Neuronal Representation of Sound Source Location in the Auditory Cortex during Active Navigation

Diana Inês Lopes Amaro

Graduate School of
Systemic Neurosciences
LMU Munich

Munich 2021

Neuronal Representation of Sound Source Location in the Auditory Cortex during Active Navigation

Graduate School of
Systemic Neurosciences
LMU Munich

Dissertation der
Graduate School of Systemic Neurosciences der
Ludwig-Maximilians-Universität München

submitted by

Diana Inês Lopes Amaro

January 22nd, 2021

First reviewer and supervisor: PD Dr. Michael Pecka
Ludwig-Maximilians-Universität München

Second reviewer: Prof. Benedikt Grothe

External reviewer: Prof. Nicholas Lesica

Additional member
of the Examination Board: PD Dr. Conny Kopp-Scheinpflug

Date of submission: January 22nd, 2021

Date of defense: March 19th, 2021

"*Windows of my room, ...*
You open onto the mystery of a street continually crossed by people, ..."

– Álvaro de Campos, *Tabacaria*

Abstract

The ability to localize sounds is crucial for the survival of both predators as well as prey. The former rely on their senses to lead them to the latter, which in turn also benefit from locating a predator in the vicinity to escape accordingly. In such cases, the sound localization process typically takes place while the animals are in motion. Since the cues that the brain uses to localize sounds are head-centered (egocentric), they can change very rapidly when an animal moves and rotates. This constitutes an even bigger challenge than sound localization in a static environment. Up to now, however, both aspects have mostly been studied separately in neuroscience, thus limiting our understanding of active sound localization during navigation.

This thesis reports on the development of a novel behavioral paradigm – the Sensory Island Task (SIT) – to promote sound localization during unrestricted motion. By attributing a different behavioral meaning (associated to different outcomes) to two spatially separated sound sources, Mongolian gerbils (*Meriones unguiculatus*) were trained to forage for an area (target island) in the arena that triggered a change in the active sound source to the target loudspeaker and to report its detection by remaining within the island for a duration of 6 s. Importantly, the two loudspeakers played identical sounds and the location of the target island in the arena was changed randomly every trial. When the probability of successfully identifying the target island exceeded the chance level, a tetrode bundle was implanted in the primary auditory cortex of the gerbils to record neuronal responses during task performance.

Canonically, the auditory cortex (AC) is described as possessing neurons with a broad hemispheric tuning. Nonetheless, context and behavioral state have been shown to modulate the neuronal responses in the AC. The experiments described in this thesis demonstrate the existence of a large variety of additional, previously unreported (or underreported) spatial tuning types. In particular, neurons that were sensitive to the midline and, most intriguingly, neurons that were sensitive to the task identity of the active loudspeaker were observed. The latter comprise neurons that were spatially tuned to only one of the two loudspeakers, neurons that exhibited a large difference in the preferred egocentric sound-source location for the two loudspeakers as well as spatially untuned neurons whose firing rate changed depending on the active loudspeaker. Additionally, temporal complexity in the neuronal responses was observed, with neurons changing their preferred egocentric sound-source location throughout their response to a sound.

Corroborating earlier studies, also here it was found that the task-specific choice of the animal was reflected in the neuronal responses. Specifically, the neuronal firing rate decreased before the animal successfully finished a trial in comparison to situations in which the gerbil incorrectly left the target island before trial completion. Furthermore, the differential behavioral meaning between the two loudspeakers was found to be represented in the neuronal tuning acuity, with neurons being more sharply tuned to sounds coming from the target than from the background loudspeaker.

Lastly, by implementing an artificial neural network, all of the observed phenomena could be studied in a common framework, enabling a better and more comprehensive understanding of the computational relevance of the diversity of observed neuronal responses. Strikingly, the algorithm was capable of predicting not only the egocentric sound-source location but also which sound source was active – both with high accuracy.

Taken together, the results presented in this thesis suggest the existence of an interlaced coding of egocentric and allocentric information in the neurons of the primary auditory cortex. These novel findings thus contribute towards a better understanding of how sound sources are perceptually stable during self-motion, an effect that could be advantageous for selective hearing.

Contents

1 Introduction **1**

1.1 Circuits for horizontal sound localization in mammals 3
1.2 Auditory cortex 6
1.2.1 Learning from the anesthetized brain 6
1.2.2 Introducing natural behavior 7
1.3 Motivation 10
1.4 Publications 11
1.5 Outline 12

2 Methods **13**

2.1 Behavioral paradigm: Sensory Island Task (SIT) 13
2.1.1 Auditory SIT in the sound localization version: aSIT$_{loc}$ 13
2.1.2 Auditory SIT in the frequency version: aSIT$_{freq}$ 16
2.1.3 Behavioral setup 17
2.1.4 Animal model: Mongolian gerbils 20
2.1.5 Behavioral training 21
2.2 Electrophysiological recordings 22
2.2.1 Tetrode bundle construction 22
2.2.2 Surgery 23
2.2.3 Electrophysiological recordings during task performance 24
2.2.4 Spike sorting 25
2.2.5 Identification of the recorded brain area 25
2.3 Offline video tracking 27
2.3.1 Initial tracking 27
2.3.2 Visual inspection of flagged frames 28
2.3.3 Visual inspection of entire videos 29
2.3.4 Error estimation of the gerbil's orientation 30
2.3.5 Calculation of Bouts 30
2.4 Data analysis 32
2.4.1 Calculation of the behavioral performance level 32
2.4.2 Peristimulus time histograms (PSTH) 33
2.4.3 Latency and offset calculation 34

2.4.4 Video-frame to sound presentation to spike assignment 36
2.4.5 Spatial tuning analysis 36
2.4.6 Neuron type analysis 37
2.4.7 Decoder analysis 38

3 Results **41**
3.1 Behavioral performance analysis 41
3.1.1 Were the animals performing a sound localization task? 41
3.1.2 Catch trials 42
3.1.3 Spatial bias in the target island location 43
3.1.4 Swapped loudspeakers 45
3.2 Behavior description 46
3.2.1 Bout analysis 46
3.2.2 Locomotion state analysis 49
3.3 Neuronal responses 52
3.3.1 Neuronal latencies 52
3.3.2 Classification and characterization of neuronal types 53
3.3.3 Spatial tuning diversity 54
3.3.4 Diversity in spatial tuning formation 59
3.3.5 Temporal analysis of spatial tuning 60
3.3.6 Spatially untuned neurons with differential magnitude between loudspeakers 64
3.3.7 Animal's performance is reflected in the population response rates . 65
3.3.8 Neuronal tuning is sharper for the target loudspeaker 67
3.3.9 Spontaneous firing rate increases during engagement 68
3.3.10 Simultaneously decoding the egocentric sound-source location and the loudspeaker identity 69

4 Discussion **73**
4.1 Behavioral paradigm 73
4.2 Neuronal recordings 75
4.3 Future directions 82

A Appendix - Sampling uniformly from a circle **85**

B Appendix - Behavioral paradigm (SIT) article **87**

References **105**

Acknowledgements **123**

CHAPTER 1

Introduction

Many animals depend on the ability to localize sounds to survive. When a prey is not in sight, predators forage and rely on smell and sounds to lead them to their potential next meal. Conversely, a prey must detect the presence and locate a predator in the vicinity to escape.

Localizing sounds is not only important in the wild: when a person goes to a café with a friend and engages in a conversation, they naturally focus their attention towards the direction of their friend, helping them to better understand speech in a noisy environment [1, 2]. Crossing a busy road also demands localizing sounds since one should quickly perceive where a car is coming from to appropriately react and avoid an accident – in particular when visiting countries where cars drive opposite to the expected side of the road.

In contrast to the somatosensory and visual systems, for which there is a direct mapping between the spatial world and the topography of the neurons that represent it – sensory "homunculus" and retinotopy – the auditory system must perform computations to obtain spatial information. This is a consequence of the cochlea not being spatially organized, but by frequency – the so-called tonotopy – meaning that contiguous hair cells transmit information about similar frequencies [3], which are not necessarily coming from sounds with the same spatial origin. The brain uses two types of information for localizing sounds, namely binaural information and spectral information [4].

- Binaural information

Binaural information is based on a comparison between the two ears of the arrival time (interaural time difference – **ITD**) or of the phase of the sound wave (interaural phase difference – **IPD** – almost equivalent to ITD for pure tones), and of the sound intensity (interaural level difference – **ILD**). Low frequency sounds have a wavelength larger than a human head and consequently diffract around it without being significantly attenuated. When such a sound is presented laterally, it arrives first at one ear and only some microseconds later at the other ear – this time difference (ITD) is used by the brain to determine where the sound comes from. A human head creates a maximum ITD of around 700 μs when the sound comes directly from the side – a pure tone with this period corresponds to approximately 1.4 kHz. Above this frequency, more than one complete wave can "fit" in

between the two ears and it is no longer possible to unambiguously determine the phase difference. This corresponds to the maximum frequency for which the brain uses this cue [5]. In reality, ambiguity on the leading ear occurs already for continuous pure tones with half that frequency [6]. For higher frequencies, the head becomes an obstacle to the sound wave and part of it is reflected or absorbed. This results in a decrease in the sound intensity on the ear that is farther away from the sound source and consequently a difference in the sound level (ILD) is created between the ears. This effect is stronger for frequencies above 5 kHz and therefore the localization errors are the largest between 1.4 kHz and 5 kHz [7] – the frequency range where neither of the two cues is very informative. Due to the symmetric arrangement of the ears, both of these sound localization cues are only useful to determine a sound-source location in the horizontal plane. However, a given ITD can be created by sounds coming from a multitude of directions which share the same angle with the interaural axis (the line that passes through both ears), forming a so-called cone of confusion. Similarly, sounds resulting in the same ILD can originate from an infinity of incoming angles for a pure tone. Using broadband stimuli reduces such ambiguities and consequently improves sound localization, due not only to the contribution of spectral cues but also to the frequency-dependent character of both IPDs and ILDs. Furthermore, IPDs can be determined for the envelope of high frequency sounds [8, 9].

- Spectral information

The head, torso, pinnae and external ear act as spectral filters that enhance certain frequencies (peaks) while reduce others (notches) depending on the direction of the incident sound wave [4]. These peaks and notches are a consequence of the interference between successive reflections and diffractions of the incoming wave. This effect is essential to determine whether a sound comes from above or below (elevation) and is often regarded as a monaural cue. However, this information complements the binaural cues and is crucial to resolve their cone of confusion (and therefore to disambiguate the front from the back).

The neuronal computations performed to localize a sound are consequently egocentric in nature (i.e. are head-centered). However, those egocentric cues are constantly altered since natural sound localization behaviors such as chasing a prey require movement, during which the orientation of the animal's head with respect to the sound source changes all the time. Hence, sound localization during motion constitutes an even bigger challenge for our brains than static sound localization. Nonetheless, in neuroscience both aspects have so far been studied separately, limiting our understanding of sound localization during active navigation.

In the following sections, I will briefly describe the neuronal circuits that are involved in the processing of the binaural cues during sound localization, with emphasis on ITDs – the cue that was used in the experiments reported in this monograph. In Section 1.1, the auditory processing stages are presented in the order of the ascending auditory pathway until the auditory cortex (AC), including a brief discussion on how the information is transformed throughout. At the end of the section, the ubiquitous feedback connections

are succinctly mentioned. Subsequently, in Section 1.2, I will focus on the AC, the structure concerned in this study.

1.1 Circuits for horizontal sound localization in mammals

A sound creates oscillations in the eardrum that are amplified, impedance-matched by the ossicles and transmitted to the fluid-filled cochlea. Here, inner hair cells transduce the relative movement of the tectorial membrane and the basilar membrane caused by mechanical fluid vibrations into bioelectric signals. The latter are transmitted as action potentials by the axons of the spiral ganglion neurons, also known as auditory nerve fibers. The responses of the auditory nerve fibers to low frequency sounds exhibit "phase locking" – action potentials occur in phase with the sound oscillations [10]. This effect occurs only up to a certain species-specific frequency [11], after which the inner hair cells cannot follow the fast movement and start acting as low pass filters.

After leaving the cochlea, the auditory nerve fibers enter the brainstem where they bifurcate in the cochlear nucleus (CN) [12]. The descending pathway innervates the posterior-ventral and the dorsal cochlear nuclei (PVCN and DCN, respectively). The former is responsible for the analysis of complex tones [13] and the latter for analyzing spectral contrasts and consequently for the monaural sound localization using spectral cues [14]. The ascending branch, on the other hand, innervates the anterior-ventral cochlear nucleus (AVCN), where some auditory nerve fibers form connections with bushy cells via endbulbs of Held. These neurons are phase-locked either to the fine structure of low frequency sounds or to the envelope of high frequencies. Interestingly, they are better synchronized to the stimuli than the auditory nerves that innervate them [15].

The ipsilateral medial nucleus of the trapezoid body (MNTB) receives glutamatergic projections from the contralateral globular bushy cells (for an extensive review see [4]) via the calyx of Held – a very large, fast and efficient synapse. The glycinergic neurons in the MNTB project to the ipsilateral lateral superior olive (LSO) where they converge with the ipsilateral glutamatergic spherical bushy cells. This circuit consists of an ipsilateral excitation and a contralateral inhibition, resulting in a subtraction process. The firing rate of LSO neurons is high if the sound comes from the ipsilateral side, whereas their response is inhibited by a contralateral sound – a mechanism that works mainly for ILDs and consequently for high frequencies. The low frequency part of the LSO is sensitive to IPDs [16], yet the medial superior olive (MSO) is the main responsible for their processing. MSO neurons, similar to the LSO, receive excitatory connections from the ipsilateral spherical bushy cells and inhibitory projections from the ipsilateral MNTB; however, they get additional excitatory projections from the contralateral spherical bushy cells and inhibitory inputs from the ipsilateral lateral nucleus of the trapezoid body which is innervated by the ipsilateral globular bushy cells. Contrary to the LSO, a MSO neuron has a larger response for a contralateral sound, maximizing its dynamic range for the physiological ITD values [17], with the steepest slope at midline (where sound localization was shown to be the

most accurate [7]). This circuit consists of a coincidence detection scheme precisely timed by glycinergic inhibition [18–21].

The contralateral-favoring MSO neurons send excitatory projections to the ipsilateral dorsal nucleus of the lateral lemniscus (DNLL) and inferior colliculus (IC) [22], creating the contralateral bias observed in the IC and, at a later stage, in the AC. Further reinforcing this contralateral bias, the ipsilateral-favoring LSO neurons send excitatory projections to the contralateral DNLL and IC and send inhibitory projections to the ipsilateral DNLL and IC [23, 24]. In the IC, the ipsilateral excitatory MSO inputs are not colocalized with the contralateral excitatory LSO inputs but rather with the inhibitory ipsilateral LSO inputs. This hints that there is at this stage some integration of ITDs and ILDs although the majority remains segregated [25]. The IC also receives excitatory projections from the contralateral AVCN and inhibitory projections from the contralateral DNLL, a circuit that creates *de novo* ILD sensitivity [26].

The IC constitutes an auditory information hub, integrating various sound features, since it receives projections from many auditory processing stages in the brainstem; however, it also receives ascending inputs from non-auditory areas such as somatosensory [27] and the retina [28], and descending inputs from the AC, as well as from the somatosensory, visual, motor, and prefrontal cortices [29, 30]. In short, the IC is a region of convergence of many diverse inputs. Furthermore, the ICs from both hemispheres are very interconnected via the commissural fibers [31]. By cooling of the contralateral IC, these fibers were shown to be responsible for enhancing the representation of sound location [32]. Interestingly, such an enhancement via increase in the dynamic range had already been reported for ITDs earlier in the brainstem, where the sound location representation was compared between MSO and DNLL [33]. Similarly to the IC, the DNLL also projects to its contralateral part – mostly inhibitory projections [34] – thereby strengthening the parallelism.

The IC projects to the superior colliculus (SC), which is responsible for orienting behavior. In this structure an alignment of auditory, visual and somatosensory maps was observed [35]; however, ITDs do not contribute to this spatial representation [36, 37]. Nonetheless, the main output of the IC is the medial geniculate body (MGB) in the thalamus, which in turn also receives most of its inputs from the IC. These projections are mainly to the ipsilateral side [38] and are predominantly excitatory, with only 10-30% releasing the inhibitory gamma-Aminobutyric acid (GABA) neurotransmitter [39, 40]. Furthermore, the MGB also receives inhibition from the thalamic reticular nucleus (TRN) [41], which has been associated with selective attention [42]. The responses to sounds presented from varying azimuths were shown not to be significantly different between MGB and the AC to which MGB mostly projects [43]. However, the spatial tuning becomes sharper when ascending the auditory pathway (from the superior olivary complex – SOC – to IC and to MGB) [44], and spatial stream segregation becomes more prominent from the IC to the MGB and to the primary auditory cortex (A1) [45].

A set of experiments demonstrated that the A1 is necessary for localizing sounds. Following unilateral A1 lesions, cats could no longer localize brief noise bursts in the con-

tralateral hemifield; however, the performance did not drop when longer sounds were played [46]. Also in ferrets, unilateral A1 lesions resulted in behavioral deficits in localizing a click on the contralateral side, and bilateral lesions caused deficits on both sides but curiously not in the midline [47]. The results were identical when A1 was reversibly cooled [48] or reversibly deactivated by the GABA agonist muscimol [49]; furthermore, as expected, the head orienting behavior was not affected – since it is mediated by the SC.

The sensory information does not only travel in one direction, since in reality the auditory system contains a multitude of feedback connections. Some go all the way from the AC (mostly layer V and VI) to every sound processing stage like the MGB, IC, nucleus of the lateral lemniscus, SOC, CN [50] and even modulate the cochlea's outer hair cells [51]. The corticofugal feedback projections are typically interpreted as modulatory, adapting the neurons to the statistics of the auditory stimuli [52, 53]. The corticocollicular feedback was also shown to reshape the sensitivity to sound localization, specifically to ILDs [54] and to be crucial for plasticity during sound localization (re-)learning using broadband noise after occluding one ear (although not essential for an already learned sound localization behavior) [55]. Other feedback loops occur at the level of the brainstem, such as the olivocochlear reflex from the SOC to the hair cells, which is thought to be implicated in gain control as protection to acoustic trauma [56]. This loop is itself modulated by the AC [57] and was shown to be important for sound localization [58]. Many other feedback connections have been observed in the auditory system, contradicting in part the canonical view of a hierarchical character and demonstrating its interconnected nature.

1.2 Auditory cortex

Despite being a sensory area, the auditory cortex is highly interconnected with other brain regions [59] and is modulated by the internal and behavioral state of the animal. Anesthetized experiments are mostly used to improve the knowledge about sensory encoding and underlying properties and circuitry. However, to fully understand the capabilities of the AC, it is necessary to add another layer of complexity by recording in the awake animal while including meaningful and natural behavior.

1.2.1 Learning from the anesthetized brain

In anesthetized animals, the auditory cortex is typically very silent, with very low spontaneous firing rates [60] (the majority below 1 Hz) and typical responses to sounds being binary (either zero or one spike per sound presentation) [61]. A1 exhibits a sparse coding with a log-norm firing rate distribution [62] – a code that has been suggested to be necessary to solve the cocktail party problem [63].

Furthermore, AC neurons in anesthetized cats showed fast stimulus-specific adaptation by possessing a higher firing rate after a pure tone that is rare (for which they show hyperacuity) in comparison to when the same sound is frequent; the same effect was not observed in the thalamus [64]. This is similar to what was described in rodents, where subthreshold responses to complex sounds showed long-lasting stimulus history dependence with a decay time of 1 s, which is one order of magnitude slower than what was observed in the MGB [65].

Regarding the encoding of horizontal auditory space, A1 neurons are predominantly spatially tuned to contralateral sound-source locations, with a smaller subset tuned to ipsilateral or frontal positions both in the ferret [60] as well as in the cat [66]. In the latter, the spatial tuning was shown to be the steepest at the midline [66], supporting the opponent-channel coding model [17]. In the rat, despite finding qualitatively the same phenomenon of the midline being represented by the steepest slope of the spatial tuning, all neurons were reported to be exclusively contralaterally tuned [67]. Interestingly, ILDs and spectral cues were shown to be processed linearly in A1 [68]; however, such a linear model did not predict as well neuronal responses of EE neurons (which predominantly receive excitatory inputs from both ears), associated with ITD encoding [60]. This suggests that these possess non-linearities in their neuronal responses.

In a typical experiment, only one auditory feature is modulated at a time; however, in reality, neurons respond to a combination of several features. This was shown for the ferrets' A1 neurons, whose firing rates simultaneously encoded changes in pitch, timbre and egocentric location of sounds in a non-linear fashion [69], a potential neuronal basis for the formation of auditory objects. Furthermore, AC neurons robustly encoded one of these features while the others were varied via multiplexing: several features were unambiguously represented by the same neuron, each feature encoded at a different time response window [70] – a possible explanation for perceptual stability.

A1 is also a region of multisensory integration, arising from its interconnection with other sensory areas, such as the visual cortex. These projections are probably responsible for the enhancement of sound localization when paired with concurrent visual stimuli [71].

1.2.2 Introducing natural behavior

The auditory cortex and specifically its encoding of azimuthal sound location were often studied in anesthetized animals [72–75]. However, several studies have shown changes relative to the anesthetized condition in neuronal responses of the AC in awake humans [76], monkeys [77, 78], cats [79] and rodents [80].

Already in the 1950s, an anecdotal article reported the existence of "attention units" that only fire if the "cat pays attention to the sound source" [81]. Accordingly, since then, it was repeatedly shown that not only anesthesia influences firing rates but also the behavioral and internal state of the animal. For example, in the mouse, motherhood induced plasticity in the parvalbumin-expressing inhibitory neurons, whose preferred frequency changed by one octave, resulting in stronger responses of the pyramidal neurons to pups vocalizations [82]. Moreover, sleeping decreased the firing rate of A1 neurons in monkeys [77, 78], as well as passively listening to sounds in comparison to performing a simple sound detection task [77, 83].

Interestingly, the modality to which attention is directed also shapes neuronal activity, since attending to auditory information differently influenced the magnetoencephalography (MEG) responses of the human AC in comparison to when attending to visual information [84]. Not only the attended modality can influence neuronal responses, but also the sound feature within the auditory modality to which one is attending. A study that included behavioral relevance showed that pre-stimulus activity in the A1 of rats encoded the selection rule between two possible tasks: pitch discrimination and sound localization – suggesting pre-activation of different task-specific neuronal networks [85].

The dynamic temporal character of AC must also be taken into account during its study: rapid changes in the order of seconds to a few minutes have been observed in the neuronal responses for the attended frequency during a task [86]. These changes corresponded either to an increase or to a decrease in the firing rate depending on whether the task was associated to an avoidance or an approach behavior, respectively [87]. This could be related to projections from the amygdala to the TRN, which were shown to amplify cortical sound responses [88]. Alternatively, the direct glutamatergic projections from the orbitofrontal cortex (OFC) [89] could be responsible for such changes, since pairing OFC stimulation with sounds resulted in rapid frequency-specific enhancement of A1 activity [90].

The stimulus-specific adaptation reported in the anesthetized condition [64] also occurred for awake non-trained monkeys; however, curiously, it was not as evident for trained ones [83] and it was also reduced for mice engaged in a sound detection task in comparison to passive listening [91], suggesting that top-down modulation reduces

such effect. Interestingly, the monkeys' AC neurons in a passive situation also showed adaptation in the context of sound localization, by shifting their best IPD depending on the previous one – a mechanism that could be at the origin of perception of sound-source motion [92].

Similar to results from the anesthetized AC, multisensory responses are also present in the AC of behaving animals. The auditory stimuli were enhanced in the AC of alert monkeys when presented with matched visual stimuli, whereas the neuronal information decreased when there was a mismatch between the two [93]. Furthermore, sensorimotor encoding was observed in a sizable fraction of AC neurons in a monkey performing an auditory discrimination task [94]. Intriguingly, multisensory integration in AC appears to be gated by attention to the auditory modality given that specific task-related somatosensory and visual cues were encoded in the AC of monkeys exclusively when the task was auditory and not in its visual version (despite the cues being present in both versions) [95].

Sensory perception is an active process that involves movement [96]. Hence, the AC also processes information when an animal is moving. Since sounds can be a consequence of self-movement, it would be advantageous for the AC to suppress those signals. Accordingly, movement was reported to decrease the firing rate in the AC of monkeys, including "very small" movements, such as "postural adjustments, stretching, yawning, chewing, grooming" [78]. Later, several other studies confirmed this initial observation. In mice, both the excitation as well as the inhibition in layers 2/3 were reduced when the mouse was active [97]. Additionally, a corollary discharge circuit involving the secondary motor cortex was proposed to be responsible for the firing rate reduction in the A1 excitatory neurons [98]. This phenomenon was also observed in the human AC, where the MEG responses decreased to sounds that were self-triggered [99]. Similar firing rate reductions were also observed in the IC [100] and in the MGB [101]. These could be a consequence of feedback connections from the AC [102] or direct inputs from motor areas [30].

The neuronal spatial tuning curves in awake monkeys maintained the same distribution as in anesthetized, with the majority of the neurons being contralaterally tuned and a minority tuned to the front or ipsilaterally [103]. In a study where monkeys had to lateralize IPDs [104], no shift in the best IPD was observed during task performance relative to passively listening; however, a small subset of neurons shifted the midpoint of the neurometric curve without a net change in the steepness of their slope. Furthermore, the firing rate of most neurons increased during behavior in relation to a passive situation. In this study, however, the conclusions cannot be generalized into a sound localization situation since the animals were performing a relatively easier lateralization task. Another study investigated the change of the spatial tuning curves during a switch between a simple sound detection task and an actual sound localization task [103]. They observed that a small subset of neurons showed significantly different firing rates between the two tasks – most of which corresponded to a location-specific increase during sound localization. A similar study performed in the AC of cats, showed that a significant proportion of neurons sharpened their sensitivity during a sound localization task in comparison to a passive

condition and to a periodicity detection task [105]. A stark difference between these experiments is that in the first two studies the spatial tuning curves were constructed during the task performance, whereas in the latter the cat had to localize in the elevation and only afterwards the azimuthal space was probed, a time period during which the animal was out of task.

Sound localization has been the subject of several decades of research in various species of bats, especially in the context of echolocation [106]. The latter involves active sound production by the animal, which is then reflected differently by different objects, thereby allowing the perception of the surrounding environment. Recording these emitted sounds constitutes an opportunity to investigate to what the animal is attending and therefore to better understand auditory scene analysis. Some species, such as the mouse-eared bats, besides the echolocation, also take advantage of prey-produced sounds to localize them by reducing the amplitude of their own echolocation sound when approaching the prey [107]. Neuronally, this duality in sound localization is also present: AC neurons of the pallid bat were responsive to both ultrasounds (30-40 kHz - corresponding to the echolocation frequencies) and to lower frequencies (10-15 kHz - for passive listening). Moreover, the neuronal spatial tuning was suggested to depend on whether the animal is echolocating or listening passively [108]. Interestingly, in the AC of the *Phyllostomus discolor*, a small percentage of neurons were shown to be object-size invariant [109], hinting that auditory object formation starts to occur at the level of AC also in bats.

In sum, these effects indicate that the AC is highly modulated by behavioral state, meaning and context.

1.3 Motivation

Information about the position of sensory objects is vital to navigate the environment. Furthermore, localizing the associated sounds contributes to the holistic perception of those objects. A large number of studies performed on sound localization were conducted on anesthetized animals [72, 73, 75, 110] or when they were head-fixed and passively being exposed to auditory stimuli [79, 111, 112]. Nonetheless, movement is an integral part of real-life localization behavior [96]; yet this component is often neglected. Strikingly, despite self-movement constantly altering the egocentric sound-source location, the perception of the source position remains stable relative to the world coordinates, i.e. is allocentric. A recent study provided first hints that allocentric representation components are present in a small subset of A1 neurons [113]. In this study, freely moving ferrets were simply exposed to sounds while foraging for water, with the sound source location being irrelevant for their behavior in this context. However, engagement in localizing sounds was shown to change the corresponding neuronal responses in the auditory cortex [103–105]. Thus, how sound-sources are spatially represented in the A1 during unrestricted movement, selective listening and active sound localization remains elusive.

To shed light on this, this thesis reports on experiments in which neuronal activity was recorded in freely exploring animals that were selectively listening to different sound sources. For these experiments, we have developed a novel paradigm – the Sensory Island Task (SIT) for sound localization. In this paradigm, freely moving animals are trained to actively localize sounds from sources with distinct task identities (i.e., associated with distinct behavioral outcomes) that only differ in their allocentric locations. Importantly, the egocentric information on its own is insufficient to solve the task of identifying the reward-associated loudspeaker (see Fig. 2.1 below). We hypothesize that besides the canonical egocentric contralateral spatial tuning, there also exists a sound-source specific tuning, i.e. an allocentric and reward modulated tuning.

1.4 Publications

During the course of this PhD, I was first (co-)author in the following two articles:

- D. N. Ferreiro*, **D. Amaro***, D. Schmidtke, A. Sobolev, P. Gundi, L. Belliveau, A. Sirota, B. Grothe and M. Pecka (2020) *Sensory Island Task (SIT): A New Behavioral Paradigm to Study Sensory Perception and Neural Processing in Freely Moving Animals.* Front. Behav. Neurosci. **14**, 576154. doi:10.3389/fnbeh.2020.576154
- **D. Amaro**, D. N. Ferreiro, B. Grothe and M. Pecka (2020) *Diverse spatial representations in primary auditory cortex during active localization simultaneously code source location and identity.* bioRχiv (preprint - submitted) doi:10.1101/2021.01.05.425444

*These authors have contributed equally to this work

In the first article (appendix B) I was responsible for the development of the paradigm in its sound localization version, including programming the script to run the task, developing the training protocol and gathering the data. In the frequency single and multiple islands versions, I programmed the code for the task. I analyzed all the data presented in the paper, with the exception of the LFP data and the effect in performance of the position of the target island in the previous trial. Furthermore, I was part of the team that prepared the figures and wrote the manuscript.

In the second article, I performed the experiments and gathered the data, analyzed the results and generated the figures. I also played a key role in writing the manuscript.

1.5 Outline

This thesis reports both on the development of a sound localization behavioral paradigm as well as on the behavioral and neuronal findings acquired by its implementation. Inspired by previous studies from other groups [114, 115] and by preliminary data acquired in a frequency discrimination task by Lucile Belliveau [116] during her PhD in our group, I developed and established a sound localization behavioral paradigm during my PhD. The paradigm is described extensively in the Chapter 2 – Methods, together with the algorithms used for the task and for video-tracking and with the data analysis procedures developed for this paradigm.

In the Chapter 3 – Results, the performance during the behavioral task is analyzed (Section 3.1), including controls to test the nature of the reported perception. Next, I describe quantitatively the locomotive behavior during task performance and compare these measures between background area and target island (Section 3.2). In this chapter, I also report on some general properties of the neuronal data acquired such as latencies and types of neurons observed and corresponding firing characteristics (Section 3.3). Subsequently, I introduce the large variety of spatial tunings of the recorded neurons and provide insight about how some of these spatial tuning curves arise. After analyzing the temporal complexity that some neurons exhibited, I show some examples of neurons with specific spatial tuning curves throughout time. Afterwards, population response rates are analyzed, specifically during comparisons between active loudspeakers, behavioral decisions when in the target island and spontaneous firing rate in engaged and passive situations. At last, an artificial neural network is implemented combining all the observed phenomena into a common framework. Furthermore, the decoding performance is analyzed in the temporal domain and its dependence on specific neuronal spatial tuning classes determined.

In the Chapter 4 – Discussion, I summarize the most important findings and discuss the most curious and interesting phenomena observed. Moreover, I present future perspectives for this line of investigation.

CHAPTER 2

Methods

2.1 Behavioral paradigm: Sensory Island Task (SIT)

SIT is a novel behavioral paradigm for probing sensory perception in unrestrained self-moving and actively engaged animals. In SIT, the freely moving animal is first exposed to a specific background stimulus containing the property of interest to be investigated [117]. The animal is trained to search for an unknown area in the arena (target island) that prompts a change into the target stimulus, and to report the detection of the target stimulus by remaining in the target island for a determined time interval (sit-time). The position of the target island is random and changes every trial, making the stimulus change the only useful cue to finish a trial (and not the memorization of the position of the target island). After correctly reporting the detection of the target stimulus, the animal is rewarded with a food pellet that is automatically dropped in the arena from an overhead food dispenser. Due to the impossibility of predicting where a food pellet bounces to after hitting the arena's floor, no specific region in the arena is associated with the reward. Every trial is limited in time, after which the position of the target island is changed. Several versions of SIT were developed using different cues of interest, specifically: sound frequency, sound-source allocentric (world-based reference frame) location and grating orientation (visual SIT).

2.1.1 Auditory SIT in the sound localization version: aSIT$_{loc}$

My thesis focused primarily on the SIT version where the varying cue was the allocentric location of the active sound-source.

A previous study [118] showed a reduction in the spontaneous activity of neurons in the AC of gerbils exclusively when a trial was voluntarily initiated by the animal, resulting in an increase in the signal-to-noise ratio and thus an improvement in signal detection. A result which was later confirmed in the rat [119]. Furthermore, spatial tuning in the AC sharpens when the animal is actively engaged in an auditory task [105]. With this in mind, and since we intended to combine the behavior with electrophysiological recordings in the AC, an initiation platform was added to the paradigm, allowing the gerbil to self-initiate a trial by staying on top of it for 1 s.

After trial initiation, in aSIT$_{loc}$, one of two loudspeakers was active at a given moment and both played exactly the same sound, corresponding the background stimulus to the sound being played from one loudspeaker and the target stimulus to the sound being played from the other loudspeaker (Fig. 2.1a). Importantly, the animal must use allocentric information to solve the task, since the egocentric information is not sufficient to find which is the rewarded-associated (target) loudspeaker (Fig. 2.1b). The loudspeakers were placed as farther apart as possible to facilitate the task (diametrically opposed — 180° angle separation from the center of the arena). Furthermore, the animal was allowed 60 s to finish a trial. After this limit and if the animal had not yet been successful, a 10 s time-out was triggered during which a low-pass filtered (<1.5kHz) noise was played to the animal and this could not initiate a new trial.

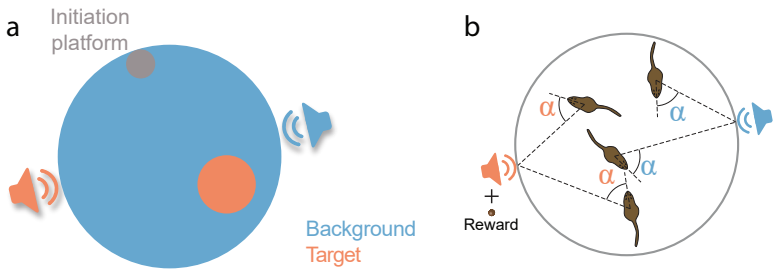


Figure 2.1 – **aSIT**$_{loc}$ **a** Schematic representation. Figure adapted from [117] **b** Egocentric information is not sufficient to solve the task – representation of two possible locations of the animal per active loudspeaker, all four depictions correspond to the same egocentric location of the active loudspeaker

In accordance with the aim of introducing ecological elements to the paradigm, such as free movement and distinct behavioral relevance for different stimuli, we chose the stimuli to possess a naturalistic character, namely we opted for harmonic complex sounds. These are very common in nature, in particular during vocalizations. Moreover, harmonic complexes were shown to elicit response facilitation in A1 neurons with harmonically related multi-peaks if these were stimulated at the frequencies of the multi-peaks, in the marmoset [120]. We low-pass filtered the harmonic complex sound below 1.5 kHz to ensure that ITDs were the only sound localization cue used. We chose 147±4 Hz as fundamental frequency since it allows for enough harmonics to be played as stimulus (10 harmonics below 1.5 kHz, in the ITD regime) and for still being within the hearing range of gerbils [121]. A roving of 4 Hz was added to the fundamental frequency to prevent the animals from picking up possible slight differences between the loudspeakers and using them to solve the task instead of the sound localization cue. For the same reason, we also rove the sound intensity by 5 dB. Because of the great importance of onset ITDs for sound localization [122], sounds played during the task were pulsed. Furthermore, if they were to be continuous, other cues such as a discontinuity in the sound when it switched to the opposite loudspeaker could be used to solve the task, which was not desirable. Each sound was 57 ms long with a 5 ms on- and offset cosine ramp to avoid spectral splatter [123]. A repetition rate of 4 Hz — a new sound onset every 250 ms — was considered fast enough

to enable the animal to complete the task (and not run through the target island without hearing once the target loudspeaker) while allowing enough time for analysis of the spikes relative to each sound.

A green LED was attached to each loudspeaker to signal to the gerbil where the target loudspeaker was located, which we later concluded not to be relevant for the behaving gerbils (see Section 3.1.1). This cannot be assigned to a possible poor vision of gerbils since their visual system is considered well adapted to diurnal conditions [124]. Furthermore, green was chosen as the LEDs' color because the gerbils' spectral sensitivity has a peak at $\approx$500 nm [125], which corresponds to green.

The allocentric location of both loudspeakers was kept constant throughout the experiments, as well as the identity of each loudspeaker (rewarded loudspeaker) in the training phase. However, for some animals during the test phase, the identity of the loudspeakers was swapped for some trials (catch trials - see Section 3.1.2).

Choice of spatial cue – ITDs

The neuronal representations of ITDs and ILDs at the level of the AC are only partially independent of each other [126, 127]. Due to the high complexity of the experiment and posterior analysis and to reduce co-varying cues, the scope of this study was therefore limited to one of the sound localization mechanisms.

Understanding the cocktail party effect – a phenomenon in which people can selectively listen to a conversation – has been a longstanding objective of auditory neuroscience. Localizing the conversation in space is essential to solve this problem [128]. Additionally, during human speech the power is concentrated at low frequencies ($<$4 kHz) [3] where ITDs ($<$2 kHz) are mainly used; therefore, we focused our investigation on this sound localization mechanism. Moreover, we often move our head or inclusively our body when listening to a conversation, thus making this sound localization mechanism relevant to be studied in the context of self-movement.

Choice of animal model

Despite not vocalizing in the ITD regime, gerbils use ITDs to localize low frequency sounds ($<$1.5 kHz) [129]. This is in contrast to most rodents which only hear in the high frequency regime and therefore use ILDs for sound localization. Furthermore, gerbils were already shown to be able to report localization of low frequency sounds masked in noise [130] and lateralization of sounds based on ITDs [129]. The acuity in localizing low frequency sounds in the horizontal plane was determined to be 14°, which corresponds to a 21 μs ITD [131]. Consequently, for their small size, for being easily trained and for having a similar audiogram to that of humans, gerbils were considered the adequate model organism to answer our proposed questions.

2.1.2 Auditory SIT in the frequency version: aSIT$_{freq}$

In aSIT$_{freq}$, the property of interest that was changed with the position of the animal was the frequency of the played sound, since we were interested in assessing frequency-change detection. A trial was started at a random time within 15 s of the end of the previous one. Afterwards, the active loudspeaker – always the same throughout all sessions – started playing the background pure tone (20 kHz) until the animal entered the randomly located target island, which triggered a change to the target frequency (660 Hz). Similarly to aSIT$_{loc}$, the animal had to remain within this island for 6 s to be rewarded. Furthermore, also in this version of SIT the sounds lasted 57 ms and were repeated at a rate of 4 Hz being roved 5 dB in amplitude. The two major differences in relation to aSIT$_{loc}$ besides the property of interest were the non-existence of an initiation platform (given that we did not intend to implant animals in this frequency task) and the implementation of the final values of island size and sit-time since the very beginning of training (compare to Section 2.1.5). Perhaps the combination of both aspects together with the higher salience of a change in frequency allowed a much shorter training time. The animals were already performing above chance at the end of the first session.

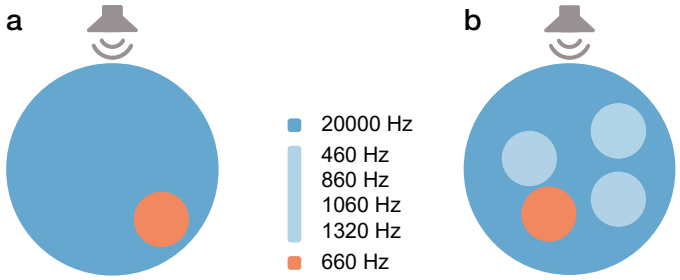


Figure 2.2 – Schematic representation of aSIT$_{freq}$ in the **a** single island version; **b** multiple island version. Figure adapted from [117]

To investigate frequency perception and how specifically the animals associated a frequency with reward, we next introduced the multiple island aSIT$_{freq}$. In this version, three extra islands corresponding to different frequencies (460, 860, 1060 and 1320 Hz) were simultaneously present in the arena. Also here, the position of the islands was pseudo-randomly distributed and differed between trials. Importantly, only the target frequency (660 Hz) was rewarded and terminated the trial when the animal remained in it for longer than the sit-time (6 s).

The training in this version of SIT was performed by Dardo Ferreiro and Paula Gundi, whereas I was responsible to program the training script and for analyzing the data. These experiments are thoroughly described in [117] (see appendix for complete paper).

2.1.3 Behavioral setup

The behavioral experiments were conducted in a custom-made setup (Fig. 2.3), which included an elevated circular arena with 92 cm diameter inside a sound-attenuating chamber. The wooden floor of the arena was painted with matte black paint to minimize light reflections that could compromise the online tracking of the animal. The arena was surrounded by perforated metal black walls (height: 16 cm) to minimize possible wave interference related to sound reflections on the walls. To prevent gerbils from jumping out of the arena, PVC walls covered with matte black vinyl sticker were firmly fixed on top of the perforated walls up to a 75 cm height around the arena with the exception of a section in which the PVC wall was detachable to allow the experimenter to easily access the interior of the arena. To achieve homogeneous lighting conditions, essential for a good tracking performance, a white light LED-stripe was glued to the top of the PVC walls covering the full 360° around the arena. This minimized shadows, which would difficult the process of tracking the animal's position.

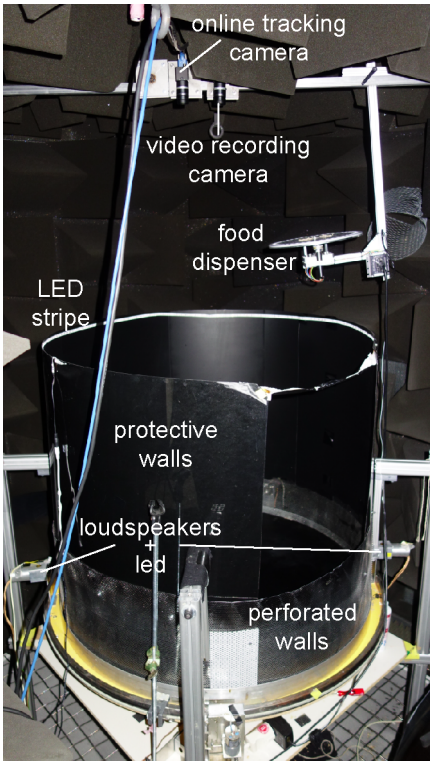


Figure 2.3 – Behavioral arena

Both loudspeakers (Aurasound NSW1-205-8A 1" Extended Range) and corresponding green LEDs were located just outside of the perforated walls (∼5 cm from the wall), opposite to each other. The loudspeakers were calibrated in a way to have similar intensities for all the frequencies of the complex sounds and equal for both loudspeakers when measured at the center of the arena.

The initiation platform was located peripherally near the wall and equidistant to both loudspeakers. This circular structure was ≈1 cm high and had a 12 cm diameter, similar to the length of a gerbil's body.

A custom-made food dispenser located 1 m above the arena was used to reward the animal (half of a sunflower seed or 20 mg, TestDiet LabTab AIN-76A). The dispenser consisted of two circular metal plates, the lower one was fixed and had one small circular opening, whereas the upper one had 45 similar openings, one of which aligned with the one in the lower plate allowing the pellet to be dropped when the upper plate was automatically rotated. This system allowed for a maximum of 44 pellets to be dropped consecutively without the need to restock. Consequently, most of the behavioral sessions corresponded to either 44 or 88 successful trials depending on whether the dispenser was in the meantime refilled.

Two video cameras (FL3-U3-13Y3M-C, Point Grey Research Inc.) were fixed to the ceiling of the chamber 1.3 m distant from the floor of the arena and approximately centered in relation to it, allowing the visualization of the whole arena. One of the cameras was used for tracking the position of the animal during the experiment, whereas the other recorded the video used for *post-hoc* analysis. The 1280×960 px grayscale images captured by the former were acquired at a frame rate of 4 fps, since they were used to calculate which stimulus should be presented to the animal and were therefore associated to the stimuli presentation, which occurred every 250 ms. The frame rate used for video recording was either 15 or 20 fps, since a better time resolution was required for the analysis of the gerbil's orientation and behavior.

Behavioral tracking

The custom-made online tracking scripts were developed in MATLAB. The main script terminated after a determined number of seeds or a certain time period — whichever condition was first fulfilled — both variables set at the beginning of the session. The possibility of manual termination was also implemented for an unplanned situation.

Before each session three background images were captured: one with the target loudspeaker-associated LED on, the following with the background loudspeaker-associated LED on and at last, one with exclusively the LED-stripe on. After image acquisition, one of these background-images was subtracted from the newly acquired frame. The appropriate background-image was chosen to match the situation at each moment: either the image without loudspeaker-associated LEDs on was used for the time periods between trials, or the image with the target/background loudspeaker-associated LED on when the animal was in/outside the target island during trials. This was necessary because if the lightning of the background-image had not perfectly matched the one of the newly acquired frame, these different lightning conditions would have made the tracking less reliable.

After background subtraction, the resulting image was filtered to eliminate the area outside of the circular arena and then it was luminance-thresholded and converted into a

binary image. The value for the threshold was chosen so as to maximize the area covered by the animal while still minimizing the inherent image acquisition noise.

Often the fur of the gerbil contained darker areas with similar luminance to the background, which resulted in holes or the splitting of the area corresponding to the animal in two or more disconnected components. To fill in the holes and/or connect adjacent disconnected components, a morphological closing operation was applied to the whole image with a 9 px radius disk as the structuring element. Subsequently, the image was morphologically opened with the same structuring element. This operation resulted in the removal of most of the noise from the image as well as of the animal's tail, whose position would have undesirably influenced the tracked position of the animal. All objects (i.e. adjoining horizontal, vertical and diagonal pixels with value 1 were considered the same object) were detected and the one with the largest area was considered corresponding to the gerbil. The centroid of this object was then relayed as the position of the gerbil.

The main script was divided in two main sections, each corresponding to a while-loop: the first ran during the inter-trials and the second during the trials. During each inter-trial loop, the position of the animal was determined as explained in the previous paragraph. Afterwards it was verified whether the gerbil was on top of the initiation platform and if so, whether for longer than 1 s — if that was the case, the program would go out of the loop and continued into the next section, otherwise the process was repeated. This section was not bound to the stimulus presentation and therefore an image was acquired at a rate as fast as a loop was processed (7.5Hz).

The target island position was then pseudo-randomly chosen from a uniform distribution as explained in appendix A and not overlapping with the initiation platform. This non-overlap was chosen due to the behavioral bias of gerbils, which during a trial spent more time on the initiation platform than in all other areas.

At the beginning of the trial section, the target associated-LED was turned on as well as another LED whose terminals' voltage drop was read by an Arduino Uno connected to another computer. This signal triggered the video recording via Bonsai [132] which lasted until the LEDs were turned off at the end of the trial.

During each trial loop, after determining the position of the animal as previously described, it was verified whether this position coincided with the target island; if that was the case, a sound from the target loudspeaker would be played, otherwise that same sound was played from the background loudspeaker. Subsequently, a waiting period was implemented to fill in up to the 250 ms desired time interval between sound onsets, ensuring jitter of the sound could not be used as a cue to finish the task.

The estimated time difference between image acquisition and the stimulus being played is 53 ms, which corresponds to a maximum animal displacement of 3 cm (considering the upper bound of the speed distribution — see Fig. 3.9).

At the end of every loop, all the image frames in the memory buffer were deleted to ensure that the processed image frame, which would result in the stimulus during trials, was the most recent one and therefore that the actual (and not the past) position of the animal translated into the correct loudspeaker being played.

In case the animal was in the target, the time that condition was fulfilled was calculated and if longer than the pre-determined sit-time, the motor of the food dispenser would be activated and the trial considered complete. Given that at the beginning of the training the sit-time was very short, the same sound was played during another 2 s (8 repetitions) after the correct termination of a trial, for the animal to get exposed to sound being played from the target loudspeaker.

As soon as the time inside the trial while-loop reached the maximum trial duration, the trial was terminated and a noise sound (low-pass filtered below 1.5 kHz) was played for 10 s by the target loudspeaker, during which a new trial could not be initiated.

At the end of each trial, the information on the position of the gerbil, timestamps, which loudspeaker was active at each time point and general information on the arena, sit-time and target island were recorded for posterior analysis.

2.1.4 Animal model: Mongolian gerbils

All experiments were performed exclusively on male Mongolian gerbils (*Meriones unguiculatus*) as a result of being larger and heavier than females and consequently being able to carry more weight on their heads, corresponding to the implant (2 g), headstage (3 g) and battery (6–10 g). Electrophysiological experiments were conducted on 5 animals, while 9 other animals also took part on the behavior experiments.

The gerbils were from the wild-type breeding colony of the Biocenter of the Ludwig-Maximilians-Universität München. Animals used in behavioral experiments were housed in groups of 3 to 4 males in a 60x40x25 cm transparent cage with nesting material such as wood shavings as bedding, a small plastic opaque house, and paper, cardboard and cotton pellets as environment enrichment. After the implantation surgery, the animals were kept in individual cages to prevent others from damaging the implant.

The animal facility room where the cages with the animals were stored was kept at a temperature of 22.4°C and 66% humidity with 12 h light/dark cycles. The experiments were conducted during the light phase of the cycle. The animals were required to be at least 8 weeks old to begin the training. Gerbils had unrestricted access to water with the exception of when they were in the setup. Food was provided as pellets *ad libitum* until the training phase started, after which animals were only allowed food as rewards for correct trials (half of a sunflower seed or 20 mg, TestDiet LabTab AIN-76A) and during weekends. The weight of every gerbil was measured daily to ensure it did not to drop more than 5% between consecutive training days and to maintain it within the desired range: 60–80 g.

All procedures were approved in accordance with the stipulations of the German animal welfare law (Tierschutzgesetz - AZ 55.2-1-54-2532-74-2016).

2.1.5 Behavioral training

To reduce the stress and increase the exploration behavior since the beginning of the training, I initially handled the animals for 3–5 days (≈45 min each session), during which they stayed in the animal facility room inside their cage. During this period, the animals got used to being held by humans and were introduced to sunflower seeds — which was not part of their diet until that point, and was from then on used as a reward. This process was followed by a one-day habituation period where they were transported to the lab and placed for 10-20 min inside the arena, where previously several sunflower seeds had been scattered.

Furthermore, the animals had to be taught how to self-initiate a trial. With this objective, during 4-5 sessions as soon as the animal stayed on top of the initiation platform for 1 s, a reward was administered and a harmonic complex sound was played from the target loudspeaker for 3 s.

When the animal started to go consecutively to the initiation platform, the aSIT$_{loc}$ task was introduced (as described in Section 2.1.1). At first, the target island was very large (42 cm diameter, ≈ 21% of the arena surface) and the sit-time required very short (0.5 s). While the target island was successively reduced (up to 25 cm diameter, ≈ 7% of the arena surface), the sit-time was simultaneously increased (up to 6 s - 8 s for gerbil 3) until the final parameters were reached after ≈ 15 sessions.

2.2 Electrophysiological recordings

2.2.1 Tetrode bundle construction

Each tetrode bundle which was later implanted in the auditory cortex consisted of four tetrodes glued together (Fig. 2.4). Each tetrode was composed of a very thin insulated tungsten wire (12.7 μm diameter, tungsten 99.95%, California Fine Wire), which was twice folded, mechanically twisted and the tips cut with a sharp scissor. The loose tips were then briefly burnt to remove the electrical insulation. This resulted practically in four very thin wires twisted around each other on the end that entered the brain (twisted part $\approx$16mm) and 4 non-twisted non-isolated loose wires on the other end which were then inserted and connected (silver conductive paint, SCP03B, Electrolube) to a custom-made printed circuit board (PCB) with an Omnetics connector (Axona) previously attached to a lightweight microdrive (250 μm/turn, Axona). Later, a ground wire was soldered to the PCB. The loose wires together with the PCB and the connection of the ground wire were covered with dental cement (Paladur, Kulzer) for protection. The twisted part of the tetrodes which was not inserted into the brain was covered and protected by an inner cannula. The later was surrounded by an outer cannula, which during the surgery was cemented to the skull and allowed the inner cannula to move inside it together with the tetrodes whenever the depth of the recordings was changed. On the day previous to the surgery, the twisted tips of the tetrodes were cut with very sharp scissors to the desired length (5mm after the end of the outer cannula) and gold plated (NanoZ™, White Matter LLD; ADPT-nanoZ-NN-16, Multi Channel Systems; Non-Cyanide Gold Plating Solution, Neuralynx) to reach the desired impedance of 100–150 kOhm (at 1 kHz). On the day of the surgery a drop of DiI was applied to the tip of the tetrodes (this step was not performed in every surgery) for posterior reconstruction and the implant was kept in the dark until the implantation.

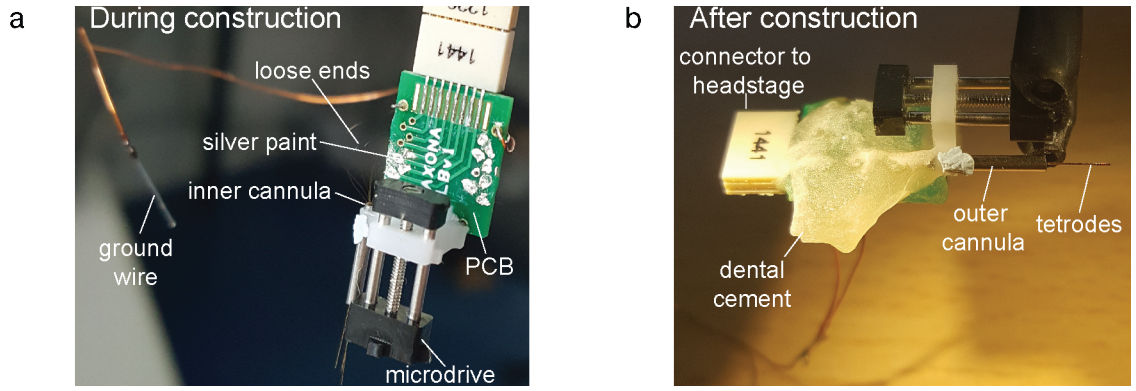


Figure 2.4 – **Images of an implant tetrode bundle a** during its construction; **b** after its construction and before the implantation.

2.2.2 Surgery

Initially a drop of analgesic (meloxicam) was placed on top of a sunflower seed, which was then given to the gerbil. Afterwards, the gerbil was anesthetized with an intraperitoneal injection of a mixture of midazolam (7.5 mg/kg), metedomidin (0.15 mg/kg) and fentanyl (0.03 mg/kg). This mixture was subcutaneously re-injected (one third of the initial dosage) every 90 min to maintain the level of anesthesia, together with 0.5 ml of saline to keep the animal hydrated. Additionally, the paw pinch and eye lid reflexes were regularly verified to ensure an adequate depth of anesthesia. To maintain its body temperature, the animal was placed on top of a microwave-heated heating pad while its head was shaved and disinfected. In the meantime, the eyes were covered with an ophthalmic gel (Thilo-Tears SE, Alcon Pharma Gmbh), which kept them protected and hydrated. One leg was shaved to enable the monitorization of the heart rate and of the oxygenation level through a pulsoximeter during the entire surgery. Bupivacain (150μl) was next subcutaneously injected below the skin on top of the skull along the midline, where the skin would be cut, and below the skin near the ears because of the pressure that the ear bars usually inflict in that region.

Subsequently, the animal was transferred to the stereotactic apparatus where its head was fixed through ears and bite bars and its temperature read by a rectal thermometer and kept constant at 37°C using a feedback controlled electric heat pad (Harvard Apparatus). The animal was provided with extra oxygen during the whole surgery (except during drilling, for safety reasons) since fentanyl causes respiratory depression. To maintain the environment sterile, all the reusable instruments and consumables had been previously autoclaved and I wore hair net, gown, mask and gloves throughout the surgery. After cutting the scalp through the midline with a scalpel, some skin was removed on the side of the implant, making sure to leave enough to cover the borders of the implant at the end of the surgery. With a bone curette, the periosteum was removed and the left temporal muscle detached and retracted downwards with the help of biodegradable gel foam and skin retractors, ensuring a dry lateral bone surface. The neck muscle was detached from the occipital bone where the ground screw was inserted, given that this muscle can create artifacts in the electrophysiological recordings. Afterwards, 35% phosphoric acid (iBOND® etch 35 gel, Kulzer) was carefully applied to the bone in order not to touch any skin or muscle and thoroughly washed away. Following stereotactic alignment, the most rostromedial point of the craniotomy was marked 4mm rostral and 4.5mm lateral to lambda and a 3×3 mm craniotomy contour was drawn with a surgical pen. A basin was then built of UV cured two-component cement (3M™ RelyX™ Unicem 2 Self-Adhesive Resin Cement cured with BA Optima 10 LED curing light, B.A. International) around the left lateral bone to avoid saline to laterally drain from the craniotomy and thus prevent the brain from drying.

Before starting drilling, 100 μl lidocain 2% was applied on the skull. At least two small structural screws were fixated to the left frontal and right parietal bones and the ground screw to the occipital bone, gently touching the brain. After performing the craniotomy

and durotomy, the tetrode bundle was lowered by an automatic micromanipulator (Scientifica IVM Triple) at a velocity of 2 μm/s either vertically from 2.9 mm rostral and 6.2 mm lateral to lambda up to a depth of 0.9 mm or at a 25° angle from 3 mm rostral and 4.6 mm lateral to lambda up to a depth of 1.2 mm. Next, the craniotomy was filled with an antiseptic biologically inert lubricant (KY-Jelly). The craniotomy, the drive base, the outer tetrode-protecting cannula and the screws were then carefully covered with dental cement (Paladur, Kulzer), making sure not to let it drain to the skin or muscles. If necessary, a suture was performed to the skin on the most caudal part of the implant. After quickly soldering the ground screw wire to the implant, the anesthesia was antagonized with a subcutaneous injection of a mixture of atipamezol (0.375 mg/kg), flumazenil (0.4 mg/kg) and naloxone (0.5 mg/kg), together with 0.5 ml of glucose 5% for hydration and faster recuperation and meloxicam (0.2 mg/kg) as an analgesic.

The animal was subsequently transferred to an infrared light-heated surgery-prepared cage without a grating, since the implant could get stuck in it. The cage was covered with surgical tissue instead of wood shavings, and soft food and water were available.

During the five days following the surgery, the animal was observed twice a day and the implantation score-sheet filled in daily. Antibiotics (7.5 mg/kg, enrofloxacin) and analgesic (20 mg/kg, metamizol) were orally administered by placing a drop on top a sunflower seed which was fed to the gerbil. During this recovery period, the animals were not trained and had food and water *ad libitum*.

2.2.3 Electrophysiological recordings during task performance

After allowing 5 days for the animal to recover from the surgery, the recording of the electrophysiological data started. Initially, a battery (W2100-B-200mAh-BB or W2100-B-300mAh-BB, depending on how long the recordings were expected to last) was attached to the wireless headstage (W2100-HS16, Multichannel Systems), which was then plugged to the matching connector of the gerbil's implant. The physiological signals were at first 100× amplified between 1 Hz and 5 kHz and digitized (16-bit resolution) in the headstage (W2100-HS16, Multichannel Systems). The headstage then wirelessly transmitted the signals to the receiver (W2100-RE-AO, Multichannel Systems) at a sampling rate of 25 kHz. The receiver passed the physiological information to the interface board (MCS-IFB 3.0 Multiboot, Multichannel Systems), which after integrating the signals with the ones from the analog and digital inputs, sent them to the computer where they were recorded via a commercial software (Multi Channel Experimenter, Multichannel Systems). Simultaneous with the onset of sound presentation, a 9 ms signal was sent via the sound card to the analog input of the interface board, allowing the synchronization of the physiological recordings with the sound presentation. This signal was positive when the usual target loudspeaker got activated and negative whenever the usual background loudspeaker got activated (in the case of swapped trials, the signal's sign still encoded the identity of the loudspeaker, and thus, in these cases, positive meant background and negative target). Together with the signal that was sent to start and end the video recording, a digital sig-

nal was transmitted to the interface board indicating the beginning and end of the trial, which was later used to align the video information.

Electrophysiological information was recorded every day until the signal deteriorated or no more auditory information was encoded in the data. At the end of a recording session, the tetrodes were lowered to allow stable recordings the following day. The tetrodes were lowered 1/8 or 1/16 of a turn (which corresponds to 31 or 16 μm, respectively) if the tetrodes were in the desired area and showed auditory responses; otherwise the tetrodes were lowered 1/4 or 1/2 of a turn (which correspond to 62.5 or 125 μm), depending on how deep the tetrodes were located. To identify whether there was any auditory response during a certain session, I heard the high-pass filtered signal from one tetrode (MC Rack, Multichannel Systems) while presenting repetitive noise or pure tones at a chosen frequency; if there was an auditory response, I could hear the multi-units activity with the same repetition rate as the sound presented.

2.2.4 Spike sorting

Initially, the raw electrophysiological signals were high-pass filtered above 300Hz to eliminate slow fluctuations. Next, a common median referencing [133] was performed to remove common information among all channels and large artifacts – which could interfere with the spike sorting algorithm – were removed. The signals were low-pass filtered below 5 kHz and fed to a spike sorting algorithm based on template matching (Kilosort [134]). Afterwards, the automatically sorted spikes were manually inspected and the corresponding clusters refined with the graphical user interface phy [135] (`https://github.com/cortex-lab/phy`). After removing noise and performing the required merging and splitting of clusters, a unit was further considered whenever the waveform was distinct and the refractory period clear. Only the units with an isolation distance larger than 20, more than 200 spikes and less than 2% of the spikes within the 2 ms refractory period were considered single units [136, 137] – often referred to as neurons throughout this monograph. All the units that passed the manual curation but not the requirements for being single units were considered multi-units.

2.2.5 Identification of the recorded brain area

After overdosing the animal with a 350μl intraperitoneal injection of pentobarbital sodium, the animal was perfused with 4% paraformaldehyde (PFA) and the brain carefully removed and stored in PFA in the fridge. Afterwards, the brain was twice washed in PBS and the frontal part cut with a sharp blade creating a plane surface that was later glued to a holder which could be fixated to the vibrotome's tissue bath container. Before gluing the brain, liquid agarose was poured on top of it and let solidify, creating a structure that prevents the brain from changing its shape during cutting. After cutting the target area in 70μm thick slices, these were placed individually with a paintbrush in microwells filled with a PBS solution with 1:300 green Nissl concentration (NeuroTrace™ 500/525

green fluorescent Nissl stain). The microplate was left overnight in the shaker at 50 rpm, after which the slices were washed 3× in PBS with 20 min interval in between washes. The slices were carefully placed in gelatin-coated slides, preserved with Vectashield® and sealed with transparent nail polish.

The stained brain slices were then compared to the gerbil brain atlas for confirming the location of the recording sites (Fig. 2.5), reassuring the recordings were performed in the A1. However, the dorsoposterior field (DP) is not represented in the brain atlas. The DP borders the A1 in its most caudal part, corresponding to a low frequency region, our aim-frequency. According to literature [138], this border lies at around 2 mm rostral to lambda – the most caudal implantation performed was at 2.9 mm rostral to lambda – making me confident the tetrodes were implanted in the A1 and not in the DP. Furthermore, the latencies in the DP are much longer [139] than the ones here described (Section 3.3.1).

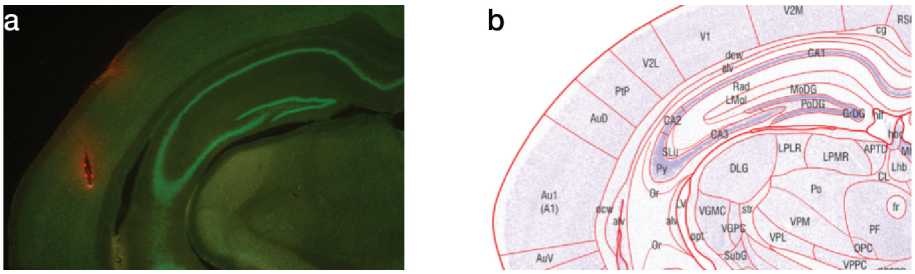


Figure 2.5 – **Comparison for the reconstruction of the recorded brain area** between the **a** stained slice with recording site in red; and the **b** gerbil brain atlas [140].

2.3 Offline video tracking

2.3.1 Initial tracking

The sampling rate of the online tracking script was much lower than the recorded videos. Furthermore, the orientation of the gerbil was not calculated in the online tracking to allow a fast online determination of the position of the gerbil. However, the orientation of the animal is an essential parameter to calculate spatial tuning curves. Therefore, the recorded trial videos were offline analyzed with custom-made tracking scripts that I developed in Python (with OpenCV library [141]).

At first, all videos were filtered with a circular mask to show only the arena region. Subsequently, one or two long videos where the animal was active in a given session were chosen for the background-image calculation. A gaussian mixture-based background/foreground segmentation algorithm (BackgroundSubtractorMOG2 based on [142] and [143]) was applied to these videos with a low learning rate, creating the background-image to be extracted from every frame of all videos of the session. This algorithm was not directly applied to every video because the shape of the gerbil would be integrated into the background-image for videos with many similar consecutive frames, and the tracking would break down. Therefore, the algorithm was applied to a couple of videos where the animal did not succeed in the task (as reporting the target involves a very long sitting time and corresponds to many identical frames) and was active. The learning rate was low to maximize the number of frames included in the calculation of the background-image.

After background subtraction, the shadows were eliminated and the image was morphologically closed and opened (for details see Section 2.1.3). Next, the largest contour (findContours algorithm based on [144]) of the morphologically closed image (where the tail is still depicted) was determined and its centroid C_{closed} calculated. An ellipse was fitted to the largest contour of the morphologically opened image (without tail) and the end points of its major axis determined. However, which of them corresponded to the nose of the animal was at this point still unknown and likewise the animal's orientation. If the distances of each of these end points to the centroid C_{closed} were very different from each other, it meant that the tail was closer to one of the end points and therefore the nose of the animal corresponds to the farthest point from the centroid C_{closed}. However, if the distances were similar, the tail could not have been detected or laid next to the body of the animal and therefore could not be used to determine the orientation of the animal. In this case, the orientation (from the two possible) of the fitted ellipse that minimized the orientation difference to the previous frame was chosen, since spatio-temporal continuity was predicted and it was not expected that the animal rotated more than 90° in consecutive frames.

At the border of the arena, the animals tended to scratch the floor while not being engaged in the task. This behavior created artifacts in the neuronal physiological recordings, as well as hindered the orientation tracking, since often in these situations their head was

positioned very low making them look circular to the algorithm. Furthermore, when the animals were at the border in front of a loudspeaker (located 5 cm away from the perforated wall), the sound was approximately 6dB louder than when they were 5 cm away from the border and those large differences in sound intensity could interfere with the experimental results on spatial tuning. Thus, the frames in which the animal was closer than a body length (≈10 cm) to the border were not considered for posterior analysis and only when the animal was located within the central 70 cm diameter circle, the frame was further analyzed. Since the temporal continuity of the animal's orientation could therefore not be assured at the borders, each first frame after the animal reentered the central circle was flagged for me to later confirm the orientation. Also frames where the orientation difference between two consecutive frames was larger than 90° were flagged for posterior manual inspection (Fig. 2.6).

2.3.2 Visual inspection of flagged frames

Visual inspection was a compulsory step and performed exclusively for the flagged frames (Fig. 2.6). The image of the arena was shown superimposed with the ellipse that was previously fitted to the gerbil's contours and with an arrow indicating the determined orientation of the gerbil in that frame. I then decided whether the tracking was correct. If so, the frames following the flagged frame remained unaltered until the next flagged frame or until the animal left the central circle. If the tracking was incorrect, I could either decide whether the orientation had to be flipped (180° rotation) or whether the ellipse had to be manually selected. In the first case, all the orientations were flipped until the next flagged frame or until a frame where the animal was no longer in the central circle, which I then evaluated to confirm the flip. In the case the ellipse had to be manually selected, the following frame also had to be visually inspected and iteratively so until one of the frames was either correct or simply flipped.

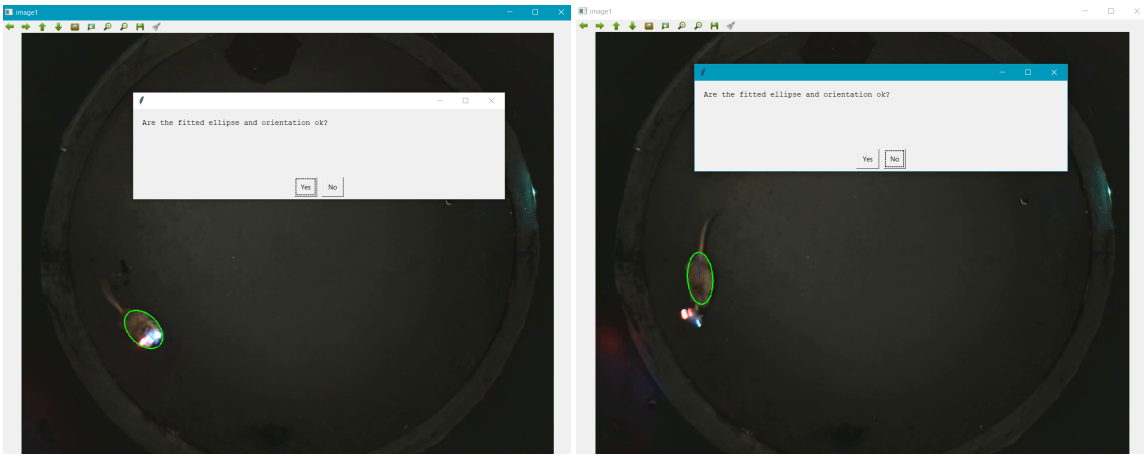


Figure 2.6 – **Examples of flagged frames** because the animal reentered the central circle (left panel) or because there was a break in the orientation continuity (right panel)

2.3.3 Visual inspection of entire videos

Whenever I was unsure about the quality of the tracking, I could next verify a complete video (Fig. 2.7). By scrolling through all the frames, I observed whether there was any mistake in the tracking and to correct it either by flipping the orientation, selecting a new ellipse or copying the ellipse and corresponding orientation from previous frames (in the case the mistake happened when the animal was immobile).

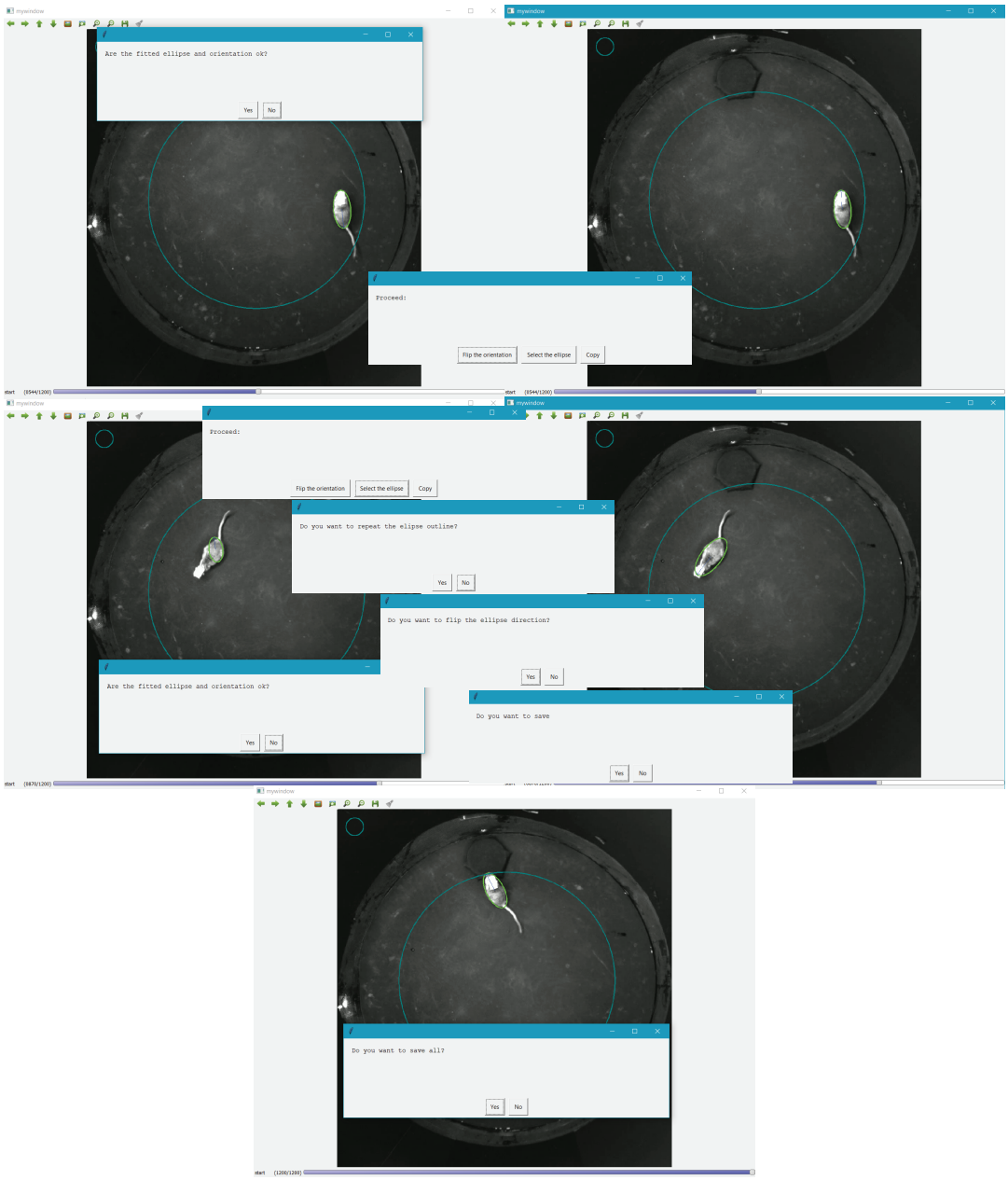


Figure 2.7 – **Manual tracking correction** in a situation where the orientation must be flipped (top panels), where the ellipse must be redrawn (middles panels), when the tracking correction is complete and the data must be saved (lower panel).

2.3.4 Error estimation of the gerbil's orientation

Despite the tracking algorithm and posterior visual inspection, some frames exhibited errors in the orientation of the animal for various reasons (Fig. 2.8). The parallax error was mitigated by eliminating the borders from the analysis; however, not completely.

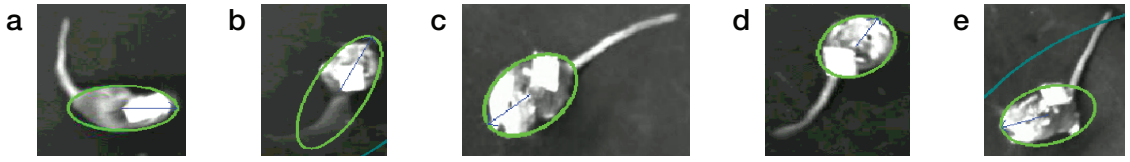


Figure 2.8 – **Tracking problems** sometimes occurred when **a** the animal's body bent; **b** the tail's movement caused motion blur which was integrated into the tracking; **c** the animal stood while looking to the side **d** the animal stood causing the ellipse to be nearly circular, thus impairing the determination of the orientation; **e** the camera was not exactly on top of the animal, causing a parallax error.

To investigate how large the orientation errors were, I manually determined (blindly) which orientation the animal had in 100 random frames (20 videos from several sessions of 3 different animals) and compared to what was determined by the offline tracking procedure. The median orientation error was relatively small: 6°; yet the distribution extends up to a maximum of 70° (being above 45° solely in 3 frames – Fig. 2.9).

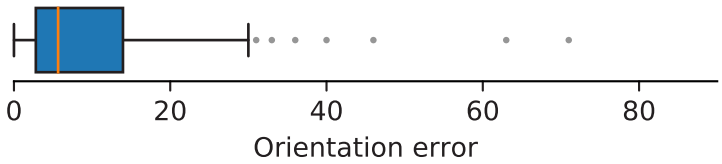


Figure 2.9 – **Orientation error of the offline tracking procedure**

Tracking pinnae movements was not considered necessary for this experiment since the use of low frequency sounds greatly reduced their influence on the neuronal responses: the wavelength of the sounds were at least one order of magnitude larger that the gerbils' pinnae and therefore reflection did not occur [145, 146].

2.3.5 Calculation of Bouts

During the exploration behavior, the animals moved in bouts interleaved with periods when they stayed still. Both the angular as well as the translational speeds were calculated to determine whether the gerbil was moving. Since there was intrinsic noise in the video recording and subsequent tracking, before calculating the speeds, a running median was applied to the angle or position, therefore reducing the noise and possible single peaks that could appear in the speeds. As seen in Section 2.3.4, the minimum detectable change in angle was around 5°. Furthermore, by visual inspection of the videos, the minimum detectable position change was 0.5 cm. Therefore, to establish whether the animal was

rotating or moving in space, thresholds were applied on the angular and translational speeds corresponding to these minimum detectable changes between consecutive frames.

A bout was bounded by frames without movement and was defined as the consecutive frames in which either the angular and/or the translational speed crossed the threshold while the gerbil was located in the central circle of the arena (70 cm diameter). The bouts were comprised either of rotatory behavior (in the azimuth), of a translation or of a combination of the two.

A stop period was bounded by bouts of movement and was defined as the consecutive frames in which both the angular and the translational speed did not cross the respective thresholds while the gerbil was located in the central region.

Note that the combination of rotation with translation can comprise bouts where the animal rotated at the beginning of the bout and then walked in a straight trajectory, bouts that corresponded to smooth curves where the orientation of the animal was constantly changing, and bouts in which the animal rotated while giving one step only.

The Markov chain between the locomotion states was performed on all frames in which the animals were in the central region, separated by target and background areas.

2.4 Data analysis

Custom-made Python or MATLAB (Mathworks) scripts were used for analyzing the data.

2.4.1 Calculation of the behavioral performance level

The percentage of finished trials was calculated as a function of the relative time passed from the beginning of the trial. For the construction of the corresponding 95% confidence interval, a bootstrapping analysis was performed in which, at each time point, the percentage of finished trials was 1000 times calculated based on trials that were randomly drawn with replacement from all the trials of the session, with the number of trials as the size of the sample. To assess the behavioral performance level, the natural pauses in movement – which sometimes can be longer than the sit-time – must be considered since the animal could thus complete a trial by chance. For that, the analysis of the typical behavior of the gerbil does not suffice, since the dynamic patterns of locomotion can be influenced by the task itself. This paramount chance level was hence calculated using the real trajectories of the animal.

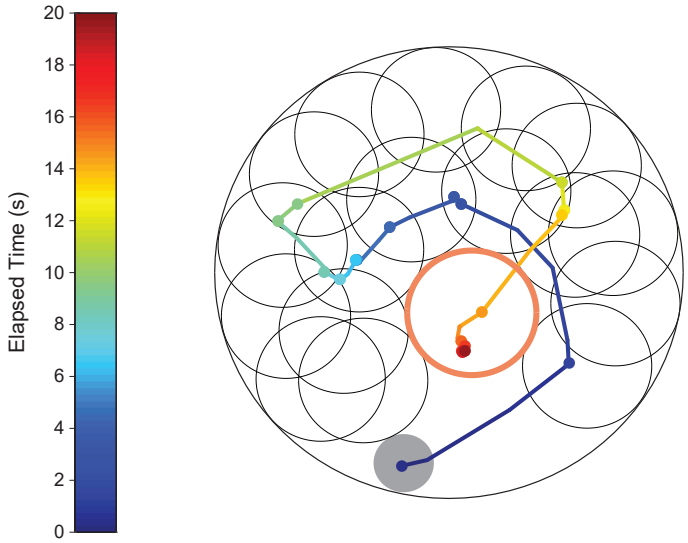


Figure 2.10 – Depiction of the surrogate island analysis for computing the performance level in the task The filled line corresponds to a real trajectory during a trial, color-coded with the time that had passed since the beginning of the trial. The animal started in the initiation platform (filled gray circle), each colored dot corresponds to the position of the animal every 1s and the gerbil successfully ended the trial in the target island (open orange circle). In between, it passed by several surrogate islands (open black circles – used for the chance level calculation – only a few of the 1000 randomly distributed islands are depicted) but it did not stay the 6s in any of them. Figure adapted from [117]

For each trial (offline, *a posteriori*), 1000 surrogate (non-real) islands were randomly distributed across the arena (not overlapping with the target one nor with the initiation platform – Fig. 2.10). The percentage of these in which the animal would have stayed for

the sit-time (6s) and therefore finished the trial was determined. For the calculation of the median chance performance and corresponding confidence interval at each time point, a bootstrapping analysis was applied on the surrogate data from trials that were at the time point of the calculation still incomplete – and consequently the target island had not yet been found – and from trajectories of already completed trials but where the animal had stayed longer than the sit-time in the surrogate island prior to finishing the trial. A trial that was correctly finished before the considered time point relative to the beginning of trial and in which the animal had not found the surrogate island cannot be used for the calculation of the chance level for that and subsequent time points, since it is unknown whether the animal would have stayed the sit-time in the surrogate island had the trial been as long as the considered time points. Therefore, for this chance level bootstrapping, it was 1000 times computed the success percentage based on the trajectories that were randomly drawn with replacement from the aforementioned trajectories, with the total number of trials as the size of the sample.

2.4.2 Peristimulus time histograms (PSTH)

The peristimulus time histogram was calculated at several steps of the analysis of the data, either for the global description of the response modulation with sound, or separately for each loudspeaker or even more specifically for different angles and states of movement. The method used for all these situations was identical and was based on bootstrapping data. From the sounds of which I wanted to determine the neuronal response, the algorithm sampled 500 times with replacement, with the total number of sounds as the sample size. For each cycle, it was calculated the histogram of the time after sound onset of occurrence of each spike that fired within 250 ms of the sound onset with binning width 8 ms. Later, the median and 95% confidence intervals were determined for the total of bootstrapping cycles normalized by the total number of analyzed sounds.

For calculating the spontaneous (engaged) firing rate, the PSTH was calculated for the last second before the beginning of a trial, during which the animal was on the initiation platform. For each trial, 4 time points were randomly picked from this 1 s time period and the spikes that were within 250 ms of those instances were assigned to it. The PSTH was then calculated on the $4 \times \#trials$ total time points and corresponding spikes as described in the previous paragraph.

At the end of the session, for 3 gerbils an extra 2 min recording time period was added to determine the spontaneous firing rate when the animal was not in the task (out-of-task). This period was binned in 250 ms time bins. All the spikes within each bin were assigned to it. The PSTH was calculated as previously described for all the time bins during which the animal was not moving.

Mean firing rates were calculated as the mean of the PSTH across all time bins.

For the comparison between the first and last 2 s inside the target island and whenever the animal wrongly left it, only situations in which the animal stayed at least 2 s in the island were considered. Furthermore, if the last 2 s overlapped in time with the first 2 s

(the animal stayed less than 4 s in the target island), only the non-overlapping part was used for the calculation of the last 2 s.

2.4.3 Latency and offset calculation

For the calculation of neuronal latencies, several methods were initially contemplated. Studies on anesthetized animals showed that the first-spike latency was the most informative cue for localizing a sound-source, inclusively more informative than spike mean counts [147]. However, in awake and behaving animals, there is considerably more background spontaneous spiking [79]. Therefore, considering the latency as the first spike after sound presentation was soon disregarded, as neurons have a baseline that would contaminate the result.

As an alternative, calculating a peristimulus time histogram would constitute a possibility, while for example considering the latency as the first bin that deviates from the baseline by more than one standard deviation. Here, once more, we faced a problem, as we had to decide on an appropriate bin width. Furthermore, this decision included a trade-off between resolution – for this, ideally the bin width would be as small as the resolution of the recording equipment – and fluctuations – which would be very large for a high resolution, as a consequence of too few spikes in each bin. Even if we found an approximate value to optimize this trade-off (such as by using Scott's or Freedman–Diaconis' rules), the latency would still be influenced by the bin width – as this would be fixed and could only be described by discrete values – and by the location of the bins – as the "true" latency could lie at the beginning or at the end of the bin.

To circumvent these problems and challenges, a Bayesian blocks algorithm was implemented [148] – a non-parametric method that identifies statistically significant variation in a time series to optimally segment the data. In the implementation of the algorithm, the false positive rate (p_0) is manually set and corresponds to the probability that a detected change in the time series is actually wrong. Therefore, the smaller this number is, the more conservative the algorithm and the fewer bin edges are created. Given that our data was actually circular (a sound follows the previous one in a periodic way) and that the algorithm prevents the creation of bins at the edges, as suggested in the original research article, a copy of the data was added at the beginning and at the end of our data. In practice this means that a spike that occurred at a time x after sound presentation, was added to the array as $[x - 250;\ x;\ x + 250]$, being 250 ms the inter-sound-onset interval. A unit was considered non-modulated if no edges were created in the interval [0:250] ms for a false positive rate of 0.05. If, on the other hand, edges were created, the degree to which the unit was modulated was analyzed by successively decreasing the order of magnitude of the p_0 until a minimum of 10^{-5}. The latency of the unit is considered as the first created edge in the interval [0:250] ms. This method has the advantage that the bin widths are not fixed and therefore the latency truly corresponds to the first created bin edge (Fig. 2.11).

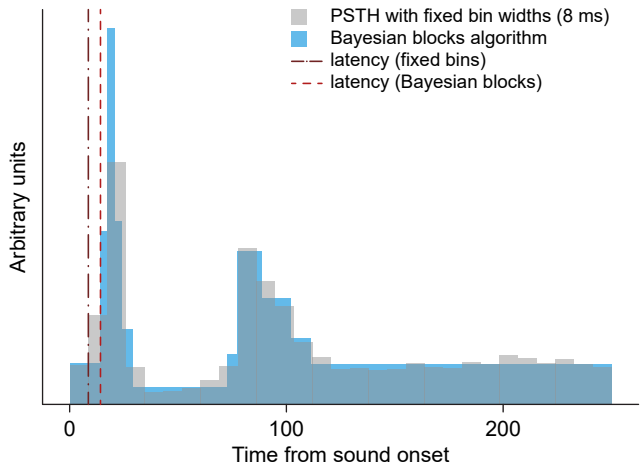


Figure 2.11 – Comparison of latency determination using the Bayesian blocks algorithm to determine the bin edges locations (blue, solid) or with a pre-determined size for every bin, in this case 8 ms (gray, translucent). The histograms were normalized to have the same area.

To determine whether the onset corresponded to an increase or a to decrease in the firing rate, the maximum displacement of the median values of the firing rate in the 15 ms after latency was compared to the baseline (last 50 ms before latency). If it was lower than the mean of the lower bound of the baseline's 95% confidence interval, the onset firing rate was considered to be decreasing. Conversely, if it was higher than the mean of the upper bound of the baseline's 95% confidence interval, the onset firing rate was considered to be increasing. Additionally, the firing rate in the time period between the onset and offset response (middle section) was also analyzed and its mean compared to the same baseline.

The same Bayesian blocks method was employed to calculate whether a unit had an offset. However, the analysis of whether new edges were created was limited to the interval [latency+50:latency+70] ms. In case an edge was created in that time interval, the maximum displacement of the firing rate in relation to the baseline in the 15 ms after the offset edge was compared to the baseline in the same manner as the onset. A unit was considered to have an offset response if it was significantly different from the baseline and if the sign of the displacement (whether the firing rate increased or decreased with respect to baseline) was different from the sign of the middle section response or, in the case it had the same sign, the maximum displacement during offset was larger than the one in the middle section.

Both the latency determination as well as the offset response were calculated using neuronal responses to all the sounds (excluding when the animal was not in the central region) from both loudspeakers. These were the values adopted for establishing the time periods used in all the spatial tuning analyses, independently of whether the analysis was limited (e.g. when the animal was not moving) or not – for the comparison between different conditions to be performed on the same time periods.

2.4.4 Video-frame to sound presentation to spike assignment

Each sound presentation was assigned to the video-frame that immediately preceded it and each spike was assigned to the sound that immediately preceded it. All the information (position, angle in the arena) about the gerbil in the video-frame was relayed to the sound presentation information. Depending on which loudspeaker was active, the egocentric sound-source-location (relative to the animal's body axis) was then calculated (Fig. 2.12) and later used for the spatial tuning analysis. Thus there is a correspondence between each spike and the position the animal had in the video-frame previous to the sound that evoked the spike.

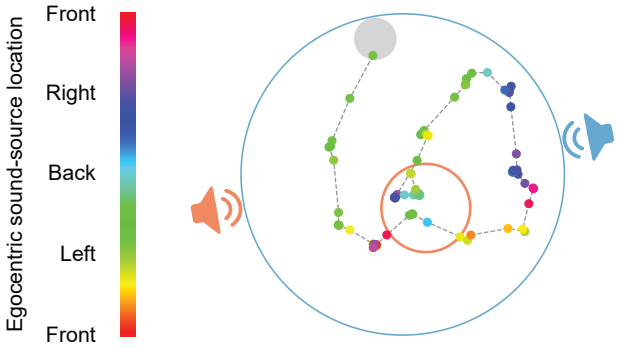


Figure 2.12 – Trajectory of the animal during a successful trial The colored dots indicate the position of the animal at the moment of sound presentation and corresponding loudspeaker angle.

2.4.5 Spatial tuning analysis

For the spatial tuning analysis, the 360° angular space around the animal was divided in 8 bins, each with 45°. This angular bin width is a compromise between being small enough – still allowing to determine spatial tuning – and large enough - including enough sound presentations at each angular bin, necessary for a meaningful statistical analysis. Furthermore, by being larger than the tracking error (see Section 2.3.4), it mitigates the potential small tracking errors. A minimum of 10 sound presentations per angular bin was a primary condition for the spatial tuning analysis. Additionally, for a unit to be spatially analyzed, the ratio between the angular bins with the most and the least number of sound presentations was limited to 10. The spatial tuning was calculated in three periods of the PSTH separately:

- Onset period: from latency until offset or, in case this was non-existent, lasting the duration of the sound (57 ms);
- Offset period: from the end of the onset period, lasting the duration of the sound;
- Late response period: until the response to the next sound presentation; it includes the time between the next sound presentation onset and the latency.

Furthermore, a minimum number of 20 spikes was required for each analyzed period.

A 1000× cycle bootstrapping method was implemented to calculate the spatial tuning. In each cycle, the presented sounds were chosen with replacement, with the number of sounds as the sample size. The 8 binned angular histogram was calculated for the corresponding spikes and normalized by the number of sounds at each egocentric location of the chosen loudspeaker. In other words, the algorithm determined the number of spikes per sound presentation that occurred when the chosen loudspeaker was at a particular location relative to the animal. The vector strength and corresponding direction tuning angle were then calculated for every bootstrapping cycle:

$\rho = \frac{\sum_{j=1}^{8} \alpha_j . e^{\imath \omega_j}}{\sum_{j=1}^{8} \alpha_j}$ where $\imath = \sqrt{-1}$, $\alpha_j = \frac{\#spikes\ in\ angular\ bin_j}{\#sound\ presentations\ in\ angular\ bin_j}$ and $\omega_j = \frac{2\pi\ middle\ of\ angular\ bin_j}{T}$.

The vector strength was calculated as $VS = |\rho|$ with $T = 360°$ and the corresponding direction tuning angle $ang_{dir} = \arg(\rho)$.

As some units revealed a more complex spatial tuning (namely encoded an orientation), the folded vector strength (in which responses to angles opposite to each other are summed [149]) was additionally calculated together with the corresponding orientation angle.

The orientation vector strength was calculated as $VS_{folded} = |\rho|$ with $T = 180°$ and the corresponding orientation tuning angle $ang_{ori} = \arg(\rho)/2$. Moreover, the direction closest to the largest response was chosen from the two possible directions an orientation could have.

From the bootstrapped data, the 95% confidence interval of the normalized response for each angular bin was calculated, as well as the 68% confidence interval of the vector strength and of the orientation vector strength. A unit was considered spatially tuned at a particular period if the vector strength (or the orientation vector strength, whichever was the largest) was larger than 0.2 and the lower bound of the 68% confidence interval of the vector strength (or orientation vector strength) was larger than 0.15. If there were less than 75 spikes, the lower bound of the vector strength was compared to a look-up table, to verify whether it was significantly tuned (since a large vector strength is more likely when calculated from a very low number of spikes [150]).

Units were considered untuned if the median vector strength was smaller than 0.15 in all three periods of the PSTH for both loudspeakers. The corresponding firing rate in response to the two loudspeakers was considered significantly different if there was no overlap in the 95% confidence interval of the calculation of the mean firing rate for each loudspeaker.

2.4.6 Neuron type analysis

After high-pass filtering all the channels above 300 Hz and common median referencing (see Section 2.2.4), all the spikes of a specific neuron were aligned in relation to the trough

time. The spike waveform considered for further analysis was the median value at a specific time point of all the aligned spikes in the channel with the largest response. After upsampling this median spike waveform – increasing 100 times the number of data points using a cubic spline interpolation to increase the resolution – several parameters were determined: full width at half maximum of the peak, full width at half minimum of the trough and trough-to-peak time. These were then clustered in 3 groups by using a k-means clustering algorithm. Based on literature [102], the two groups with the largest trough-to-peak time were combined corresponding to regular spiking neurons (RS), which are putative excitatory neurons (pyramidal neurons). The neurons with shorter trough-to-peak time correspond to fast spiking neurons and are putative inhibitory neurons (GABAergic, thought to be basket and chandelier cells [151–153]).

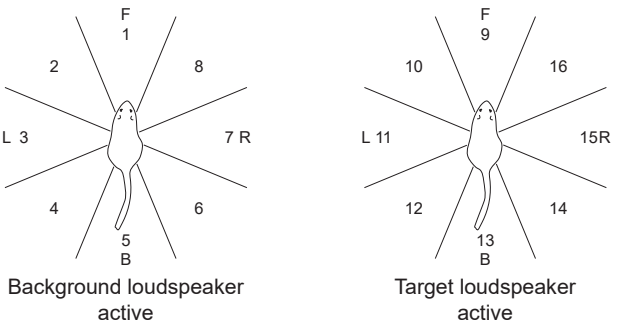
2.4.7 Decoder analysis

For each session, each sound was categorized with respect both to the active loudspeaker as well as to the egocentric sound-source location at the time of sound onset. For the latter, the 360° angular space around the animal was divided in 8 angular bins. Therefore, each sound corresponded to one of the 16 possible loudspeaker-angular bin classes ((Fig. 2.13 – upper panel)). Simultaneously, the 250 ms after each sound onset were divided in 10 bins, each with a 25 ms duration. For each unit in the session, it was counted how many spikes occurred during each time bin.

To reliably decode the spatial egocentric and loudspeaker identity information, a large set of units must be used, such as it is available to the higher order areas in the brain. However, due to the low yield of neurons in each session (maximum of 18 units in one session – single and multi-units), several sessions had to be combined to achieve such a large set. For that, across-sessions class-specific population response ensembles were created – corresponding to a combination of sounds from different sessions that share the loudspeaker-angular bin class (Fig. 2.13 – middle panel). Yet for an individual class, in each session were presented a different number of sounds. Therefore, for the creation of the sound ensembles, for each individual class, the sound identities of each session were randomly under-sampled (without replacement) to the number of sound presentations of the session with the least sound presentations for that specific class. Additionally, some classes were under-represented in relation to others, which would result in a class imbalance that could lead to certain categories being more likely decoded than others. To prevent this situation, random under-sampling (without replacement) was implemented to feed the decoder with exactly the same number of sound ensembles of each class. This class sample size was chosen to be 75% of the number of sound ensembles in the class with the least sound ensembles. All the other sound ensembles (with a minimum of 25% of sound ensembles per class) were used in testing the decoder.

The features provided to the decoder were the number of spikes created by each unit for a particular sound ensemble and in a specific 25 ms time interval bin. Consequently, the number of features per sound ensemble were $10 \times$total number of units. Occasionally,

only one or a selection of time interval bins of the PSTH was used for decoding (during the temporal analysis of the decoder).

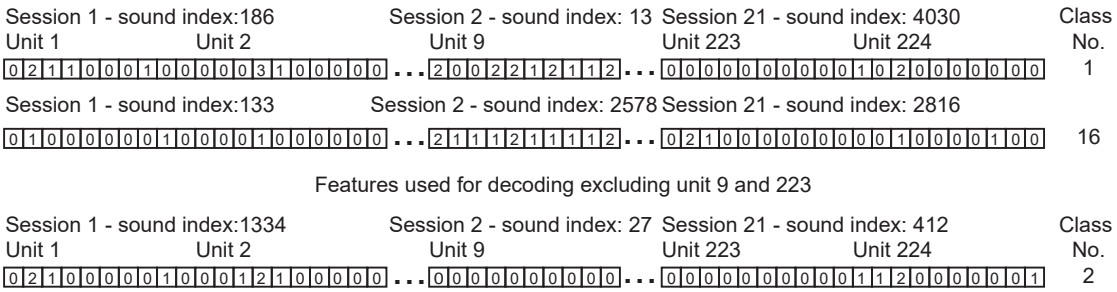


Each sampling cycle:

Across-sessions class-specific population response ensembles

Ensemble Index	Class No.	Session 1 - Sound index	...	Session 21 - Sound index
1	1	2011		1489
2	**1**	**186**		**4030**
...	...			
139	**2**	**3054**		**13**
140	2	1334		412
⋮	⋮			
2547	**16**	**133**		**2816**
2548	**16**	**1613**		**3054**
...	...			

Features used for training and decoding



Features used for decoding excluding unit 9 and 223

Figure 2.13 – **Pre-processing steps for decoding the egocentric sound-source position and the loudspeaker identity** Top panel: division of the angular space around the animal in classes. Middle panel: Example of cross-sessions class-specific population response ensembles formation (training data in bold). Lower panel: Features that are fed to the algorithm to train and later to test the decoding performance, including the situation of units' exclusion.

Before training the decoder, each feature was normalized to lie between 0 and 1 (using the mix-max scaler algorithm) to avoid that units that are inherently more active dominate

the results. The training data was fitted by a multi-layer perceptron classifier (MLPC) [154].

The MLPC is a feedforward artificial neural network that uses backpropagation during training for optimization of the parameters in a supervised learning manner, and therefore was interpreted as a more biologically inspired decoder. In the MLPC, one hidden layer was implemented with the number of nodes as the mean between the number of features used and the total number of classes (16). As suggested in the sklearn package for small datasets, the optimization algorithm used was the "lbfgs" from the family of the quasi-Newton methods. The biologically-inspired rectified linear unit was the activation function for the hidden layer.

The decoder was trained on 224 units (single and multi-units) from two animals in a total of 21 sessions. Only sessions that allowed a minimum of 40 sound ensemble per class for training were used in decoding.

The whole process was repeated 20 times to estimate errors, since both the sound ensembles' formation as well as the choice of the training datasets were based on random data sampling and were inherently part of a distribution that must be several times assessed. The accuracy of identifying the active loudspeaker was determined per loudspeaker for each sampling cycle. Later the accuracy per repetition was considered as the mean between the two loudspeakers and the total accuracy and standard deviation calculated across all 20 sampling cycles. The accuracy of predicting the right class and the egocentric sound-source location was similarly calculated. For the construction of the confusion matrices, all the predictions from all the sampling repetitions were simultaneously used. The normalized accuracy was calculated as $\frac{accuracy - 100/\#possibilities}{100 - 100/\#possibilities}$, being 0% if equal to chance level and 100% if it predicted all correctly – $\#possibilities = 2$ for loudspeaker identity, $\#possibilities = 8$ for angular bin class and $\#possibilities = 16$ for loudspeaker-angular bin class.

To evaluate the influence of spatial tuning classes, units were eliminated in the test data by setting to zero all the bins corresponding to the chosen units (Fig. 2.13 – lower panel). In practice, this means that these neurons were set to have produced no spikes during the task and thus were not informative about the location or identity of the loudspeakers. The accuracy of the decoder (trained on the complete training dataset - without eliminating units) when some units were eliminated from the test data was then compared to its accuracy when the test data was complete. The units with differential spatial tuning for both loudspeakers that were eliminated corresponded to all the units that at some temporal response period had a difference in egocentric sound-source angle between loudspeakers larger than 80°. The eliminated canonical units were randomly selected for each sampling cycle from the units that showed clear ipsi or contralateral tuning during the onset period (±45°). The randomly selected units were randomly selected for each sampling cycle from all the units, meaning that different units could have been eliminated in different sampling cycles. This analysis was based on 100 sampling cycles per eliminated spatial tuning class.

CHAPTER 3

Results

3.1 Behavioral performance analysis

3.1.1 Were the animals performing a sound localization task?

The gerbils were successfully trained in the aSIT$_{loc}$ and finished significantly more trials in the target island than it would be expected by random locomotive behavior, evidenced by the non-overlap between the confidence intervals for the target and by chance (Fig. 3.1a), calculated as explained in Section 2.4.1. This is valid for all animals when calculated for the loudspeaker reward configuration that was used during the training phase (Fig. 3.1b).

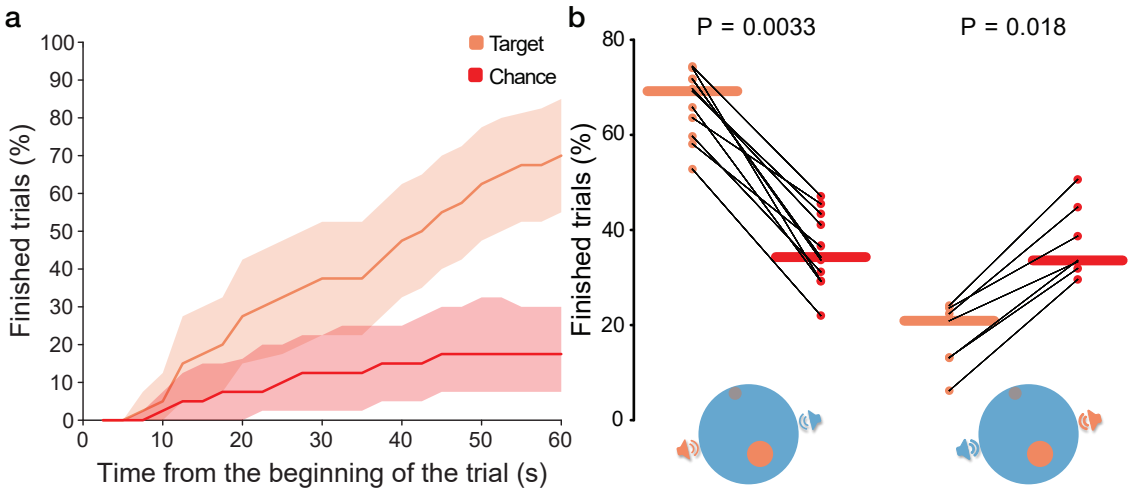


Figure 3.1 – **Behavioral performance** in aSIT$_{loc}$ **a** Percentage of finished trials in the target and by chance as a function of time from the beginning of the trial in a session **b** Comparison of the percentage of finished trials in the target island and in a random surrogate island (chance level) at the maximum time allowed for a trial (60s) in both the training loudspeaker configuration (left - 11 animals) and in the swapped loudspeaker configuration (right - 7 animals), Wilcoxon signed-rank test. Figure adapted from [117]

Nonetheless, since the gerbils could be reporting any change in the properties of the sound (odd-ball), during the test phase the rewarded loudspeaker (including the green LED) was swapped in 1/8 of the trials in a random order. If the odd-ball hypothesis were correct, the effect seen in this swapped configuration would have been similar to the one in the normal configuration, as the animals would still be reporting an odd-ball. However, we observe that the performance drops below chance level. Since the only cue that could have allowed the animals to differentiate a normal and a swapped loudspeaker reward configuration was the allocentric (world-reference based) location of the rewarded and the non-rewarded sound-sources, this suggests they were actively localizing where the sound was coming from to perform in this task and not pursuing an odd-ball strategy.

This result also shows that in reality the gerbils did not associate the green LED to the target loudspeaker, contrary to what was initially intended.

3.1.2 Catch trials

The behavior in these catch trials was further investigated, since it was unclear whether the animals greatly reduced their locomotion during the trial – resulting in them not finding the target – or whether they had found the target island, but left it before the end of the sit-time.

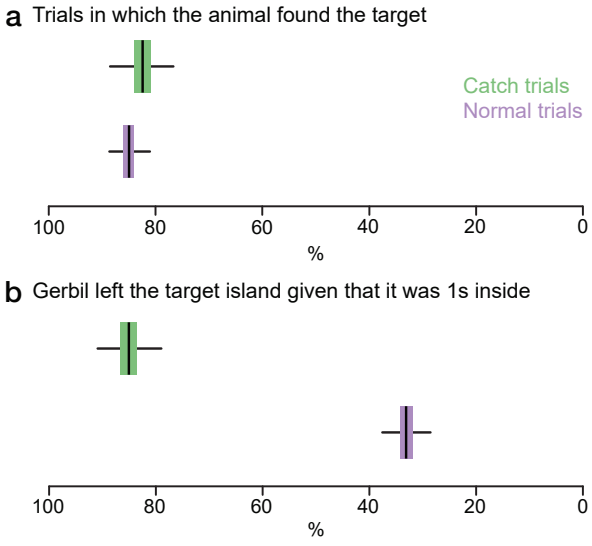


Figure 3.2 – **a** Gerbils found the target as often in catch trials as in normal trials. **b** Gerbils actively avoided remaining in the target island in catch trials. Number of sessions: 39; Number of trials with a normal loudspeaker reward configuration: 1784; Number of catch trials: 285. The thick lines correspond to the median, the filled boxes to the 1st and 3rd quartile and the whiskers to $\pm 2.7\sigma$, calculated using a bootstrapping method. Figure adapted from [117]

When a catch trial started, the active loudspeaker at the beginning of the trial corresponded to the usual (in the normal configuration) target loudspeaker. This could thus have resulted in a halt of the movement of the animal, which would be awaiting for the reward. Consequently, the animals would have found the target island less often than in a normal loudspeaker reward configuration trial. However, as the confidence intervals between catch and normal trials overlap (Fig. 3.2a), this option is refuted.

To understand whether in fact the animal purposefully left early the target island in catch trials, I plotted the percentage of times in which the animal left the target island given that it was at least 1 s inside (to ensure that the animal did not just run past the target island). In a catch trial, an animal was more than twice as likely to leave the target island before the sit-time in comparison to a normal trial (Fig. 3.2b), and was consequently avoiding to hear sound from the typical background loudspeaker. This hints that the animals associated their own spatial position to a change in the active loudspeaker, an association that we further investigated.

3.1.3 Spatial bias in the target island location

A spatial bias in the location of the target island was introduced to dissect whether the gerbils associated the target island and corresponding stimulus change with an actual physical space or if they were merely stopping whenever they heard that change.

Usually the target islands were drawn from a uniform distribution (Fig. A.1b). This means that in the central region of the arena – corresponding to a circle with a radius double of the target island radius – lie 59% of the target island centers (Fig. 3.3 left panel). After introducing a bias in the spatial location of the target islands (Fig. A.1a), this percentage increased to 78% (Fig. 3.3 right panel).

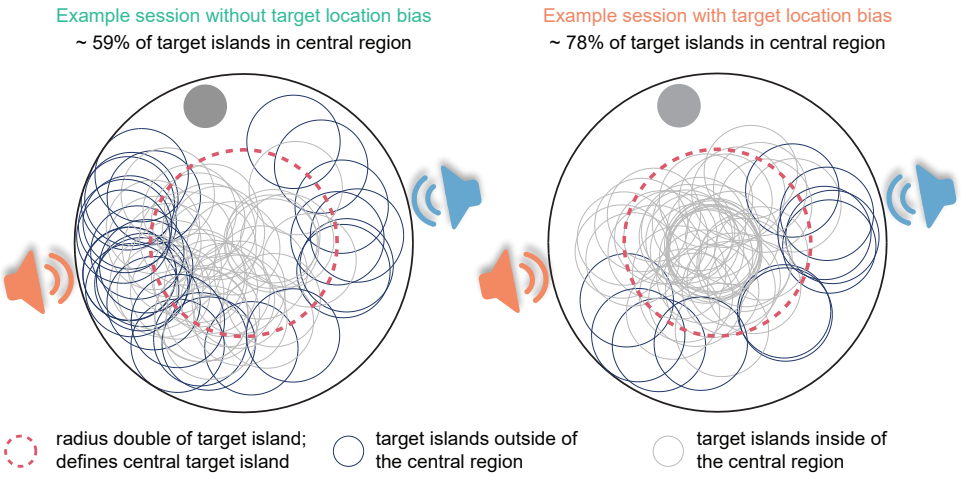


Figure 3.3 – Introduction of a spatial bias in the position of the target island Two sessions that depict a situation where there is no spatial bias (left) and another where the spatial bias was implemented (right). Figure adapted from [117]

In a non-biased situation, there is no advantage in going to any particular area in the arena, since all areas are equally likely to contain the target island. Thus, a difference in the performance between trials where the target was centrally located and non-centrally located is not expected. However, if a spatial bias is introduced towards the center, it is advantageous to go directly to the center and therefore the likelihood of finishing a trial in the center should increase in comparison to the outer area.

Two gerbils were initially trained in a non-biased situation and after they performed successfully in the task, the spatial bias was introduced. Another gerbil was initially trained in the biased situation and only after good performance, the bias was removed. A change in the behavior of the animals was not expected on the session following to the change of the bias situation, since there were not enough repetitions for the animals to notice it. However, if the animals associated the target island with an actual physical location, after one month of training they would already have adapted their target search strategy.

The gerbils were able to incorporate the spatial bias into their exploration strategy (Fig. 3.4), as shown in the gerbil i and ii data, and to adapt their strategy when the spatial bias was removed, as shown in the gerbil iii data.

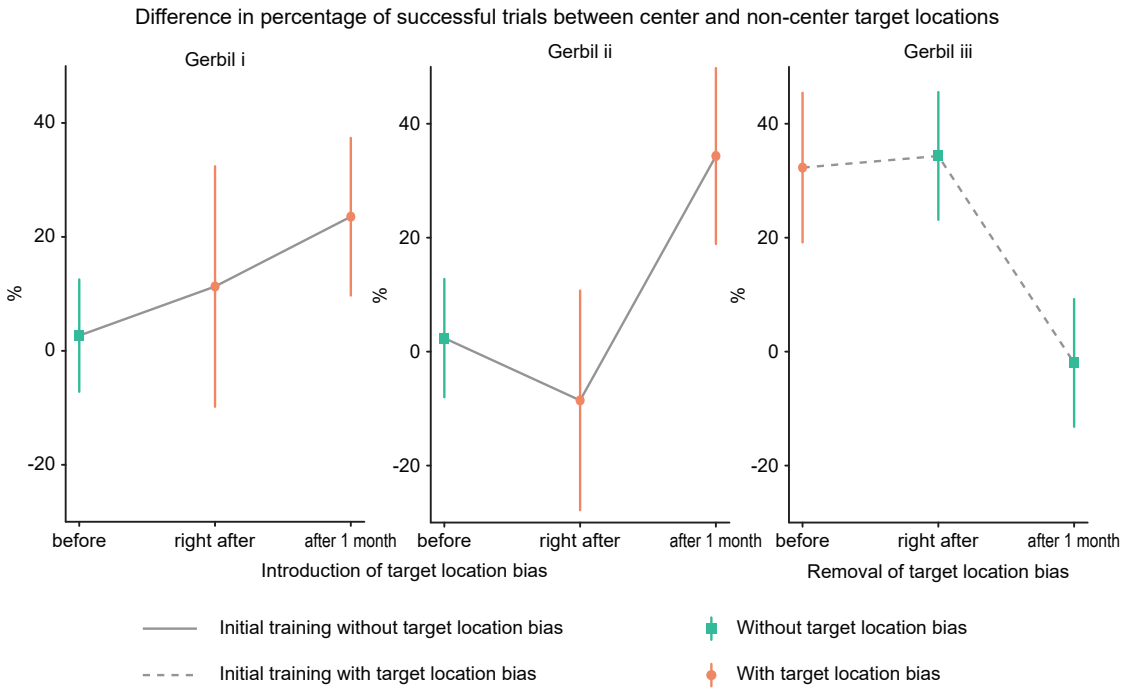


Figure 3.4 – Gerbils optimized their search strategy based on the target island distribution likelihood Difference in the percentage of successful trials between trials in which the target island was in the center and trials in which the target was not in the center (error bars correspond to the 95% confidence interval, calculated using a bootstrapping method followed by error propagation for the subtraction). Figure adapted from [117]

This demonstrates that the gerbils associated the change in stimulus and consequently the reward expectancy with a specific physical location in the arena and that this task does not merely consist of a GO/NOGO paradigm.

3.1.4 Swapped loudspeakers

The success in the task could still be explained by some slight frequency or amplitude difference in the frequencies that compose the harmonic complex sound, which could be used to solve the task (despite the roving in fundamental frequency and amplitude). To ensure that was not the case, for three animals the two loudspeakers were physically swapped. The performance did not drop after the swap (Fig. 3.5), showing that it was not a specificity of the loudspeakers what was being used to solve the task, but indeed their allocentric location.

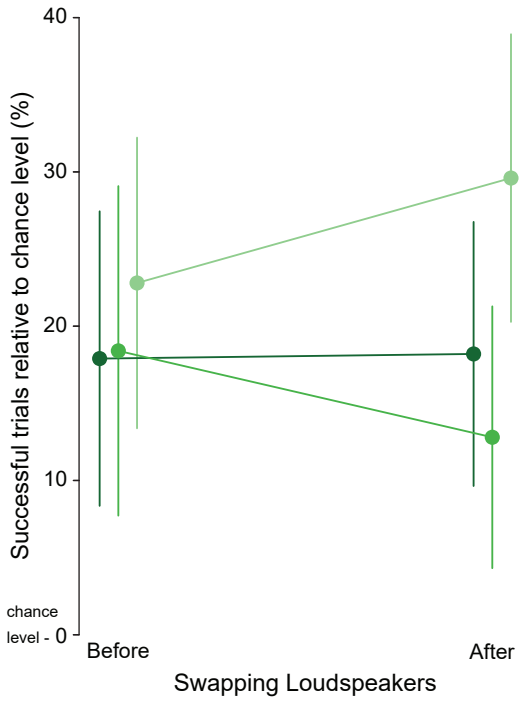


Figure 3.5 – No influence of swapping loudspeakers in task performance The performance of three animals in the task before and after physically swapping the loudspeakers is similar, showing that the animals were not using specific spectral cues from the loudspeakers to solve the task. The data is presented as median performance relative to chance at maximum trial length (60s). The error bars correspond to the 95% confidence interval. Number of trials before swapping: $N_{Gerbil\ 1}$=206 $N_{Gerbil\ 2}$=163 $N_{Gerbil\ 3}$=177 After swapping: $N_{Gerbil\ 1}$=237 $N_{Gerbil\ 2}$=234 $N_{Gerbil\ 3}$=172. (Combination of the 2 sessions prior to the swapping for the “Before swapping” data and of the 2 sessions following the swapping for the “After swapping” data.)

3.2 Behavior description

3.2.1 Bout analysis

During aSIT$_{loc}$, all animals moved typically in short bouts, corresponding to a few cm distance (see Section 2.3.5). This behavior was expected for such a paradigm, given that a small change in the position could have led to the desired change in the stimulus property. Besides, since the stimuli were only repeated every 250 ms, the animals should have moved at a pace slow enough that allowed them to sample the area of the arena with auditory stimuli.

If the animal did not rotate during a bout trajectory, the likelihood that such a bout was short was higher than for bouts where there was a combination of rotation and translation (Fig. 3.6, median bout length for translation only: 1.2-2.7 cm compared to 3.8-7.9 cm for the combination of rotation with translation). However, the bout length distributions for the latter appear to be a combination of two distributions, one with very short length values (below 5cm) – which by visual inspection corresponds approximately to one or two steps of the animal – and another distribution with longer length values that go up to 1 m (not shown in Fig. 3.6, as the figure was truncated for better visualization; median for bout lengths larger than 5cm: 11-20 cm).

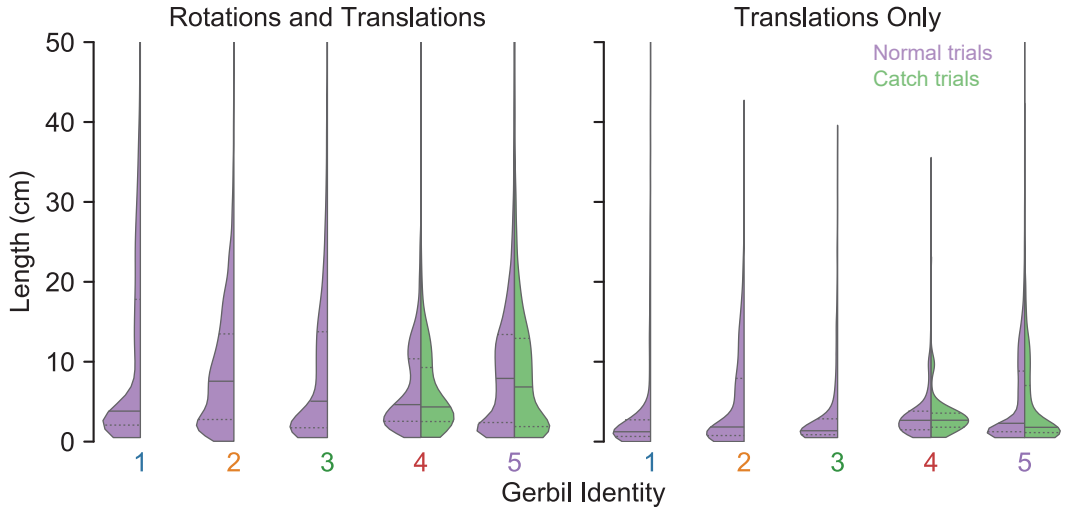


Figure 3.6 – **Distribution of bout lengths** per animal separated by bout type and whether the trial is a catch trial. (Median value – horizontal solid line; 1st and 3rd quartiles – horizontal dashed lines)

The maximum linear distance within the central area was only 70 cm (remember that all the analyzed data lied in this central area - see Section 2.3.5 for bout calculation details), which explains the distribution difference between the bout types. This maximum corresponded to a situation where the animal started the bout at the edge of this central area (note that this area was unknown to the gerbil) and walked in a straight line to the diametrically opposed limit of this area. In practice, this means that the animals could cover a shorter distance in a straight line than if their trajectory was curved. Additionally,

the circular geometry of the arena could have induced more rotatory behavior than what usually happens in an unbounded territory, resulting in a change in the azimuth whenever the trajectory was longer.

A difference in the distribution of bout lengths between normal and catch trials is apparent in the situation of an accompanying rotation, for which the bout lengths were significantly shorter during catch trials (median 4.6 cm and 7.9 cm for normal trials compared to 4.3 cm and 6.8 cm for catch trials, $p=0.03$ and $p=3\times10^{7}$, Mann-Whitney rank test, for gerbil 4 and 5, respectively).

As already hinted by the distribution of bout lengths, the bouts lasted on average for only one or two frames (Fig. 3.7), especially in the case of rotations/translations only. Additionally, the bouts that were a combination of rotation and translation lasted usually below 1s, inclusively when only bouts with a length longer than roughly the gerbils' body length (>10 cm) were considered (median: 0.48-0.6 s - not shown).

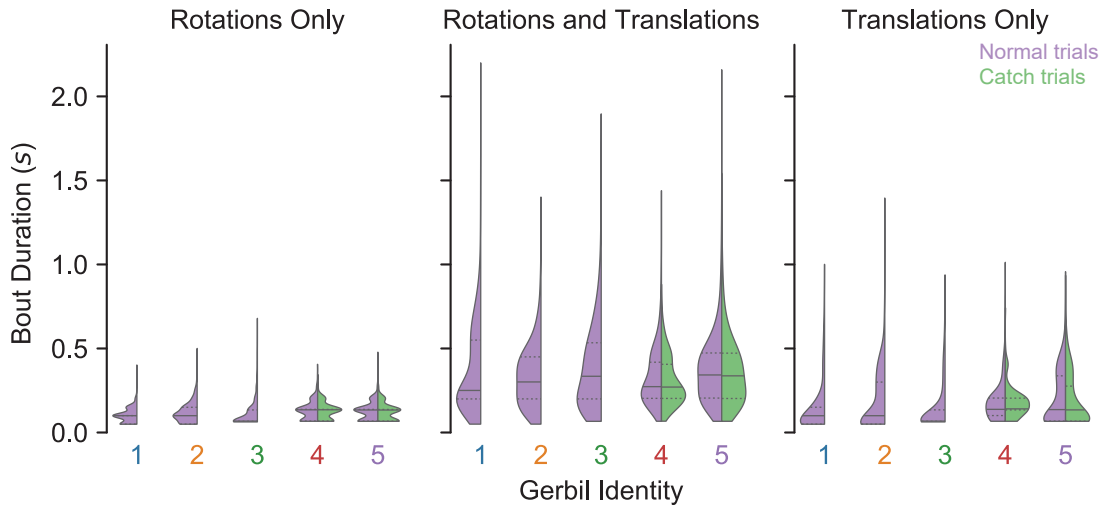


Figure 3.7 – **Distribution of bout durations** per animal separated by bout type and whether the trial is a catch trial.

If the bout consisted exclusively of a rotatory movement, the cumulative angle throughout the bout was generally smaller than if it was in combination with a translation (Fig. 3.8 median: 10°-19° for rotations vs 29°-43° for the combination), which is in agreement with the rotation only bouts being generally shorter in duration. If only bouts longer than 10 cm are considered, the median of the cumulative angle increases to 48°-69°.

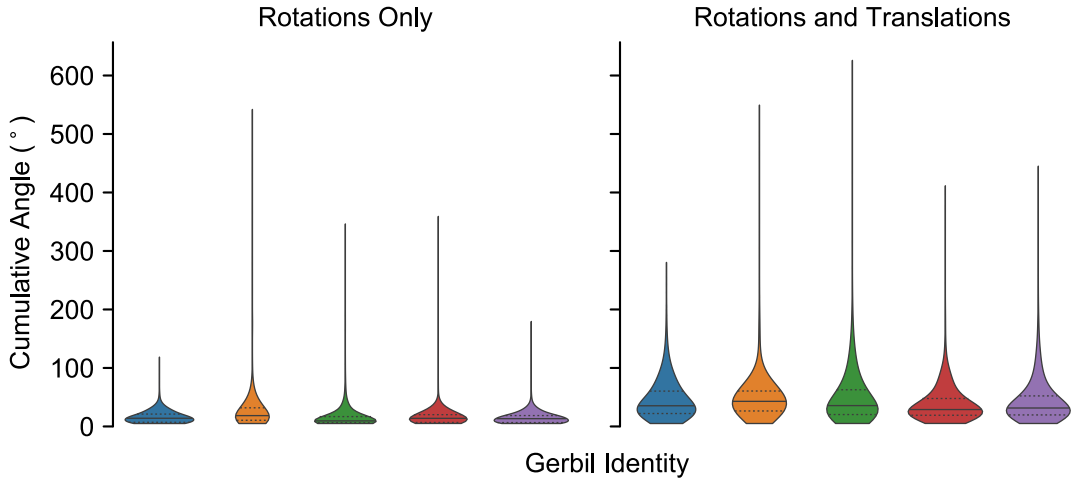


Figure 3.8 – **Distribution of the cumulative rotation of the gerbil during a bout** per animal and separated by bout type. (Median value – horizontal solid line; 1st and 3rd quartiles – horizontal dashed lines)

To describe how fast a gerbil usually moved, the analysis was focused exclusively on movements with a substantial magnitude – with lengths longer than roughly the body size of the animal (10 cm) or an angle larger than 20°.

During $aSIT_{loc}$, all gerbils moved usually with a speed between 20 and 60 cm/s independently of whether the motion was purely translational or also had a rotational component (Fig. 3.9a).

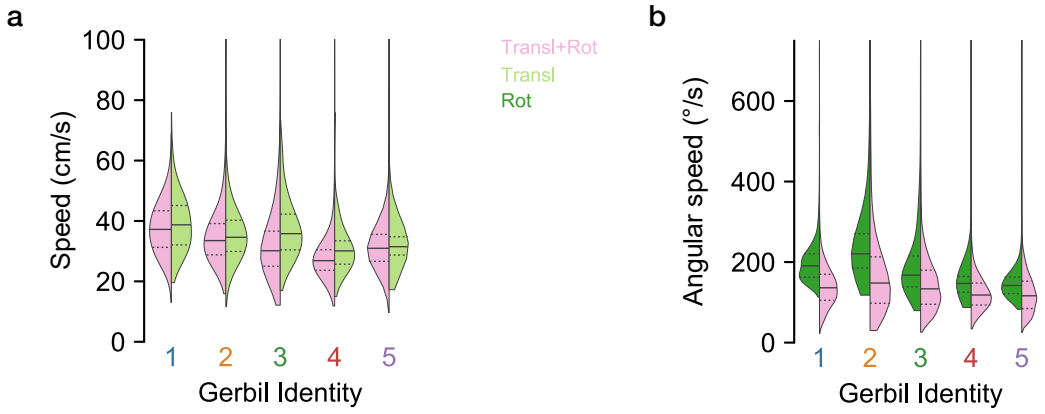


Figure 3.9 – **Distribution of bout speeds** per animal and separated by bout type **a** speed calculated for bouts with a length larger than the gerbil's body size (>10 cm) **b** angular speed calculated for bouts with an accumulated rotation larger than 20°. (Median value – horizontal solid line; 1st and 3rd quartiles – horizontal dashed lines)

For motions in which the azimuth changed, the angular speed was typically between 100°/s and 300°/s for purely rotational bouts (Fig. 3.9b). These were consistently faster than the ones where translation was present, which mostly were up to 200°/s.

Independently of being in the target island or not, the animals' stop periods were typically short (below 1 s - Fig. 3.10). Yet for some animals, as expected, it is possible to observe longer stop periods when the animal was in the target island, comparable to the sit-time.

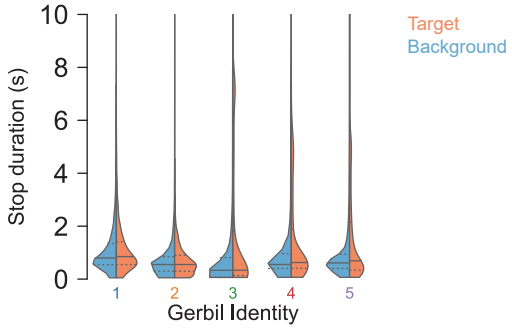


Figure 3.10 – Distribution of the durations of the time periods during which the animal remained stopped per animal and separated by whether the animal was in the target island. (Median value – horizontal solid line; 1st and 3rd quartiles – horizontal dashed lines)

All but one animal (gerbil 1) turned more often towards the target loudspeaker when in the target island (Fig. 3.11). However, on average the turning angle was small (with the exception of gerbil 2) but significant when considering turning angles of over 20° (except for gerbil 4).

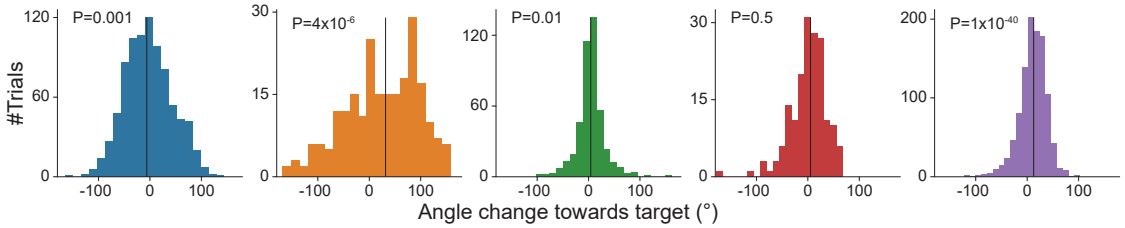


Figure 3.11 – Change of angle relative to the target louspeaker per animal. Comparison of the first frame after entering the target island with the last frame before correctly finishing a trial. Positive for turning towards the target loudspeaker and negative for turning away from it. The vertical line corresponds to the median value. All p-values are based on a binomial test which was performed only on the trials where the difference between beginning and end frames' angle was larger than 20° (a substantial change in the orientation angle).

3.2.2 Locomotion state analysis

The great majority of frames corresponded to periods of pause in the movement (no measurable translation or rotation – see Section 2.3.5) both when the animal was in the target island, as well as when it was outside (Fig. 3.12). As expected, for all animals, being inside the target corresponded to more frames without any movement.

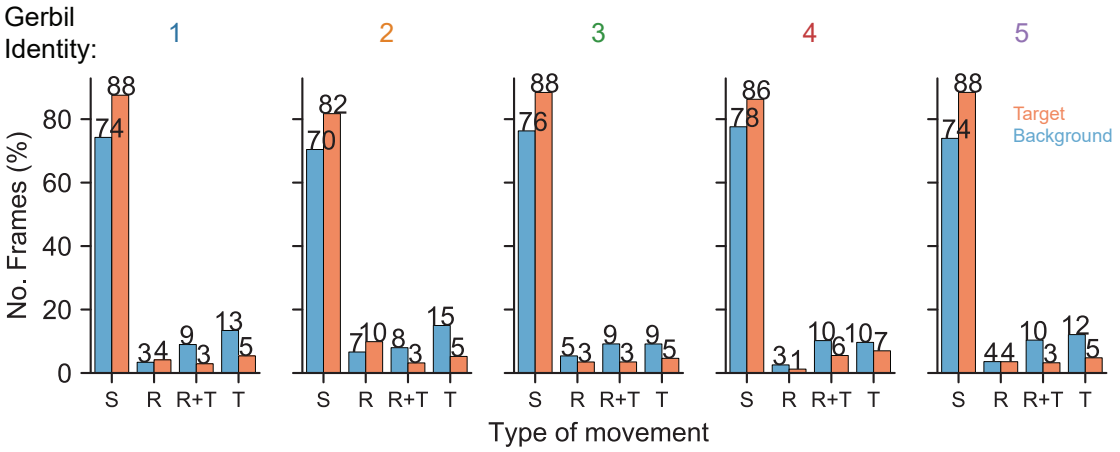


Figure 3.12 – **Distribution of state of movement** separated by animal and by it being in the target during the respective video frame (for normal reward configuration trials) S: stopped R: rotation only R+T: simultaneous rotation and translation T: translation only

By calculating the state of movement of the animal during the whole time period a sound lasted plus the associated intersound interval (in total 250 ms), we also observed that in the majority of instances the animal was stationary, especially if we consider only the situations when the animal was inside the target at the moment of sound onset (Fig. 3.13).

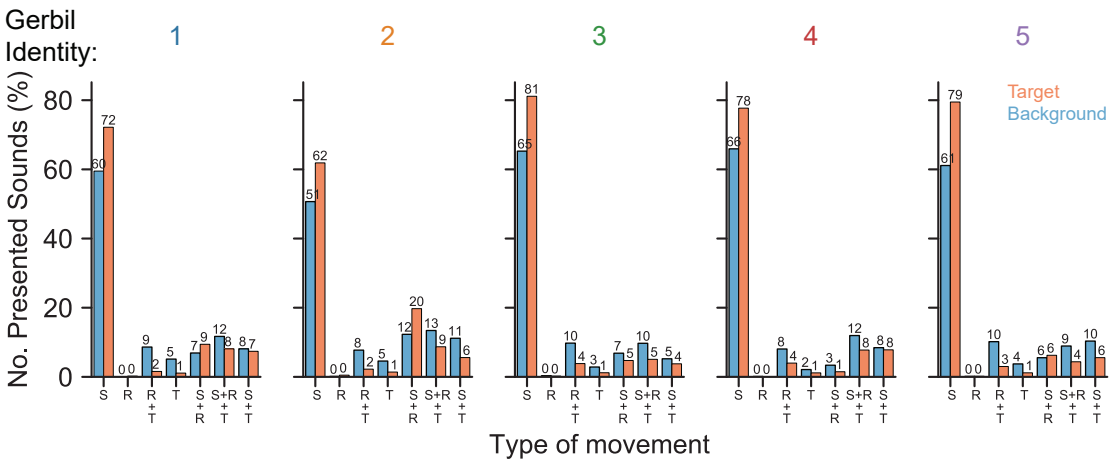


Figure 3.13 – **Distribution of the state of movement during the 250 ms time period (sound plus the associated intersound interval)** separated by animal and by it being in the target at the moment of sound onset (for normal trials) S: stopped R: rotation T: translation

To better understand the dynamics associated with the various locomotion states and how these varied between target and background areas, the probabilities of transitioning to another specific locomotion state or maintaining in the same state were calculated for both when the active loudspeaker was the target as well as the background and represented in a Markov chain (see Section 2.3.5).

The animals tended to remain in the same locomotion state (Fig. 3.14), with the exception of the rotatory behavior, which was more likely to transition to a stationary behavior. In the background area, the animals were almost equally likely to start a bout in either of the three locomotion states; whereas in the target island, the beginning of a bout did not usually involve translation. All the locomotion states were more likely to transition to "stop" when in the target island than in the background area and the "R+T" was more likely to transition to a purely rotatory behavior.

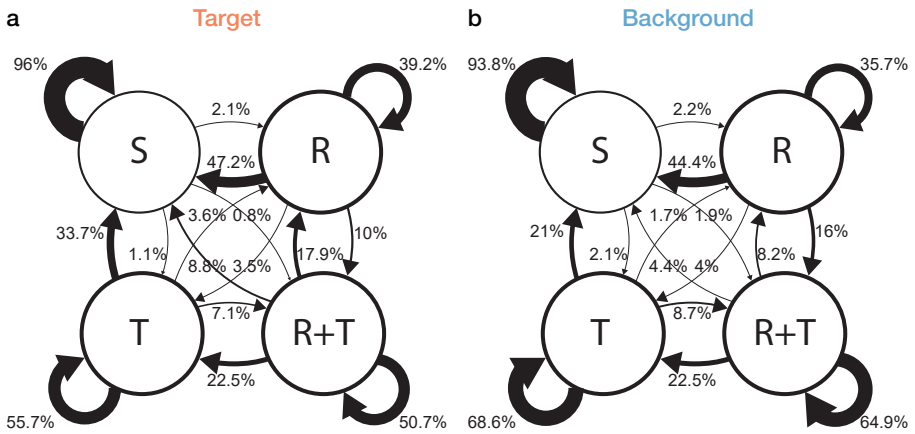


Figure 3.14 – **Markov chain** representing transition probabilities between different locomotion states when the active loudspeaker was the **a** target **b** background. S: stopped R: rotation only R+T: simultaneous rotation and translation T: translation only

3.3 Neuronal responses

3.3.1 Neuronal latencies

The latency of the majority of neurons was shorter than 20 ms (193/364), which is in line with the latencies reported for recordings in the A1 [139]. There are clearly two latency clusters (Fig. 3.15): the first being right after sound onset (Fig. 3.16a,b) and the second after sound offset, meaning that some neurons are only responsive to the offset (Fig. 3.16c,d), as already shown in earlier studies [77, 155, 156].

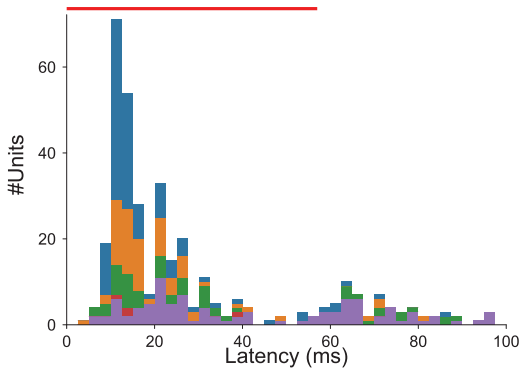


Figure 3.15 – **Neuronal latencies distribution** for all single units (neurons) color-coded by animal as in Fig. 3.8. The red solid line represents the sound stimulation.

The onset response can be characterized by either an increase of the firing rate (Fig. 3.16a,c) or a decrease (Fig. 3.16b,d), corroborating earlier studies [77, 79, 155]. However, most studies concerned with sound localization in cortex disregard the neurons whose firing rate decrease with sound stimulation, which is thought to be most prevalent in layer V and VI of the AC and related to intracortical inhibition [156].

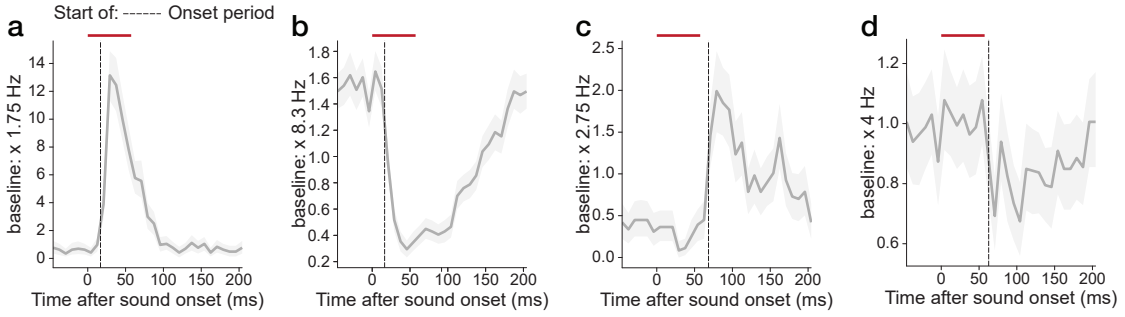


Figure 3.16 – **Representative PSTH of responses** to the **a,b** sound stimulus onset; **c,d** sound stimulus offset. These neuronal responses correspond to **a,c** an increase in the firing rate; **b,d** a decrease in the firing rate. The solid gray line corresponds to the median and the shaded area to the 95% confidence interval. The red solid line represents the sound stimulation. The baseline corresponds to the engaged spontaneous firing rate (without sound - see Section 2.4.2).

Interestingly, the latencies of the response to sounds coming from the background loudspeaker are significantly shorter than for the target (not shown, P=0.006, Wilcoxon signed-rank test, 239 neurons, median latency for background 14 ms vs. 16 ms for the target).

3.3.2 Classification and characterization of neuronal types

The spike waveforms of single units were analyzed and divided into two types – as previously described in the literature – using a clustering algorithm (for details see Section 2.4.6): the regular-spiking neurons (RS - putative excitatory) and the fast-spiking neurons (FS - putative inhibitory). Similar to earlier studies [151, 152], the RS neurons were the most abundant neuronal type – corresponding to 72% of the total of neurons (N=262, Fig. 3.17a); whereas the FS corresponded to 28% (N=102). Confirming earlier research [151, 157], the latencies of the FS neurons were in general shorter than those of the RS (Fig. 3.17b, median: 17 ms vs 20 ms). Contrary to [151], but in line with [157], the FS neurons had higher values than RS neurons both for the spontaneous (Fig. 3.17c, median spontaneous: 7.76 Hz vs 5.35 Hz) as well as for the evoked firing rates for both loudspeakers (Fig. 3.17d,e, median target: 7.25 Hz vs 4.65 Hz; median background: 6.47 Hz vs 4.84 Hz).

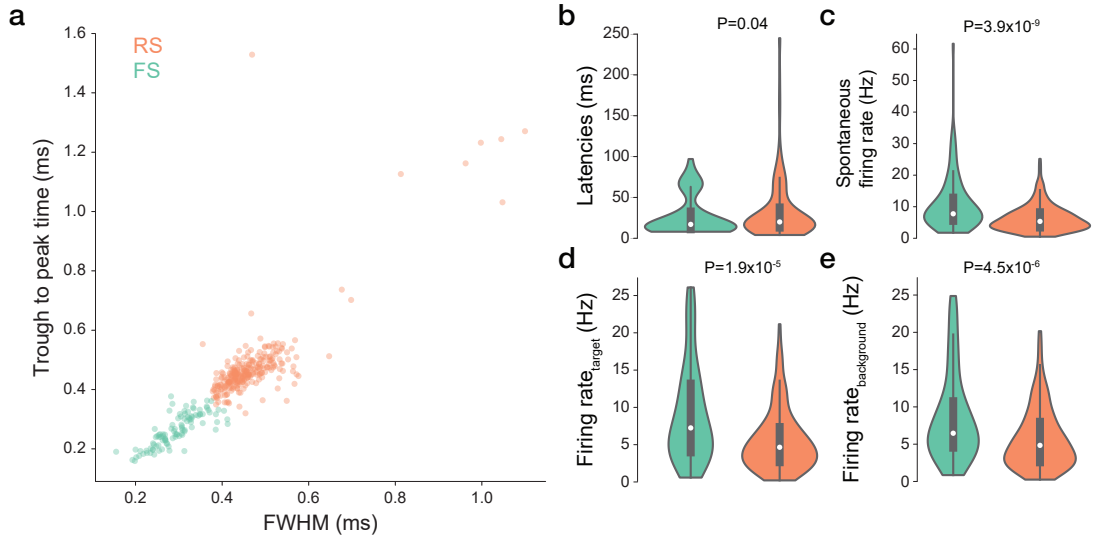


Figure 3.17 – **a**, Clustering of the two different neuronal types (FWHM – full width at half maximum of the peak, 364 neurons: 102 FS and 262 RS). Comparison between neuronal types of **b** latency (364 neurons) **c** spontaneous firing rate (364 neurons) **d** evoked firing rate as response to the target loudspeaker (244 neurons) **d** evoked firing rate as response to the background loudspeaker (333 neurons) All p-values are based on the Mann-Whitney U-test.

3.3.3 Spatial tuning diversity

The egocentric location of the loudspeaker at the moment of sound onset was shown to modulate throughout time the neuronal responses to the sound presentation (Fig. 3.18).

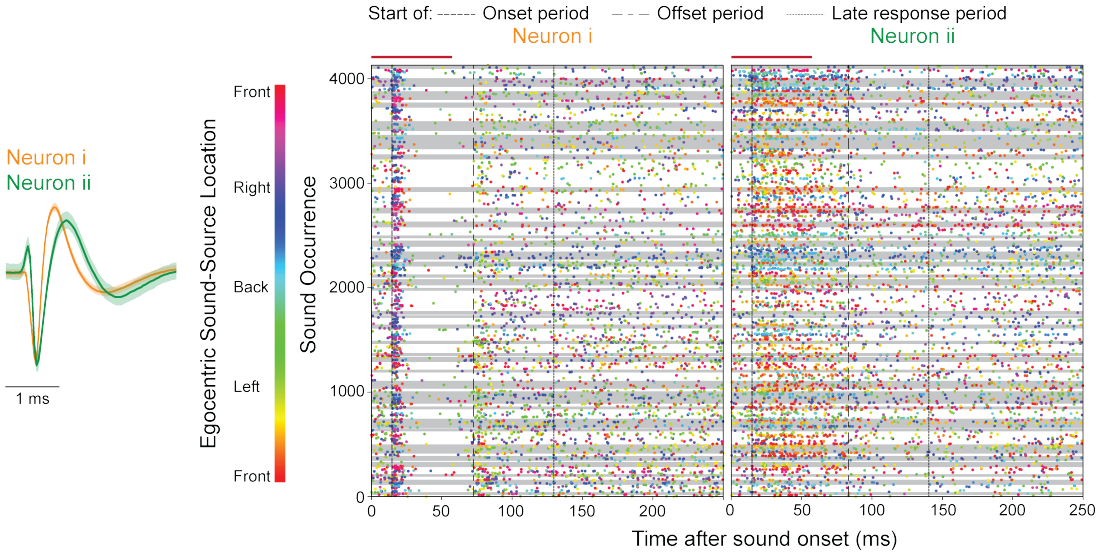


Figure 3.18 – **Spike waveforms and corresponding raster plots** of neuronal activity in two AC neurons during task performance. Periods of stimulation by the target loudspeaker are highlighted by gray areas. The loudspeaker location relative to the animal at the moment of the sound occurrence is color-coded (see Section 2.4.5 for details). The period of sound presentation is represented by the red line and the start of the different analysis periods by the dashed lines. Both neurons are regular-spiking neurons.

The neuronal responses were separated between the two loudspeakers and between the analyzed time periods (see Section 2.4.5 for details). For each of these conditions, a spatial plot was constructed by further analyzing the responses with respect to where the loudspeaker of interest was located relative to the animal's body axis (egocentric reference frame). Neurons exhibited a remarkable variety of spatial tuning: while some showed the canonical hemispherical contra- or ipsilateral preferred egocentric sound-source location (Fig. 3.19 i and iii, respectively), identical when calculated for either loudspeaker, other neurons possessed an orientation-sensitivity, with a predominant bias to the front/back (Fig. 3.19 ii). This last neuron type can be interpreted as favoring 0 μs ITD, since a sound coming from the front or from the back arrives simultaneously at both ears. Furthermore, the pinnae – which could help to discriminate back and frontal loudspeaker's location – has minimal influence for low frequency sounds [146]. Neurons with a frontal preferred egocentric sound-source location were already described in previous studies [79, 155].

Remarkably, some neurons were sensitive to the identity of the loudspeaker (differential behavioral meaning and different world-based reference frame location): both neurons that were spatially tuned for one loudspeaker only (Fig. 3.19 iv), as well as neurons that showed a preferred egocentric sound-source location that depended on the active loudspeaker (Fig. 3.19 v) were found.

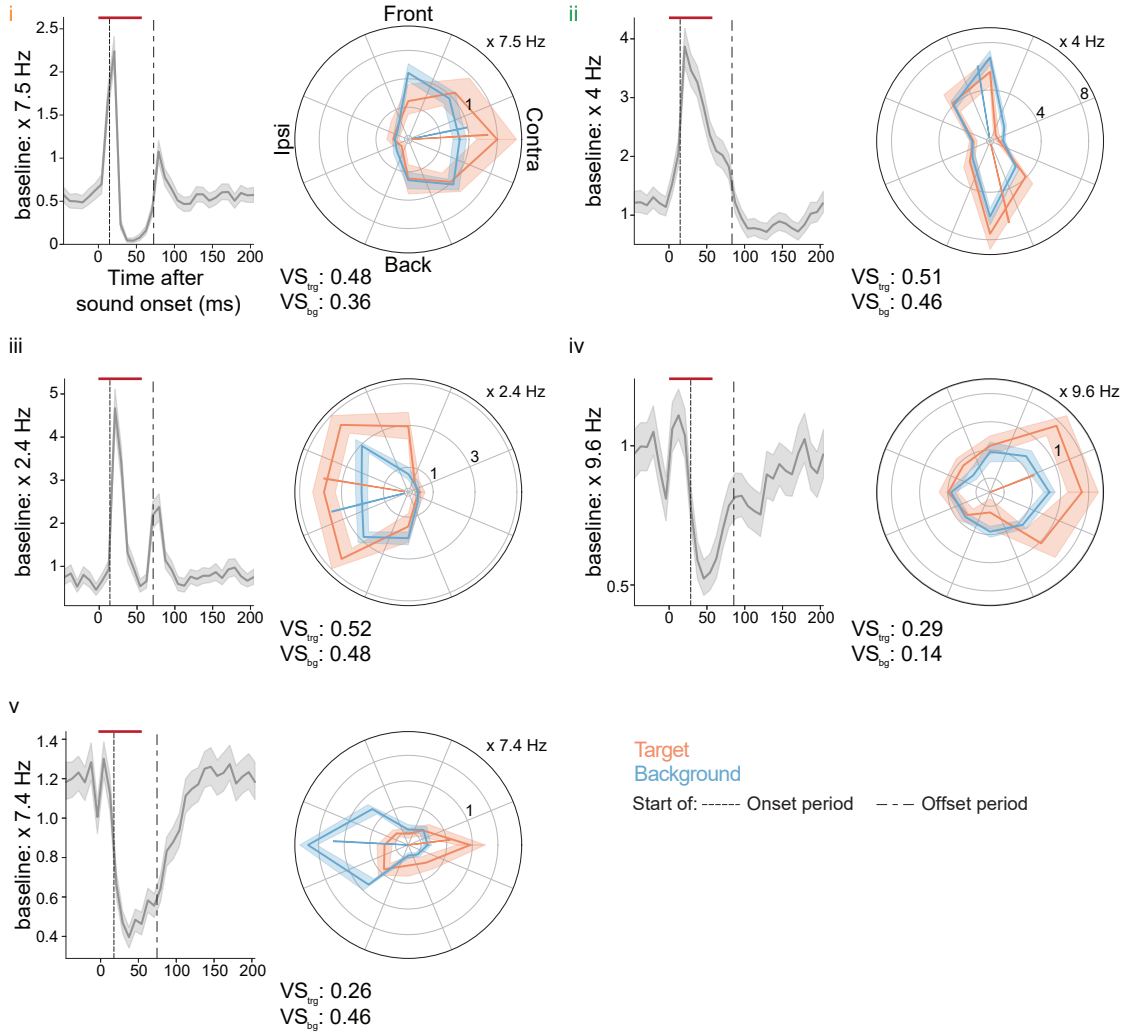


Figure 3.19 – **PSTH and corresponding spatial tuning** during the onset period for 5 representative neurons. The solid lines correspond to the median and the shaded areas to the 95% confidence interval. The red solid line represents the sound stimulation. The baseline corresponds to the engaged spontaneous firing rate. The polar plots correspond to the spikes that occurred between the dashed lines in the PSTH. The colored lines in the polar plots indicate the preferred tuning angle, with the length scaling with the vector strength. Neurons i and ii correspond to the ones depicted in Fig. 3.18.

The number of neurons that were spatially tuned for at least one loudspeaker decreased with the analyzed time from the latency of each neuron (Fig. 3.20), which suggests that most of the spatial information was transmitted by the onset response. Whereas in the onset period a relatively large proportion of neurons encoded the canonical egocentric sound-source location for both loudspeakers (37/118, contra- or ipsi-, $\pm 45°$), during the offset and the late response periods this proportion was reduced (7/58 and 0/42, respectively). Orientation-selective neurons to both loudspeakers were present during all time periods, being more predominant during the onset period (onset period: 16/118; offset period: 1/58; late period: 2/42), reinforcing that the onset period is the most informative period about the egocentric sound-source location.

Surprisingly, a large fraction of neurons showed differential spatial tuning to both loudspeakers (65/118, during the onset period), either by being only spatially tuned to one of them (47/65) – mainly to the target – or by exhibiting a large tuning difference between the two loudspeakers (difference in preferred egocentric angle $> 90°$, 18/65). Interestingly, the proportion of neurons that revealed a differential response to both loudspeakers was larger during the late response period (41/42), suggesting that this time period might be the most informative with respect to the identity of the active loudspeaker.

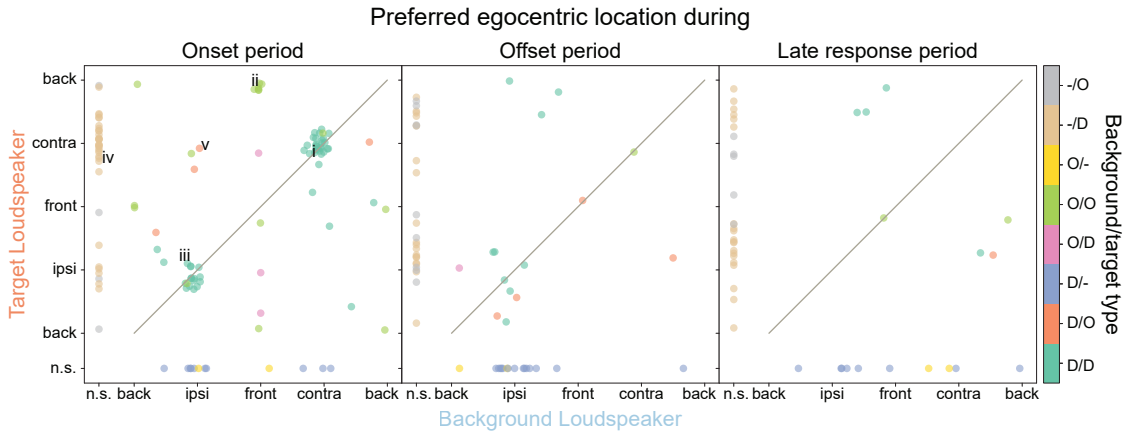


Figure 3.20 – **Preferred egocentric sound-source location** correspondence between the target and the background loudspeakers during the onset (left - 118 neurons), the offset (center - 58 neurons) and the late response periods (right - 42 neurons) color-coded with the type of spatial tuning for each loudspeaker. D:direction-selective O:orientation-selective

The latencies of the spatially tuned neurons were in general very short (<25 ms, Fig. 3.21a), in agreement with the reported latencies in the A1 [139]. The very few neurons with slightly longer latencies possessed differential spatial tuning for the two loudspeakers – which hints that this differential behavior might be a consequence of top-down modulation. However, other neurons that showed differential spatial tuning had comparably short latencies to the hemispherically tuned neurons.

A dependency of tuning preferences on the putative neuronal type (regular or fast spiking) was not evident (Fig. 3.21b) during the onset period.

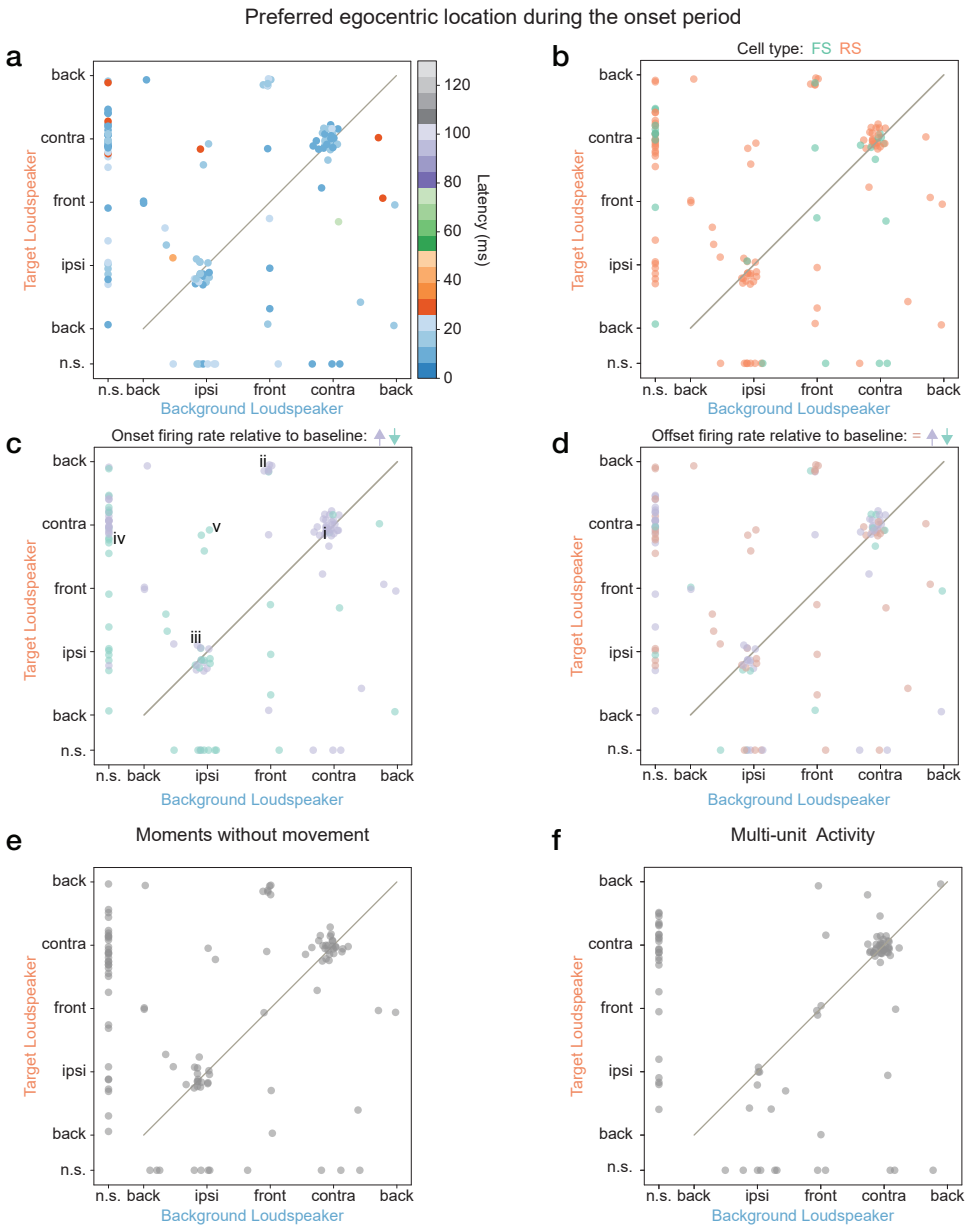


Figure 3.21 – **Preferred egocentric sound-source location** correspondence between the target and the background loudspeakers during the onset period **a** color-coded with the neurons' latency; **b** color-coded with the neuronal response type (FS: fast spiking neurons; RS: regular spiking neurons); **c** color-coded with whether the onset firing rate increases or decreases in relation to the baseline; **d** color-coded with whether there is an offset and if so, if whether the offset firing rate increases or decreases; **e** calculated for moments in which the animal was immobile; **f** calculated for the multi-units.

The firing rate of most contralaterally and front/back tuned neurons increased during onset (Fig. 3.21c). However, a decrease in the onset firing rate was apparent for the majority of neurons with a differential preferred egocentric location. A clear trend was not visible for the onset firing behavior of the neurons with an ipsilateral preference,

possessing this group of neurons a mixture of onset behaviors. It is noteworthy to observe that an ipsilaterally tuned neuron that decreased its firing rate during onset actually was most responsive to the egocentric contralateral sound-source location, since a stimulation on this location led to the largest reduction in the firing rate.

Contrary to the onset firing rate, the offset firing rate behavior did not show any particular pattern in the preferred egocentric location (Fig. 3.21d).

Given that movement is known to reduce the neuronal activity in the AC [98, 158], to ensure that the spatial tuning behavior observed was not due to this kind of modulation, the analysis was restricted to periods in which the animal was stationary (Section 2.3.5). This hypothesis of movement-induced (i.e. unspecific) modulation was refuted, as a similar qualitative behavior was observed for this re-analysis (Fig. 3.21e).

Multi-units also showed very similar tuning types with respect to the preferred egocentric sound-source locations compared to the neurons (i.e. single units) results (Fig. 3.21f). However, a larger fraction of units were contralaterally tuned in comparison to the single units. This might result from the combination of more neurons being contralaterally tuned (relative to other egocentric sound locations) and the neuronal responses being time locked to the auditory stimulation, making the process of separating the neurons harder in the case of contralateral tuning and therefore combining the responses from several neurons into a multi-unit during the spike sorting process.

The geometry of the setup – with the two loudspeakers opposite to each other – left open the question of whether the orientation-sensitive neurons were a consequence of that geometry. To test this hypothesis, the preferred egocentric location was re-calculated for these units exclusively when the animal was not in the center of the arena in between the loudspeakers. The observation that also at such extreme angles most neurons maintain their orientation-sensitivity (Fig. 3.22) is evidence that this neuronal type is not an effect of the setup's geometry.

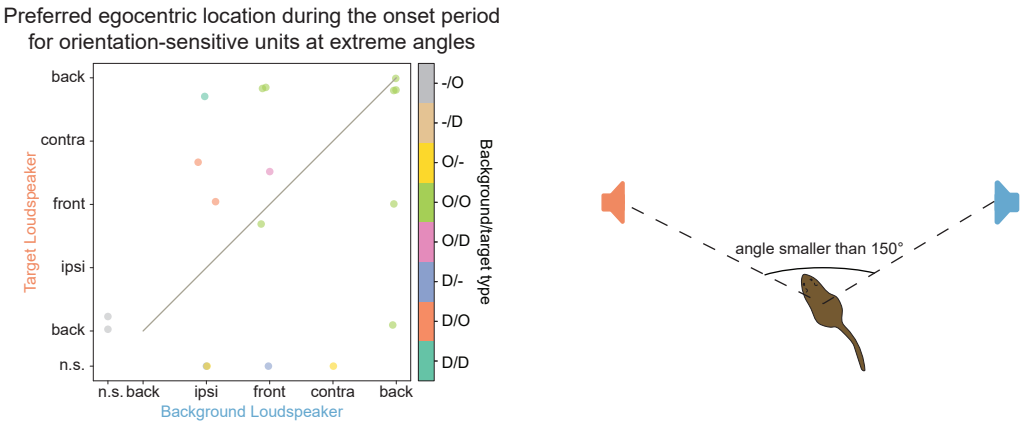


Figure 3.22 – Preferred egocentric sound-source location correspondence between the target and the background loudspeakers during the onset period calculated for orientation-sensitive units only at extreme angles (left) and respective explanation of what is considered an extreme angle (right)

3.3.4 Diversity in spatial tuning formation

A large fraction of neurons were canonically tuned to contra- or to ipsilateral sound-source locations, as explained in the last section. Yet this spatial tuning was originated from different firing rate's relationships between situations in which the sound-source was ipsi- or contralaterally located (Fig. 3.23). In Fig. 3.23a the neuron was strongly modulated at both contra- and ipsilateral sound-source locations; however, while for the first sound-source location the firing rate dropped, it increased for the latter sound-source location. Therefore, we must be reminded that to classify neurons as onset-increase or -decrease is a simplification of the actual situation, since the same neuron can show excitation and suppression depending on where the sound-source is located.

As earlier mentioned, the spatial tuning can also arise from a larger decrease in the firing rate for one location than for the other, as shown by the neuron in Fig. 3.23b. Despite this neuron being ipsilaterally tuned during the onset (as the firing rate is larger for this location), the largest response (suppression) occurred when the sound-source was contralaterally located. Interestingly, the ipsilateral suppression occurred for the offset response, which resulted in an untuned offset response.

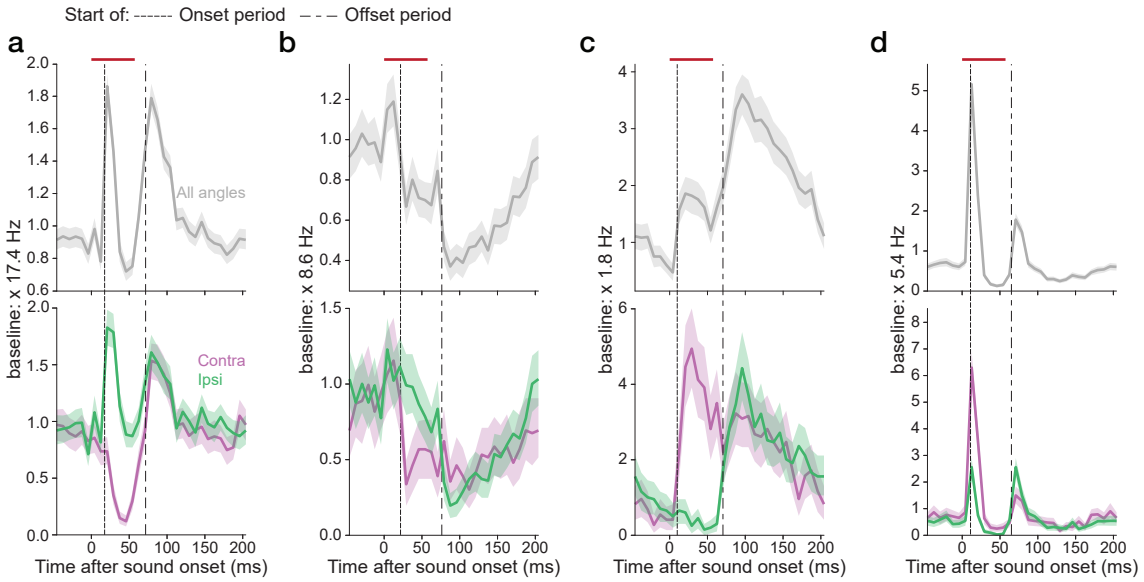


Figure 3.23 – Representative PSTH of neurons with offset responses and a canonical ipsi-/contralateral tuning during the onset period and untuned during the offset period for the background loudspeaker **a, b** ipsilaterally tuned; **c, d** contralaterally tuned; The upper panel represents the PSTH calculated for all angles and loudspeakers. The lower panel represents the PSTH for the background loudspeaker separated by location of sound-source location (ipsi/contra$\pm$45°). The solid lines correspond to the median and the shaded area to the 95% confidence interval. The red solid line represents the sound stimulation. The baseline corresponds to the engaged spontaneous firing rate.

Similarly, in Fig. 3.23c the neuron also exhibited an offset response for the ipsilateral location after an onset response for the contralateral location. However, because in this case the response was excitatory, the preferred egocentric location was considered to be contralateral.

Despite showing both an onset and offset responses, during the onset period the neuron in Fig. 3.23d showed a larger contralateral excitation, whereas during the offset period this difference in excitation was no longer present.

3.3.5 Temporal analysis of spatial tuning

A subset of neurons was spatially tuned in more than one time period (Fig. 3.24, Fig. 3.25), some inclusively changed their preferred sound-source location throughout time.

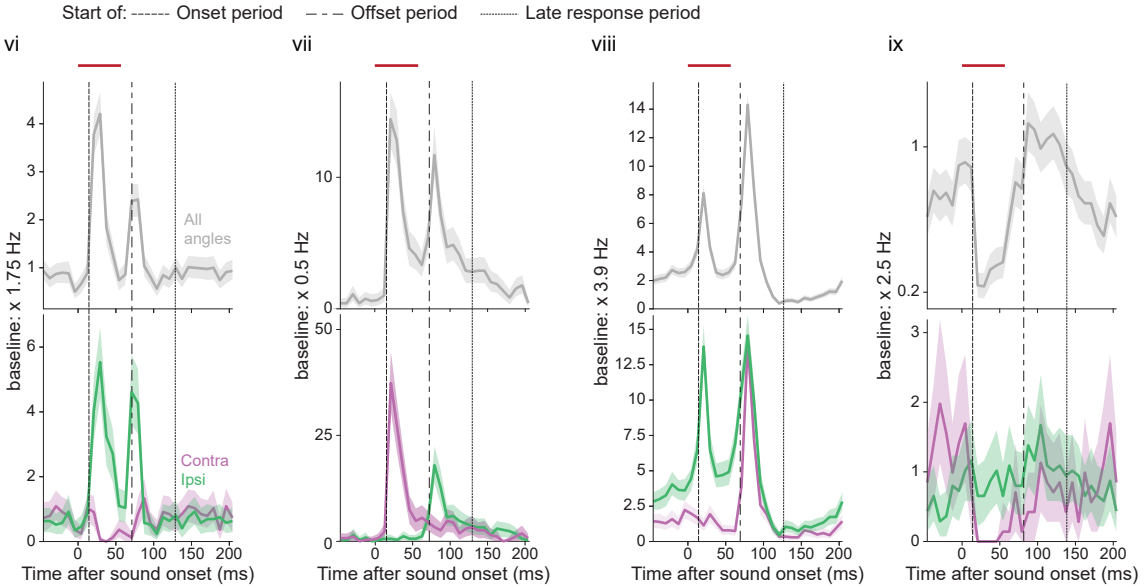


Figure 3.24 – Representative PSTH of neurons that are spatially tuned in several analyzed time periods **vi, vii** spatially tuned during onset and offset periods; **viii, ix** spatially tuned during onset and late response periods; **vi, viii** maintained the preferred egocentric sound-source location; **vii, ix** changed the preferred egocentric sound-source location. The upper panel represents the PSTH calculated for all angles and loudspeakers. The lower panel represents the PSTH separated by location of sound-source location (ipsi/contra±45°). The solid lines correspond to the median and the shaded area to the 95% confidence interval. The red solid line represents the sound stimulation. The baseline corresponds to the engaged spontaneous firing rate.

A few neurons showed both an onset and an offset response to an ipsilaterally located sound-source, not showing any response to when the sound-source was contralaterally located (neuron vi). Other neurons exhibited an onset response when the sound-source was contralaterally located and an offset response when ipsilateral (neuron vii). This is in accordance with the observation that synapses responsible for the onset responses do not correspond to the ones causing the offset responses [159].

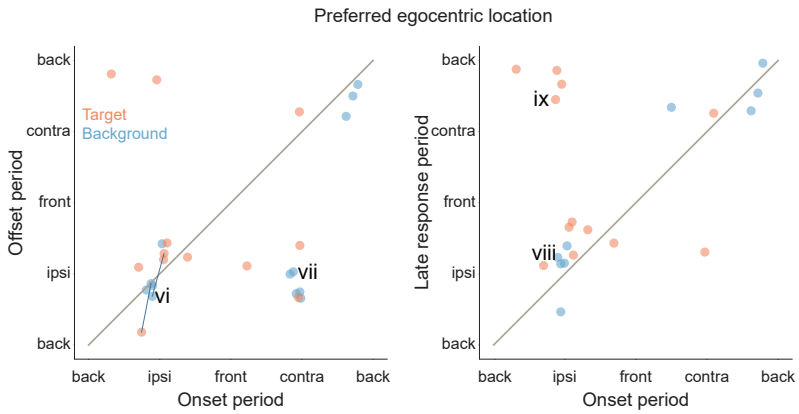


Figure 3.25 – Temporal analysis of spatial tuning Preferred egocentric sound-source location comparison between the onset and the offset time periods (left panel, 23 neurons) and between the onset and the late response periods (right panel, 21 neurons). The dark blue lines connect background and target preferred egocentric locations that correspond to the same neuron.

Similarly interesting are the differences between the responses to both loudspeakers during the late response period (which blend into the effective in-task baseline) depending on the egocentric location of the sound-source. Such differences are present both in neurons that increase as well as decrease their firing rate during onset (neuron viii and ix, respectively).

Both neuron vii and viii are in line with earlier studies that observed that the offset responses at the cortical level are unrelated to preceding inhibition [160]. Furthermore, in macaques, the onset was very similar to the offset spatial tuning [161]. Contrarily, in anesthetized ferrets, the onset and the offset spatial responses were negatively correlated both for ILDs and for ITDs [162]. Our study captured both neuronal behaviors as it is shown in the left panel of Fig. 3.25 (neuron vi and vii, respectively).

Example neuron with opposite spatial tuning for both loudspeakers

Most neurons that exhibited a differential spatial tuning to both loudspeakers decreased their firing rate during onset (Fig. 3.21c), being this time period not the most informative regarding spatial information for these neurons. Contrarily, the late response period, which can be considered as the baseline while in task since the sounds are successively repeated, changed not only with the egocentric location of the loudspeaker but also with the active loudspeaker at each sound presentation (Fig. 3.26). The fact that this in-task baseline is tuned suggests that this modulation is somewhat slower and most likely originated in higher order areas.

The opposite preferred egocentric sound-source location for the two loudspeakers could be a consequence of the geometry of the setup, given that the loudspeakers were opposite to each other, being the animal most of the time in the middle. To investigate whether this particular spatial tuning was related to the animal being in the middle be-

tween the two loudspeakers, the spatial analysis was recalculated only for extreme positions of the animal, in which the loudspeakers were not on each side of the animal (similar to Fig. 3.22). If the preferred egocentric locations followed the loudspeakers position relative to the animal, the difference in the preferred egocentric location should be smaller than 150°. Because that change in the spatial tuning was not observed in this situation (not shown), it is possible to conclude that this differential spatial tuning is not due to the loudspeakers being located on each side of the animal opposite to each other.

Furthermore, to ensure that this differential spatial tuning was not a consequence of a combination of a biased orientation of the gerbil in the arena and neuronal rate adaptation, the spatial analysis was recalculated exclusively for the first second after (re-)entering the target/background areas. Also in this situation no changes in the spatial tuning were observed (not shown).

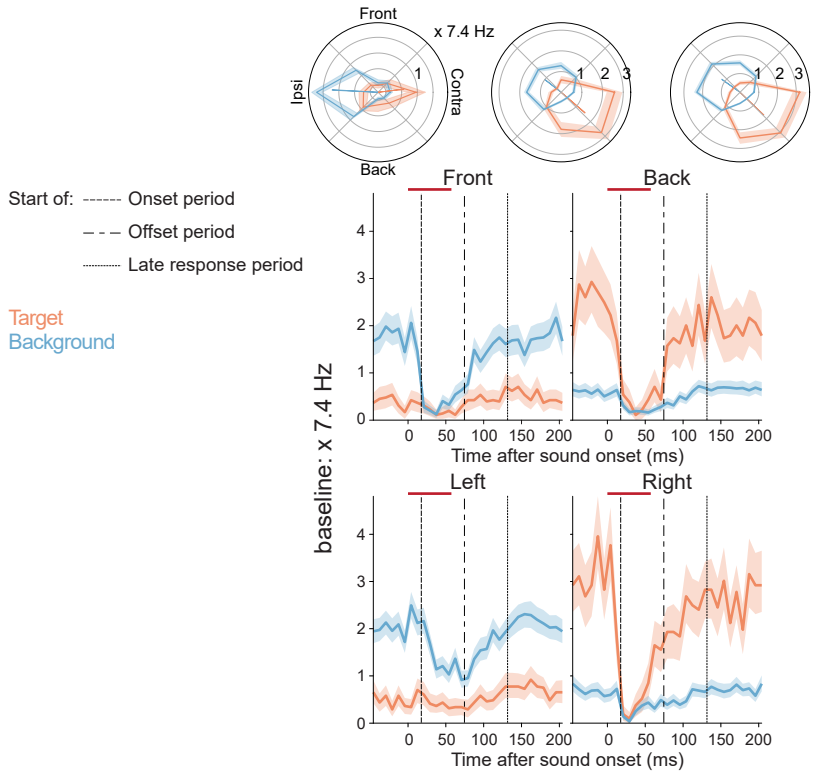


Figure 3.26 – Example neuron with differential spatial tuning Polar plots calculated for the onset, offset and late response periods (upper panel). PSTH separated by egocentric location of the sound-source (corresponding to the angular bins represented in the polar plots) and by active loudspeaker (lower panel). The solid lines correspond to the median and the shaded area to the 95% confidence interval. The red solid line represents the sound stimulation. The baseline corresponds to the engaged spontaneous firing rate. The colored lines in the polar plots indicate the preferred tuning angle, with the length scaling with the vector strength. This neuron corresponds to the neuron v in Fig. 3.19.

Example neuron with non-overlapping spatial tuning curves during the late response period

Many onset- spatially tuned neurons exhibited non-overlapping spatial tuning curves during the late response period, either by firing more to one egocentric location for a particular loudspeaker than for the other (Fig. 3.27) or by firing overall more to one loudspeaker than to the other during this time period. However, because often the vector strength is not large enough to consider the responses spatially modulated, the difference in the responses to both loudspeakers is not easily represented. It is nonetheless crucial to remember that also these small differences can be used by the brain to encode loudspeaker allocentric location or identity.

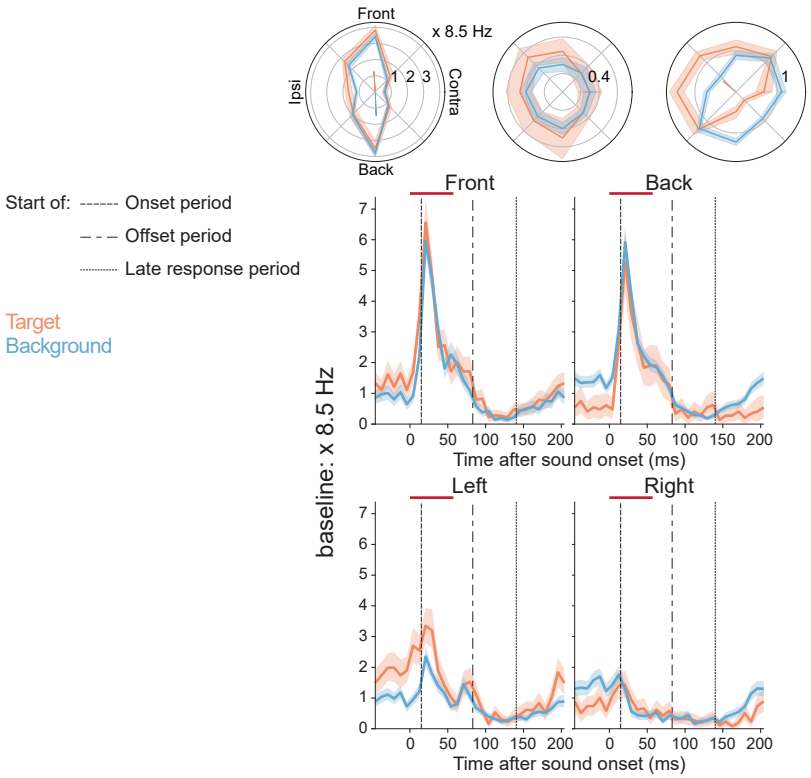


Figure 3.27 – Example neuron with non-overlapping spatial tuning curves during the late response period Polar plots calculated for the onset, offset and late response periods (upper panel). PSTH separated by egocentric location of sound-source (corresponding to the angular bins represented in the polar plots) and by active loudspeaker (lower panel). The solid lines correspond to the median and the shaded area to the 95% confidence interval. The red solid line represents the sound stimulation. The baseline corresponds to the engaged spontaneous firing rate. The colored lines in the polar plots indicate the preferred tuning angle, with the length scaling with the vector strength.

3.3.6 Spatially untuned neurons with differential magnitude between loudspeakers

Notably, in a sizable fraction of neurons that classified as spatially untuned to either loudspeaker in all analyzed time periods (see Section 2.4.5), the response magnitude differed significantly between the two loudspeakers (16/40, 40%, Fig. 3.28a,b), indicative of "pure" allocentricity, i.e. sound-source angle independent coding. These neurons differ from the previously presented type of neurons in the fact that while the neuron in Fig. 3.27 is spatially tuned during a time period, the untuned neurons show no spatial tuning in all of the analyzed periods. Population-wide the amount of untuned neurons that responded the most to the target loudspeaker did not significantly differ from those that fired the most to the background loudspeaker (Fig. 3.28b). Furtheremore, whereas the latencies were not distinct between the untuned neurons that showed no magnitude difference and those that did, the latter ones exhibited both a larger spontaneous as well as evoked firing rates (Fig. 3.28c).

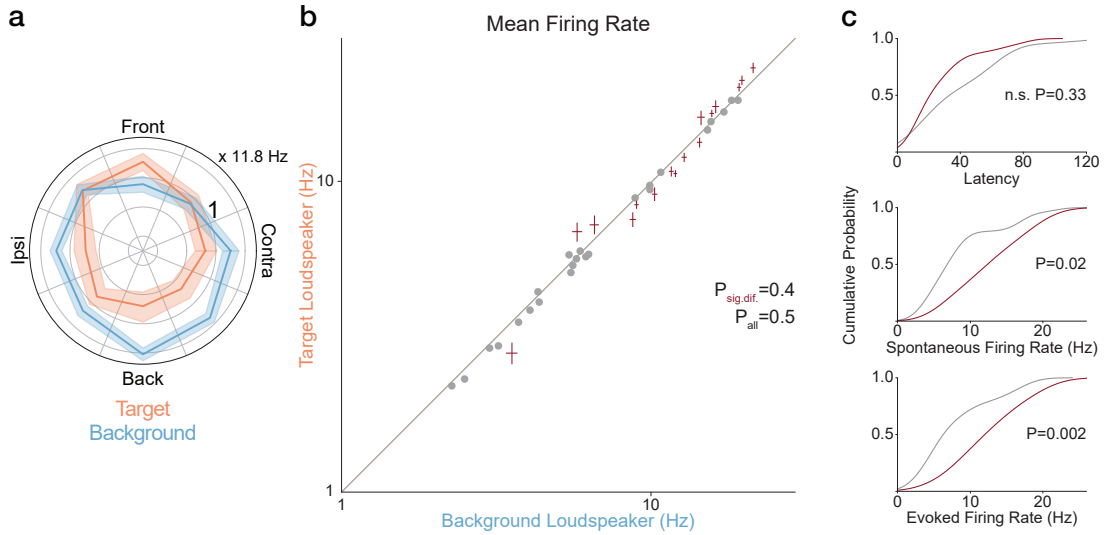


Figure 3.28 – Spatially untuned units with differential magnitude to both loudspeakers a Polar plot representing one untuned neuron with differential magnitude to both loudspeakers during the onset period. The solid lines correspond to the median and the shaded area to the 95% confidence interval. **b** Comparison of the mean firing rate of the spatially untuned neurons in all time periods between the target and the background loudspeakers (Wilcoxon signed-rank test). The crosses correspond to the 95% confidence interval of the mean firing rate of the neurons with differential magnitude (16 neurons) and the gray dots correspond to the ones without significantly different firing rate between responses to both loudspeakers (24 neurons). **c** Cumulative probability distribution comparison of the latencies (upper panel), the spontaneous firing rates (middle panel) and the evoked firing rates (lower panel) between untuned neurons with (red) and without (gray) significantly different responses to both loudspeakers. All p-values based on the two sample Kolmogorov–Smirnov test.

3.3.7 Animal's performance is reflected in the population response rates

The neuronal responses in the AC reflected the behavior of the animal in the task, since the mean firing rates to the target loudspeaker were significantly lower during the last 2 s before the animals correctly finished the trial as opposed to during the last 2 s before they wrongly left the island (Fig. 3.29a). As expected, this difference in firing rate was enhanced relative to the first 2 s after entering the target island (Fig. 3.29b).

Neuronal rate adaptation to the sounds could explain this difference, given that the animals certainly stayed longer in the island when they correctly finished a trial, as opposed to when the animals left (for this analysis only situations where the animal stayed for 2 s in the target island were considered – see Section 2.4.2). To tackle this question, the neuronal responses when finishing a trial were compared to the responses to the background loudspeaker (Fig. 3.29c). Whereas the mean firing rate did not differ in the first 2 s after (re-)entering the island/background area, the difference was evident when considering the time period from 4 to 6 s, being the responses to the target considerably smaller than to the background loudspeaker. Therefore, one can conclude this phenomenon is not caused by neuronal rate adaptation.

Movement was shown to reduce activity in the AC [98, 158], and since the animals have longer moments without movement before finishing a trial, the animal's locomotion state should not contribute to the observed effect. However, to ensure the differences were not consequence of differential levels of locomotion in the two conditions, the analysis was redone for moments in which the animal did not move (Fig. 3.29d). Also in this case the effect was present, proving that indeed it was not a consequence of motor activity.

As shown in earlier sections, the neuronal activity was modulated by the egocentric sound-source location. If the animal showed a particular turning behavior while in the target island before finishing a trial, this would be visible in the neuronal responses and it would be possible to produce a reduction in firing rate. To test this possibility, the same comparison was calculated exclusively for moments when the loudspeaker was ipsilaterally located ($\pm 45°$). Also this re-analysis showed the same effect (Fig. 3.29e), demonstrating it is also not caused by a spatial bias in the orientation of the animal.

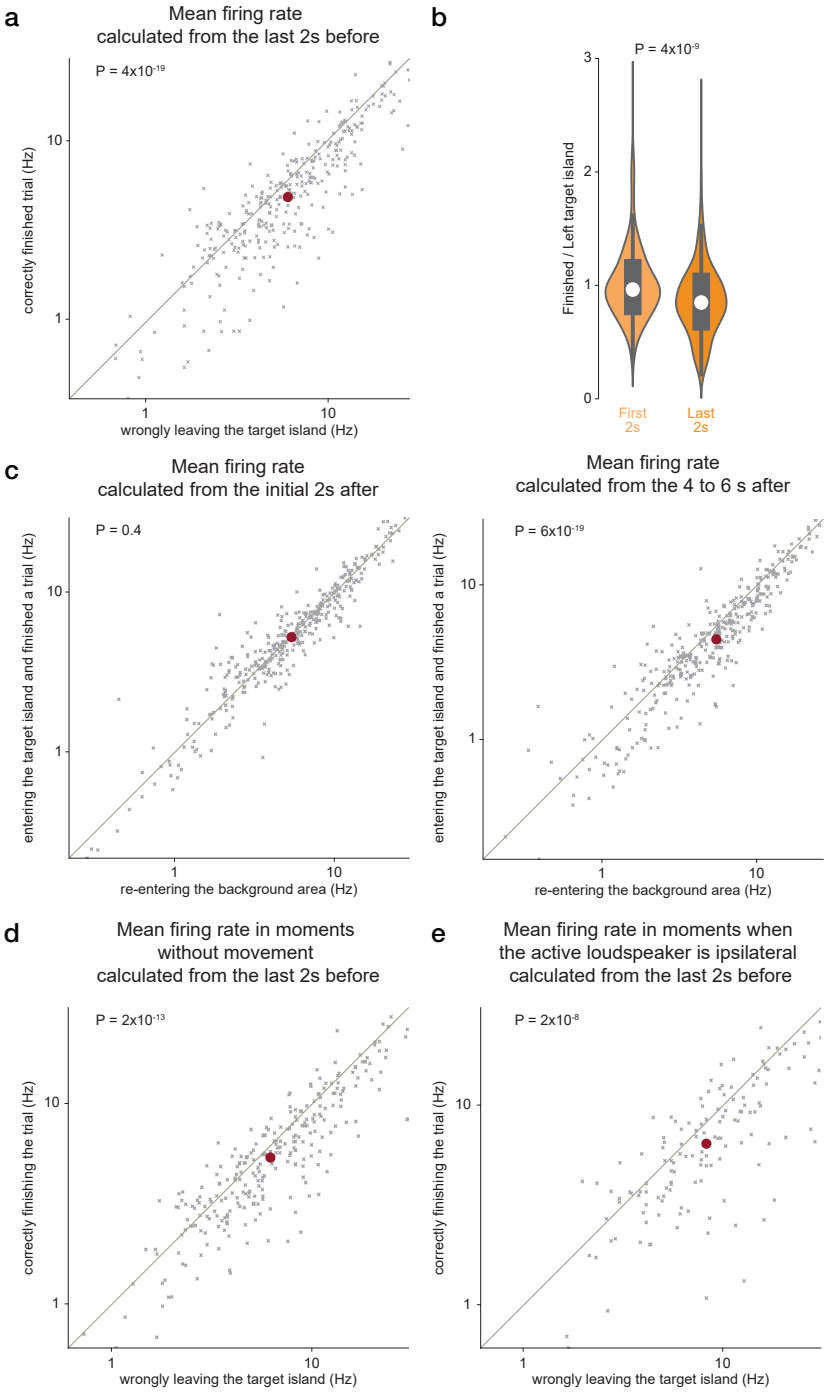


Figure 3.29 – **Changes in firing rate are specific to trial-outcome a** Firing rates were significantly lower in the last 2s when the animal finished a trial correctly compared to when it wrongly abandons the island (321 neurons) **b** This decrease in firing rate was enhanced with more time spent in the island (Mann-Whitney U-test, first 2s: 346 neurons, last 2s: 321 neurons). Violin plots surrounding boxplots (the median is represented by a white dot) depict the distribution of the ratios between the mean firing rate when the animal correctly finished a trial and when it wrongly left the island. The decrease in mean firing rate when the animal finishes the trial is not due to **c** unspecific neuronal rate adaptation (360 neurons); **d** movement (267 neurons); **e** a spatial bias in the orientation of the animal (137 neurons). The red dots correspond to the median values. All the mean firing rate plots possess log-log scale axis. The Wilcoxon signed-rank test was used in all non-specified analyses.

3.3.8 Neuronal tuning is sharper for the target loudspeaker

Peculiarly, responses to the target loudspeaker were generally more egocentric location-specific than to the background, demonstrated by their larger vector strength (Fig. 3.30a). This signifies that the responses to the target sharpened in comparison to the background loudspeaker.

In line with what was found in the visual [163] and somatosensory cortices [164], also here the fast-spiking neurons showed smaller tuning than the regular spiking neurons (Fig. 3.30b), as demonstrated by the latter's larger vector strength.

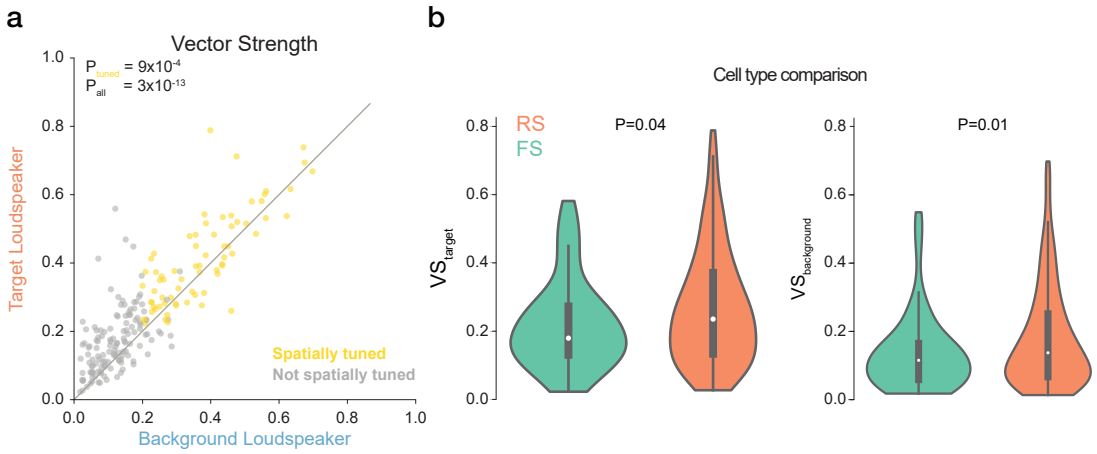


Figure 3.30 – **Vector strength a** was larger for the target than for the background loudspeaker (Wilcoxon signed-rank test). Yellow data points represent neurons spatially tuned to both loudspeakers (71 neurons) and data from neurons that were not significantly spatially modulated for at least one of the loudspeakers are shown in gray (133 neurons). **b** was larger for the regular than for the fast spiking neurons for both loudspeakers (204 and 354 neurons for target and background, respectively, median $VS_{FS_{target}}$:0.18 median $VS_{RS_{target}}$:0.24 median $VS_{FS_{background}}$:0.12 median $VS_{RS_{background}}$:0.14, Mann-Whitney U-test).

3.3.9 Spontaneous firing rate increases during engagement

The effect of attention or task engagement in the spontaneous firing rate is still a point of debate in the field with studies that show a decrease [118, 119], a maintenance [165, 166] and an increase [104, 167–169] in the firing rate with task engagement. To address this question, a population-wide comparison of the spontaneous firing rate was calculated between passive and engaged situations. The passive situation corresponds to an out-of-task 2 min period at the end of the training (see Section 2.4.2), calculated for periods of time when the animal was stationary. The engaged situation corresponds to the 1 s during which the animal was on top of the initiation platform to self-initiate a trial. As in [104, 167–169], the spontaneous firing rate in an engaged situation increased relative to a passive situation (Fig. 3.31).

The potential causes for the observed difference between our study and [119] may relate to the fact that, in the latter, the calculation of the spontaneous firing rate was performed for the time immediately before the tone onset when the animal was expecting a sound. Also in [118], the spontaneous firing rate that decreased relative to the disengaged situation corresponded to a time when the animal was expecting a sound that was not presented, a completely different situation from ours when the animal's expectations were met.

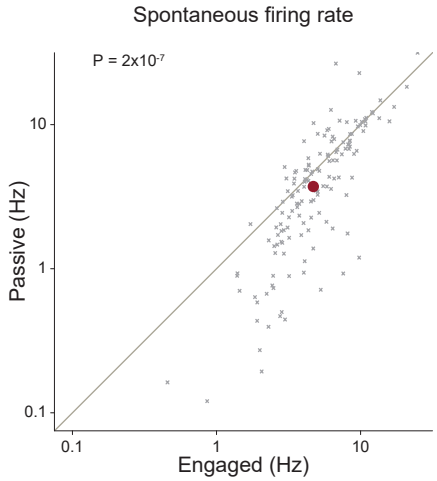


Figure 3.31 – **Spontaneous firing rate** comparison between engaged and passive conditions (153 neurons, Wilcoxon signed-rank test). The red dot corresponds to the median value.

The spontaneous firing rate values observed were similar to what was previously reported in awake animals [62].

3.3.10 Simultaneously decoding the egocentric sound-source location and the loudspeaker identity

An artificial neural network (ANN, see Section 2.4.7 for details) was implemented to better understand how the observed diversity of neuronal spatial tuning could be combined into useful information about the identity and egocentric location of the different loudspeakers – which is essential to solve the task.

The ANN was trained to decode a combination of egocentric sound-source location and loudspeaker identity (world-based reference frame). By merging the data from both loudspeakers it was possible to calculate the accuracy of the decoder on the egocentric location of the active loudspeaker. The ANN classified the egocentric locations with high accuracy (Fig. 3.32a). By merging the data from all egocentric locations to each loudspeaker, the accuracy of the algorithm in identifying the active loudspeaker irrespective of the egocentric location was calculated. Remarkably, the decoding accuracy for identifying the sound-source was also highly significant (Fig. 3.32b), demonstrating the coexistence of both subject-based and world-based reference frames (identity) in the neuronal representations (Fig. 3.32c).

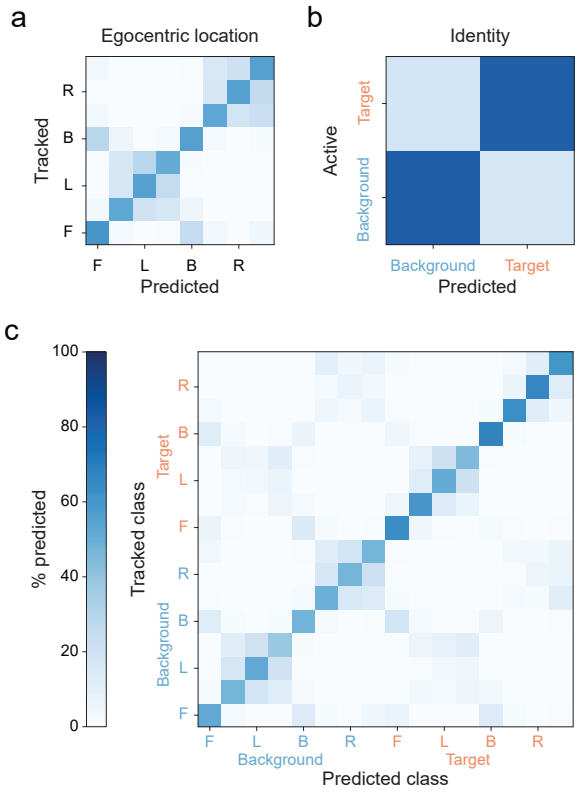


Figure 3.32 – **Confusion matrix** for decoding **a**, the egocentric loudspeaker location; **b**, the identity of the active loudspeaker; **c**, the combination of the loudspeaker location and identity. F: front, L: left, B: back, R: right. For the construction of all the confusion matrices, 38980 values were used across 20 sampling cycles (1949 values per cycle).

In Fig. 3.32c, the accuracy to the target loudspeaker (diagonal values) is higher than the one to the background (60±8% for the target loudspeaker and 47±4% for the background – mean±std calculated across classes), meaning that the target loudspeaker was better localized than the background.

A temporal analysis of the decoding performance indicates that the egocentric location of the active loudspeaker was largely encoded within the first 100 ms after sound onset, whereas the information about the identity of the loudspeaker increased monotonically throughout the analyzed period (Fig. 3.33) and reached behavioral performance levels (Fig. 3.1). This result confirms what was already suggested by the comparison between the different analyzed time periods of the preferred egocentric location to the target and the background loudspeakers (Fig. 3.20).

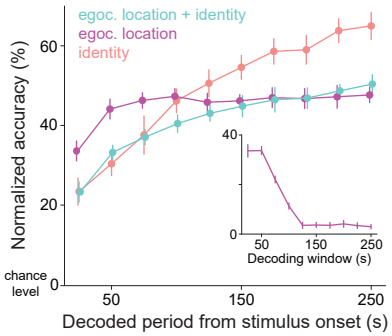


Figure 3.33 – **Decoder accuracy** as a function of the cumulative time after sound onset used for decoding. Inset: Accuracy of decoding the egocentric sound-source location as a function of the time after onset of the 25 ms bin used for decoding. The normalized accuracy of the egocentric location drops linearly from over 30% until 50 ms after sound onset to almost chance level after 125ms after sound onset. All error bars correspond to the standard deviation of the mean normalized accuracy across 20 sampling cycles. The 0% of the normalized accuracy corresponds to the chance level accuracy and the 100% normalized accuracy to 100% accuracy.

To explore to what extent the diverse spatial tunings across the population of neurons contributed to the egocentric and identity decoding accuracy, units with specific tuning characteristics were selectively eliminated from the test data. For that, the corresponding spike counts were set to zero throughout all the time bins (Fig. 2.13) and the decoding accuracy was re-assessed. Elimination of either all units with differential spatial tuning for the two sound-sources (18 units, compare Fig. 3.26) or exclusion of the same number of units with either diverse response profiles (randomly selected – different units were selected in each sampling cycle) or exclusively canonic hemispherically tuned units (compare Fig. 3.20, type “i” and “iii”) resulted in only mild reductions in the decoding accuracy of the egocentric location and the identity (Fig. 3.34a), demonstrating the robustness towards these interventions.

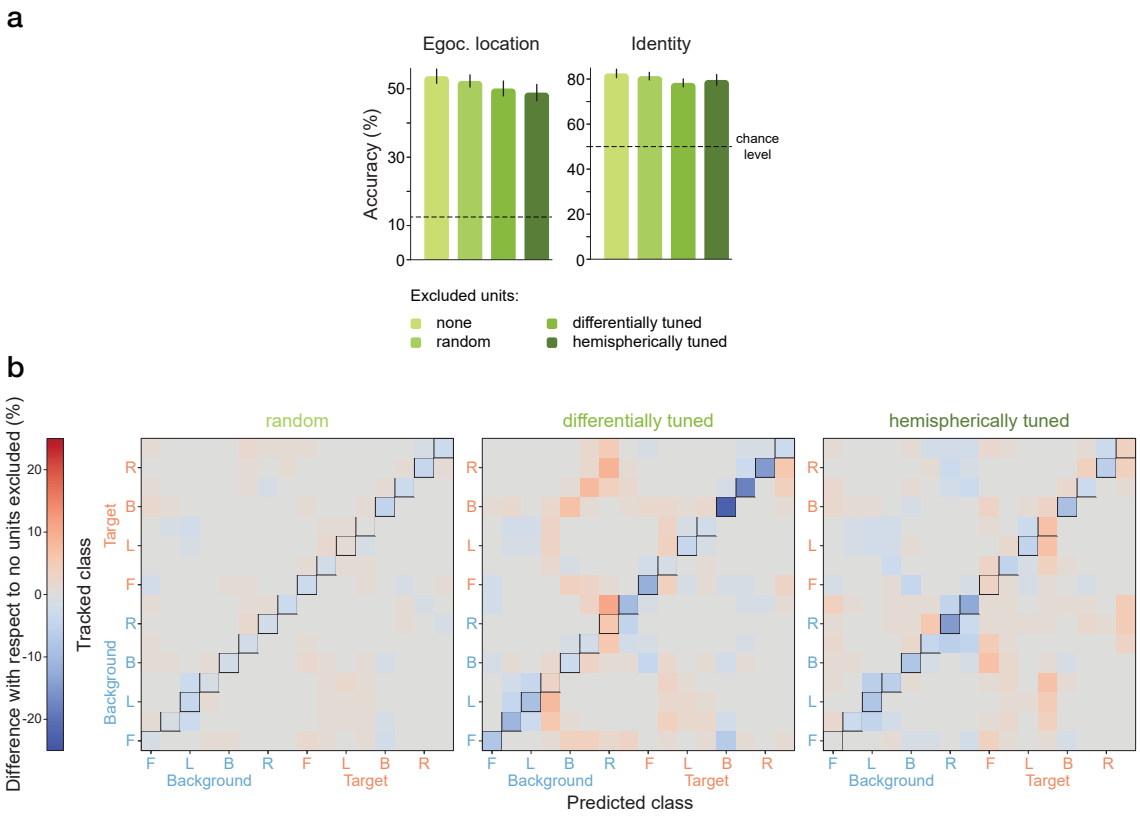


Figure 3.34 – Influence on the decoder performance of eliminating from the test dataset 18 units belonging to different spatial tuning classes **a** Comparison between eliminated classes of the accuracy of the decoder in determining the egocentric sound-source location (left) and the loudspeaker identity (right). All error bars correspond to the standard deviation of the mean accuracy across 100 sampling cycles. **b** Difference in the confusion matrix relative to the original test dataset when 18 randomly selected units (left), 18 units with differential spatial tuning (middle) or 18 canonic hemispherically tuned units (right) are excluded from the test dataset. Number of sampling cycles: 100.

As expected, the drop in the accuracy for decoding the egocentric sound-source location was the largest when eliminating the canonical hemispherically tuned units ($-4.8 \pm 2.4\%$, mean$\pm$std). Accordingly, this drop is most evident for the left and the right sound-source locations (Fig. 3.34b - right panel) - the preferred egocentric sound-source locations of these units. The drop in accuracy by eliminating random units was minimal ($-1.4 \pm 1.8\%$) and, similarly to eliminating differentially tuned units ($-3.6 \pm 2.2\%$), was not significantly different from not excluding units ($0 \pm 2.1\%$) - which can be observed by the overlap of the standard deviations.

Also not surprising is the fact that the largest drop in the accuracy for decoding the identity of the active loudspeaker occurred when the units with a differential spatial tuning to both loudspeakers were excluded (-4.2 ± 1.8). Fig. 3.34b (middle panel) evidences the confusion between loudspeakers' identities since the drop in the correctly predicted values (diagonal) is accompanied by an increase of the percentage decoded in the same egocentric sound-source location yet by the other loudspeaker. Once more, for decoding

the identity of the sound-source, elimination of random units produced a non-significantly different accuracy from the condition in which no units were excluded ($-1.2 \pm 1.7\%$), as well as when hemispherically tuned units were eliminated ($-2.9 \pm 2.4\%$ – compare to $0 \pm 1.9\%$).

CHAPTER 4

Discussion

4.1 Behavioral paradigm

In this thesis, I reported on the development and implementation of a new behavioral paradigm – the Sensory Island Task (SIT). In its sound localization version, animals must forage in the arena for an area that prompts a change in the active loudspeaker to a specific target loudspeaker and appropriately report its detection to get rewarded. In most of the earlier studies, animals were either anesthetized [170], must hold a constant head position [127] or had their movements restricted [79]. In contrast to these more conventional approaches, in SIT animals freely move and engagement is required for a successful task performance. By allowing voluntary self-motion and taking advantage of the animals' natural exploratory and goal-oriented behavior, SIT can be considered an ecologically relevant paradigm for the study of sound localization. Furthermore, the egocentric and allocentric representations of the sound-sources are decoupled in SIT, enabling the study of how static sound-sources are encoded in the brain of moving animals. This question is often neglected but constitutes a major challenge for a system that is not topographically represented in the sensory epithelium. Importantly, egocentric information is not sufficient for the animals to be able to succeed in this task, which requires the use of a combination of ego- and allocentric cues.

I showed that gerbils could learn the task and performed well above chance level (Section 3.1.1). By introducing catch trials it was possible to verify that their behavior was sound-source/allocentric location specific (Section 3.1.2). Moreover, a bias in the spatial location of the target island was implemented for some animals, which incorporated this bias into their exploration strategy (Section 3.1.3). This suggests that the task is not a mere GO/NOGO paradigm, but that it includes the active search of target island locations by the subjects. Furthermore, the animals performed well in the task even after the physical swap of the loudspeakers, showing that they were not reporting a potential specificity of a loudspeaker (Section 3.1.4).

Gerbils in SIT moved typically in short bouts with short stationary moments in between. The duration of these stationary moments did not significantly differ between target and background areas. This demonstrates that the animals did not completely halt their movement in the target, further reinforcing the difference in nature between SIT and a GO/NOGO task (Section 3.2.1). Interestingly, most animals turned towards the direction of the target loudspeaker before finishing a trial (Fig. 3.11). This suggests that this behavior is advantageous to localize the auditory stimuli, which is in agreement with sound localization being the most accurate in the midline [7]. In the background area, the animals were equally likely to start a bout in a locomotion state with or without translation; whereas in the target island, they tended to avoid translational movement (Section 3.2.2). This again points towards an association by the gerbils between a stimulus change and a physical spatial location – the target island.

4.2 Neuronal recordings

After achieving a good performance in the task, a tetrode bundle was implanted in the left A1 of the trained gerbils. The observed neuronal latencies were typically short (<20 ms, Section 3.3.1), as expected for recordings performed in the A1 [139]. Nonetheless, a few neurons responded only to the sound offset, visible by the temporal alignment of their latencies with the sound termination – in line with findings reported in the literature [77, 155, 156]. Curiously, the latency calculated for sounds coming from the target loudspeaker was slightly longer than the one for the background loudspeaker. A study where monkeys were performing a frequency discrimination task [171] reported on latency differences between a rewarded and an unrewarded condition for the same sound, with neurons exhibiting a longer latency in the rewarded condition; this indicates that such a latency difference could be caused by reward expectation. Alternatively, it could be a consequence of different internal states of the animal since in [155] longer latencies in the A1 were reported when cats were localizing in comparison to a passive condition. This suggests that the gerbils in the experiments reported here might have been localizing sounds more vigorously during presentations by the target loudspeaker.

A response to a sound did not always correspond to an increase in the firing rate, since some neurons decreased their firing rate relative to the baseline. As reported in [156], this is typically observed in layer V or VI of the AC and related to intracortical inhibition.

Characterization of the neuronal types

Based on the spike waveform, two types of neurons were found: the fast spiking neurons and the regular spiking neurons, putative inhibitory and excitatory neurons, respectively [151–153] (Section 3.3.2). The first exhibited shorter latencies and higher firing rates than the latter, corroborating earlier research [157]. Interestingly, the fast-spiking neurons were not as sharply tuned as the regular spiking ones, a similar result to what was observed in the visual [163] and somatosensory [164] cortices. There, the weaker tuning of the inhibitory PV+ neurons is thought to be a consequence of them receiving inputs from several neurons with diverse tuning, possessing a gain control function of the ascending projections, whereas the excitatory pyramidal neurons receive mostly inputs from cells with a similar tuning. Such a process could also occur in the AC.

Spatially tuned neurons for sounds from both loudspeakers

A large variety of spatial tuning curves was observed (Fig. 3.19), including the canonic broadly tuned hemispheric neurons [172], both with contra- and ipsilateral preference. Interestingly, some neurons showed orientation sensitivity with a preference for the midline, which corresponds to a 0 μs ITD preference. These neurons have been described in awake animals [111, 112, 155] as well as in anesthetized ones [60, 173] for ILDs, where they were described as receiving excitatory inputs from both ears [60]. They were already reported to exist in the MGB [43] where the experimenters also stimulated these neurons

monaurally and described them as excitatory from the contralateral side and inhibitory from the ipsilateral or as both sides being excitatory. This signifies that the same spatial tuning can arise from different computations. Also the IC exhibits these neuronal spatial tuning type [174], suggesting that the computation could be primarily performed at this stage.

The formation of the preferred egocentric sound-source location was also briefly discussed (Section 3.3.4). Some neurons exhibited an increased firing rate relative to the baseline for one egocentric sound-source location, whereas the response was decreased for the opposite location. Other neurons simply exhibited a larger firing rate for one egocentric sound-source location than for the opposite. Nonetheless, the results of what is the contribution of each ear to the response was not possible to pinpoint in this paradigm due to the stimulation always being binaural.

The largest majority of the spatially tuned neurons exhibited a tuning preference for a cardinal direction (front/back/ contra/ipsi), arising from the existing symmetry of the diametrically opposed ears that gerbils in particular and mammals in general possess. Moreover, this observation is in line with the proposed opponent-channel coding model [17] since our results can be explained by the MSO neurons being more active for the contralateral side. An "excitatory" contralaterally tuned neuron could be a direct reflection of an MSO neuron response, whereas an "excitatory" ipsilaterally tuned neuron could be the result of an excitation from a contralateral MSO neuron. Also a "suppressed" ipsilateral favoring neuron could be computed via some stages that transform the excitatory input from MSO into inhibition. At last, a midline favoring neuron could be a consequence of the sum of two MSO neuron responses, each favoring the opposite hemisphere, with the inflexion point being slightly off from zero. All of these computations are likely performed by several different processing stages (midbrain and thalamus) until reaching the AC and certainly with the output of more than one or two MSO neurons. Nonetheless, although the MSO neurons' tuning can explain the observed A1 egocentric spatial tuning curves, it cannot explain the existence of loudspeaker identity-specific neurons.

Loudspeaker identity-specific neurons

Surprisingly, a large proportion of neurons were sensitive to which loudspeaker was active (loudspeaker identity-specific neurons – Fig. 3.20). The largest subset of those were only spatially tuned to one of the two loudspeakers, typically to the target; whereas other neurons were spatially tuned to both loudspeakers and yet exhibited a large difference in the preferred egocentric sound-source location to sounds from the two loudspeakers. Furthermore, some spatially untuned neurons exhibited a magnitude difference between responses to both loudspeakers (Section 3.3.6), corresponding to "pure" allocentric neurons, i.e. sound-source angle independent. First hints of the existence of allocentric units were already found in one earlier study [113], where however the animals were not actively localizing sounds but merely looking for water. Future studies should clarify for our paradigm whether the observed magnitude difference arises from the allocentric position

of the loudspeakers or from their different behavioral meaning to the animal (one loudspeaker is associated with a reward whereas the other is not) since in this study the two factors are correlated.

The spatially tuned neurons that showed an opposite tuning (Fig. 3.26) for both loudspeakers could be interpreted as being sound-modulated head-direction neurons, given that they maintained their tuning also at extreme angles with the loudspeakers. They could help in combining a sound presentation with the world reference frame. The importance of head movements for sound localization was already mentioned in the 1930s [175]. There are some possible explanations for the existence of neurons with such information in the AC. The DCN principal neurons are known to show sensitivity to head-direction [176]. This information could be maintained throughout the ascending auditory pathway and arrive at the AC. Alternatively, this representation could arise from a connection with the retrosplenial cortex (RSC) [59], where 10% of neurons are reported to be head-direction selective in the rat [177]. In the visual system, head-motion information was observed in the layer 6 of the mouse V1, where it is integrated with visual signals [178] – a similar effect could also occur in the A1.

Another possible explanation for the existence of loudspeaker identity-specific neurons is spatial selective attention. In the visual system, several studies provide evidence for this effect. In [179], humans could better detect stimuli at attended locations. In the areas V2 and V4 of the macaque, baseline shifts occurred with higher firing rates when the attended location corresponded to the receptive field of the neuron [180]. Furthermore, the membrane potential of mice V1 neurons was reported to be depolarized at a behaviorally relevant location [181]. Such a phenomenon could also explain what we observed. In fact, also in the auditory modality, humans were better at identifying birdsongs if they knew where the songs would be played from [2]. Neuronally, the firing rate of some neurons in the AC of monkeys that were performing a lateralization task was shown to increase when the sound was presented at the attended ear [182]; however, without a corresponding change in the spontaneous activity. This spatial attention could be driven by the cingulate cortex, where spatial information was reported to be allocentrically represented [183], connects to A1 [59] and was referred to as having an analogous function to the primate frontal eye field (FEF), known for its role in spatial selective attention [184]. The FEF exhibited auditory information [185] with very short latencies and inclusively offset responses were represented [186]. Furthermore, the FEF was involved in auditory spatial selective attention not only in people with closed eyes but also in congenitally blind subjects [187]. In mice, the cingulate cortex was shown to enhance visual spatial selective attention via its projections to V1 inhibitory interneurons [188]; a similar circuit could potentially occur for the auditory modality.

Temporal analysis of spatial tuning

The spatial tuning was calculated for three different periods of the PSTH: onset, offset and late response (Fig. 3.20). Whereas during the onset period, the majority of spatially tuned neurons preferred the canonical contra-, ipsilateral and midline egocentric sound locations, during the offset and the late response periods that proportion was reduced. Furthermore, during the late response period, the majority of spatially tuned neurons exhibited a differential response to both loudspeakers. Altogether, this suggests that the encoding of the egocentric location occurs preferentially during the onset period and of the loudspeaker identity during the late response period. While most spatially tuned neurons exhibited very short latencies (Fig. 3.21a), the few that had longer latencies corresponded to loudspeaker identity-specific neurons. This could be indicative of top-down modulation. Nonetheless, the latencies of other loudspeaker identity-specific neurons were as short as the ones of the canonically tuned neurons. The loudspeaker identity-specific neurons typically exhibited a decrease in the firing rate during the onset period (Fig. 3.21c), whereas the contralateral and midline favoring neurons increased their firing rate. The neurons that preferred ipsilateral egocentric locations could either increase or decrease their firing rate during onset. To rule out movement confounding responses, the preferred egocentric sound-source location was also determined for stationary moments, which did not result in any qualitative differences (Fig. 3.21e).

Interestingly, some neurons were spatially tuned in more than one time period (Section 3.3.5). Whereas a subset of those maintained the same preferred egocentric sound-source location, others changed theirs throughout time. Some neurons possessed an offset response oppositely tuned to the onset spatial tuning. The same effect was already observed in ferrets both for ILDs as well as for ITDs, strengthening its general character [162], and is in line with a study that showed that the synapses responsible for the onset responses do not correspond to the ones encoding the offset responses [159]. Contrarily, a study in macaques reported that neurons did not change their spatial tuning between onset and offset periods [161]. Here, both neuronal behaviors were captured.

Changes in the spatial tuning between onset and late response periods were also observed in this study. Furthermore, for many onset-spatially tuned neurons, the spatial tuning during the late response period was not significant (based on the vector strength) for the neuron to be considered spatially tuned; nonetheless, the spatial tuning curves did not completely overlap between the responses to both loudspeakers (Section 3.3.5). Both cases can be informative about the identity of the loudspeaker and this information could be relayed to higher order brain regions.

Late responses have already been suggested to be crucial for the creation of auditory objects [189]. In fact, they were observed in neuronal responses to bird chirps in the A1 of anesthetized cats [190] and passively listening monkeys [191], where these late responses were hypothesized to be relevant for echoic memory or for temporal integration. Importantly, such late responses were also reported to exist in behaving animals. When monkeys performed a task where they had to compare two frequencies with an intersound interval

of 1 s, the firing rate after that interval was increased relative to non-performing in more than half of the AC neurons [192]. In this study, the late response was interpreted as the basis for short term memory; however, attention and expectation were also referred to as possible causes. In [193], ferrets were trained to discriminate pitch in a task where they had to compare a target pitch to a reference and report whether the target had a higher or a lower pitch. Interestingly, despite the fundamental frequency being well encoded in the onset responses, the late responses predicted the reported pitch better, especially responses recorded in deeper layers. The anatomical evidence from our work (Fig. 2.5) and the fact that in our study many neurons exhibited late responses and decreased their firing rate in response to a sound (characteristic of deeper layers) suggest that our recordings were also mainly performed in deeper layers (however, the recording depth was lowered throughout the experiments).

Performance encoded in the population response

Similar to [193–198], the perception or choice of the gerbil in our task was also encoded in the neuronal recordings since the firing rate was significantly lower during the last 2 s before correctly finishing a trial (which in this paradigm corresponds to reporting a decision) in comparison to the last 2 s before wrongly leaving the target island (Section 3.3.7). This effect could not be explained by neuronal rate adaptation, nor by movement or egocentric sound-source location. Also in [171], a decrease in the onset excitation of AC neurons was observed when the animal was rewarded in comparison to an unrewarded situation for the same sound.

Neuronal spatial tuning is sharper for sounds from the target loudspeaker

The recorded neurons exhibited a sharper spatial tuning to sounds coming from the target loudspeaker than from the background loudspeaker (Section 3.3.8). This effect was also observed in the amplitude of the LFP evoked response, which similarly was larger for the target loudspeaker (data analyzed by Dardo Ferreiro and shown in [117] – see appendix B). Earlier studies showed that the need to localize a sound sharpened the neurons' spatial response [105], as well as when competing sound-sources were added [127, 199, 200]. This sharpening effect, combined with the latencies being longer for the target loudspeaker and the animals turning towards the loudspeaker in the target island, suggests that the animals localized sounds more vigorously in the target than in the background area: perhaps they were more engaged/attentive in the island due to reward expectation. An alternative hypothesis for this neuronal sharpening could be related to the pairing of the target loudspeaker with an LED light, which could have led to multisensory integration – already shown to enhance sound localization [69]. However, because the results indicate that the animals never did that association (Section 3.1.1), I would consider this an unlikely possibility.

Simultaneous decoding of the egocentric location and of the identity of the sound source

The implementation of an artificial neural network allowed for the combination of the neuronal responses almost in their entirety to form a framework that enabled a better understanding of the data acquired. This algorithm was capable not only of predicting with high accuracy the egocentric location of the active loudspeaker but also of identifying which of the two loudspeakers was active (Section 3.3.10). This proves that the interconnection of subject- and world-based information is present in the neuronal representations in the A1.

Interestingly, the decoding algorithm also exhibited phenomena that were already observed in psychophysics. Specifically, that the frontal sound-source location was mainly mistaken with the back [7, 201], an effect that is stronger for low frequencies and that was already observed in earlier neuronal recordings [74, 202]. This decoding property was probably influenced by the spatial tuning of the orientation-sensitive neurons (such as the neuron ii in Fig. 3.19) as well as by the broad hemispherically tuned neurons that have a sharp transition at the midline. This poor front/back discrimination effect is a consequence of the binaural cues being identical for the front and back and the spectral cues not being very effective at low frequencies. Nonetheless, the decoder attributed with high accuracy the sounds presented in the midline to the midline orientation (front/back) and was very good in lateralization – the sounds presented from both left and right hemifields were correctly assigned to the correct hemifield. Only sometimes were these incorrectly assigned to another location in the same hemifield (e.g. a left-back egocentric sound-source location could be confused with left or left-front locations). This effect is probably a consequence of the ITD's rate change decreasing for larger azimuths – which means that for the same change in azimuth, the ITD value changes less when the sound-source is laterally located than when it is located around the midline [203, 204]. This is in agreement with the slope of the neuronal response curve being the largest in the midline [17] and therefore the most informative at those azimuths. This effect of the decrease in the localization accuracy (and consequent increase in the minimum audible angle) for lateral azimuths was already behaviorally shown in humans [7, 201, 205, 206] and in cats [207, 208]. It was also earlier observed in a neuronal population in the AC of cats [209].

The accuracy in predicting the correct egocentric loudspeaker location was larger for the target than for the background, which indicates that the target loudspeaker was better localized than the background. This might be a consequence of the observed higher vector strength to the target loudspeaker (Section 3.3.8) and points towards a more engaged state of the animal when in the target island.

The temporal analysis of the decoder revealed that the onset response encoded mostly the egocentric location of the active loudspeaker, whereas the identity of the active loudspeaker was encoded throughout the inter-onset period, reaching behavioral levels. The comparison between the preferred egocentric sound-source locations during the onset and the late response periods for all the spatially tuned neurons already hinted at such a decoding property. Despite the neuronal input signaling the identity of the loudspeaker

presumably coming from higher order regions, the decoder already predicts the identity of the active loudspeaker above chance since the beginning of the stimulus. However, since the sounds are repeated every 250 ms, this might also be influenced by the beginning of a stimulation in reality following to the late response period of the neuron's response to the previous sound.

To better understand the influence of specific spatial tuning classes in the decoding, a subset (or in the case of the differentially tuned units, the whole set) of neurons of either hemispherically tuned, differentially tuned or randomly chosen neurons was eliminated. Overall, the artificial neural network demonstrated remarkable robustness towards these interventions, as decoding accuracy for both reference frames remained far above chance level regardless of the eliminated neuronal tuning class. This robustness suggests that distributed interactions in population responses underlie an interlaced coding of egocentric and allocentric information.

Conclusion

Taken together, this study signifies that sound localization in the AC can no longer be assumed to be static, hard-wired and completely egocentric. In contrast, it is highly modulated by context, reward expectation, choice, and even where in the world the sound sources are located during active navigation. Specifically, different areas in the world reference frame where sound sources are located can be of different relevance to the animal. Here, we show that such differences impact the coding of sound localization, perhaps by mechanisms of spatial selective attention. This study demonstrates the necessity of incorporating active navigation, engagement and behavioral relevance in further sound localization research to better understand everyday life auditory scene analysis.

4.3 Future directions

Based on the results presented here, there are many potential experiments that could further deepen our understanding of how sound localization is performed in an active and natural environment.

One open question, for example, is whether the need to better localize sounds results in a sharpening of the neuronal responses. In [105] it was shown that the neuronal tuning sharpens between an auditory task where the animal does not need to localize and one where localization is crucial. However, it did not explore whether there are various degrees of sharpening depending on the difficulty of the sound localization task. This could for example be studied in the SIT paradigm by gradually bringing together in space the target and the background – which would increase the task difficulty. Increased task difficulty was previously shown to result in a larger neuronal signal-to-noise ratio in a sound detection task in ferrets [210].

In our experiments, the gerbils were highly trained before the recordings were performed. This training could influence the neuronal responses observed and therefore we could be reporting on phenomena that are not prevalent in nature. An experiment that could easily be implemented is the comparison of our results with those of naive animals. Even more interesting would perhaps be to record neuronal responses during learning. This would allow for observing hypothetical changes in the encoding of sound source location throughout different learning stages. A potential caveat is that naive animals would most likely not actively localize sounds, and any observed change could be interpreted as a result of the lack of engagement. Hopefully, intermediate learning stages where animals just learned the importance of localizing sounds for task performance would be insightful regarding which factors influence the neuronal responses.

A possibility to investigate the nature of the inputs that provide allocentric information (e.g. the retrosplenial or cingulate cortex) to the AC would be to optogenetically identify the neurons that receive projections from those areas for later recording their signals in the AC.

The corticostriatal pathway was shown to encode choice and expected size of reward [195]. Moreover, its optogenetic activation biases animals' choices [211]. It would be interesting to observe whether this would also be the case in the SIT paradigm, given that the period of time during which the animal is choosing is much longer. Furthermore the sit-time could be reduced after the animal has learned the task. I hypothesize that the effect is stronger for shorter sit-times, for which there is less time for evidence accumulation. In a related experiment, the role of the corticostriatal pathway could be investigated in the observed decrease of neuronal firing before correctly finishing a trial by silencing the pathway, given that this reduction could signal reward expectation.

Another intriguing question that remains is how different sound-source allocentric locations are encoded in the brain. In the A1 of mice, there is evidence that sounds are categorized by the activation of a few discrete response modes [212]. Each of these modes corresponds to a specific small local neuronal network. Some modes are common between

sounds, specifically the sounds that are perceptually classified into the same category. Interestingly the transition between modes is highly non-linear, suggesting attractor-like dynamics. This could potentially be the basis for auditory object formation. I propose to test whether this phenomenon can also be observed in SIT. I hypothesize that at least one response mode is specific to a certain allocentric position/sound-source, irrespective of other varying sound features. To test this hypothesis, I would suggest recording neuronal activity using miniscopes [213] instead of electrophysiology, due to evidence that a complete sound representation occurs at a global scale and not in one local network [212]. Moreover, the sound varying features that would not be task-relevant could potentially be the fundamental frequencies of the harmonic complexes played to the animal.

In the paradigm presented here the reward and spatially relevant area for the animal is confounded with the sound source location and therefore it is not possible to investigate their individual contributions. To decouple the two and actually understand if spatial selective attention is an observed mechanism, I would suggest performing the same experiment with more than two loudspeakers, where the target island would correspond to the activation of one specific loudspeaker and another island (or more) would activate other specific loudspeaker(s) – such as the multiple islands in the frequency version of SIT [117].

Many other experiments will have to be performed until we finally understand how the brain localizes sounds. This study shows us that active navigation in the real world and selective listening must be taken into account to have a complete picture of the reality.

Appendix - Sampling uniformly from a circle

To sample uniformly from a circle with radius R, one cannot simply follow the intuition and pick from a uniform distribution an angle $\theta \in [0, 2\pi)$ and a radius $r \in [0, R]$, since that would lead to many more points in the center than at the borders (Fig. A.1a).

Dots uniformly distributed along a circle with radius r have a certain density of dots/length. To maintain this density along a circle with radius ar, the number of dots must also be multiplied by a, since the perimeter is proportional to r. The number of dots must then also be proportional to the radius of the circle and consequently the radial probability density function is of the type mr. The integral over all possible radius of the radial probability density function is by definition of probability density function 1: $\int_0^R mr\,\mathrm{d}r = 1 \Leftrightarrow m = \frac{2}{R^2}$. The cumulative distribution function (CDF) will then be $\int \frac{2}{R^2} r\,\mathrm{d}r = \frac{r^2}{R^2}$. The CDF has as limits per definition 0 and 1 and as our goal is to have a uniform distribution, any value of the CDF must always be a number drawn from a uniform distribution between 0 and 1: u. By calculating the inverse CDF, we arrive to the r to which each u corresponds: $u == \frac{r^2}{R^2} \Leftrightarrow r = R\sqrt{u}$ (Fig. A.1b).

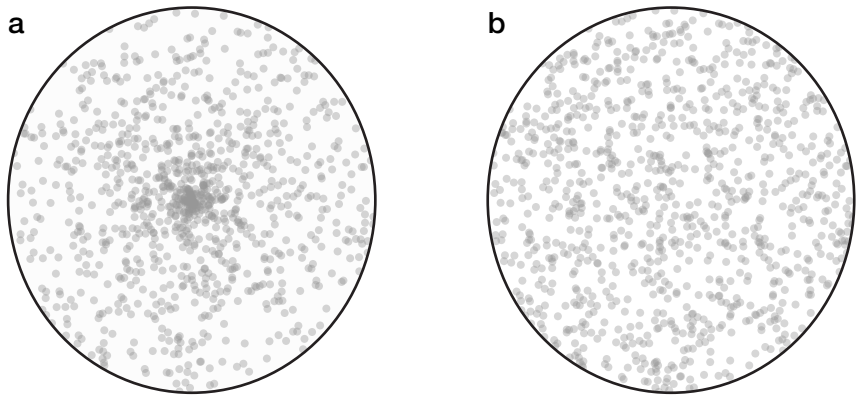


Figure A.1 – **Sampling from a circle a** using a non-uniform distribution **b** using a uniform distribution

B

Appendix - Behavioral paradigm (SIT) article

Part of my PhD was dedicated to developing and improving a behavioral paradigm that can be used to assess sensory perception in a more naturalistic context, namely in freely moving animals. However, not all aspects of this study were presented in this monograph. For all these reasons, this article is here presented as an appendix.

frontiers in Behavioral Neuroscience

METHODS
published: 25 September 2020
doi: 10.3389/fnbeh.2020.576154

Sensory Island Task (SIT): A New Behavioral Paradigm to Study Sensory Perception and Neural Processing in Freely Moving Animals

*Dardo N. Ferreiro[1,2†], Diana Amaro[1,3†], Daniel Schmidtke[4], Andrey Sobolev[3], Paula Gundi[1,3], Lucile Belliveau[1], Anton Sirota[5], Benedikt Grothe[1,3] and Michael Pecka[1]**

[1] Division of Neurobiology, Department Biology II, Ludwig-Maximilians-Universität München, Munich, Germany, [2] Department of General Psychology and Education, Ludwig-Maximilians-Universität München, Munich, Germany, [3] Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, Munich, Germany, [4] Institute of Zoology, University of Veterinary Medicine Hannover, Hanover, Germany, [5] Faculty of Medicine, Bernstein Center for Computational Neuroscience Munich, Munich Cluster of Systems Neurology (SyNergy), Ludwig-Maximilians-Universität München, Munich, Germany

OPEN ACCESS

Edited by:
Ralf Heinrich,
University of Göttingen, Germany

Reviewed by:
Xiao-Dong Wang,
Zhejiang University, China
Sharlen Moore,
Johns Hopkins University,
United States

***Correspondence:**
Michael Pecka
pecka@bio.lmu.de

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: 25 June 2020
Accepted: 27 August 2020
Published: 25 September 2020

Citation:
Ferreiro DN, Amaro D, Schmidtke D, Sobolev A, Gundi P, Belliveau L, Sirota A, Grothe B and Pecka M (2020) Sensory Island Task (SIT): A New Behavioral Paradigm to Study Sensory Perception and Neural Processing in Freely Moving Animals. Front. Behav. Neurosci. 14:576154. doi: 10.3389/fnbeh.2020.576154

A central function of sensory systems is the gathering of information about dynamic interactions with the environment during self-motion. To determine whether modulation of a sensory cue was externally caused or a result of self-motion is fundamental to perceptual invariance and requires the continuous update of sensory processing about recent movements. This process is highly context-dependent and crucial for perceptual performances such as decision-making and sensory object formation. Yet despite its fundamental ecological role, voluntary self-motion is rarely incorporated in perceptual or neurophysiological investigations of sensory processing in animals. Here, we present the Sensory Island Task (SIT), a new freely moving search paradigm to study sensory processing and perception. In SIT, animals explore an open-field arena to find a sensory target relying solely on changes in the presented stimulus, which is controlled by closed-loop position tracking in real-time. Within a few sessions, animals are trained via positive reinforcement to search for a particular area in the arena ("target island"), which triggers the presentation of the target stimulus. The location of the target island is randomized across trials, making the modulated stimulus feature the only informative cue for task completion. Animals report detection of the target stimulus by remaining within the island for a defined time ("sit-time"). Multiple "non-target" islands can be incorporated to test psychometric discrimination and identification performance. We exemplify the suitability of SIT for rodents (Mongolian gerbil, *Meriones unguiculatus*) and small primates (mouse lemur, *Microcebus murinus*) and for studying various sensory perceptual performances (auditory frequency discrimination, sound source localization, visual orientation discrimination). Furthermore, we show that pairing SIT with chronic electrophysiological recordings allows revealing neuronal signatures of

sensory processing under ecologically relevant conditions during goal-oriented behavior. In conclusion, SIT represents a flexible and easily implementable behavioral paradigm for mammals that combines self-motion and natural exploratory behavior to study sensory sensitivity and decision-making and their underlying neuronal processing.

Keywords: psychophysics, sensory feedback, chronic recording, go no-go, freely moving, sound localization, frequency discrimination, orientation selectivity

INTRODUCTION

Understanding how specific behaviors (reflexes, motor patterns, sensory representations, subjective perception, or cognitive functions) arise from neural processing is a primary goal of neuroscience. Pioneering research on sensory processing was based on observations of organisms and their innate behavior in their natural habitats (von Frisch, 1954; Tinbergen, 1963; Lorenz, 1981). This minimal-intervention approach laid the groundwork for the study of natural behavior during ethologically adequate sensory stimulation, yet left questions regarding the underlying neuronal mechanisms and brain circuits largely unanswered. In the last decades, experimental methods to study neural activity in awake and behaving animals have been increasing in number and complexity, providing previously unreachable insights into processing capabilities of neural populations. However, the great complexity of these techniques often requires highly controlled experimental conditions, which in turn limit their ecological relevance. Thus, they are prone to underestimate the dimensionality of neuronal processing (Gao and Ganguli, 2015; Krakauer et al., 2017).

A central evolutionary driving force acting on sensory systems is the processing of environmental cues in relation to self-motion: the interdependence of a motor action and the resulting modulation of sensory information is a fundamental aspect of both neural coding and decision making (Etienne et al., 1996; Ma and Jazayeri, 2014; Case et al., 2015), because this reciprocal interaction with the outside world allows for the continuous update of the "internal framework" within which the sensory inputs are interpreted (von Holst and Mittelstaedt, 1950; review: Campbell and Giocomo, 2018). Accordingly, substantial neural resources are dedicated to gathering and interpreting sensory information in relation to one's own voluntary actions (Keller et al., 2012; Rancz et al., 2015; Vélez-Fort et al., 2018). A number of studies recently demonstrated the impact of movement on neuronal processing across sensory modalities, including somatosensation (Fu et al., 2014; Kerekes et al., 2017), vision (Chiappe et al., 2010; Niell and Stryker, 2010; Maimon et al., 2010; Dadarlat and Stryker, 2017; Clancy et al., 2019), and audition (Zhou et al., 2014; Schneider et al., 2014; for review see Schneider and Mooney, 2018). Likewise, multisensory co-modulation of the physical properties of the environment is crucial for inference and sensory object formation (Noppeney et al., 2008; Diehl and Romanski, 2014; Altieri et al., 2015; Atilgan et al., 2018) and, thus, highlights the importance of active task engagement of the experimental animals. This informational framework is highly plastic and subject to context-dependent modulation (Chabrol et al., 2015; Deneux et al., 2019).

However, despite the fundamental role of self-movement during goal-oriented behavior and the resulting multisensory co-modulation in complex sensory scenes, experimental investigations including these aspects are still underrepresented in the literature (Krakauer et al., 2017). While reports on psychophysical measurements involving decision-making are recently increasing (Carandini and Churchland, 2013; Saleem et al., 2018; The International Brain Laboratory et al., 2020), to this date, a flexible experimental paradigm to study sensory processing during goal-oriented behavior in freely moving animals is lacking. Here, we modified and expanded the existing concept of using closed-loop free navigation assays (Polley et al., 2004; Whitton et al., 2014). We present the Sensory Island Task (SIT), a novel experimental paradigm to study sensory processing of variable modalities during unrestricted self-movement in actively engaged animals that also allows for simultaneous neural recordings.

MATERIALS AND METHODS

In SIT, animals freely explore an arena in the presence of sensory background stimulation. They are trained to search for a hidden target island (a small circular sub-space in the arena, see below). Upon entering the target island, the background stimulus switches to the target stimulus. The animals are trained to report the detection of the target stimulus by staying at this position in the arena (i.e., within the target island). The position of the target island is altered in each trial and, thus, can only be found by detection of the change in sensory stimulation. A trial is considered correct when the animal stays within the target island for a specific duration ("sit-time," typically 5–6 s). After a correct trial, a food reward is dropped in the arena via an overhead food dispenser. Trials have a time limit (typically 60 s) after which they are considered incorrect. Additionally, in some experiments (multi-island, see section "Results" for details), non-target islands were introduced simultaneously with the target island. These islands triggered a different change of stimulation than the target and no reward was provided when sit-time was achieved. This design of the task renders it a natural implementation of the NO-GO sensory change detection task, which are typically used in head-fixed experiments (Carandini and Churchland, 2013) and here is replaced by a sit-in-place condition.

Animals and Housing: Gerbils

Here, SIT was used in two sensory modalities (auditory and visual) and in two species. Mongolian gerbils (*Meriones*

unguiculatus) were used to probe auditory frequency discrimination and identification ($aSIT_{freq}$) and sound source localization ($aSIT_{loc}$) as well as visual orientation discrimination ($vSIT_{ori}$). All procedures involving gerbils were approved in accordance with the stipulations of the German animal welfare law (Tierschutzgesetz) (AZ 55.2-1-54-2532-74-2016 and AZ 55.2-1-54-2532-70-2016). The animals were from the breeding colony of the Biocenter of the Ludwig-Maximilians University Munich. Animals were housed in groups of 3–4 individuals with 12 h light/dark cycles.

Animals and Housing: Mouse Lemurs

Additionally, $aSIT_{freq}$ was conducted with two gray mouse lemurs (*Microcebus murinus*). The non-invasive experiments were in accordance to the NRC Guide for the Care and Use of Laboratory Animals, the European Directive 2010/63/EU, and the German Animal Welfare Act. They were approved by the Animal Welfare Committee of the University of Veterinary Medicine and approved and licensed by the Animal Welfare Committee of the LAVES (AZ 33.19-42502-04-18/3050). The animals were from the breeding colony of the Institute of Zoology of the University of Veterinary Medicine Hannover. Maintaining and breeding were permitted by the Landeshauptstadt Hannover and the Landesamt für Verbraucherschutz und Lebensmittelsicherheit (LAVES; AZ 42500/1H).

Setup and Stimulation During $aSIT_{freq}$ and $aSIT_{loc}$ With Gerbils

The $aSIT_{freq}$ and $aSIT_{loc}$ (freq: sound frequency as target indicator; loc: sound source location as target indicator) tests with gerbils were conducted in a custom-made setup consisting of a circular arena (diameter = 92 cm) within a sound attenuated chamber (**Figure 1A**). The arena floor consisted of a black-painted wood or PVC surface surrounded by perforated metal walls (height: 16 cm). Additionally, PVC walls were mounted on top of the metal wall around the entire arena up to a height of 75 cm.

Stimuli were computer generated and transmitted through an amplifier (AVR 445 Harman/Kardon, Germany). Stimulus presentation was delivered through loudspeakers (Aurasound NSW1-205-8A 1″ Extended Range) mounted externally of the arena (∼5 cm distance to the metal walls). Auditory stimuli during $aSIT_{freq}$ were 57 ms long pure tones with frequency according to task structure. Trial initiation elicited the playback of the background frequency (20 kHz), and animal entrance into an island triggered the switch of the frequency played to the target frequency of 660 Hz or non-target frequencies of 460, 860, 1060, or 1320 Hz (**Figure 2** - see section "Results" for details). Stimuli during $aSIT_{loc}$ were 57 ms long harmonic complex sounds with a fundamental frequency of 147±4 Hz and low-pass filtered below 1.5 kHz. Trial initiation triggered the playback of the above-mentioned harmonic complex by the background loudspeaker, and animal entrance into the island triggered the switch of the playback to the target loudspeaker. Stimuli in either aSIT version were played at a repetition rate of 4 Hz and their amplitude was 70 dB SPL roved ±5 dB, which rendered a stimulation of about 55 dB above background noise. The animal's position was tracked via images captured every 250 ms with a Flea3 camera (FL3-U3-13Y3M-C, Point Grey Research Inc.), centered over the arena at a height of 130 cm from the arena floor. Stimulation parameters (i.e., sound frequency or source location) were updated online according to the animals' position within the arena (see section "Results" for details). Custom-made software for animal tracking, stimuli generation and food reward delivery was developed in MATLAB. A custom-made overhead rotating food dispenser positioned 100 cm over the arena was used for automatic reward administration by dropping a food pellet (∼20 mg, TestDiet LabTab AIN-76A) or part of a sunflower seed after every correct trial. If the animal did not correctly report the target island within the time limit, a low-pass filtered noise was presented to the animal for 10 s, during which no new trial could be initiated.

Behavioral Training During $aSIT_{freq}$ and $aSIT_{loc}$ With Gerbils

Two gerbils were used for the behavioral testing of the $aSIT_{freq}$ paradigms, and 11 gerbils were tested in the $aSIT_{loc}$ version of the task. Training of gerbils began at a minimum of 8 weeks of age. All gerbils within this study were male. Water and food (pellets) were provided *ad libitum* until training started, at which point food was only available during training sessions as reward for correct trials. No more than two training sessions were carried out per day, lasting up to 60 min each for $aSIT_{freq}$ and up to 90 min for $aSIT_{loc}$. Final parameters of island size (diameter = 25 cm, ∼7% of the arena surface) and sit-time (6 s) were identical for both $aSIT_{loc}$ and $aSIT_{freq}$. For $aSIT_{freq}$, animals were presented with the final parameters from the beginning of training. For $aSIT_{loc}$ the training of the animals was performed by gradually reducing island size (starting at diameter = 42 cm, ∼21% of the arena surface) and increasing sit-time (starting with 2 s) over the course of the training sessions. Additionally, for $aSIT_{loc}$, animals were initially trained in a protocol with one slightly elevated, peripheral, circular initiation platform (diameter = 12 cm), which the animals had to visit in order to initiate a trial. For $aSIT_{freq}$, an additional configuration with multiple islands was tested, where three non-target islands were available in the arena alongside the aforementioned target island (see section "Results" for details). All gerbils in the aSIT tasks underwent a general habituation period in the SIT setup for 15 min per day for 5 days.

Setup and Stimulation During $aSIT_{freq}$ With Mouse Lemurs

The $aSIT_{freq}$ experiments with mouse lemurs were conceptually identical, yet with adapted parameters to accommodate to species-specific exploration behaviors. Experiments were conducted in a circular open field arena with a diameter of 80 cm and a height of 70 cm (**Figure 1F**). For online animal tracking, a camera (Logitech C500 Webcam) with removed infrared filter was positioned above the center of the maze and at a distance of 92 cm from the floor plate, so that the arena floor optimally fitted the vertical dimensions of the video picture. For acoustic stimulation, a single broadband speaker (Visaton B200, VISATON GmbH & Co., KG, Haan, Germany) was mounted

above the arena at a distance of 165 cm from the arena floor. The floorplate was made of frosted light-conducting acrylic glass (Plexiglas® LED, Evonik Industries, Darmstadt, Germany) and illuminated with infrared diodes (peak wavelength at 940 nm) from below to provide optimal contrast between background and experimental animal during tracking. The sidewall of the circular arena was made of opaque, dark-gray acrylic glass (Zimmermann + Collegen Kunststoff-Technik GmbH, Hannover, Germany). Food rewards used as positive reinforcement during the learning experiments (see below) were provided on-top of the regular, *ad libitum* diet. To provide food rewards (small peanut pieces of approximately 15 mg) for correct behavioral responses during training, commercially available aquarium feeders (Rondomatik 400, Grässlin GmbH, St. Georgen, Germany) were modified to be controllable with Arduino Uno microcontrollers via Arduino Uno motor shields (v1). Two of these modified feeders were installed at opposing positions on the arena wall (i.e., at a distance of approximately 70 cm from the floor) and their positions could easily be shifted between sessions to reduce predictability of the reward location. For online animal tracking as well as sound stimulation and hardware control based on the animal's behavior, we used self-coded Python scripts, running on windows machines with Windows 7 and Python 3.7.

Behavioral Training During $aSIT_{freq}$ With Mouse Lemurs

Training of mouse lemurs was conducted in male individuals aged 5 and 6 years that had previously participated in non-auditory behavioral experiments unrelated to SIT. To avoid stress, subjects were transported to the setup in their sleeping boxes and experiments were conducted under low-light conditions (1–5 lux). Each animal was trained once per day during workdays in a single session of 60 min or 50 completed trials (depending on which limit was reached first). Animals were trained in a protocol with one slightly elevated, peripheral, circular initiation platform, which the respective animal had to visit in order to initiate a trial, and one circular target island. Once a trial had been initiated, a background sound (pure tone of 10 kHz, 57 ms duration, sound pressure level = 67.5 ± 2.5 dB) was played back at a repetition rate of approximately 5 Hz while the geometric center of the animal remained outside of the target island (pseudo-randomly generated position without overlap with the initiation platform). As soon as the animal entered the target island during a given trial, stimulation switched to the target sound (pure tone of 4 kHz, all other properties were identical to the background sound). The frequency of the stimuli was chosen to lie within the range of optimal hearing described for mouse lemurs (Schopf et al., 2014). If the animal failed to find the target island or to remain within it for the desired sit-time within a pre-defined trial duration, the trial stopped, as did the acoustic stimulation, and the animal had to revisit the initiation platform to start a new trial. During the experiments, the setup was illuminated with dim red light, comparable to the illumination of the housing rooms during the daily activity phase of the nocturnal mouse lemurs. While the location and size (diameter = 18 cm, 5% of arena surface) of the initiation platform were fixed values, the size of the target island, the sit-time, and the trial duration could vary between sessions. In the first session, the size of the target island was set to a diameter of 32 cm (∼16% of arena surface), the target duration to 1 s, and the trial duration to 120 s. To increase the difficulty with increasing training and to better differentiate behavioral responses to the target sound from chance-level performance, these variables were changed between sessions, depending on the animal's performance on the preceding training days. Values for the final sessions were a target island diameter of 24 cm (∼9% of arena surface) and a sit-time of 5 s. Animals were trained until performance in three consecutive sessions under these conditions was above chance level.

Setup and Stimulation During $vSIT_{ori}$ With Gerbils

The $vSIT_{ori}$ experiments were conducted in a 3D virtual reality setup called ratCAVE (Del Grosso et al., 2017), which was designed for behavioral experiments in freely moving animals. To this end, a large rectangular arena (dimensions 162 cm × 72 cm and walls of 60 cm height, placed with a 70 degrees angle to accommodate the visual projection), was used. A set of 7 cameras (Prime 13W 240 fps, OptiTrack, NaturalPoint Inc., United States) served to record the 3D position of reflective markers fixed on the head of the animal. A projector with 240 fps frame rate (VPixx Technologies Inc., Canada), mounted to the ceiling, was used to project the image of the virtual environment on the walls of the arena depending on animal position (**Figure 3A**). A food dispenser (Campden Instruments Ltd.) positioned above the arena served for automatic reward administration by dropping a food pellet (∼20 mg, TestDiet LabTab AIN-76A) after every correct trial. A custom-written python-based software was used to manage the projection, animal rewarding, positioning, and data logging.

The virtual environment for the $vSIT_{ori}$ experiment consisted of black and white square-wave grating patterns with stripes of 10 cm width, projected on all four walls of the arena. When animals entered the target island, the projected grating pattern on the walls changed its orientation from vertical to horizontal (**Figure 3B**). A non-target island was additionally implemented for one of the animals which, upon animal entrance, triggered change from the vertical grating projection to oblique (45 degrees). Each successful trial was followed by an inter-trial period of 15 s with only light projected on the arena floor (no patterns on the walls) to allow the animal to find the rewarded pellet. After the inter-trial interval, the new trial started automatically.

Behavioral Training During $vSIT_{ori}$ With Gerbils

Two male gerbils were trained in this version of $vSIT_{ori}$. No habituation was required, as they had previously participated in another study within the same arena. Animals were food restricted and kept at a minimum weight of 85% of the *ad libitum* condition. Similar to the $aSIT_{loc}$ experiments, training of the animals was performed by gradually reducing island size (starting

at ∼10% of the arena surface) and increasing sit-time (starting at 2 s) over the course of the training sessions. At the end of the training (15 and 24 sessions), a trial was considered correct when the animal stayed within the target island of minimal size (∼6% of arena surface area) for a sit-time of 6 s. For one of the gerbils, the non-target island was introduced to the trials after performance reached a level significantly different from chance (see section "Results").

Source Code Availability

Protocols to perform $aSIT_{freq}$ experiments are freely accessible for download at https://gin.g-node.org/asobolev/runsit/.

Surgical Procedures and Chronic Electrophysiological Recordings

One adult male Mongolian gerbil (∼70 g) that was trained in $aSIT_{loc}$ underwent tetrode implantation surgery. At the beginning of the surgery, the animal was anesthetized with an intraperitoneal injection of a mixture of metedomidin (0.15 mg/kg), midazolam (7.5 mg/kg), and fentanyl (0.03 mg/kg). The depth of the anesthesia was verified by lack of paw pinch or eye lid reflexes. To maintain it at a constant level, the same mixture was subcutaneously re-injected every 90 min. After shaving and disinfecting the head, a local anesthetic (50 μl, 2% xylocaine) was injected under the scalp skin and below the skin near the ears. For protection and to prevent dehydration, the eyes were covered with an ophthalmic gel (Thilo-Tears SE, Alcon Pharma GmbH). The animal was then transferred to the stereotactic apparatus, where its head was securely fixed via a bite and ear bars. Its internal temperature was monitored with a rectal thermometer and kept constant at 37°C throughout the experiment by a feedback controlled electric heating pad (Harvard Apparatus). After disinfection, a midline scalp incision was performed to expose the skull. Subsequently, the connective tissue on the skull was removed with a bone curette and the skull was treated with 35% phosphoric acid (iBOND etch gel, Kulzer), which was promptly washed away. Structural screws were placed on top of the left frontal and right parietal bones and the ground screw on the occipital bone, so that it gently touched the brain. After stereotactic alignment, a 3 × 3 mm craniotomy and durotomy were performed on top of the left auditory cortex, followed by a very slow lowering (2 μm/s) of a tetrode bundle to a maximum depth of 0.9 mm into the cortex, using a micromanipulator (Scientifica). The craniotomy was carefully filled with KY-jelly and immediately sealed with dental cement (Paladur, Kulzer), which also fixated the bottom of the microdrive and the outer cannula that protected the tetrodes. 1 ml of Ringer's solution was subcutaneously injected at the end of the surgery and the anesthesia was reversed via subcutaneous injection of the antagonist mixture composed of naloxone (0.5 mg/kg), flumazenil (0.4 mg/kg), and atipamezol (0.375 mg/kg). Analgesics (0.2 mg/kg, meloxicam) and antibiotics (7.5 mg/kg, enrofloxacin) were orally administered post surgically for five subsequent recovery days. During this time, the animals had food and water *ad libitum* and were not trained.

The implant used in this experiment was a tetrode bundle consisting of four tetrodes glued together, which, on their turn, consisted of four insulated tungsten wires (12.7 μm diameter each, tungsten 99.95%, California Fine Wire) twisted around each other. Each wire was connected to a custom-made printed circuit board with Omnetics connector (Axona), which was attached to a lightweight microdrive (0.25 mm/turn, Axona). The tetrodes were glued together and protected by an inner and outer cannula that could slide by each other. On the day prior to the surgery, the tip of all electrodes were cut with sharp scissors and gold plated (Non-Cyanide Gold Plating Solution, Neuralynx) to reach a desired impedance of 100–150 kOhm (at 1 kHz). The tetrode bundle was implanted vertically in the following coordinates from lambda: 6.2 mm lateral, 2.6 mm anterior. The recording depicted in **Figure 4** occurred at an electrode depth of ∼1.3 mm.

Recorded signals were amplified and digitized (16-bit resolution) in the wireless headstage (W2100-HS16, Multichannel Systems), and transmitted to the receiver. Through an interface board (W2100-System, Multichannel Systems), the signal was then sent to the computer where it was acquired with a sampling rate of 25 kHz via commercial software (Multi Channel Experimenter). A digital signal for posterior alignment of the sounds and video with the neural signal was simultaneously sent to the interface board.

Data Analysis

All data analyses were performed in MATLAB (Mathworks) and Python using custom scripts. To test the performance of the animals, we compared the percentage of correct trials in each session with surrogate runs based on random target island shuffling. That is, for each trial (offline, *a posteriori*), 1000 surrogate (non-real) islands, non-overlapping with the target one, were randomly set and the real trajectory of the animal was used to calculate in how many of these islands the trial would have been correct given the required sit-time (**Supplementary Figure S1**). At each time point, we determined how many trials were already finished and the respective uncertainty (95% confidence interval) was calculated based on bootstrapping (random sampling with replacement from all the trials of the session). The median chance performance and confidence interval at each time point was calculated based on bootstrapping from the random target island shuffling data (random sampling with replacement from the 1000 surrogate trials with number of trials as size of the sample). The chance performance calculation was based on trajectories from trials which were incomplete up to the considered time point (real target island not yet found) and trajectories in which the animal stayed longer than the sit-time in the surrogate island before that time point. A trial which had been finished by that time point and in which the animal did not find the surrogate island cannot be used in the bootstrapping of the time points posterior to the finishing time because it is unknown whether the animal would have found the island if the trial had been longer. This method allows obtaining an estimate of the proportion of correct trials the animal would have gotten just by chance given their locomotion trajectory and dynamics.

In the multi-islands version of SIT, the sit-time incidence was calculated by assigning an island to each trial. This assignment

corresponds to the first island in which the animal stayed longer than the sit-time. For example, if the animal correctly finished a specific trial but had been sitting for longer than the sit-time in a non-target island prior to finishing, this trial is assigned to the respective non-target feature and not to the target one, even though the animal also remained sufficiently long in the target island later. A trial in which the animal never remained for longer than the sit-time in any island is assigned to "None." In the $aSIT_{freq}$ multi-island configuration, the target frequency is always present but the non-target frequencies are not, as there are 4 non-target frequencies and only 3 non-target islands in each trial. Therefore, for each session, we calculated the percentage of trials assigned to each frequency, normalized by the total amount of trials in which the respective frequency was available. As a measure of uncertainty, the 95% confidence interval was calculated by bootstrapping (the percentage calculation was done on 1000 random samples with replacement from the assignment to each frequency with number of trials as size of the sample). The chance level (calculated per animal with data from the last session of the single island version) was subtracted from this percentage and the 95% confidence interval was calculated using error propagation.

For the construction of the psychometric function, in each session all the events in which the animal stayed at least 1s in the island were identified. For those events, the percentage of times the animal stayed in a specific frequency island for the designated sit-time (6 s) was calculated. This allows the construction of a perception curve by fitting a logistic function $\frac{max}{1+e^{-slope(x-x_0)}} + offset$ to these percentage values, with the frequency distance in octaves of each island to the target frequency as x; the offset in relation to zero describes the recurrent behavior of stopping randomly, which occasionally can last longer than the sit-time.

For the analysis of the local field potential (LFP), the recorded signal was low-pass filtered at 600 Hz. Auditory evoked potential (AEP) was calculated per trial, by loudspeaker active. Amplitude of the AEPs was calculated from peak to peak, that is, the difference between the maximum and minimum voltage recorded in the time window corresponding to the first 100 ms after stimulus onset.

Statistics

Binomial tests were used to compare, on a given experimental session, the percentage of correct trials with the ones expected by chance, as calculated using the surrogate runs analysis.

All error bars correspond to the 95% confidence interval as calculated via bootstrapping, except for the boxplots in **Figure 4D**.

For the investigation of possible linear relationships between the distance between islands in consecutive trials and the time to completion in the latter trial (**Figure 1E**), we used Pearson correlation analyses.

For comparisons of central tendencies on the group level, we used two-tailed non-parametric tests: Wilcoxon signed-rank tests for paired samples and Mann-Whitney U-test for independent samples.

All hypotheses were tested at an alpha level of 0.05.

RESULTS

The Sensory Island Task (SIT) is an operant conditioning foraging task in an open-field arena (**Figure 1**). We designed SIT to allow for high flexibility regarding the implementation of sensory modalities and parameters to address the desired specific research question. Animals can roam freely in the arena, in search for a sensory "target island" (in auditory versions of SIT, we used a circular target area within the arena, area ∼5–9% of the arena surface), relying solely on changes in the presented stimulus, which is controlled in real-time via closed-loop position tracking. They are trained via positive reinforcement to discover the target island by detecting a change in stimulation from a "background" to a "target" stimulus. Animals report this detection of the target stimulus by remaining within the island for a defined time (sit-time). Upon correct reporting, a food reward is administered by dropping from an overhead dispenser, which ensures that any association of the reward consumption with a specific location in the arena is prevented (since the reward bounces unpredictably on the arena floor). The location of the target island is randomized across trials, making the stimulus feature under investigation the only informative cue for task completion. Multiple "non-target islands" (areas where the relevant stimulus feature is changed into neither the target nor the background and where the animal is not rewarded) can be incorporated in SIT to test identification performance. Furthermore, SIT can readily be adapted to the species and sensory system under investigation. To demonstrate this high flexibility, here we present data from Mongolian gerbils (*Meriones Unguiculatus*, rodents) and gray mouse lemurs (*Microcebus murinus*, small primates) trained in SIT to perform auditory frequency discrimination and identification ($aSIT_{freq}$). We further demonstrate the suitability of SIT to study sound source localization ($aSIT_{loc}$), as well as visual orientation identification and discrimination ($vSIT_{ori}$).

Auditory Frequency Discrimination ($aSIT_{freq}$)

We trained animals to detect a change in the presented stimulus frequency upon entering the target island. Throughout a trial, a "background" frequency was played in repetitive pulses (duration 57 ms, repetition rate 4 Hz in rodents, 5 Hz in mouse lemurs) through a single loudspeaker as long as the animal was outside of the target island. Once (and if) the animal entered the target island, the stimulation (played from the same loudspeaker) switched to the "target" frequency (**Figure 1B**). Two gerbils and two mouse lemurs (see below) were trained to perform this task in this configuration.

For gerbils, background and target frequencies of 20,000 and 660 Hz were chosen, respectively (see **Supplementary Video 1**). Both gerbils reached similarly high proportions of correct trials within three training sessions (**Figure 1D**; see figure legend for trial numbers). The percentage of successful trial completion highly exceeded chance performance levels (i.e., random stopping in the arena for $>$ 6 s, **Figures 1C,D**, P = 6E-17 for gerbil 1 and P = 2E-27 for gerbil 2, binomial test, calculated for the last session). Chance performances were calculated by the use

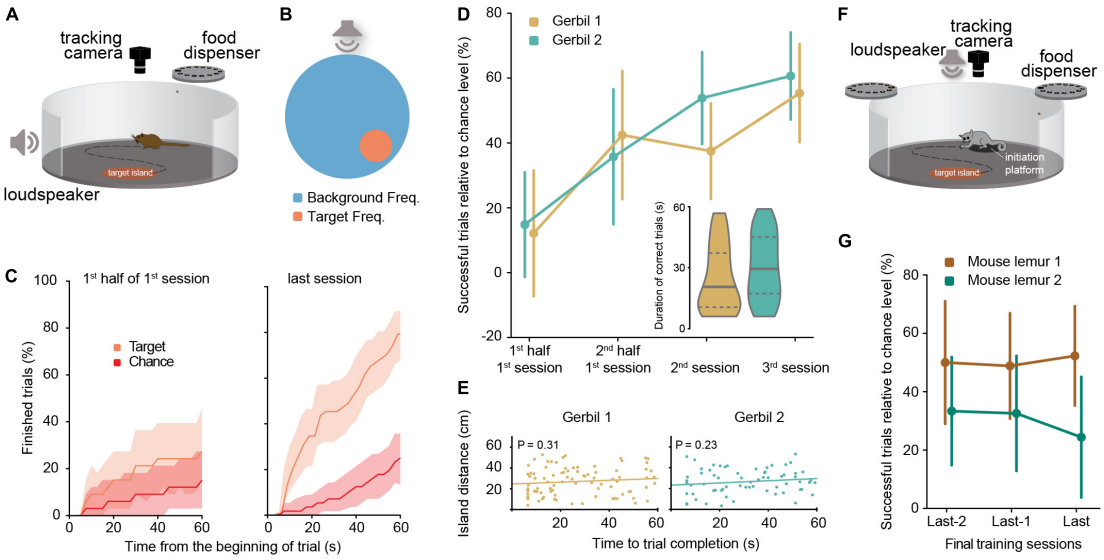


FIGURE 1 | (A) Schematic of the experimental aSIT setup for gerbils. **(B)** Schematic representation (top view) of the $aSIT_{freq}$ arena in the single island version for gerbils (background and target frequencies for gerbils were 20 kHz and 660 Hz). **(C)** Comparison for gerbil 1 of the percentage of trials finished with the percentage of trials which would have been finished by chance at each time point after the beginning of a trial (shadow areas correspond to 95% confidence interval); Left panel: 1st half of the 1st session; Right panel: 3rd session. **(D)** Percentage of successful trials relative to the chance level (as calculated in **C** at 60 s) for each gerbil (error bars correspond to the 95% confidence interval). Session 1: $N_{Gerbil\ 1}$ = 66 trials, $N_{Gerbil\ 2}$ = 55 trials; Session 2: $N_{Gerbil\ 1}$ = 72 trials, $N_{Gerbil\ 2}$ = 65 trials; Session 3: $N_{Gerbil\ 1}$ = 56 trials, $N_{Gerbil\ 2}$ = 61 trials. Inset: duration of successful trials for each gerbil in the two last training sessions, horizontal lines denote median (solid) and quartiles (dashed) of the distribution. Durations of correct trials per session are available in **Supplementary Figure S2** (both for gerbils and for mouse lemurs). **(E)** Time to success in two consecutive successful trials was not correlated with geometric island distance in either gerbil. Pearson correlation, $N_{Gerbil\ 1}$ = 89 pairs of trials, $N_{Gerbil\ 2}$ = 68 pairs of trials. **(F)** Schematic of the experimental aSIT setup for mouse lemurs. Background and target frequencies for lemurs were 10 and 4 kHz, respectively. **(G)** Performance of two mouse lemurs in three consecutive days at the end of the training: percentage of successful trials relative to the daily chance level (as calculated in C; error bars correspond to the 95% confidence interval). Session 1: $N_{Lemur\ 1}$ = 32 trials, $N_{Lemur\ 2}$ = 48 trials; Session 2: $N_{Lemur\ 1}$ = 43 trials, $N_{Lemur\ 2}$ = 43 trials; Session 3: $N_{Lemur\ 1}$ = 44 trials, $N_{Lemur\ 2}$ = 41 trials. For performance levels during intermediate training sessions (see **Supplementary Figure S3**).

of bootstrapping methods with surrogate target locations and the actual animal locomotion trajectories (see **Supplementary Figure S1A** and section "Materials and Methods"). Thus, the animals stopped and remained significantly longer in the portion of the arena that triggered the appearance of the target frequency compared to any other location. This behavior was independent of the relative location of the target island position, within the arena as the animals explored the arena uniformly (i.e., no center avoidance was observed, **Supplementary Figures S1B,C**). Indeed, performance levels of both gerbils was significantly above chance level already for the second half of trials in the very first session of exposure to the task, and further increased with more training (**Figure 1D**, significance is denoted by the lower bound of the confidence interval not extending to chance level).

In both animals, more than half of the correct trials had durations of less than 30 s (half of the maximally allowed duration, inset in **Figure 1D**), suggesting that the chosen maximum trial length was adequate for the animals to complete the task. As rodents may exhibit history-biased behavior in operant conditioning paradigms (Busse et al., 2011), it raises the question if the gerbils might preferentially re-visit (or alternatively avoid) the locations in the arena which triggered the target stimulus in the previous correct trial. To test if they employed specific spatial bias in their search strategy based on the successful detection of the target island location in the prior trial, we plotted the linear distance between the target islands in two consecutive successful trials as a function of the time to completion in the latter of the two trials (**Figure 1E**). Across the two animals, no significant correlation was observed (**Figure 1E**, Pearson correlation, details in figure legend), demonstrating that the animals' exploration behavior was not influenced by the short-term history of task success.

The results so far demonstrated the suitability of $aSIT_{freq}$ for assessing frequency-change detection (discrimination) in gerbils. Next we asked to what extent these results are qualitatively specific to the innate locomotion behavior and learning capabilities of the species/clade we used (gerbils/rodentia) or generalizable across clades. To this end, we also trained two gray mouse lemurs on $aSIT_{freq}$. Gray mouse lemurs are primates, yet comparable in size to gerbils. Notably, they exhibit a quite distinct innate exploration behavior compared to gerbils, as they usually show low levels of spontaneous exploration in an open

field setting (Picq, 2016). Only once they learnt that active exploration of the setup was occasionally rewarded, exploration rate increased. Therefore, we adapted some SIT parameters accordingly and started the training with a large target island size (diameter = 32 cm), a short sit-time (1 s) and a long maximum trial duration (120 s) to increase the initial likelihood of rewarded trials. Once exploration activity of a given individual had increased, parameters were successively changed toward the target values (target diameter = 24 cm, SIT-time = 5 s, maximum trial duration = 60 s). We further introduced an initiation platform for the mouse lemurs, which allowed the animals to decide when to start a trial by visiting the platform (**Figure 1F**, see **Supplementary Video 2**, section "Materials and Methods" and section on sound localization below). Mouse lemur 1 reached the final target parameters in session 14 (after 438 trials), mouse lemur 2 in session 19 (after 567 trials). Under these conditions, both mouse lemurs achieved highly significant performance levels in $aSIT_{freq}$ (**Figure 1G**, P = 2E-3 for mouse lemur 1 and P = 1E-12 for mouse lemur 2, binomial test, calculated for the last training session). Note that our surrogate island bootstrapping method to obtain chance levels and to determine significant performances (see section "Materials and Methods") is sensitive to a subject's moving velocity as well as the specific parameter settings of each trial and, thus, provides an objective evaluation. Hence, SIT can readily be adapted to different species.

Multiple Island $aSIT_{freq}$

The results so far imply that the animals' behavior in SIT serves to seek out the target sound. However, it is unclear whether this behavior is based simply on change-detection (i.e., simply stopping whenever the stimulation changed) or if SIT can also be utilized to test the animals' sensitivity for identification of the target stimulus. To test this hypothesis in more detail, we extended the paradigm design of SIT.

We implemented a version of $aSIT_{freq}$ with several islands simultaneously offered in the arena (**Figure 2A**, see also **Supplementary Video 3**). The same two gerbils that were tested in the single-island task were used in this task. Four islands were simultaneously and pseudo-randomly positioned in each trial corresponding to different stimulus frequencies, including the original target frequency (660 Hz). The frequencies of the non-target islands were 460, 860, 1060, and 1320 Hz. The background frequency "outside" of islands remained as before (20,000 Hz). Importantly, in this SIT version, animals again only received a reward for sit-time stays in the actual target island (no reward was provided for sit-time stays in the non-target islands, and trials were allowed to continue). Overall, the animals showed high success rates (**Figure 2B**, comparable to those in $aSIT_{loc}$, **Figure 4B**) already from session 1, yet because non-target island sit-time stays did not trigger trial termination, the animals could have stopped in any of the non-target islands for 6 s before entering the target island and finishing the trial. Such behavior would still correspond to a non-selective searching behavior based on detection of a change from the background frequency. Note that in this multiple island configuration of SIT, it is not possible to compute the chance level as the surrogate islands would overlap with the non-target ones which correspond to a change in frequency. To address the specificity of island preferences (and therefore the possibility of oddball strategies) directly, we calculated "sit-time incidences" *a posteriori,* that is, we determined the first island in which the animal remained for longer than the sit-time for each trial. Each recorded trial was assigned to only one island (if any at all), namely the one where the animal first stayed for longer than the sit-time. Afterward, we computed the proportion of trials that corresponded to each island frequency relative to the animal's recurrent random sitting behavior calculated as the chance level in the last single island session (i.e., a proxy for the sit-time incidences outside of islands, see section "Materials and Methods"). Notably, significantly high sit-time incidence percentages for the target island were observed already after the first session of exposure to the multi-island $aSIT_{freq}$ (**Figure 2C**, significance is given by the fact that chance level lies outside the 95% confidence interval for the target). Likewise, sit-time incidences for non-target islands dropped in prevalence after the first training session and reached baseline level for most non-target frequencies besides 860 Hz (see below). These results strongly indicate that the animals learn to specifically associate the target island frequency with the reward. It is further evidence that the animals were actively searching for the location of the target island (i.e., the arena location that induces the appearance of the target stimulus) and not simply awaiting a change in stimulation that is independent of their own spatial behavior. This assessment is further corroborated by the finding that gerbils adapted their arena occupancy during exploration according to target island location biases (see section on sound localization and **Supplementary Figure S5**).

Interestingly, the proportion of sit-time incidences in non-target islands was not uniform. We observed that sit-time incidences for the 860 Hz island were significantly increased relative to baseline for either animal for some of the training sessions (for gerbil 1, the lower bound of the confidence interval remained above chance level on all sessions, while for gerbil 2 it only did so on the second session). Gerbils are generally capable of discriminating even smaller frequency differences than used here (0.4 octaves) when presented in succession (Klinge and Klump, 2009). However, Chen et al. (2019) have recently shown that when confronted with a memory-based frequency discrimination task, mice generalize auditory stimuli. Therefore, one plausible explanation to the increased sit-time incidences for 860 Hz is that the gerbils generalized the new presented stimulus initially after introduction of the non-target islands.

The data, thus, suggest that multi-island SIT might represent an adequate behavioral readout of perceptual thresholds. This premise is further supported by the observation that the sit-time incidence percentage for the 860 Hz island of gerbil 2 decreased to baseline at later training sessions, which is indicative of increased frequency identification ability with experience (**Figure 2C**, lower panel), which could be explained by an extinction of the prior generalization (Chen et al., 2019). The reason why generalization (and extinction) is seen at 860 Hz, but not 460 Hz might be related to asymmetrical filter broadening and/or the closer logarithmic spacing (Schnupp et al., 2011).

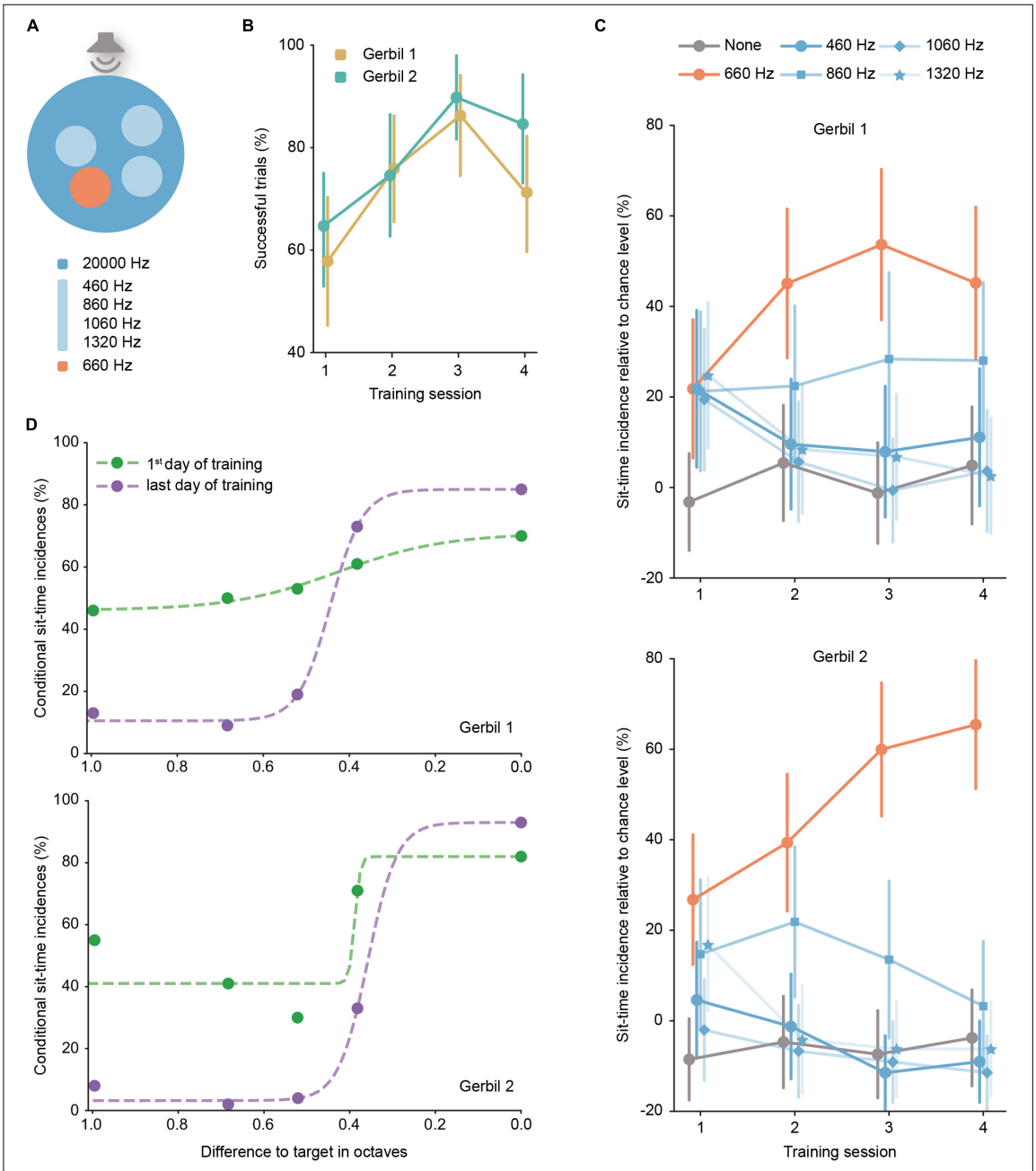


FIGURE 2 | (A) Schematic representation (top view) of the $aSIT_{freq}$ in the multiple island version. Note that on a given trial, only three of the four possible non-target frequencies were offered. **(B)** Performance of each gerbil per training session (error bars correspond to the 95% confidence interval). Session 1: $N_{Gerbil\ 1}$ = 64 trials, $N_{Gerbil\ 2}$ = 68 trials; Session 2: $N_{Gerbil\ 1}$ = 58 trials, $N_{Gerbil\ 2}$ = 59 trials; Session 3: $N_{Gerbil\ 1}$ = 51 trials, $N_{Gerbil\ 2}$ = 49 trials; Session 4: $N_{Gerbil\ 1}$ = 62 trials, $N_{Gerbil\ 2}$ = 52 trials. **(C)** Incidence of sit-time across sessions, relative to chance level per island (error bars correspond to the 95% confidence interval). **(D)** Psychometric function: comparison between the first and last training session of the percentage of events the animal stayed the sit-time in each island depending on the frequency distance in octaves of the island to the target frequency; results were fit with a logistic function (dashed line).

To directly describe performance levels and their change across training sessions, we next calculated the "conditional sit-time incidences" for each of the tested island frequencies (expressed in octave distance to the target frequency - **Figure 2D**). For this analysis, we only considered trials where the animal encountered at least 1 s of sound exposure in the respective island, to ensure that the animal had the opportunity to evaluate the nature of the frequency change (see section "Materials and Methods"). The results of this analysis revealed two findings: first, a clear dependence of the conditional sit-time incidences on the octave-distance to the target frequency is apparent; second, the peak performance values increased, while conditional sit-time incidences of non-target frequencies decreased over the training sessions. These results indicate that learning occurred, which resulted in better identification of the different frequencies. Hence, multi-island SIT in combination with sit-time incidence analyses allows constructing psychometric functions to determine perceptual learning progress.

So far, we established that SIT allows the investigation of auditory frequency discrimination and identification in rodents and in primates. Next, we tested the suitability of SIT to study another sense, namely vision.

Visual Grating Orientation Discrimination ($vSIT_{ori}$)

Here, SIT was incorporated into an existing free-navigation visual stimulation setup (from Del Grosso et al., 2017, 2019) and two gerbils were trained to report when the orientation of the grating projected on the walls of the arena changed from vertical to horizontal (**Figure 3B** and **Supplementary Video 4**). Both gerbils achieved a performance above chance level (**Figure 3C**, $P = 2E\text{-}28$ for gerbil 3 and $P = 1E\text{-}4$ for gerbil 4, binomial test, calculated for the last training session) at the end of the training (gerbil 3 was trained in a total of 24 sessions – 672 trials – and gerbil 4 in 15 sessions – 384 trials).

Gerbil 3 was additionally tested for stimulus feature specificity by introducing a non-target island. The non-target island corresponded to a 45° orientation of the grating (**Figure 3B** and **Supplementary Video 5**). As in previous versions of SIT, this island was not rewarded if the gerbil spent longer than the sit-time inside and the trial continued. To analyze the specificity of the gerbil's behavior, we again calculated the sit-time incidence percentage and assigned each trial to the island in which the animal stayed first for the duration of the sit-time. Already in the first session in which the non-target island was introduced, the animal exhibited high selectivity for the target stimulus and stayed for the sit-time almost exclusively in the target island (**Figure 3D**). The sit-time incidence percentage for the non-target island is not different from chance, which supports the hypothesis that the gerbil learned that a specific grating orientation is associated with reward and not any change in orientation. Thus, SIT is readily adaptable to other sensory modalities, suggesting that it is suitable for multi- or cross-modal investigations.

Next, we examined how SIT can be utilized to study another fundamental auditory computation – sound localization – and to what extent employing SIT (hence introducing its inherent ecological relevance by allowing free exploration) in chronically implanted animals may facilitate the identification of new neural processing signatures.

Sound Localization ($aSIT_{loc}$)

We applied SIT to study sound localization in freely behaving and engaged animals. Traditionally used paradigms to study spatial sensitivity require a constant head position during sound presentation (Wood et al., 2019), often in naïve or anesthetized animals (Middlebrooks and Knudsen, 1984). In contrast, $aSIT_{loc}$ allows investigations in the locomoting animal during active localization, providing more naturalistic conditions and, thus, higher ecological relevance. We used the single-island configuration, yet here the target island cue was a change in the sound source location (i.e., the active loudspeaker). The arena was equipped with two diametrically opposed loudspeakers (180° angle separation from the center of the arena), from which a short (57 ms) harmonic complex sound (see section "Materials and Methods") was presented at 4 Hz repetition rate. Upon trial initiation (see below), the sound was played by one of the two loudspeakers (the background) until the animal entered the target island, at which moment the stimulation switched to the second loudspeaker (target) (**Figure 4A** and **Supplementary Video 6**). The identity of the target and background loudspeaker was maintained throughout training and testing yet catch-trials with swapped identities were introduced in a subset of the animals (see below). Since we combined this paradigm with neural recordings in the auditory cortex (AC), we added an initiation platform (∼1 cm in height) for the animals during training and testing on $aSIT_{loc}$ (similar to the mouse lemur paradigm in $aSIT_{freq}$). Voluntary trial initiation has been shown to reduce spontaneous discharge and improve the detection of thresholds (Buran et al., 2014) and task engagement sharpens spatial tuning of neurons in AC in cats (Lee and Middlebrooks, 2011). The platform was positioned near the wall of the arena and animals were required to stay on it for one second to start a trial.

Locomotion and Sitting Behavior Are Specific to Target Loudspeaker and to Target Island Distribution Likelihood

We tested 11 gerbils in $aSIT_{loc}$, all of which reached highly significant success rates (**Figure 4B**, $P = 0.0033$, $N = 11$ gerbils, Wilcoxon signed-rank test). Swapping the identity of the target and background loudspeakers in 1/8 of trials during the testing phase (the identities of target and background loudspeakers remained fixed during training) resulted in performance levels that were significantly lower than chance level (**Figure 4B**, $P = 0.018$, $N = 7$ gerbils, Wilcoxon signed-rank test). Given that these catch-trials started with the presentation of the usual target stimulus, the animals could potentially have just stopped moving immediately after initiating a trial in anticipation of the reward, which could explain the extremely low success rate. However, further analysis revealed that the animals indeed encountered the target-islands with similar prevalence in catch-trials as in normal trials, but rarely remained in the island for the required sit-time in catch-trials (**Supplementary Figure S4**). Thus, the animals

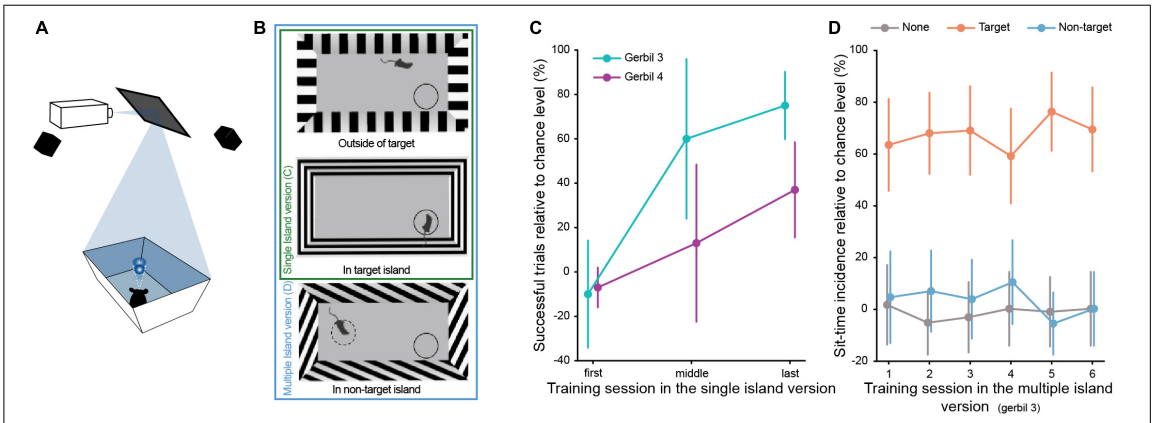


FIGURE 3 | (A) Schematic representation of the ratCAVE setup (adapted from Del Grosso et al., 2017). **(B)** Schematic representation (top view) of the $vSIT_{ori}$ arena, both in the single and in the multiple island version. Animal entrance to a target island and non-target island triggered the change of grating orientation from vertical to horizontal or oblique, respectively. All gratings used only differed in orientation angle. Differences in appearance is due to visual angle from above. **(C)** Performance of each gerbil in example sessions from the beginning, middle, and end of the training, in the single island task (error bars correspond to the 95% confidence interval). First session: $N_{Gerbil\ 3}$ = 20 trials, $N_{Gerbil\ 4}$ = 42 trials; middle session: $N_{Gerbil\ 3}$ = 12 trials, $N_{Gerbil\ 4}$ = 20 trials; last session: $N_{Gerbil\ 3}$ = 40 trials, $N_{Gerbil\ 4}$ = 32 trials. **(D)** Incidence of sit-time across sessions relative to chance level per island, in the multiple island task (error bars correspond to the 95% confidence interval). Gerbil 3 $N_{Session\ 1}$ = 34 trials; $N_{Session\ 2}$ = 41 trials; $N_{Session\ 3}$ = 43 trials; $N_{Session\ 4}$ = 39 trials; $N_{Session\ 5}$ = 44 trials; $N_{Session\ 6}$ = 39 trials.

actively avoided staying in the target island in these catch-trials, revealing that they indeed associated the identity of the active loudspeaker (target or background) with reward predictability. Since the spatial location of the active loudspeaker was the only parameter that allowed determination of loudspeaker identity, these data validate that the animals were actively localizing the sound source to achieve task performance. Hence, similar as for frequency discrimination, the gerbils did not follow an oddball strategy but specifically searched for the target stimulus.

We also tested to which extent the animals associate their locomotive searching behavior with target detection success. To this end, we employed a biased distribution likelihood of target island locations in the arena. We found that after the animals were trained on one specific distribution likelihood, their arena occupancy was specific to this distribution (**Supplementary Figure S5**). That is, the animals predominantly visited locations in the arena that were most likely to contain the target island. Thus, a clear association existed between the animals' locomotive behavior and their reward expectancy, i.e., they actively searched for the target island position. Together, these data validate that SIT allows the interrogation of different cues based on the concept of a locomotive search for a target stimulus (i.e., island).

Electrophysiological Recording of Neural Activity During SIT Performance

We were interested in combing SIT with chronic electrophysiological recording techniques. Specifically, we asked to what extent the unrestricted self-movement and task relevance that are provided by SIT might facilitate exploring neural signatures of spatial processing in AC. Therefore, we implanted a tetrode bundle in AC of a previously trained gerbil (see section "Materials and Methods"), and recorded brain activity during task performance in $aSIT_{loc}$. We collected local field potential (LFP), from which we calculated Auditory Evoked Potentials (AEPs). Remarkably, although the acoustic stimulation was identical from both loudspeakers (sound intensity was roved throughout trials), AEPs were different between the two sound sources (**Figures 4C,D**). Specifically, AEP amplitudes were significantly larger during stimulation by the target loudspeaker (P = 0.000049, Mann-Whitney U-test). A plausible reason for this difference in AEP amplitude could be differences in the intensity of the sounds presented from each loudspeaker, due to the animal being closer to the target loudspeaker than to the non-target, at the moment of respective sound presentation. This does not seem to be the case, as the histograms of animal position for target and non-target loudspeaker sound presentations do not show such a bias (**Supplementary Figure S6**). More likely, these data suggest that the learned relevance of each specific sound source modulates neural response amplitude. Such differences in sound-source-specific responses have – to our knowledge – not previously been reported in studies on spatial processing and thus demonstrate that the use of SIT may be beneficial to reveal neuronal signatures of sensory processing under ecologically relevant conditions.

DISCUSSION

SIT is a novel experimental paradigm for freely moving animals that are actively engaged in a sensory processing task and can be combined with simultaneous neural recordings. It exploits voluntary exploratory self-motion – and its cessation upon detection of a change in the sensory stimulation – for testing psychophysical sensitivity in a variety of cues and

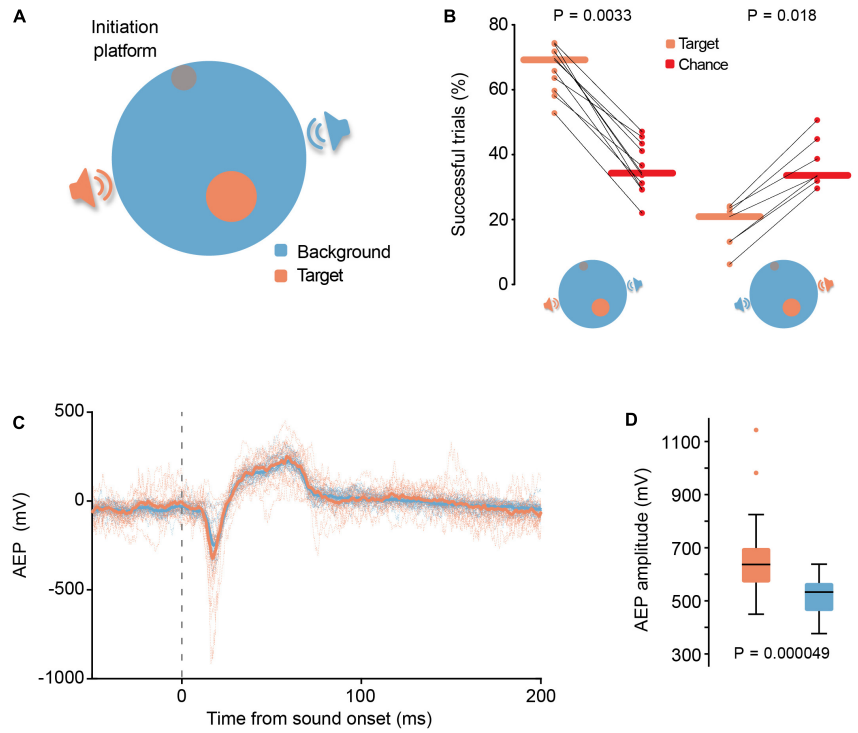


FIGURE 4 | (A) Schematic representation (top view) of $aSIT_{loc}$, a sound source localization version of SIT. **(B)** Left-hand panel: Reporting of sound location was highly significant ($P = 0.0033$, $N = 11$ gerbils, Wilcoxon signed-rank test) compared to their chance level given their actual locomotion behavior, calculated by surrogate island computation. Right-hand panel: In 1/8 of the trials the identity of the target and background loudspeakers was swapped for a subset of the animals. A significant decrease in the performance for the "swapped" trials below chance level ($P = 0.018$, $N = 7$ gerbils, horizontal lines depict the median, Wilcoxon signed-rank test) suggests the animals were actively avoiding the target island under these conditions. **(C,D)** Response magnitude differences in auditory evoked potential (AEP) recordings of auditory cortex neural populations. **(C)** Single session example traces. Dotted lines represent AEP per trial and active loudspeaker identity. Thick traces represent the median of all trials. **(D)** Quantification of AEP amplitude. In this example session the AEP amplitude was significantly larger during target loudspeaker activity ($P = 0.000049$; $N_{target} = 19$, $N_{background} = 23$; Mann-Whitney U-test). Boxplots depict the median (black line), 1st and 3rd quartile (filled boxes), ± 2.7 σ (whiskers) and outliers (crosses).

sensory modalities. Self-motion occurs constantly under natural conditions and, throughout evolution, neural processing has adapted to the resulting continuous modulation of the sensory input (Niell and Stryker, 2010; Zhou et al., 2014; McGinley et al., 2015; Williamson et al., 2015; Willett et al., 2019). SIT consequently captures ethologically relevant behavior that is crucial for sensory processing and decision making. SIT was inspired by existing closed-loop free navigation assays (Polley et al., 2004; Whitton et al., 2014), but differs significantly in a number of aspects. Most importantly, the introduction of discrete sensory islands instead of a gradient fundamentally changes the locomotion behavior toward free exploration of the entire arena. Moreover, the introduction of multiple islands allows the interrogation of animals about perception thresholds and the construction of psychometric functions.

The last decade has seen a rise in the study of perceptual decision making, particularly in rodents. Data from established and commonly used paradigms, such as go/no-go tasks (G/NG) and two alternative forced choice tasks (2AFC), can be difficult to interpret. For example, in 2AFC designs, the animals are forced to give an answer on every trial, which renders the disentanglement between real decisions and guesses difficult (Carandini and Churchland, 2013). The sensory environment in which rodents are immersed while performing these tasks has been increasing in complexity in recent years, from lever operation, to full 360° virtual reality with online locomotive update. However, animals require substantial training to learn how to use and navigate these setups. Moreover, a major drawback of many virtual reality setups is a lack of vestibular feedback (due to head fixation) that is naturally present during self-movement.

In contrast, SIT is characterized by shorter training periods than many traditional behavioral paradigms or techniques involving virtual reality (e.g., as little as one training session for gerbils in $aSIT_{freq}$), high flexibility to readily adapt parameters to both the constraints of the scientific question at hand and to the behavioral characteristics of the animal clade used. If required

(e.g., depending on complexity and species), the motivational state of the animals can be controlled by addition of an initiation platform, which assures the willingness of the individual to perform a trial. In essence, SIT represents a refined version of a G/NG task. Nonetheless, the possibility to add multiple non-target islands allows testing of cue identification and determining psychometric functions. In its currently presented form with pseudo-randomized island locations, SIT does not represent a spatial association nor a long-term memory task. Nonetheless, SIT can be easily transformed into such a task by maintaining the target island location constant across trials or switching between a limited number of target locations; e.g., a recent study by Rossato et al. (2018) which used electromagnets to switch between available islands in the Morris water maze could be performed in SIT, with greater flexibility due to the amount and position of the islands depending on software rather than hardware. In addition, the lack of water in SIT facilitates maintenance of the setup and coupling of experiments with interventions such as electrophysiology. Although dry versions of the water maze already exist, such as in Bast et al. (2005), where animals forage for food in hidden compartments, SIT provides an easier, more versatile alternative in which the search for food can be replaced by the search for target island (to receive food reward). Thus, spatial learning and memory studies in relation to sensory cuing could be performed, a task of high ecological relevance in many species (Sherry, 1985; Collett et al., 1986).

In any of its potential variants, combining SIT with specific time points of electrode implantation (e.g., before/during training), opens exciting possibilities to study aspects of learning and plasticity of sensory processing during voluntary self-motion and active sensing. We have exemplified some of this potential here, as our AC recording during $aSIT_{loc}$ revealed previously unreported response modulation of spatial sensitivity based on sound source identity. Previous reports had established that neuronal responses in auditory cortex can be modulated by "attention" (Hubel et al., 1959; Evans and Whitfield, 1964). Our findings are related, but potentially more profound, as the difference in responses to both loudspeakers is unlikely to be due to the attentive state of the animal, but rather the relative relevance of the two sound sources regarding reward expectancy and experimental design. Multiple studies in AC have found relevance-specific response modulation in animals if engaged in the experimental task (Miller et al., 1972; Fritz et al., 2003, 2007; Atiani et al., 2009; Otazu et al., 2009; Lee and Middlebrooks, 2011; Guo et al., 2019). Moreover, a recent study with macaque monkeys that were trained to respond differentially to the same auditory stimulation depending on the context reported larger auditory cortex responses to the same stimulus when it required a no-go response (Huang et al., 2019). Likewise, greater neural responses during $aSIT_{loc}$ were observed for target sounds that required the animal to remain sitting.

In summary, SIT is a flexible and easily implementable behavioral paradigm that uniquely incorporates self-motion and natural exploratory behavior, which are essential for ecological sensory processing. SIT is readily applicable across species and sensory modalities and extendable to use for neurophysiological investigations. Beyond the options we have exemplified here, SIT is widely adaptable to a large variety of neuroscientific and ecological fields. For example, besides the auditory and visual cues probed here, we suggest that somatosensory cues can be studied by dynamically changing the floor texture, or olfactory sensitivity could be tested collocating the target island and odor release valves beneath the arena. Similarly, decision-making based on congruent or ambiguous combinations of different sensory modalities is ecologically important and could readily be applied in SIT. In the future, it would be particularly interesting to use high yield recording devices, such as neuropixel electrodes (Juavinett et al., 2019), to sample a wide range of brain areas. Moreover, the ongoing miniaturization of technology will allow precise stimulus control in various sensory modalities and combinations (e.g., through wireless miniature cameras or microphones). These new technologies coupled with SIT should garner unprecedented insights to unravel ecologically relevant sensory neural processes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

All procedures involving gerbils were reviewed and approved by the Regierung von Oberbayern AZ 55.2-1-54-2532-74-2016 and AZ 55.2-1-54-2532-70-2016. All procedures involving mouse lemurs were reviewed and approved by the Animal Welfare Committee of the LAVES (AZ 33.19-42502-04-18/3050).

AUTHOR CONTRIBUTIONS

MP, DNF, DA, and LB conceived SIT. MP, DNF, and DA designed the experiments. BG, DS, and ASi contributed to paradigm refinement. DA, DNF, DS, ASo, and PG performed the experiments. LB, DA, DS, and ASo contributed to programming SIT code. DA, DNF, DS, and ASo analyzed the results. DNF, MP, and DA designed and generated the figures and wrote the manuscript. All authors provided comments and approved the manuscript.

FUNDING

This study was supported by the Deutsche Forschungsgemeinschaft DFG (PE2251/2-1 to MP, and SFB 870, project B02 to MP and BG, and RTG 2175 to ASi), the Munich Cluster for Systems Neurology (SyNergy, EXC 1010 to ASi), Bundesministerium für Bildung und Forschung via grant no. 01GQ0440 (Bernstein Centre for Computational Neuroscience Munich, to ASi), European Union Horizon

2020 FETPOACT program via grant agreement no. 723032 (BrainCom, to ASi) and by the International Max Planck Research School for Molecular Life Sciences (to DA).

ACKNOWLEDGMENTS

We thank C. Leibold for valuable comments and F. Felmy for support.

DEDICATION

This work is dedicated to Lutz Wiegrebe, whose expertise and help was invaluable for the development of SIT.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbeh.2020.576154/full#supplementary-material

FIGURE S1 | (A) Schematic representation of the surrogate island random permutation. Colored line depicts a real trajectory of an animal in a trial color coded with the time at which the animal was at each position, starting from the initiation platform (filled gray circle). The real target island is where the animal ends (open orange circle). The dots correspond to the position of the animal with 1s interval between them. The chance level of task completion was calculated using *a posteriori* surrogate island locations (open black circles, only a few shown here from the 1000 actually used for each trial). **(B)** Trajectories of gerbil 1 and 2 during the 2nd training session separated by correct and incorrect trials: no apparent change in pattern of locomotion is seen when the animal did not succeed in the task. **(C)** Comparison of the distance of the target island's center to the center of the arena between correct and incorrect trials for gerbil 1 and 2 in the same session as in B. Gerbil 1: $N_{correct}$ = 44, $N_{incorrect}$ = 28, P = 0.12; Gerbil 2: $N_{correct}$ = 44, $N_{incorrect}$ = 21, P = 0.63 (Mann-Whitney U test). Boxplots depict the median (black line), 1st and 3rd quartile (filled boxes), ± 2.7 σ (whiskers) and outliers (cross).

FIGURE S2 | Duration of correct trials in $aSIT_{freq}$ for gerbils (left panel) and for mouse lemur (Right panel).

FIGURE S3 | Mouse lemur performance at intermediate training sessions, relative to chance level. Target island diameter = 26.7 cm. For mouse lemur 1, x = 6 and sit-time = 4 s. For mouse lemur 2, x = 5 and sit-time = 2 s.

FIGURE S4 | Comparison in the $aSIT_{loc}$ version between the trials in which the target loudspeaker was the one from the training, with catch-trials (1/8 of total trials) in which the opposite loudspeaker was the target one. **(A)** The gerbils found the target island as often in catch-trials as in normal target trials. **(B)** The gerbils left the target island much more often (∼85% trials) in catch-trials than in normal target trials (∼35% trials). Only situations where the gerbils stayed in the target island for at least 1 s were used to assure the gerbil listened to the sound and did not just run through the island. Number of sessions: 39; Number of normal target trials: 1784; Number of catch trials: 285. Uncertainty was determined using a bootstrapping method.

FIGURE S5 | Association between spatial position and stimulus change in the $aSIT_{loc}$. **(A)** Distribution of the target islands for all the trials in a session where there was not a target location bias (left) and in a session where there was a target location bias (right). The filled gray circle corresponds to the initiation platform. The dashed magenta circle radius is twice as large as that of a target island and divides the target islands which were considered to be in the center (light gray circles) from the target islands considered not to be in the center (dark blue circles). In sessions without target location bias ∼59% of the islands occurred in the center whereas, in sessions with target location bias, ∼78% occurred in the center. **(B)** Difference in percentage of successful trials between trials in which the target was in the center and trials in which the target was not in the center (error bars correspond to the 95% confidence interval, calculated using a bootstrapping method). Gerbil 1 and 2 (these are not the same gerbils that were trained in $aSIT_{freq}$) were first trained in an unbiased condition and the bias condition was later introduced. Gerbil 3 was first trained in a biased condition, and the bias was later removed. When the target location was biased to the center, the animals spent more time in that region and their performance increased in relation to when the target was outside the center.

FIGURE S6 | Histograms of gerbil position at sound presentation times for the session during which LFP was recorded, reported on main (**Figure 4**). Left panel shows the histogram for target stimulus presentations (orange loudspeaker). Right panel shows the histogram for background stimulus presentations (blue loudspeaker).

REFERENCES

Altieri, N., Stevenson, R. A., Wallace, M. T., and Wenger, M. J. (2015). Learning to associate auditory and visual stimuli: behavioral and neural mechanisms. *Brain Topogr.* 28, 479–493.

Atiani, S., Elhilali, M., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron* 61, 467–480.

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., et al. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron* 97, 640.e4–655.e4. doi: 10.1016/j.neuron.2017.12.034

Bast, T., da Silva, B. M., and Morris, R. G. M. (2005). Distinct contributions of hippocampal NMDA and AMPA receptors to encoding and retrieval of one-trial place memory. *J. Neurosci.* 25, 5845–5856. doi: 10.1523/JNEUROSCI.0698-05.2005

Buran, B. N., von Trapp, G., and Sanes, D. H. (2014). Behaviorally gated reduction of spontaneous discharge can improve detection thresholds in auditory cortex. *J. Neurosci.* 34, 4076–4081. doi: 10.1523/JNEUROSCI.4825-13.2014

Busse, L., Ayaz, A., Dhruv, N. T., Katzner, S., Saleem, A. B., Scholvinck, M. L., et al. (2011). The detection of visual contrast in the behaving mouse. *J. Neurosci.* 31, 11351–11361. doi: 10.1523/JNEUROSCI.6689-10.2011

Campbell, M. G., and Giocomo, L. M. (2018). Self-motion processing in visual and entorhinal cortices: inputs, integration, and implications for position coding. *J. Neurophysiol.* 120, 2091–2106. doi: 10.1152/jn.00686.2017

Carandini, M., and Churchland, A. K. (2013). Probing perceptual decisions in rodents. *Nat. Neurosci.* 16, 824–831. doi: 10.1038/nn.3410

Case, L. K., Pineda, J., and Ramachandran, V. S. (2015). Common coding and dynamic interactions between observed, imagined, and experienced motor and somatosensory activity. *Neuropsychologia* 79, 233–245. doi: 10.1016/j.neuropsychologia.2015.04.005

Chabrol, F. P., Arenz, A., Wiechert, M. T., Margrie, T. W., and DiGregorio, D. A. (2015). Synaptic diversity enables temporal coding of coincident multisensory inputs in single neurons. *Nat. Neurosci.* 18, 718–727. doi: 10.1038/nn.3974

Chen, C., Krueger-Burg, D., and de Hoz, L. (2019). Wide sensory filters underlie performance in memory-based discrimination and generalization. *PLoS One* 14:e0214817. doi: 10.1371/journal.pone.0214817

Chiappe, M. E., Seelig, J. D., Reiser, M. B., and Jayaraman, V. (2010). Walking modulates speed sensitivity in *Drosophila* motion vision. *Curr. Biol.* 20, 1470–1475. doi: 10.1016/j.cub.2010.06.072

Clancy, K. B., Orsolic, I., and Mrsic-Flogel, T. D. (2019). Locomotion-dependent remapping of distributed cortical networks. *Nat. Neurosci.* 22, 778–786. doi: 10.1038/s41593-019-0357-8

Collett, T. S., Cartwright, B. A., and Smith, B. A. (1986). Landmark learning and visuo-spatial memories in gerbils. *J. Comp. Physiol.* 158, 835–851. doi: 10.1007/BF01324825

Dadarlat, M. C., and Stryker, M. P. (2017). Locomotion enhances neural encoding of visual stimuli in mouse V1. *J. Neurosci.* 37, 3764–3775. doi: 10.1523/JNEUROSCI.2728-16.2017

Del Grosso, N. A., Graboski, J. J., Chen, W., Blanco-Hernández, E., and Sirota, A. (2017). Virtual reality system for freely-moving rodents. *bioRxiv [preprint]* doi: 10.1101/161232. Available online at: https://www.biorxiv.org/content/10.1101/161232v1.full

Del Grosso, N. A., and Sirota, A. (2019). Ratcave: a 3D graphics python package for cognitive psychology experiments. *Behav. Res. Methods* 51, 2085–2093. doi: 10.3758/s13428-019-01245-x.

Deneux, T., Harrell, E. R., Kempf, A., Ceballo, S., Filipchuk, A., and Bathellier, B. (2019). Context-dependent signaling of coincident auditory and visual events in primary visual cortex. *eLife* 8:e44006. doi: 10.7554/eLife.44006

Diehl, M. M., and Romanski, L. M. (2014). Responses of prefrontal multisensory neurons to mismatching faces and vocalizations. *J. Neurosci.* 34, 11233–11243. doi: 10.1523/JNEUROSCI.5168-13.2014

Etienne, A. S., Maurer, R., and Séguinot, V. (1996). Path integration in mammals and its interaction with visual landmarks. *J. Exp. Biol.* 199, 201–209.

Evans, E. F., and Whitfield, I. C. (1964). Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat. *J. Physiol.* 171, 476–493. doi: 10.1113/jphysiol.1964.sp007391

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223. doi: 10.1038/nn1141

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hear. Res.* 229, 186–203. doi: 10.1016/j.heares.2007.01.009

Fu, Y., Tucciarone, J. M., Espinosa, J. S., Sheng, N., Darcy, D. P., Nicoll, R. A., et al. (2014). A cortical circuit for gain control by behavioral state. *Cell* 156, 1139–1152. doi: 10.1016/j.cell.2014.01.050

Gao, P., and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* 32, 148–155. doi: 10.1016/j.conb.2015.04.003

Guo, L., Weems, J. T., Walker, W. I., Levichev, A., and Jaramillo, S. (2019). Choice-selective neurons in the auditory cortex and in its striatal target encode reward expectation. *J. Neurosci.* 39, 3687–3697. doi: 10.1523/JNEUROSCI.2585-18.2019

Huang, Y., Heil, P., and Brosch, M. (2019). Associations between sounds and actions in early auditory cortex of nonhuman primates. *eLife* 8:e43281. doi: 10.7554/eLife.43281

Hubel, D. H., Henson, C. O., Rupert, A., and Galambos, R. (1959). Attention. Units in the auditory cortex. *Science* 129, 1279–1280. doi: 10.1126/science.129.3358.1279

Juavinett, A. L., Bekheet, G., and Churchland, A. K. (2019). Chronically implanted Neuropixels probes enable high-yield recordings in freely moving mice. *eLife* 8:e47188. doi: 10.7554/eLife.47188

Keller, G. B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* 74, 809–815. doi: 10.1016/j.neuron.2012.03.040

Kerekes, P., Daret, A., Shulz, D. E., and Ego-Stengel, V. (2017). Bilateral discrimination of tactile patterns without whisking in freely running rats. *J. Neurosci.* 37, 7567–7579. doi: 10.1523/JNEUROSCI.0528-17.2017

Klinge, A., and Klump, G. M. (2009). Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans. *J. Acoust. Soc. Am.* 125, 304–314. doi: 10.1121/1.3021315

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., and Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron* 93, 480–490. doi: 10.1016/j.neuron.2016.12.041

Lee, C.-C., and Middlebrooks, J. C. (2011). Auditory cortex spatial sensitivity sharpens during task performance. *Nat. Neurosci.* 14, 108–114. doi: 10.1038/nn.2713

Lorenz, K. (1981). *The Foundations of Ethology. This Engl. ed. is a Rev. and enl. Version*. New York Wien: Springer.

Ma, W. J., and Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annu. Rev. Neurosci.* 37, 205–220. doi: 10.1146/annurev-neuro-071013-014017

Maimon, G., Straw, A. D., and Dickinson, M. H. (2010). Active flight increases the gain of visual motion processing in *Drosophila*. *Nat. Neurosci.* 13, 393–399. doi: 10.1038/nn.2492

McGinley, M. J., David, S. V., and McCormick, D. A. (2015). Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron* 87, 179–192. doi: 10.1016/j.neuron.2015.05.038

Middlebrooks, J., and Knudsen, E. (1984). A neural code for auditory space in the cat's superior colliculus. *J. Neurosci.* 4, 2621–2634. doi: 10.1523/JNEUROSCI.04-10-02621.1984

Miller, J. M., Sutton, D., Pfingst, B., Ryan, A., Beaton, R., and Gourevitch, G. (1972). Single cell activity in the auditory cortex of rhesus monkeys: behavioral dependency. *Science* 177, 449–451. doi: 10.1126/science.177.4047.449

Niell, C. M., and Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65, 472–479. doi: 10.1016/j.neuron.2010.01.033

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., and Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cereb. Cortex* 18, 598–609. doi: 10.1093/cercor/bhm091

Otazu, G. H., Tai, L.-H., Yang, Y., and Zador, A. M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nat. Neurosci.* 12, 646–654. doi: 10.1038/nn.2306

Picq, J.-L. (2016). "The gray mouse lemur (*Microcebus murinus*): a novel cognitive primate brain aging model," in *The Dwarf and Mouse Lemurs of Madagascar*, eds S. M. Lehman, U. Radespiel, and E. Zimmermann (Cambridge: Cambridge University Press), 381–404. doi: 10.1017/CBO9781139871822.021

Polley, D. B., Heiser, M. A., Blake, D. T., Schreiner, C. E., and Merzenich, M. M. (2004). Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16351–16356. doi: 10.1073/pnas.0407586101

Rancz, E. A., Moya, J., Drawitsch, F., Brichta, A. M., Canals, S., and Margrie, T. W. (2015). Widespread vestibular activation of the rodent cortex. *J. Neurosci.* 35, 5926–5934. doi: 10.1523/JNEUROSCI.1869-14.2015

Rossato, J. I., Moreno, A., Genzel, L., Yamasaki, M., Takeuchi, T., Canals, S., et al. (2018). Silent learning. *Curr. Biol.* 28, 3508.e5–3515.e5. doi: 10.1016/j.cub.2018.09.012

Saleem, A. B., Diamanti, E. M., Fournier, J., Harris, K. D., and Carandini, M. (2018). Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature* 562, 124–127. doi: 10.1038/s41586-018-0516-1

Schneider, D. M., and Mooney, R. (2018). How movement modulates hearing. *Annu. Rev. Neurosci.* 41, 553–572. doi: 10.1146/annurev-neuro-072116-031215

Schneider, D. M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* 513, 189–194. doi: 10.1038/nature13724

Schnupp, J., Nelken, I., and King, A. (2011). *Auditory Neuroscience: Making Sense of Sound*. Cambridge, Mass: MIT Press.

Schopf, C., Zimmermann, E., Tünsmeyer, J., Kästner, S. B. R., Hubka, P., and Kral, A. (2014). Hearing and age-related changes in the gray mouse lemur. *JARO* 15, 993–1005. doi: 10.1007/s10162-014-0478-4

Sherry, D. F. (1985). "Food storage by birds and mammals," in *Advances in the Study of Behavior*, (Amsterdam: Elsevier), 15, 153–188. doi: 10.1016/S0065-3454(08)60489-1

The International Brain Laboratory, Aguillon-Rodriguez, V., Angelaki, D. E., Bayer, H. M., Bonacchi, N., and Carandini, M. (2020). A standardized and reproducible method to measure decision-making in mice. *bioRxiv [preprint]* doi: 10.1101/2020.01.17.909838

Tinbergen, N. (1963). On aims and methods of Ethology. *Z. Tierpsychol.* 20, 410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x

Vélez-Fort, M., Bracey, E. F., Keshavarzi, S., Rousseau, C. V., Cossell, L., Lenzi, S. C., et al. (2018). A circuit for integration of head- and visual-motion signals in layer 6 of mouse primary visual cortex. *Neuron* 98, 179.e6–191.e6. doi: 10.1016/j.neuron.2018.02.023

von Frisch, K. (1954). *The Dancing Bees: An Account of the Life and Senses of the Honey Bee*. Wien: Springer-Verlag, doi: 10.1007/978-3-7091-4697-2

von Holst, E., and Mittelstaedt, H. (1950). Das Reafferenzprinzip: wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften* 37, 464–476. doi: 10.1007/BF00622503

Whitton, J. P., Hancock, K. E., and Polley, D. B. (2014). Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2606–E2615. doi: 10.1073/pnas.1322184111

Willett, S. M., Groh, J. M., and Maddox, R. K. (2019). "Hearing in a 'Moving' visual world: coordinate transformations along the auditory pathway," in *Multisensory Processes Springer Handbook of Auditory Research*, eds A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, and R. R. Fay (Cham: Springer International Publishing), 85–104. doi: 10.1007/978-3-030-10461-0_5

Williamson, R. S., Hancock, K. E., Shinn-Cunningham, B. G., and Polley, D. B. (2015). Locomotion and task demands differentially modulate thalamic audiovisual processing during active search. *Curr. Biol.* 25, 1885–1891. doi: 10.1016/j.cub.2015.05.045

Wood, K. C., Town, S. M., and Bizley, J. K. (2019). Neurons in primary auditory cortex represent sound source location in a cue-invariant manner. *Nat. Commun.* 10:3019. doi: 10.1038/s41467-019-10868-9

Zhou, M., Liang, F., Xiong, X. R., Li, L., Li, H., Xiao, Z., et al. (2014). Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. *Nat. Neurosci.* 17, 841–850. doi: 10.1038/nn.3701

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] T. L. Arbogast and G. Kidd. *Evidence for spatial tuning in informational masking using the probe-signal method*. J. Acoust. Soc. Am. **108**, 1803–1810 (2000). See page: 1

[2] V. Best, E. J. Ozmeral, and B. G. Shinn-Cunningham. *Visually-guided Attention Enhances Target Identification in a Complex Auditory Scene*. JARO **8**, 294–304 (2007). See pages: 1, 77

[3] J. Schnupp, I. Nelken, and A. J. King. *Auditory Neuroscience: Making Sense of Sound*. The MIT Press (2010). See pages: 1, 15

[4] B. Grothe, M. Pecka, and D. McAlpine. *Mechanisms of Sound Localization in Mammals*. Physiol. Rev. **90**, 983–1012 (2010). See pages: 1, 2, 3

[5] A. Brughera, L. Dunai, and W. M. Hartmann. *Human interaural time difference thresholds for sine tones: The high-frequency limit*. J. Acoust. Soc. Am. **133**, 2839–2855 (2013). See page: 2

[6] B. M. Sayers. *Acoustic-Image Lateralization Judgments with Binaural Tones*. J. Acoust. Soc. Am. **36**, 923–926 (1964). See page: 2

[7] S. S. Stevens and E. B. Newman. *The Localization of Actual Sources of Sound*. Am. J. Psychol. **48**, 297 (1936). See pages: 2, 4, 74, 80

[8] D. McFadden and E. G. Pasanen. *Lateralization at high frequencies based on interaural time differences*. J. Acoust. Soc. Am. **59**, 634–639 (1976). See page: 2

[9] G. B. Henning. *Detectability of interaural delay in high-frequency complex waveforms*. J. Acoust. Soc. Am. **55**, 84–90 (1974). See page: 2

[10] R. Galambos and H. Davis. *The Response of Single Auditory-Nerve Fibers to Acoustic Stimulation*. J. Neurophysiol. **6**, 39–57 (1943). See page: 3

[11] J. E. Rose, J. F. Brugge, D. J. Anderson, and J. E. Hind. *Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey.* J. Neurophysiol. **30**, 769–793 (1967). See page: 3

[12] C. M. Hackney, K. K. Osen, and J. Kolston. *Anatomy of the cochlear nuclear complex of guinea pig*. Anat. Embryol. **182** (1990). See page: 3

[13] N. Golding, D. Robertson, and D. Oertel. *Recordings from slices indicate that octopus cells of the cochlear nucleus detect coincident firing of auditory nerve fibers with temporal precision*. J. Neurosci. **15**, 3138–3153 (1995). See page: 3

[14] D. Oertel and E. D. Young. *What's a cerebellar circuit doing in the auditory system?* Trends Neurosci. **27**, 104–110 (2004). See page: 3

[15] P. X. Joris, L. H. Carney, P. H. Smith, and T. C. Yin. *Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency*. J. Neurophysiol. **71**, 1022–1036 (1994). See page: 3

[16] P. G. Finlayson and D. M. Caspary. *Low-frequency neurons in the lateral superior olive exhibit phase-sensitive binaural inhibition*. J. Neurophysiol. **65**, 598–605 (1991). See page: 3

[17] D. McAlpine, D. Jiang, and A. R. Palmer. *A neural code for low-frequency sound localization in mammals*. Nat. Neurosci. **4**, 396–401 (2001). See pages: 3, 6, 76, 80

[18] B. Grothe and D. H. Sanes. *Bilateral inhibition by glycinergic afferents in the medial superior olive*. J. Neurophysiol. **69**, 1192–1196 (1993). See page: 4

[19] A. Brand, O. Behrend, T. Marquardt, D. McAlpine, and B. Grothe. *Precise inhibition is essential for microsecond interaural time difference coding*. Nature **417**, 543–547 (2002).

[20] M. Pecka, A. Brand, O. Behrend, and B. Grothe. *Interaural Time Difference Processing in the Mammalian Medial Superior Olive: The Role of Glycinergic Inhibition*. J. Neurosci. **28**, 6914–6925 (2008).

[21] M. Roberts, S. Seeman, and N. Golding. *A Mechanistic Understanding of the Role of Feedforward Inhibition in the Mammalian Sound Localization Circuitry*. Neuron **78**, 923–935 (2013). See page: 4

[22] C. K. Henkel and K. M. Spangler. *Organization of the efferent projections of the medial superior olivary nucleus in the cat as revealed by HRP and autoradiographic tracing methods*. J. Comp. Neurol. **221**, 416–428 (1983). See page: 4

[23] J. K. Brunso-Bechtold, G. C. Thompson, and R. B. Masterton. *HRP study of the organization of auditory afferents ascending to central nucleus of inferior colliculus in cat*. J. Comp. Neurol. **197**, 705–722 (1981). See page: 4

[24] K. K. Glendenning, B. N. Baker, K. A. Hutson, and R. B. Masterton. *Acoustic chiasm V: Inhibition and excitation in the ipsilateral and contralateral projections of LSO*. J. Comp. Neurol. **319**, 100–122 (1992). See page: 4

[25] W. C. Loftus, D. C. Bishop, R. L. Saint Marie, and D. L. Oliver. *Organization of binaural excitatory and inhibitory inputs to the inferior colliculus from the superior olive*. J. Comp. Neurol. **472**, 330–344 (2004). See page: 4

[26] G. D. Pollak. *Circuits for processing dynamic interaural intensity disparities in the inferior colliculus*. Hear. Res. **288**, 47–57 (2012). See page: 4

[27] A. M. H. Lesicko, T. S. Hristova, K. C. Maigler, and D. A. Llano. *Connectional Modularity of Top-Down and Bottom-Up Multimodal Inputs to the Lateral Cortex of the Mouse Inferior Colliculus*. J. Neurosci. **36**, 11037–11050 (2016). See page: 4

[28] S. K. Itaya and G. W. Van Hoesen. *Retinal innervation of the inferior colliculus in rat and monkey*. Brain Res. **233**, 45–52 (1982). See page: 4

[29] M. H. Cooper and P. A. Young. *Cortical projections to the inferior colliculus of the cat*. Exp. Neurol. **51**, 488–502 (1976). See page: 4

[30] B. M. Olthof, A. Rees, and S. E. Gartside. *Multiple Nonauditory Cortical Regions Innervate the Auditory Midbrain*. J. Neurosci. **39**, 8916–8928 (2019). See pages: 4, 8

[31] E. Saldaña and M. A. Merchań. *Intrinsic and commissural connections of the rat inferior colliculus: Intrinsic and Commissural Collicular Connections*. J. Comp. Neurol. **319**, 417–437 (1992). See page: 4

[32] L. D. Orton, C. A. Papasavvas, and A. Rees. *Commissural Gain Control Enhances the Midbrain Representation of Sound Location*. J. Neurosci. **36**, 4470–4481 (2016). See page: 4

[33] M. Pecka, I. Siveke, B. Grothe, and N. A. Lesica. *Enhancement of ITD Coding Within the Initial Stages of the Auditory Pathway*. J. Neurophysiol. **103**, 38–46 (2010). See page: 4

[34] A. Shneiderman, D. L. Oliver, and C. K. Henkel. *Connections of the dorsal nucleus of the lateral lemniscus: An inhibitory parallel pathway in the ascending auditory system?* J. Comp. Neurol. **276**, 188–208 (1988). See page: 4

[35] A. R. Palmer and A. J. King. *The representation of auditory space in the mammalian superior colliculus*. Nature **299**, 248–249 (1982). See page: 4

[36] R. A. A. Campbell, T. P. Doubell, F. R. Nodal, J. W. H. Schnupp, and A. J. King. *Interaural Timing Cues Do Not Contribute to the Map of Space in the Ferret Superior Colliculus: A Virtual Acoustic Space Study*. J. Neurophysiol. **95**, 242–254 (2006). See page: 4

[37] S. Ito, Y. Si, D. A. Feldheim, and A. M. Litke. *Spectral cues are necessary to encode azimuthal auditory space in the mouse superior colliculus*. Nat. Commun. **11**, 1087 (2020). See page: 4

[38] M. Kudo and K. Niimi. *Ascending projections of the inferior colliculus onto the medial geniculate body in the cat studied by anterograde and retrograde tracing techniques*. Brain Res. **155**, 113–117 (1978). See page: 4

[39] J. A. Winer, R. L. Saint Marie, D. T. Larue, and D. L. Oliver. *GABAergic feedforward projections from the inferior colliculus to the medial geniculate body.* Proc. Natl. Acad. Sci. U.S.A. **93**, 8005–8010 (1996). See page: 4

[40] J. G. Mellott, N. L. Foster, A. P. Ohl, and B. R. Schofield. *Excitatory and inhibitory projections in parallel pathways from the inferior colliculus to the auditory thalamus*. Front. Neuroanat. **8** (2014). See page: 4

[41] B. Clarke and C. Lee. *Inhibitory Projections in the Mouse Auditory Tectothalamic System*. Brain Sciences **8**, 103 (2018). See page: 4

[42] K. McAlonan. *Attentional Modulation of Thalamic Reticular Neurons*. J. Neurosci. **26**, 4444–4450 (2006). See page: 4

[43] F. K. Samson et al. *Directionality Derived From Differential Sensitivity to Monaural and Binaural Cues in the Cat's Medial Geniculate Body*. J. Neurophysiol. **84**, 1330–1345 (2000). See pages: 4, 75

[44] D. C. Fitzpatrick, R. Batra, T. R. Stanford, and S. Kuwada. *A neuronal population code for sound localization*. Nature **388**, 871–874 (1997). See page: 4

[45] J. D. Yao, P. Bremen, and J. C. Middlebrooks. *Emergence of Spatial Stream Segregation in the Ascending Auditory Pathway*. J. Neurosci. **35**, 16199–16212 (2015). See page: 4

[46] W. M. Jenkins and M. M. Merzenich. *Role of cat primary auditory cortex for sound-localization behavior*. J. Neurophysiol. **52**, 819–847 (1984). See page: 5

[47] G. L. Kavanagh and J. B. Kelly. *Contribution of auditory cortex to sound localization by the ferret (Mustela putorius)*. J. Neurophysiol. **57**, 1746–1766 (1987). See page: 5

[48] S. Malhotra and S. G. Lomber. *Sound Localization During Homotopic and Heterotopic Bilateral Cooling Deactivation of Primary and Nonprimary Auditory Cortical Areas in the Cat*. J. Neurophysiol. **97**, 26–43 (2007). See page: 5

[49] F. R. Nodal, V. M. Bajo, and A. J. King. *Plasticity of spatial hearing: behavioural effects of cortical inactivation*. J. Physiol. **590**, 3965–3986 (2012). See page: 5

[50] J. A. Winer. *Decoding the auditory corticofugal systems*. Hear. Res. **207**, 1–9 (2005). See page: 5

[51] Z. Xiao and N. Suga. *Modulation of cochlear hair cells by the auditory cortex in the mustached bat*. Nat. Neurosci. **5**, 57–63 (2002). See page: 5

[52] Y. A. Ayala et al. *Differences in the strength of cortical and brainstem inputs to SSA and non-SSA neurons in the inferior colliculus*. Sci. Rep. **5**, 10383 (2015). See page: 5

[53] B. L. Robinson, N. S. Harper, and D. McAlpine. *Meta-adaptation in the auditory midbrain under cortical influence*. Nat. Commun. **7**, 13442 (2016). See page: 5

[54] K. T. Nakamoto, S. J. Jones, and A. R. Palmer. *Descending Projections From Auditory Cortex Modulate Sensitivity in the Midbrain to Cues for Spatial Position*. J. Neurophysiol. **99**, 2347–2356 (2008). See page: 5

[55] V. M. Bajo, F. R. Nodal, D. R. Moore, and A. J. King. *The descending corticocollicular pathway mediates learning-induced auditory plasticity*. Nat. Neurosci. **13**, 253–260 (2010). See page: 5

[56] S. F. Maison and M. C. Liberman. *Predicting Vulnerability to Acoustic Injury with a Noninvasive Assay of Olivocochlear Reflex Strength*. J. Neurosci. **20**, 4701–4707 (2000). See page: 5

[57] C. D. Dragicevic et al. *The Olivocochlear Reflex Strength and Cochlear Sensitivity are Independently Modulated by Auditory Cortex Microstimulation*. JARO **16**, 223–240 (2015). See page: 5

[58] K. N. Darrow, S. F. Maison, and M. C. Liberman. *Cochlear efferent feedback balances interaural sensitivity*. Nat. Neurosci. **9**, 1474–1476 (2006). See page: 5

[59] E. Budinger and H. Scheich. *Anatomical connections suitable for the direct processing of neuronal information of different modalities via the rodent primary auditory cortex*. Hear. Res. **258**, 16–27 (2009). See pages: 6, 77

[60] T. D. Mrsic-Flogel, A. J. King, and J. W. H. Schnupp. *Encoding of Virtual Acoustic Space Stimuli by Neurons in Ferret Primary Auditory Cortex*. J. Neurophysiol. **93**, 3489–3503 (2005). See pages: 6, 75

[61] M. R. DeWeese, M. Wehr, and A. M. Zador. *Binary Spiking in Auditory Cortex*. J. Neurosci. **23**, 7940–7949 (2003). See page: 6

[62] T. Hromádka, M. R. DeWeese, and A. M. Zador. *Sparse Representation of Sounds in the Unanesthetized Auditory Cortex*. PLoS Biol. **6**, e16 (2008). See pages: 6, 68

[63] H. Asari, B. A. Pearlmutter, and A. M. Zador. *Sparse Representations for the Cocktail Party Problem*. J. Neurosci. **26**, 7477–7490 (2006). See page: 6

[64] N. Ulanovsky, L. Las, and I. Nelken. *Processing of low-probability sounds by cortical neurons*. Nat. Neurosci. **6**, 391–398 (2003). See pages: 6, 7

[65] H. Asari and A. M. Zador. *Long-Lasting Context Dependence Constrains Neural Encoding Models in Rodent Auditory Cortex*. J. Neurophysiol. **102**, 2638–2656 (2009). See page: 6

[66] G. C. Stecker, I. A. Harrington, and J. C. Middlebrooks. *Location Coding by Opponent Neural Populations in the Auditory Cortex*. PLoS Biol. **3**, e78 (2005). See page: 6

[67] J. D. Yao, P. Bremen, and J. C. Middlebrooks. *Rat primary auditory cortex is tuned exclusively to the contralateral hemifield*. J. Neurophysiol. **110**, 2140–2151 (2013). See page: 6

[68] J. W. H. Schnupp, T. D. Mrsic-Flogel, and A. J. King. *Linear processing of spatial cues in primary auditory cortex*. Nature **414**, 200–204 (2001). See page: 6

[69] J. K. Bizley, K. M. M. Walker, B. W. Silverman, A. J. King, and J. W. H. Schnupp. *Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex*. J. Neurosci. **29**, 2064–2075 (2009). See pages: 6, 79

[70] K. M. M. Walker, J. K. Bizley, A. J. King, and J. W. H. Schnupp. *Multiplexed and Robust Representations of Sound Features in Auditory Cortex*. J. Neurosci. **31**, 14565–14576 (2011). See page: 6

[71] J. K. Bizley and A. J. King. *Visual–auditory spatial processing in auditory cortical neurons*. Brain Res. **1242**, 24–36 (2008). See page: 7

[72] J. Middlebrooks, A. Clock, L. Xu, and D. Green. *A panoramic code for sound location by cortical neurons*. Science **264**, 842–844 (1994). See pages: 7, 10

[73] J. F. Brugge, R. A. Reale, and J. E. Hind. *The Structure of Spatial Receptive Fields of Neurons in Primary Auditory Cortex of the Cat*. J. Neurosci. **16**, 4420–4437 (1996). See page: 10

[74] G. C. Stecker and J. C. Middlebrooks. *Distributed coding of sound locations in the auditory cortex*. Biol. Cybern. **89**, 341–349 (2003). See page: 80

[75] L. A. C. Belliveau, D. R. Lyamzin, and N. A. Lesica. *The Neural Representation of Interaural Time Differences in Gerbils Is Transformed from Midbrain to Cortex*. J. Neurosci. **34**, 16796–16808 (2014). See pages: 7, 10

[76] M. H. Dueck et al. *Propofol attenuates responses of the auditory cortex to acoustic stimulation in a dose-dependent manner: A FMRI study*. Acta Anaesthesiol Scand **49**, 784–791 (2005). See page: 7

[77] B. Pfingst, T. O'Connor, and J. Miller. *Response plasticity of neurons in auditory cortex of the rhesus monkey*. Exp. Brain. Res. **29-29** (1977). See pages: 7, 52, 75

[78] J. F. Brugge and M. M. Merzenich. *Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation*. J. Neurophysiol. **36**, 1138–1158 (1973). See pages: 7, 8

[79] B. J. Mickey and J. C. Middlebrooks. *Representation of Auditory Space by Cortical Neurons in Awake Cats*. J. Neurosci. **23**, 8649–8663 (2003). See pages: 7, 10, 34, 52, 54, 73

[80] M. Banks et al. *Altered stimulus representation in rat auditory cortex is not causal for loss of consciousness under general anaesthesia*. Br. J. Anaesth. **121**, 605–615 (2018). See page: 7

[81] D. H. Hubel, C. O. Henson, A. Rupert, and R. Galambos. *"Attention" Units in the Auditory Cortex*. Science **129**, 1279–1280 (1959). See page: 7

[82] L. Cohen and A. Mizrahi. *Plasticity during Motherhood: Changes in Excitatory and Inhibitory Layer 2/3 Neurons in Auditory Cortex*. J. Neurosci. **35**, 1806–1815 (2015). See page: 7

[83] J. M. Miller et al. *Single Cell Activity in the Auditory Cortex of Rhesus Monkeys: Behavioral Dependency*. Science **177**, 449–451 (1972). See page: 7

[84] M. Chait, C. C. Ruff, T. D. Griffiths, and D. McAlpine. *Cortical responses to changes in acoustic regularity are differentially modulated by attentional load*. NeuroImage **59**, 1932–1941 (2012). See page: 7

[85] C. Rodgers and M. DeWeese. *Neural Correlates of Task Switching in Prefrontal Cortex and Primary Auditory Cortex in a Novel Stimulus Selection Task for Rodents*. Neuron **82**, 1157–1170 (2014). See page: 7

[86] J. Fritz, S. Shamma, M. Elhilali, and D. Klein. *Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex*. Nat. Neurosci. **6**, 1216–1223 (2003). See page: 7

[87] S. V. David, J. B. Fritz, and S. A. Shamma. *Task reward structure shapes rapid receptive field plasticity in auditory cortex*. Proc. Natl. Acad. Sci. U.S.A. **109**, 2144–2149 (2012). See page: 7

[88] M. Aizenberg et al. *Projection from the Amygdala to the Thalamic Reticular Nucleus Amplifies Cortical Sound Responses*. Cell Reports **28**, 605–615.e4 (2019). See page: 7

[89] D. E. Winkowski et al. *Orbitofrontal Cortex Neurons Respond to Sound and Activate Primary Auditory Cortex Neurons*. Cereb. Cortex **28**, 868–879 (2018). See page: 7

[90] D. E. Winkowski, S. Bandyopadhyay, S. A. Shamma, and P. O. Kanold. *Frontal Cortex Activation Causes Rapid Plasticity of Auditory Cortical Processing*. J. Neurosci. **33**, 18134–18148 (2013). See page: 7

[91] H. Kato, S. Gillet, and J. Isaacson. *Flexible Sensory Representations in Auditory Cortex Driven by Behavioral Relevance*. Neuron **88**, 1027–1039 (2015). See page: 7

[92] B. J. Malone, B. H. Scott, and M. N. Semple. *Context-Dependent Adaptive Coding of Interaural Phase Disparity in the Auditory Cortex of Awake Macaques*. J. Neurosci. **22**, 4625–4638 (2002). See page: 8

[93] C. Kayser, N. K. Logothetis, and S. Panzeri. *Visual Enhancement of the Information Representation in Auditory Cortex*. Curr. Biol. **20**, 19–24 (2010). See page: 8

[94] E. Vaadia, Y. Gottlieb, and M. Abeles. *Single-unit activity related to sensorimotor association in auditory cortex of a monkey.* J. Neurophysiol. **48**, 1201–1213 (1982). See page: 8

[95] M. Brosch. *Nonauditory Events of a Behavioral Procedure Activate Auditory Cortex of Highly Trained Monkeys*. J. Neurosci. **25**, 6797–6806 (2005). See page: 8

[96] C. E. Schroeder, D. A. Wilson, T. Radman, H. Scharfman, and P. Lakatos. *Dynamics of Active Sensing and perceptual selection*. Current Opinion in Neurobiology **20**, 172–176 (2010). See pages: 8, 10

[97] M. Zhou et al. *Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex.* Nat. Neurosci. **17**, 841–850 (2014). See page: 8

[98] D. M. Schneider, A. Nelson, and R. Mooney. *A synaptic and circuit basis for corollary discharge in the auditory cortex.* Nature **513**, 189–194 (2014). See pages: 8, 58, 65

[99] M. H. Martikainen. *Suppressed Responses to Self-triggered Sounds in the Human Auditory Cortex*. Cereb. Cortex **15**, 299–302 (2004). See page: 8

[100] Y. Yang, J. Lee, and G. Kim. *Integration of locomotion and auditory signals in the mouse inferior colliculus*. eLife **9**, e52228 (2020). See page: 8

[101] R. Williamson, K. Hancock, B. Shinn-Cunningham, and D. Polley. *Locomotion and Task Demands Differentially Modulate Thalamic Audiovisual Processing during Active Search*. Curr. Biol. **25**, 1885–1891 (2015). See page: 8

[102] K. K. Clayton et al. *Auditory Corticothalamic Neurons Are Recruited by Motor Preparatory Inputs*. Curr. Biol. p. S0960982220315311 (2020). See pages: 8, 38

[103] D. Benson, R. Hienz, and M. Goldstein. *Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: Spatial tuning and behavioral dependency*. Brain Res. **219**, 249–267 (1981). See pages: 8, 10

[104] B. H. Scott, B. J. Malone, and M. N. Semple. *Effect of Behavioral Context on Representation of a Spatial Cue in Core Auditory Cortex of Awake Macaques*. J. Neurosci. **27**, 6489–6499 (2007). See pages: 8, 68

[105] C.-C. Lee and J. C. Middlebrooks. *Auditory cortex spatial sensitivity sharpens during task performance*. Nat. Neurosci. **14**, 108–114 (2011). See pages: 9, 10, 13, 79, 82

[106] Moss. *Probing the natural scene by echolocation in bats*. Front. Behav. Neurosci. (2010). See page: 9

[107] D. Russo, G. Jones, and R. Arlettaz. *Echolocation and passive listening by foraging mouse-eared bats Myotis myotis and M. blythii*. J. Exp. Biol. **210**, 166–176 (2007). See page: 9

[108] K. A. Razak, Z. M. Fuzessery, and T. D. Lohuis. *Single Cortical Neurons Serve Both Echolocation and Passive Sound Localization*. J. Neurophysiol. **81**, 1438–1442 (1999). See page: 9

[109] U. Firzlaff, M. Schuchmann, J. E. Grunwald, G. Schuller, and L. Wiegrebe. *Object-Oriented Echo Perception and Cortical Representation in Echolocating Bats*. PLoS Biol. **5**, e100 (2007). See page: 9

[110] J. J. Eggermont. *Azimuth Coding in Primary Auditory Cortex of the Cat. II. Relative Latency and Interspike Interval Representation*. J. Neurophysiol. **80**, 2151–2161 (1998). See page: 10

[111] T. M. Woods, S. E. Lopez, J. H. Long, J. E. Rahman, and G. H. Recanzone. *Effects of Stimulus Azimuth and Intensity on the Single-Neuron Activity in the Auditory Cortex of the Alert Macaque Monkey*. J. Neurophysiol. **96**, 3323–3337 (2006). See pages: 10, 75

[112] E. D. Remington and X. Wang. *Neural Representations of the Full Spatial Field in Auditory Cortex of Awake Marmoset (Callithrix jacchus)*. Cereb. Cortex **29**, 1199–1216 (2019). See pages: 10, 75

[113] S. M. Town, W. O. Brimijoin, and J. K. Bizley. *Egocentric and allocentric representations in auditory cortex*. PLoS Biol. **15**, e2001878 (2017). See pages: 10, 76

[114] J. P. Whitton, K. E. Hancock, and D. B. Polley. *Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise*. Proc. Natl. Acad. Sci. U.S.A. **111**, E2606–E2615 (2014). See page: 12

[115] D. B. Polley, M. A. Heiser, D. T. Blake, C. E. Schreiner, and M. M. Merzenich. *Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex*. Proc. Natl. Acad. Sci. U.S.A. **101**, 16351–16356 (2004). See page: 12

[116] L. Belliveau. *The role of spatial cues for processing speech in noise*. PhD thesis, University College London and Ludwig-Maximilians-Universität München (2017). Available at: https://discovery.ucl.ac.uk/id/eprint/10027947/. See page: 12

[117] D. N. Ferreiro et al. *Sensory Island Task (SIT): A New Behavioral Paradigm to Study Sensory Perception and Neural Processing in Freely Moving Animals*. Front. Behav. Neurosci. **14**, 576154 (2020). See pages: 13, 14, 16, 32, 41, 42, 43, 44, 79, 83

[118] B. N. Buran, G. von Trapp, and D. H. Sanes. *Behaviorally Gated Reduction of Spontaneous Discharge Can Improve Detection Thresholds in Auditory Cortex*. J. Neurosci. **34**, 4076–4081 (2014). See pages: 13, 68

[119] I. Carcea, M. N. Insanally, and R. C. Froemke. *Dynamics of auditory cortical activity during behavioural engagement and auditory perception*. Nat. Commun. **8**, 14412 (2017). See pages: 13, 68

[120] S. C. Kadia and X. Wang. *Spectral Integration in A1 of Awake Primates: Neurons With Single- and Multipeaked Tuning Characteristics*. J. Neurophysiol. **89**, 1603–1622 (2003). See page: 14

[121] A. Ryan. *Hearing sensitivity of the mongolian gerbil, Meriones unguiculatis*. J. Acoust. Soc. Am. **59**, 1222 (1976). See page: 14

[122] R. L. Freyman and P. M. Zurek. *Strength of onset and ongoing cues in judgments of lateral position*. J. Acoust. Soc. Am. **142**, 206–214 (2017). See page: 14

[123] D. P. Phillips. *Effect of tone-pulse rise time on rate-level functions of cat auditory cortex neurons: excitatory and inhibitory processes shaping responses to tone onset*. J. Neurophysiol. **59**, 1524–1539 (1988). See page: 14

[124] E. Waiblinger. *The Laboratory Gerbil*. In R. Hubrecht and J. Kirkwood, editors, *The UFAW Handbook on the Care and Management of Laboratory and Other Research Animals*, pp. 327–347. Wiley-Blackwell, Oxford, UK (2010). See page: 15

[125] G. H. Jacobs and J. Neitz. *Cone monochromacy and a reversed Purkinje shift in the gerbil*. Experientia **45**, 317–319 (1989). See page: 15

[126] B. A. Edmonds and K. Krumbholz. *Are Interaural Time and Level Differences Represented by Independent or Integrated Codes in the Human Auditory Cortex?* JARO **15**, 103–114 (2014). See page: 15

[127] K. C. Wood, S. M. Town, and J. K. Bizley. *Neurons in primary auditory cortex represent sound source location in a cue-invariant manner*. Nat. Commun. **10**, 3019 (2019). See pages: 15, 73, 79

[128] A. S. Bregman. *Auditory scene analysis: the perceptual organization of sound*. A Bradford book. MIT Press, Cambridge, Mass., 2. paperback ed., repr edition (2006). OCLC: 550546350. See page: 15

[129] S. Tolnai, R. Beutelmann, and G. M. Klump. *Exploring binaural hearing in gerbils (Meriones unguiculatus) using virtual headphones*. PLoS ONE **12**, e0175142 (2017). See page: 15

[130] A. Lingner, L. Wiegrebe, and B. Grothe. *Sound Localization in Noise by Gerbils and Humans*. JARO **13**, 237–248 (2012). See page: 15

[131] N. A. Lesica, A. Lingner, and B. Grothe. *Population Coding of Interaural Time Differences in Gerbils and Barn Owls*. J. Neurosci. **30**, 11696–11702 (2010). See page: 15

[132] G. Lopes et al. *Bonsai: an event-based framework for processing and controlling data streams*. Front. Neuroinform. **9** (2015). See page: 19

[133] J. Rolston, R. Gross, and S. Potter. *Common median referencing for improved action potential detection with multielectrode arrays*. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1604–1607, Minneapolis, MN (2009). IEEE. See page: 25

[134] M. Pachitariu, N. Steinmetz, S. Kadir, M. Carandini, and H. Kenneth D. *Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels*. preprint (2016). See page: 25

[135] C. Rossant et al. *Spike sorting for large, dense electrode arrays*. Nat. Neurosci. **19**, 634–641 (2016). See page: 25

[136] N. Schmitzer-Torbert, J. Jackson, D. Henze, K. Harris, and A. Redish. *Quantitative measures of cluster quality for use in extracellular recordings*. Neuroscience **131**, 1–11 (2005). See page: 25

[137] D. N. Hill, S. B. Mehta, and D. Kleinfeld. *Quality Metrics to Accompany Spike Sorting of Extracellular Signals*. J. Neurosci. **31**, 8699–8705 (2011). See page: 25

[138] E. Budinger, P. Heil, and H. Scheich. *Functional organization of auditory cortex in the Mongolian gerbil (Meriones unguiculatus). III. Anatomical subdivisions and corticocortical connections: Gerbil auditory cortex anatomy and connections*. Eur. J. Neurosci. **12**, 2425–2451 (2000). See page: 26

[139] H. Thomas, J. Tillein, P. Heil, and H. Scheich. *Functional Organization of Auditory Cortex in the Mongolian Gerbil (Meriones unguiculatus). I. Electrophysiological Mapping of Frequency Representation and Distinction of Fields*. Eur. J. Neurosci. **5**, 882–897 (1993). See pages: 26, 52, 56, 75

[140] S. Radtke-Schuller et al. *Brain atlas of the Mongolian gerbil (Meriones unguiculatus) in CT/MRI-aided stereotaxic coordinates*. Brain Struct. Funct. **221**, 1–272 (2016). See page: 26

[141] G. Bradski. *The OpenCV Library*. Dr. Dobb's Journal of Software Tools (2000). See page: 27

[142] Z. Zivkovic. *Improved adaptive Gaussian mixture model for background subtraction*. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, pp. 28–31 Vol.2, Cambridge, UK (2004). IEEE. See page: 27

[143] Z. Zivkovic and F. van der Heijden. *Efficient adaptive density estimation per image pixel for the task of background subtraction*. Pattern Recognition Letters **27**, 773–780 (2006). See page: 27

[144] S. Suzuki and K. be. *Topological structural analysis of digitized binary images by border following*. Comput. Vis. Graph. Image Process. **30**, 32–46 (1985). See page: 27

[145] T. J. Imig, N. G. Bibikov, P. Poirier, and F. K. Samson. *Directionality Derived From Pinna-Cue Spectral Notches in Cat Dorsal Cochlear Nucleus*. J. Neurophysiol. **83**, 907–925 (2000). See page: 30

[146] K. Maki and S. Furukawa. *Acoustical cues for sound localization by the Mongolian gerbil, Meriones unguiculatus*. J. Acoust. Soc. Am. **118**, 872–886 (2005). See pages: 30, 54

[147] S. Furukawa and J. C. Middlebrooks. *Cortical Representation of Auditory Space: Information-Bearing Features of Spike Patterns*. J. Neurophysiol. **87**, 1749–1762 (2002). See page: 34

[148] J. D. Scargle, J. P. Norris, B. Jackson, and J. Chiang. *Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations*. ApJ **764**, 167 (2013). See page: 34

[149] M. Mazurek, M. Kager, and S. D. Van Hooser. *Robust quantification of orientation selectivity and direction selectivity*. Front. Neural Circuits **8** (2014). See page: 37

[150] D. N. Ferreiro et al. *Spatial clustering of orientation preference in primary visual cortex of the large rodent agouti*. iScience **24**, 101882 (2021). See page: 37

[151] C. A. Atencio and C. E. Schreiner. *Spectrotemporal Processing Differences between Auditory Cortical Fast-Spiking and Regular-Spiking Neurons*. J. Neurosci. **28**, 3897–3910 (2008). See pages: 38, 53, 75

[152] A. Hasenstaub et al. *Inhibitory Postsynaptic Potentials Carry Synchronized Frequency Information in Active Cortical Networks*. Neuron **47**, 423–435 (2005). See page: 53

[153] H. Markram et al. *Reconstruction and Simulation of Neocortical Microcircuitry*. Cell **163**, 456–492 (2015). See pages: 38, 75

[154] F. Pedregosa et al. *Scikit-learn: Machine Learning in Python*. J. Mach. Learn. Res. **12**, 2825–2830 (2011). See page: 40

[155] C.-C. Lee and J. C. Middlebrooks. *Specialization for Sound Localization in Fields A1, DZ, and PAF of Cat Auditory Cortex*. JARO **14**, 61–82 (2013). See pages: 52, 54, 75

[156] I. Volkov and A. Galazjuk. *Formation of spike response to sound tones in cat auditory cortex neurons: Interaction of excitatory and inhibitory effects*. Neuroscience **43**, 307–321 (1991). See pages: 52, 75

[157] A. K. Moore and M. Wehr. *Parvalbumin-Expressing Inhibitory Interneurons in Auditory Cortex Are Well-Tuned for Frequency*. J. Neurosci. **33**, 13713–13723 (2013). See pages: 53, 75

[158] D. M. Schneider, J. Sundararajan, and R. Mooney. *A cortical filter that learns to suppress the acoustic consequences of movement*. Nature **561**, 391–395 (2018). See pages: 58, 65

[159] B. Scholl, X. Gao, and M. Wehr. *Nonoverlapping Sets of Synapses Drive On Responses and Off Responses in Auditory Cortex*. Neuron **65**, 412–421 (2010). See pages: 60, 78

[160] L. Qin, S. Chimoto, M. Sakai, J. Wang, and Y. Sato. *Comparison Between Offset and Onset Responses of Primary Auditory Cortex Neurons in Awake Cats*. J. Neurophysiol. **97**, 3421–3431 (2007). See page: 61

[161] D. L. Ramamurthy and G. H. Recanzone. *Spectral and spatial tuning of onset and offset response functions in auditory cortical fields A1 and CL of rhesus macaques*. J. Neurophysiol. **117**, 966–986 (2017). See pages: 61, 78

[162] D. E. H. Hartley, J. C. Dahmen, A. J. King, and J. W. H. Schnupp. *Binaural sensitivity changes between cortical on and off responses*. J. Neurophysiol. **106**, 30–43 (2011). See pages: 61, 78

[163] W.-p. Ma et al. *Visual Representations by Cortical Somatostatin Inhibitory Neurons–Selective But with Weak and Delayed Responses*. J. Neurosci. **30**, 14371–14379 (2010). See pages: 67, 75

[164] R. M. Bruno and D. J. Simons. *Feedforward Mechanisms of Excitatory and Inhibitory Cortical Receptive Fields*. J. Neurosci. **22**, 10966–10975 (2002). See pages: 67, 75

[165] G. H. Otazu, L.-H. Tai, Y. Yang, and A. M. Zador. *Engaging in an auditory task suppresses responses in auditory cortex*. Nat. Neurosci. **12**, 646–654 (2009). See page: 68

[166] A. Ryan, J. Miller, B. Pfingst, and G. Martin. *Effects of reaction time performance on single-unit activity in the central auditory pathway of the rhesus macaque*. J. Neurosci. **4**, 298–308 (1984). See page: 68

[167] M. Niwa, J. S. Johnson, K. N. O'Connor, and M. L. Sutter. *Active Engagement Improves Primary Auditory Cortical Neurons' Ability to Discriminate Temporal Modulation*. J. Neurosci. **32**, 9323–9334 (2012). See page: 68

[168] C. Dong, L. Qin, Z. Zhao, R. Zhong, and Y. Sato. *Behavioral Modulation of Neural Encoding of Click-Trains in the Primary and Nonprimary Auditory Cortex of Cats*. J. Neurosci. **33**, 13126–13137 (2013).

[169] S. Bagur et al. *Go/No-Go task engagement enhances population representation of target stimuli in primary auditory cortex*. Nat. Commun. **9**, 2529 (2018). See page: 68

[170] J. Middlebrooks and E. Knudsen. *A neural code for auditory space in the cat's superior colliculus*. J. Neurosci. **4**, 2621–2634 (1984). See page: 73

[171] R. Beaton and J. M. Miller. *Single cell activity in the auditory cortex of the unanesthetized, behaving monkey: Correlation with stimulus controlled behavior*. Brain Res. **100**, 543–562 (1975). See pages: 75, 79

[172] M. Pecka and J. Encke. *Coding of Spatial Information*. In *The Senses: A Comprehensive Reference*, pp. 713–731. Elsevier (2020). See page: 75

[173] R. Rajan, L. M. Aitkin, and D. R. Irvine. *Azimuthal sensitivity of neurons in primary auditory cortex of cats. II. Organization along frequency-band strips*. J. Neurophysiol. **64**, 888–902 (1990). See page: 75

[174] G. Bock and W. Webster. *Coding of spatial location by single units in the inferior colliculus of the alert cat*. Exp. Brain Res. **21** (1974). See page: 76

[175] H. Wallach. *Über die Wahrnehmung der Schallrichtung*. Psychol. Forsch. **22**, 238–266 (1938). See page: 77

[176] E. Wigderson, I. Nelken, and Y. Yarom. *Early multisensory integration of self and source motion in the auditory system*. Proc. Natl. Acad. Sci. U.S.A. **113**, 8308–8313 (2016). See page: 77

[177] S. D. Vann, J. P. Aggleton, and E. A. Maguire. *What does the retrosplenial cortex do?* Nat. Rev. Neurosci. **10**, 792–802 (2009). See page: 77

[178] M. Vélez-Fort et al. *A Circuit for Integration of Head- and Visual-Motion Signals in Layer 6 of Mouse Primary Visual Cortex*. Neuron **98**, 179–191.e6 (2018). See page: 77

[179] H. S. Bashinski and V. R. Bacharach. *Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations*. Percept. Psychophys. **28**, 241–248 (1980). See page: 77

[180] S. J. Luck, L. Chelazzi, S. A. Hillyard, and R. Desimone. *Neural Mechanisms of Spatial Selective Attention in Areas V1, V2, and V4 of Macaque Visual Cortex*. J. Neurophysiol. **77**, 24–42 (1997). See page: 77

[181] A. Speed, J. Del Rosario, N. Mikail, and B. Haider. *Spatial attention enhances network, cellular and subthreshold responses in mouse visual cortex*. Nat. Commun. **11**, 505 (2020). See page: 77

[182] D. Benson and R. Hienz. *Single-unit activity in the auditory cortex of monkeys selectively attending left vs. right ear stimuli*. Brain Res. **159**, 307–320 (1978). See page: 77

[183] H. L. Dean. *Allocentric Spatial Referencing of Neuronal Activity in Macaque Posterior Cingulate Cortex*. J. Neurosci. **26**, 1117–1127 (2006). See page: 77

[184] E. I. Knudsen. *Neural Circuits That Mediate Selective Attention: A Comparative Perspective*. Trends Neurosci. **41**, 789–805 (2018). See page: 77

[185] G. S. Russo and C. J. Bruce. *Frontal eye field activity preceding aurally guided saccades*. J. Neurophysiol. **71**, 1250–1253 (1994). See page: 77

[186] H. Kirchner, E. J. Barbeau, S. J. Thorpe, J. Regis, and C. Liegeois-Chauvel. *Ultra-Rapid Sensory Responses in the Human Frontal Eye Field Region*. J. Neurosci. **29**, 7599–7606 (2009). See page: 77

[187] A. Garg, D. Schwartz, and A. A. Stevens. *Orienting auditory spatial attention engages frontal eye fields and medial occipital cortex in congenitally blind humans*. Neuropsychologia **45**, 2307–2321 (2007). See page: 77

[188] S. Zhang et al. *Long-range and local circuits for top-down modulation of visual cortex processing*. Science **345**, 660–665 (2014). See page: 77

[189] I. Nelken, A. Fishbach, L. Las, N. Ulanovsky, and D. Farkas. *Primary auditory cortex of cats: feature detection or something else?* Biological Cybernetics **89**, 397–406 (2003). See page: 78

[190] O. Bar-Yosef, Y. Rotman, and I. Nelken. *Responses of Neurons in Cat Primary Auditory Cortex to Bird Chirps: Effects of Temporal and Spectral Context*. J. Neurosci. **22**, 8619–8632 (2002). See page: 78

[191] J. E. Cooke, J. J. Lee, E. L. Bartlett, X. Wang, and D. Bendor. *Post-stimulatory activity in primate auditory cortex evoked by sensory stimulation during passive listening*. Sci. Rep. **10**, 13885 (2020). See page: 78

[192] Y. Gottlieb, E. Vaadia, and M. Abeles. *Single unit activity in the auditory cortex of a monkey performing a short term memory task*. Exp. Brain Res. **74** (1989). See page: 79

[193] J. Bizley, K. Walker, F. Nodal, A. King, and J. Schnupp. *Auditory Cortex Represents Both Pitch Judgments and the Corresponding Acoustic Cues*. Curr. Biol. **23**, 620–625 (2013). See page: 79

[194] M. M. Zempeltzi et al. *Task rule and choice are reflected by layer-specific processing in rodent auditory cortical microcircuits*. Commun. Biol. **3**, 345 (2020).

[195] L. Guo, J. T. Weems, W. I. Walker, A. Levichev, and S. Jaramillo. *Choice-Selective Neurons in the Auditory Cortex and in Its Striatal Target Encode Reward Expectation*. J. Neurosci. **39**, 3687–3697 (2019). See page: 82

[196] J. Tsunada, A. S. K. Liu, J. I. Gold, and Y. E. Cohen. *Causal contribution of primate auditory cortex to auditory perceptual decision-making*. Nat. Neurosci. **19**, 135–142 (2016).

[197] E. Selezneva, H. Scheich, and M. Brosch. *Dual Time Scales for Categorical Decision Making in Auditory Cortex*. Curr. Biol. **16**, 2428–2433 (2006).

[198] N. A. Francis et al. *Small Networks Encode Decision-Making in Primary Auditory Cortex*. Neuron **97**, 885–897.e6 (2018). See page: 79

[199] R. K. Maddox, C. P. Billimoria, B. P. Perrone, B. G. Shinn-Cunningham, and K. Sen. *Competing Sound Sources Reveal Spatial Effects in Cortical Processing*. PLoS Biol. **10**, e1001319 (2012). See page: 79

[200] J. C. Middlebrooks and P. Bremen. *Spatial Stream Segregation by Auditory Cortical Neurons*. J. Neurosci. **33**, 10986–11001 (2013). See page: 79

[201] J. C. Makous and J. C. Middlebrooks. *Two-dimensional sound localization by human listeners*. J. Acoust. Soc. Am. **87**, 2188–2200 (1990). See page: 80

[202] J. C. Middlebrooks, L. Xu, S. Furukawa, and E. A. Macpherson. *Book Review: Cortical Neurons That Localize Sounds*. Neuroscientist **8**, 73–83 (2002). See page: 80

[203] S. A. Gelfand. *Hearing: an introduction to psychological and physiological acoustics*. CRC Press, Boca Raton, sixth edition edition (2018). See page: 80

[204] R. Pavão, E. S. Sussman, B. J. Fischer, and J. L. Peña. *Natural ITD statistics predict human auditory spatial perception*. eLife **9**, e51927 (2020). See page: 80

[205] A. W. Mills. *On the Minimum Audible Angle*. J. Acoust. Soc. Am. **30**, 237–246 (1958). See page: 80

[206] C. Freigang, K. Schmiedchen, I. Nitsche, and R. Rübsamen. *Free-field study on auditory localization and discrimination performance in older adults*. Exp. Brain. Res. **232**, 1157–1172 (2014). See page: 80

[207] R. S. Heffner and H. E. Heffner. *Sound localization acuity in the cat: Effect of azimuth, signal duration, and test procedure*. Hear. Res. **36**, 221–232 (1988). See page: 80

[208] B. J. May and A. Y. Huang. *Sound orientation behavior in cats. I. Localization of broadband noise*. J. Acoust. Soc. Am. **100**, 1059–1069 (1996). See page: 80

[209] S. Furukawa, L. Xu, and J. C. Middlebrooks. *Coding of Sound-Source Location by Ensembles of Cortical Neurons*. J. Neurosci. **20**, 1216–1228 (2000). See page: 80

[210] S. Atiani, M. Elhilali, S. V. David, J. B. Fritz, and S. A. Shamma. *Task Difficulty and Performance Induce Diverse Adaptive Patterns in Gain and Shape of Primary Auditory Cortical Receptive Fields*. Neuron **61**, 467–480 (2009). See page: 82

[211] L. Guo, W. I. Walker, N. D. Ponvert, P. L. Penix, and S. Jaramillo. *Stable representation of sounds in the posterior striatum during flexible auditory decisions*. Nat. Commun. **9**, 1534 (2018). See page: 82

[212] B. Bathellier, L. Ushakova, and S. Rumpel. *Discrete Neocortical Dynamics Predict Behavioral Categorization of Sounds*. Neuron **76**, 435–449 (2012). See pages: 82, 83

[213] D. Aharoni and T. M. Hoogland. *Circuit Investigations With Open-Source Miniaturized Microscopes: Past, Present and Future*. Front. Cell. Neurosci. **13**, 141 (2019). See page: 83

Acknowledgements

Firstly, I would like to thank my supervisor, Michael Pecka. Thank you Misku for introducing me to an early version of SIT (way before it was called like this), a paradigm that immediately captured my attention; for welcoming me into the lab; for all the time spent discussing crazy ideas about how to train gerbils when I did not have a HINT or how to better analyze data; for always giving me hope that in the end everything would fall into place; for your trust; for showing me the bigger picture.

I would also like to thank Benedikt Grothe for his continued support and guidance, valuable input and for giving me the opportunity to do my PhD in a stimulating research environment full of interesting people and scientific questions.

I would like to thank the other members of my examination board, Nick Lesica and Conny Kopp-Scheinpflug, for the insightful and pleasant discussion as well as for their suggestions.

I am also grateful to the other members of my TAC committee, Mark Hübener and Mike Myoga, for the time they took to discuss my project, steering it into the right direction.

Thanks to everyone from my graduate schools, the International Max Planck Research School for Molecular Life Sciences: From Biological Structures to Neural Circuits and the Graduate School of Systemic Neurosciences, who made my academic and bureaucratic life in Munich much easier. Thanks also to both schools for all the workshops and traveling opportunities, which broadened my scientific and cultural horizons.

Thank you Dardo for your scientific and musical suggestions throughout the end of my PhD. Thank you Pepe for your valuable comments on the article and for going through my defense presentation in such detail together with Dardo. Thank you Lucile for all you taught me at the beginning of my PhD. Thanks to Dan Sanes and his whole lab, to Evgeny and to Elena for the initial help with the surgeries. Thanks to the histolab team, in particular to Hilde, Olga and Karin for their help with the histology.

I would like to thank everyone that collaborated with me: Misku, Benedikt, Dardo, Daniel, Andrey, Paula, Lucile and Anton.

I always felt very comfortable in the Biocenter in a big part because of all the kind and funny people with whom I shared an office. Thank you Dieter, Shreya, Steffi, Babsi, Dani and of course our three mascots: Pollux, Tolstoi and Bazi.

Thank you Eli, Ella, Ela, Sara, Alex, Delwen, Alisha, Dardo, Franzi, Mihai, and Nicola for always making my lunches an event to look forward to, where one could learn a bit of everything – from science to philosophy and politics – or just relax with good friends.

Thanks to my other fellow Löwenzähne, who always made pleasurable a casual encounter by the coffee machine: Madga, Marga, Gregory, Michael S., Leonie, Simon, Yannik, Felix, Veronika, Sara H., Matthias and Ravi. Thank you coffee and everyone who helped maintaining our fuel constant: Misku, Sven and Andrea.

Thank you Shreya and Magda and all the other helpers for the best hat ever, even in these pandemic times.

I am also very grateful to everyone who participated and gave suggestions in the Neurobio lab meeting. Thank you Laura for giving me the opportunity to join your journal club expanding my knowledge. I would also like to thank the animal caretakers and everyone that worked with me in organizing either the CNS social hour or the Talking Science, both events greatly enriched my PhD life.

I would like to thank all my friends, both the "newly" found ones from my PhD as well as those "from another planet", known as Portugal. In particular to Inês, Gonçalo, Henrique, Susana, Mariana and Zé, who made my trips to Portugal feel as if I had never left.

A big thanks to my family, in particular to my parents, Elsa and João, for their constant love and support.

Lastly, a big thanks to you, Michi, for driving through the snow to check in on an operated gerbil in the middle of the night; for knowing about the calyx of Held and the LSO; for all the discussions and suggestions; for your optimism and incentive; for always being there.

List of Publications

- D. N. Ferreiro*, **D. Amaro***, D. Schmidtke, A. Sobolev, P. Gundi, L. Belliveau, A. Sirota, B. Grothe and M. Pecka (2020) *Sensory Island Task (SIT): A New Behavioral Paradigm to Study Sensory Perception and Neural Processing in Freely Moving Animals.* Front. Behav. Neurosci. **14**, 576154. doi:10.3389/fnbeh.2020.576154

- **D. Amaro**, D. N. Ferreiro, B. Grothe and M. Pecka (2020) *Diverse spatial representations in primary auditory cortex during active localization simultaneously code source location and identity.* bioRχiv (preprint - submitted) doi:10.1101/2021.01.05.425444

*These authors have contributed equally to this work

Eidesstattliche Versicherung/Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation *Neuronal Representation of Sound Source Location in the Auditory Cortex during Active Navigation* selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation *Neuronal Representation of Sound Source Location in the Auditory Cortex during Active Navigation* is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München, den 22. Januar, 2021
Munich, 22nd January 2021

_____
Diana Amaro

Declaration of author contribution

The contributions of the authors Dardo N. Ferreiro (DNF), Diana Amaro (DA), Daniel Schmidtke (DS), Andrey Sobolev (ASo), Paula Gundi (PG), Lucile Belliveau (LB), Anton Sirota (ASi), Benedikt Grothe (BG) and Michael Pecka (MP) to the study *Sensory Island Task (SIT): A New Behavioral Paradigm to Study Sensory Perception and Neural Processing in Freely Moving Animals* included in this thesis are as follows:

MP, DNF, DA, and LB conceived SIT. MP, DNF, and DA designed the experiments. BG, DS, and ASi contributed to paradigm refinement. DA, DNF, DS, ASo, and PG performed the experiments. LB, DA, DS, and ASo contributed to programming SIT code. DA, DNF, DS, and ASo analyzed the results. DNF, MP, and DA designed and generated the figures and wrote the manuscript. All authors provided comments and approved the manuscript.

_____ _____

Diana Amaro PD Dr. Michael Pecka

_____

Dardo N. Ferreiro

All experiments, data analysis and subsequent results presented in this monograph were performed by Diana Amaro, unless specifically stated otherwise, under the supervision of PD Dr. Michael Pecka.

_____ _____

Diana Amaro PD Dr. Michael Pecka