00059	-2.01	0.045 *	-0.00233	-0.00003
StimType × Trial Number × Duration	0.00267	0.00120	2.22	0.026 *	0.00032	0.00501

Table 61. Random Effects Coefficients for LME: Task Variation 3, 4762 Hz Signal. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for reaction times in the 4762 Hz signal condition.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
Session	Intercept	0.00043	0.0207	0.0000	0.0564
Time of Day	Intercept	0.00000	0.0000	0.0000	0.0931
Ferret	Intercept	0.0163	0.1278	0.0389	0.3819
Residual	–	0.1401	0.3743	0.3579	0.3895

7.2.3.2.3.3 Signal Frequency 7559 Hz

Table 62. Fixed Effects Coefficients for LME: Task Variation 3, 7559 Hz Signal. Summary of fixed effects predicting reaction times for the 7559 Hz signal condition in Task Variation 3. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	β	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	1.669	0.076	21.880	0.0284*	1.519	1.818
StimType (Probe)	-0.028	0.025	-1.136	0.5260	-0.77	0.020

Table 63. Random Effects Coefficients for LME: Task Variation 3, 7559 Hz Signal. Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for reaction times in the 7559 Hz signal condition.

Variable	Effect	Variance (σ^2)	σ	95% CI (Lower)	95% CI (Upper)
Session	Intercept	0.0000	0.0000	0.0000	0.0403
Ferret	Intercept	0.01138	0.1067	0.0332	0.3183
Time of Day	Intercept	0.0000	0.0000	0.0000	0.0783
Residual	-	0.10769	0.3282	0.3147	0.3422

7.2.3.2.4 Task Variation 4

Table 64. Fixed Effects Coefficients for LME: Task Variation 4 (Complex Inharmonic Stimuli). Summary of fixed effects predicting reaction times in Task Variation 4, using a broadband, inharmonic probe. Asterisks denote significant effects at $p < 0.05$.

Fixed Effect	β	SE	t	p-value	95% CI (lower)	95% CI (upper)
Intercept	1.623	0.0510	31.82	<0.0001*	1.52	1.72
Duration	-0.163	0.0531	-3.07	0.0023**	-0.27	-0.06
Trial Number	-0.00042	0.00081	-0.52	0.6031	-0.0020	0.0012
StimType (Probe)	0.0130	0.1337	0.097	0.9225	-0.25	0.28
Duration x StimType	0.120	0.156	0.77	0.4416	-0.19	0.43
TrialNumber x StimType	-0.00047	0.0024	-0.19	0.8471	-0.0052	0.0043
Duration x TrialNumber	0.00417	0.0011	3.64	0.0003***	0.0019	0.0064
StimType x Duration x TrialNumber	-0.00044	0.0029	-0.15	0.8813	-0.0062	0.0053

Table 65. Random Effects Coefficients for LME: Task Variation 4 (Complex Inharmonic Stimuli). Variance components for random intercepts grouped by *FerretID, SessionID*, and *TimeOfDay* for reaction times in Task Variation 4.

Variable	Effect	Variance (σ^2)	95% CI (Lower)	95% CI (Upper)
Session	Intercept	0.00000	0.0000	0.0409
Ferret	Intercept	0.00141	0.0000	0.1359
Time of Day	Intercept	0.00068	0.0000	0.1207
Residual	-	0007614	0.2562	0.2935

Table 66. Post Hoc Comparisons for *TrialNumber × Duration Interaction* in Task Variation 4. Post hoc comparisons across trial bins for the interaction between tone duration and trial number, based on the LME model for Task Variation 4. Asterisks denote significant effects at $p < 0.05$.

TrialBin	Contrast	Estimate	SE	df	t-ratio	p-value
Bin1	0.1 - 0.2 s	0.1155	0.1120	401	1.032	0.3027
Bin2	0.1 - 0.2 s	0.0246	0.0919	399	0.267	0.7893
Bin3	0.1 - 0.2 s	-0.1026	0.0826	399	-1.242	0.2150
Bin4	0.1 - 0.2 s	-0.1901	0.0675	397	-2.814	0.0051

Table 67. rial Bin Comparisons Within Duration Conditions for Task Variation 4. Pairwise comparisons of trial bins separately within short and long durations, based on the LME model for Task Variation 4. Asterisks denote significant effects.

Duration	Contrast	Estimate	SE	df	t-ratio	p-value
short	Bin1 - Bin2	0.09512	0.1210	401	0.784	1.0000
short	Bin1 - Bin3	0.10748	0.1250	393	0.863	1.0000
short	Bin1 - Bin4	0.08319	0.1150	400	0.724	1.0000
short	Bin2 - Bin3	0.01235	0.0957	389	0.129	1.0000
short	Bin2 - Bin4	-0.01193	0.0848	402	-0.141	1.0000
short	Bin3 - Bin4	-0.02429	0.0895	355	-0.272	1.0000
long	Bin1 - Bin2	0.00421	0.0768	399	0.055	0.9563
long	Bin1 - Bin3	-0.11058	0.0641	401	-1.725	0.3360
long	Bin1 - Bin4	-0.22237	0.0621	374	-3.583	0.0023*
long	Bin2 - Bin3	-0.11478	0.0798	398	-1.438	0.3360
long	Bin2 - Bin4	-0.22658	0.0786	364	-2.883	0.0209*
long	Bin3 - Bin4	-0.11179	0.0645	401	-1.732	0.3360

7.2.4 Proportion of Long and Short Responses by Frequency and Condition

To assess whether a frequency-duration perceptual bias similar to that reported by Luthra et al. (2021) was present, the proportion of long versus short responses was examined for each frequency and task variation. These descriptive tables characterise response tendencies across frequencies and conditions, allowing evaluation of potential interactions between frequency and perceived duration.

7.2.4.1 Variation 1

Table 68. Proportion of Long and Short Responses by Frequency for Task Variation 1. Summary of the number and proportion of long versus short responses for each tested frequency in Task Variation 1. Frequencies correspond to both signal and probe tones presented during the task.

Frequency (Hz)	Trials (n)	Long Responses (n)	Proportion Long	Proportion Short
1890	262	164	0.626	0.374
2381	281	175	0.623	0.377
3000	4225	2378	0.563	0.437
3780	276	169	0.612	0.388
4762	265	145	0.547	0.453

7.2.4.2 Variation 2

Table 69. Proportion of Long and Short Responses by Frequency for Task Variation 2. Summary of the number and proportion of long versus short responses for each tested frequency in Task Variation 2. Frequencies include the signal and probe tones used in this task variation.

Frequency (Hz)	Trials (n)	Long Responses (n)	Proportion Long	Proportion Short
3150	517	338	0.654	0.346
3969	523	317	0.606	0.394
5000	8319	4072	0.489	0.511
6300	532	351	0.660	0.340
7937	526	262	0.498	0.502

7.2.4.3 Variation 3

Table 70. Proportion of Long and Short Responses by Frequency and Condition for Task Variation 3. Summary of long versus short responses by frequency and signal condition (3000 Hz, 4762 Hz, or 7559 Hz) in Task Variation 3. Each row indicates the total number of trials, the number of long responses, and calculated proportions, split by condition.

Frequency (Hz)	Trials (n)	Long Responses (n)	Proportion Long	Proportion Short	Condition (Signal Frequency)
3000	1325	596	0.45	0.55	3000
3780	86	46	0.535	0.465	3000
4762	75	37	0.493	0.507	3000
6000	105	61	0.581	0.419	3000
7559	83	36	0.434	0.566	3000
3000	79	40	0.506	0.494	4762
3780	89	50	0.562	0.438	4762
4762	1505	669	0.445	0.555	4762
6000	94	64	0.681	0.319	4762
7559	81	47	0.58	0.42	4762
3000	84	37	0.44	0.56	7559
3780	97	49	0.505	0.495	7559
4762	81	35	0.432	0.568	7559
6000	94	59	0.628	0.372	7559
7559	1477	651	0.441	0.559	7559

7.2.4.4 Variation 4

Table 71. Proportion of Long and Short Responses by Sound Type for Task Variation 4. Summary of long versus short responses for Task Variation 4, comparing the pure tone signal (4762 Hz) and the broadband, inharmonic probe stimulus. Proportions reflect the response bias across sound types.

Sound	Trials (n)	Long Responses (n)	Proportion Long	Proportion Short
4762 Hz	599	265	0.442	0.558
Complex inharmonic	128	96	0.75	0.25

7.2 APPENDIX 3 – VERTICAL SOUND LOCALISATION

7.2.1 LME Model Full Results

To assess pupil responses during vertical sound localisation, separate linear mixed-effects models were fitted for each condition (top speaker deviant, bottom speaker deviant, and control) in each cohort.

7.2.1.1 Intact Cohort

Table 72. LME Model for Pupil Dilation in the Top Speaker Deviant Condition (Pinnae-intact Cohort). Predictors include stimulus type, peak pupil time, peak wheel speed, baseline pupil size, and trial history. Leave-one-out cross-validation (LOO) approximation: 7138.26. Asterisks indicate significant effects at $p < 0.05$.

Predictor	Beta (β)	SE	z	p	95% CI
Intercept	4.726	0.551	8.758	<.0001*	[3.646, 5.806]
Stimulus Type (Deviant)	0.959	0.454	2.112	0.035*	[0.069, 1.849]
Peak Pupil Time	4.415	0.262	16.883	<.0001*	[3.903, 4.928]
Trial History (Deviant)	-0.001	0.004	-0.326	0.744	[-0.008, 0.006]
Peak Wheel Speed	89.53	6.142	14.577	<.0001*	[77.492, 101.567]
Baseline Pupil	-0.063	0.012	-5.196	<.0001*	[-0.086, -0.039]

Table 73. LME for Pupil Dilation in the Bottom Speaker Deviant Condition (Pinnae-Intact Cohort). Predictors include stimulus type, peak pupil time, peak wheel speed, baseline pupil size, and trial history. LOO approximation: 7024.99. Asterisks indicate significant effects at $p < 0.05$.

Predictor	Beta (β)	SE	z	p	95% CI
Intercept	5.043	0.434	11.620	<.0001*	[4.192, 5.893]
Stimulus Type (Deviant)	-0.899	0.434	-2.071	0.038*	[-1.750, -0.048]
Peak Pupil Time	5.15	0.261	19.748	<.0001*	[4.639, 5.661]
Trial History (Deviant)	-0.002	0.003	-0.747	0.455	[-0.008, 0.004]
Peak Wheel Speed	85.468	5.322	16.059	<.0001*	[75.037, 95.899]
Baseline Pupil	-0.069	0.01	-6.584	<.0001*	[-0.089, -0.048]

Table 74. LME model for Pupil Dilation in the Control Condition (Pinnae-intact Cohort). Predictors include stimulus type, peak pupil time, peak wheel speed, baseline pupil size, and trial history. LOO approximation: 7013.29. Asterisks indicate significant effects at $p < 0.05$.

Predictor	Beta (β)	SE	z	p	95% CI
Intercept	4.737	1.901	3.976	<.0001*	[2.402, 7.072]
Stimulus Type (Deviant)	0.72	0.26	2.772	0.006*	[0.211, 1.230]
Peak Pupil Time	4.367	0.249	17.547	<.0001*	[3.879, 4.855]
Trial History (Deviant)	0.012	0.009	1.439	0.15	[-0.004, 0.029]
Peak Wheel Speed	76.786	6.276	12.235	<.0001*	[64.485, 89.086]
Baseline Pupil	-0.047	0.011	-4.287	<.0001*	[-0.068, -0.025]

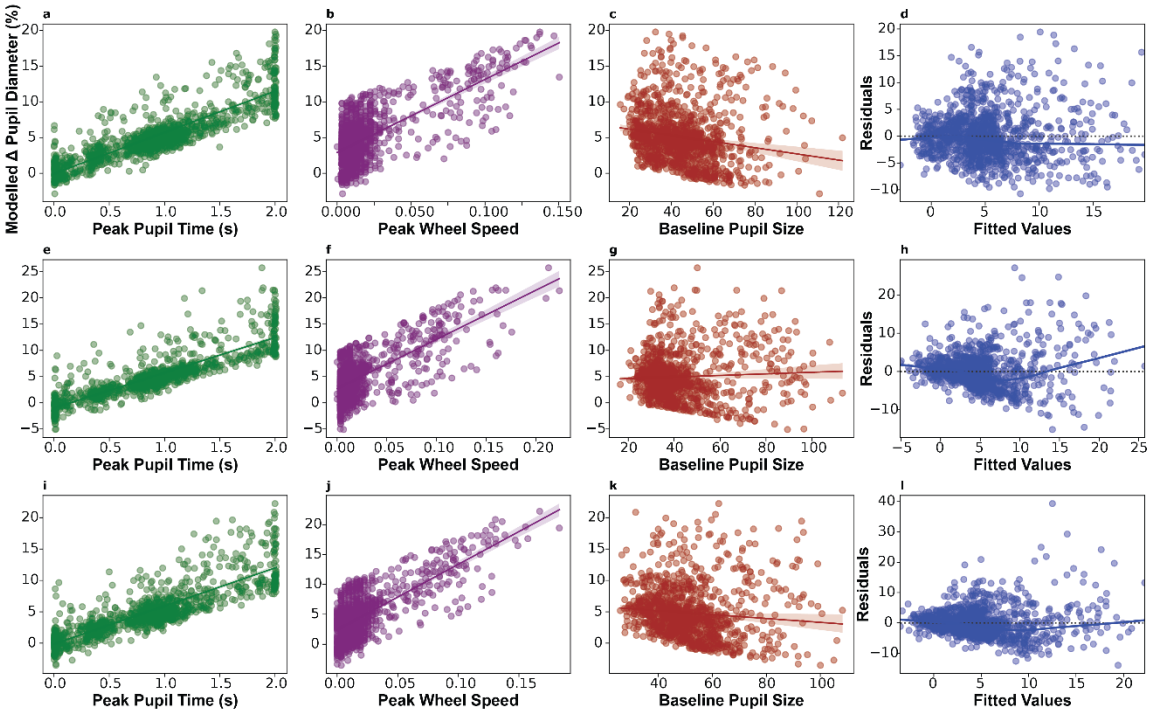

Figure 59. Diagnostic Plots for Linear Mixed Effects Models in the Pinnae-Intact Cohort. Each row corresponds to a condition (top row: Control, middle row: Bottom Speaker Deviant; bottom row: Top Speaker Deviant). **(a, e, i)** Relationship between peak pupil time and modelled Δ pupil diameter. **(b, f, j)** Relationship between peak wheel speed and modelled Δ pupil diameter. **(c, g, k)** Relationship between baseline pupil size and modelled Δ pupil diameter. **(d, h, l) Residuals** plotted against fitted values. All values reflect model-predicted pupil dilation per trial. Each dot corresponds to a single trial, and lines represent the fitted relationship estimated by the model. Residual plots are included to assess the quality of model fit and detect any systematic deviations.

7.2.1.2 Pinnae-Removed Cohort

Table 75. LME model for pupil size in the pinnae-removed cohort (Top Speaker Deviant Condition). Predictors include stimulus type (standard/deviant), peak pupil time, peak wheel speed, baseline pupil size, and trial history. LOO approximation: 5821.19. Asterisks indicate significant effects at $p < 0.05$.

Predictor	Beta (β)	SE	z	p	95% CI
Intercept	6.898	1.227	5.623	0.002*	[4.494, 9.302]
Stimulus Type (Deviant)	0.884	0.718	1.231	0.218	[-0.524, 2.291]
Peak Pupil Time	8.073	0.357	22.599	<.0001*	[7.373, 8.773]
Trial History (Deviant)	0.013	0.009	1.505	0.132	[-0.004, 0.031]
Peak Wheel Speed	34.584	6.851	5.048	<.0001*	[21.157, 48.010]
Baseline Pupil	-0.118	0.015	-7.644	<.0001*	[-0.148, -0.087]

Table 76. LME For Pupil Size in the Pinnae-Removed Cohort (Bottom Speaker Deviant Condition). Predictors include stimulus type (standard/deviant), peak pupil time, peak wheel speed, baseline pupil size, and trial history. LOO approximation: 3677.60. Asterisks indicate significant effects at $p < 0.05$.

Predictor	Beta (β)	SE	z	p	95% CI
Intercept	9.656	2.778	3.475	<.0001*	[9.808, 21.647]
Stimulus Type (Deviant)	-2.507	0.871	-2.88	0.004*	[-4.213, -0.801]
Peak Pupil Time	7.221	0.466	15.488	<.0001*	[6.308, 8.135]
Trial History (Deviant)	-0.018	0.014	-1.297	0.195	[-0.044, 0.009]
Peak Wheel Speed	110.228	13.155	8.379	<.0001*	[84.444, 136.011]
Baseline Pupil	-0.271	0.022	-12.536	<.0001*	[-0.314, -0.229]

Table 77. LME Model for Pupil Size in the Pinnae-Removed Cohort (Control Condition). Predictors include stimulus type (standard/deviant), peak pupil time, peak wheel speed, baseline pupil size, and trial history. LOO approximation: 5279.73. Asterisks indicate significant effects at $p < 0.05$.

Predictor	Beta (β)	SE	z	p	95% CI
Intercept	7.386	1.130	6.535	<.0001*	[4.543, 10.216]
Stimulus Type (Deviant)	0.511	0.431	1.186	0.235	[-0.333, 1.355]
Peak Pupil Time	7.988	0.447	17.877	<.0001*	[7.112, 8.863]
Trial History (Deviant)	-0.03	0.016	-1.887	0.059	[-0.061, 0.001]
Peak Wheel Speed	159.218	12.294	12.951	<.0001*	[135.122, 183.314]
Baseline Pupil	-0.206	0.018	-11.194	<.0001*	[-0.242, -0.170]

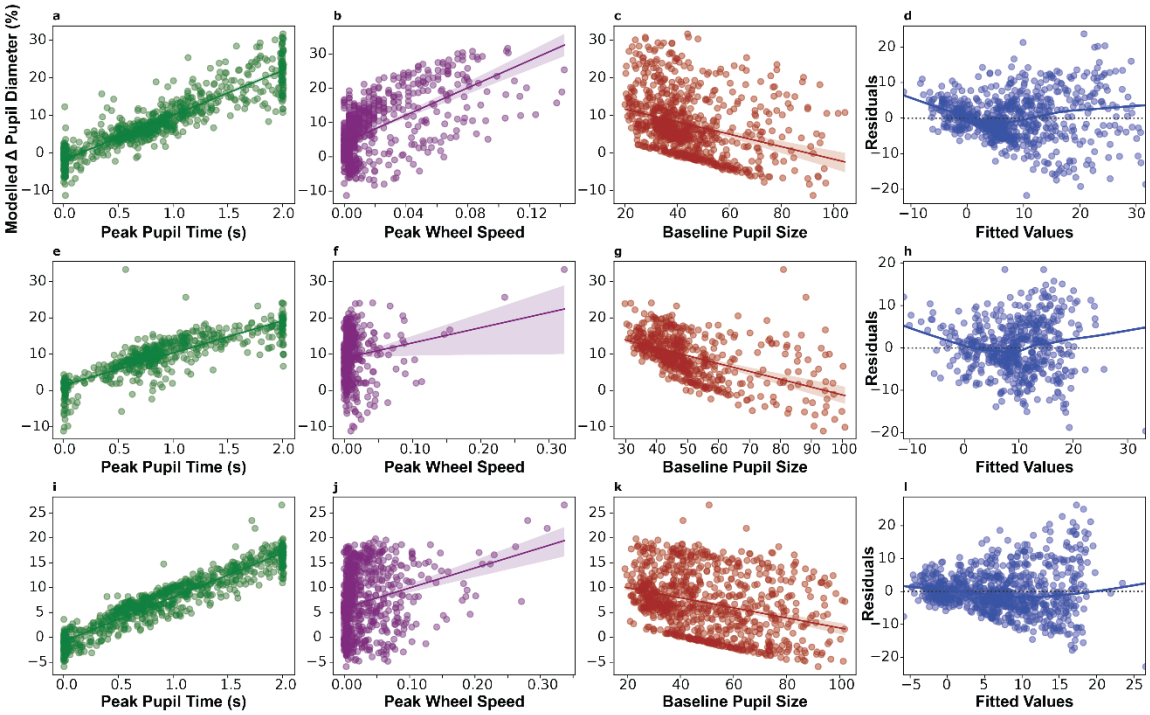

Figure 60. Model Diagnostics for the Pinnae-Removed Cohort.
Each row represents one condition (top: Control; middle: Bottom Speaker Deviant; bottom: Top Speaker Deviant). **(a, e, i)** Relationship between peak pupil time and modelled change in pupil diameter (Δ Pupil Diameter). **(b, f, j)** Relationship between peak wheel speed and modelled Δ pupil diameter. **(c, g, k)** Relationship between baseline pupil size and modelled Δ pupil diameter. **(d, h, l)** Residuals plotted against fitted values for each model. Each dot represents a trial-level datapoint, and lines represent the fitted relationship estimated by the model. Residual plots are included to assess the quality of model fit and detect any systematic deviations.

7.2.2 Group-level Learning Curve

To complement the individual learning curve analyses presented in the main text, a group-level sigmoid was fitted across all mice and sessions. Although the group model was not used for primary analyses due to substantial inter-animal variability, it is shown here for completeness.

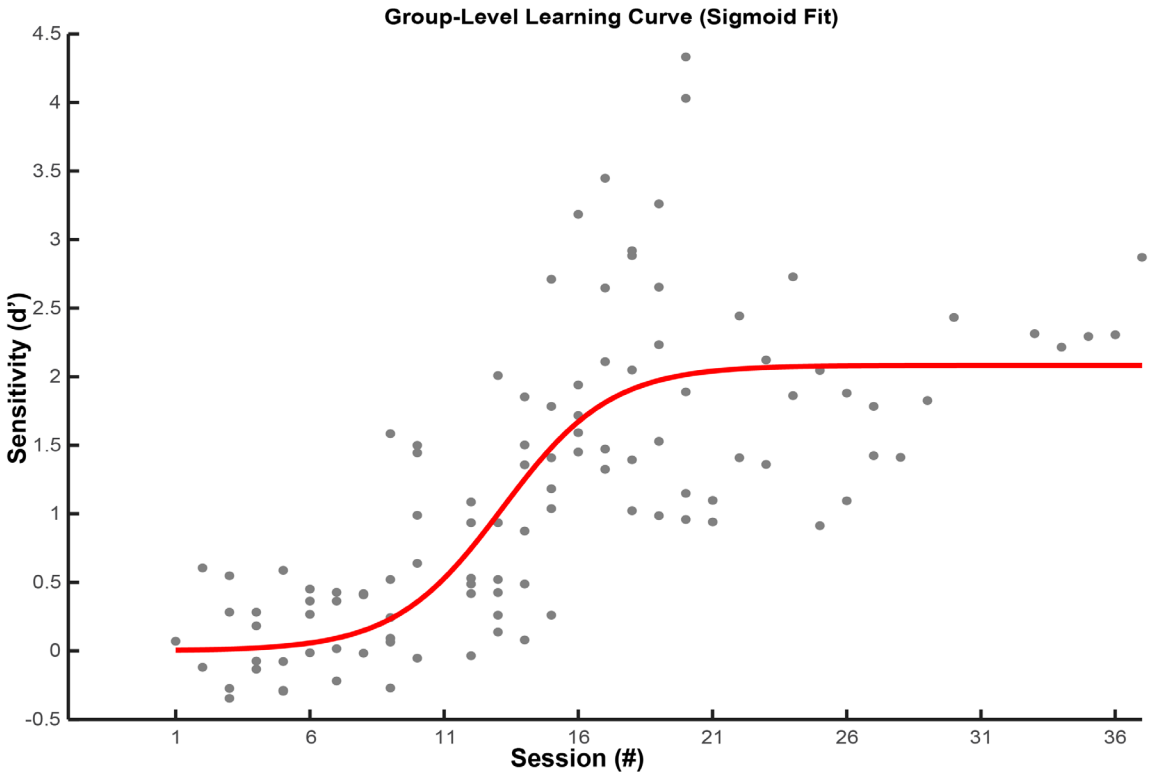

Figure 61. Group-Level Learning Curve. A sigmoidal function was fitted to *d'* values pooled across all mice to characterise the overall learning trajectory. Grey dots represent individual session d' values, and the red line shows the fitted curve. The function was parametrised as $\boldsymbol{d' = \frac{c}{1+e^{b(x-x_0)}}}$, where c represents the asymptotic performance, b the learning rate, and x0 the session at which performance reaches the midpoint of the asymptote. Estimated parameters were b=0.50, c=0.28, x0=31.17.