Nonlinear encoding of sounds in the auditory cortex

Alexandre Kempf

CENTER FOR RESEARCH AND INTERDISCIPLINARITY

DOCTORAL THESIS

Nonlinear encoding of sounds in the auditory cortex

Author:
Alexandre KEMPF

Supervisor:
Dr. Brice BATHELLIER

Thesis Committee Members :
Dr. Tania Barkat Rinaldi (reporter)
Dr. Laurent Bourdieu (reporter)
Dr. Jean-Marc Edeline
Dr. Valerie Ego-Stengel
Dr. Daniel Pressnizter

A thesis submitted in fulfillment of the requirements for the degree of Doctor in Neuroscience

in the

Unit of Neuroscience, Information, and Complexity (UNIC)
Centre National de la Recherche Scientifique (CNRS)

October 17, 2018

Declaration of Authorship

I, Alexandre KEMPF, declare that this thesis titled, "Nonlinear encoding of sounds in the auditory cortex" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"I'm singing in the brain"

Modified from Gene Kelly

CENTER FOR RESEARCH AND INTERDISCIPLINARITY

Abstract

Unit of Neuroscience, Information, and Complexity (UNIC)
Centre National de la Recherche Scientifique (CNRS)

Doctor in Neuroscience

Nonlinear encoding of sounds in the auditory cortex

by Alexandre KEMPF

Perceptual objects are the elementary units used by the brain to construct an inner world representation of the environment from multiple physical sources, like light or sound waves. While the physical signals are first encoded by receptors in peripheral organs into neuroelectric signals, the emergence of perceptual object require extensive processing in the central nervous system which is not yet fully characterized. Interestingly, recent advances in deep learning shows that implementing series of nonlinear and linear operations is a very efficient way to create models that categorize visual and auditory perceptual objects similarly to humans. In contrast, most of the current knowledge about the auditory system concentrates on linear transformations.

In order to establish a clear example of the contribution of auditory system nonlinearities to perception, we studied the encoding of sounds with an increasing intensity (up ramps) and a decreasing intensity (down ramps) in the mouse auditory cortex. Two behavioral tasks showed evidence that these two sounds are perceived with unequal salience despite carrying the same physical energy and spectral content, a phenomenon incompatible with linear processing. Recording the activity of large cortical populations for up- and down-ramping sounds, we found that cortex encodes them into distinct sets of non-linear features, and that asymmetric feature selection explained the perceptual asymmetry. To complement these results, we also showed that, in reinforcement learning models, the amount of neural activity triggered by a stimulus (e.g. a sound) impacts learning speed and strategy. Interestingly very similar effects were observed in sound discrimination behavior and could be explain by the amount of cortical activity triggered by the discriminated sounds. This altogether establishes that auditory system nonlinearities have an impact on perception and behavior.

To more extensively identify the nonlinearities that influence sounds encoding, we then recorded the activity of around 60,000 neurons sampling the entire horizontal extent of auditory cortex. Beyond the fine scale tonotopic organization uncovered with this dataset, we identified and quantified 7 nonlinearities. We found interestingly that different nonlinearities can interact with each other in a non-trivial manner. The knowledge of these interactions carry good promises to refine auditory processing model.

Finally, we wondered if the nonlinear processes are also important for multisensory integration. We measured how visual inputs and sounds combine in the visual and auditory cortex using calcium imaging in mice. We found no modulation of supragranular auditory cortex in response to visual stimuli, as observed in previous others studies. We observed that auditory cortex inputs to visual cortex affect visual responses concomitant to a sound. Interestingly, we found that auditory cortex projections to visual cortex preferentially channel activity from neurons encoding a particular non-linear feature: the loud onset of sudden sounds. As a result, visual cortex activity for an image combined with a loud sound is higher than for the image alone or combine with a quiet sound. Moreover, this boosting effect is highly nonlinear. This result suggests that loud sound onsets are behaviorally relevant in the visual system, possibly to indicate the presence of a new perceptual objects in the visual field, which could represent potential threats.

As a conclusion, our results show that nonlinearities are ubiquitous in sound processing by the brain and also play a role in the integration of auditory information with visual information. In addition, it is not only crucial to account for these nonlinearities to understand how perceptual representations are formed but also to predict how these representations impact behavior.

Acknowledgements

Je tiens à remercier Daniel Shulz, directeur de l'UNIC, ainsi que Yves Frégnac, ex-directeur de l'UNIC, de m'avoir permis de travailler à l'UNIC, dans ce cadre de travail exceptionnel et, oserai-je le dire... unique. Ils ont permis mon épanouissement scientifique à travers discussions et débats, le tout dans une chaleureuse atmosphère familiale.

Je souhaite aussi remercier Tania Barkat Rinaldi de me faire le plaisir d'évaluer mon manuscrit et d'être parmi les membres du jury. Je souhaite remercier Valérie Ego-Stengel, Jean Marc Edeline, et Daniel Pressnitzer qui me font l'honneur d'être membres du jury et d'évaluer mon travail de thèse. Je souhaite remercier tout particulièrement Laurent Bourdieu, qui me conseille sur mon parcours professionnel depuis près de 6 ans, d'abord en tant que tuteur puis en tant que membre du comité de suivi de thèse, et qui aujourd'hui me fait l'honneur d'évaluer mon manuscrit et ma thèse en tant que membre du jury.

Ce travail n'aurait pas été possible sans le Centre de Recherche Interdisciplinaire et la Fondation Bettencourt qui m'ont financé pendant trois ans. Merci d'avoir cru en mon projet et d'avoir fait en sorte que je puisse me concentrer pleinement sur ce travail de recherche. Je remercie aussi Srdjan Ostojic pour ses conseils avisés lors des réunions de suivi de thèse.

Je tiens à remercier toutes les personnes de l'UNIC pour votre bonne humeur au quotidien et votre expertise, j'ai énormément appris d'un point de vue scientifique et humain pendant ces 3 ans grâce à vous. Pour ne lister que quelques noms, merci à : Edu, Jennifer, Aamir, Dorian, Yannick, Pauline, Margot, Zuzanna, Anton, Jacques, Joanna, Lucie... Un merci tout particulier à Aline et Justine pour avoir supporté sans broncher mon armée de stagiaires sans papiers. Un grand merci à Patrick pour son efficacité, ses conseils, et son travail sur le microscope. Merci à Guillaume pour sa généreuse contribution dans ce travail et pour m'avoir appris à découper un cerveau dans les règles de l'art. Je tiens à remercier mille fois Evan pour sa disponibilité, sa gentillesse, ses blagues, son allure de cowboy, et ses nombreuses relectures sur le manuscrit de thèse. Merci énormément à Thomas pour tout ce qu'il m'a appris pendant nos discussions, tant d'un point de vue scientifique que humain. Rien de travail n'aurait été possible sans toi ! Je souhaite aussi remercier Valérie et le personnel de l'animalerie pour s'être occupés de mes petites souris, et pour nos discussions dans les couloirs. Merci à Marjolaine pour nos longues journées de stages, et pour sa bonne humeur. Merci à Aurélie pour les allers-retours dans l'ascenceurs, pour les mercredi génotypages (merci Sandra mille fois pour ton savoir faire et ta gentillesse), pour les corvées de doxycycline. Je suis triste qu'on ait plus la possibilité de travailler ensemble. Ah... et pas merci pour les chansons dans la tête, notamment pour celles qui résistent, qui prouvent que tu existes ! Je tiens à remercier Anthony pour être la cible constante de mes blagues pas drôle et d'y rigoler quand même. Merci pour ta bonne humeur, tes mouvements de gym inattendus, et pour la naïveté de ta jeunesse.

Un énorme merci à Sebastian pour ... tout ! Tu as été une de mes motivations pour venir au labo le matin. Ta détermination, ta gentillesse sans borne, tes raps, ta bonne humeur et tes conseils du quotidien ... Reste chaud patate comme ça t'es farpait !

I want to thanks all my neuro-related friends, researchers or students, for the nice moments we spent together: Clément, Benjamin, Lina, Alexis, Hind, Aurélie, Bérénice, Julie & Sophie, Marielle, Ana, Arthuro, Omer, Tom, Dan, Alejandro... In particular, thank you Tim and Bence for your amazing supervision in Boston, you opened me the doors on the amazing field of neuroscience. You were the best supervisors a student could wish! Thank you Esther for your enthusiast and for being such a close friend despite the distance and for being my confident.

Un grand merci aussi à Camille et Jérôme pour leur supervision, et pour m'avoir enseigné la rigueur scientifique pendant mon passage au MNHN.

Merci aux amis du CRI : Pierre, Max, Fanny, Jérémie, Charles... Un merci particulier à Camille pour nos discussions entre les cours autour de pizzas, pour nos potins, et sur les avertissements concernant la vie d'adulte. Merci à Élodie, Sofie, et Maria pour leur gentillesse même quand je perdais patience. Vous avez su m'aiguillez avec sourire dans les profondeurs de l'administration d'une école doctorale. Un grand merci à Pierre pour tous les bons moments passés ensemble à discuter, dessiner, ou jouer.

A big thank to the Cedevita crew : Seb, Nina, Dora, Anna-Maria, Vova, Srinath, Sindhuja, Cedric, Ming (the captain), Mario, Yosip, Petra ... and to the wonderful and brilliant students from the bee team : Julia, David, Davor, Salih, Anastasija, Francesco, Melita, Barni, Eszter. Another big up for Edu, my partner in crime during this last summerschool ! You were all amazing, and it was a pleasure to be part of the S3 adventure with you !

Un remerciement à la famille de l'ENS pour ces 3 formidables années en votre compagnie : Romain, Maxime, Guilhem, Mathurin, Louise, Mathilda, Nina, Lucile, Cadiou, Hortense, Aymeric, Alexis, Miguel, Kathleen, Adrien, Vero, Marianne, Milena... En particulier merci à la team volley : Seb, Ed, John, Marouane, Caucal, Benji, Tony, Thibaut, Marc, Joceran, Celia, Vadim, Garrel, Tiphanie... pour le temps passé à gainer ensemble et à taper dans un ballon. Un énorme merci à la coloc' Jess, Greg, et Fossat' de m'avoir adopté comme un des leurs.

Un gros merci à mes parents adoptifs Anne et Xavier ! Vous êtes géniaux à tout point de vue et j'ai eu une chance énorme de vous compter dans mes amis proches. Merci à Solène pour ces belles années d'amitiés, les brèves ambiguës, la team SMAK, les hivers à Préma, les balades en kayak... Tu as toujours été là dans le moments difficiles, merci mille fois. Merci à Alexandra, en sa qualité de confidente depuis des années. Un remerciement tout spécial à Laure pour toutes les expériences qu'on a vécu ensemble. Merci pour tout ce que tu m'as appris et apporté. Merci pour ton soutien tout au long de la thèse dans les moments de doute au début, au milieu et à la fin.

Un remerciement tout particulier à la bande de Roscoff et associés : Gilles, Jo, Fanny, Aurélien, Mathilde, Ula, Brice, Tangi, Clio, Laure, Nathalie, Morgane Raph', Aline, Ramzi, Valentin, Morgane, Oscar, Magali, Aurore, Rozenn, Anna & Zoé, Cécile, Brigitte de l'accueil... Merci à la promo pour ces moments partagés dans la station, les patiflettes en plein été, les chansons de Renaud... Merci à Stéphane pour ses conseils de vieux pirate pendant nos sessions de pêche ou pendant nos expériences sur disquettes avec Nora Jones. Merci à toi pour m'avoir donné envie de faire de la recherche dès la licence.

Merci à tous les gens qui ont contribué de près ou de loin à ce travail : Merci à Nelly pour cette aventure dans le deep side of machine learning. Merci à Thibault pour ta contribution à ce travail, pour nos veillés python jusqu'au petit matin, et pour ta détermination remarquable dans les manips. Merci à Angèle, Elena et Aurélien pour vos contributions, qui même si elles ne sont pas incluses, m'ont permis d'y voir plus clair sur le système auditif. Merci à Alexis, pour les pauses thé et les plans pour renverser l'économie dofusienne. Merci au groupe de Polytechnique: Inès, Mathieu, Charlotte et Victoria... pour leur investissement.

Un énorme merci aux membres de ma famille pour me supporter depuis bientôt 26 ans. Merci à ma soeur Pénélope, à mon frère Hadrien et merci à mes parents pour leur investissement dans mon éducation et mon travail académique. Merci à eux pour tous ! Je leur dédie ce travail car c'est en grande partie le fruit de leurs efforts.

Je tiens à remercier Laurent pour ces années exceptionnelles que j'ai passé avec lui. Tu es une belle personne Laurent ! T'es devenu l'ami que je rêvais d'avoir ! Merci de tolérer mes pizzas brésiliennes, mes kda minables, mes yoyos émotionnels, et le drapeau breton dans le salon !

Je souhaite remercier tout particulièrement Gaïa pour son support au quotidien, son dynamisme et son rire qui ont été une source de bonne humeur et qui m'ont remonté le moral en toute circonstance. Merci a tes parents aussi, Anne et Michel, pour leur gentillesse et leur hospitalité. Merci pour tes messages d'encouragements laissés dans mes carnets ! Merci pour ces moments de complicité partagés ! Merci pour les appels téléphoniques, source de réconfort au milieu de la nuit ! Merci pour tout le bonheur que tu m'apportes jour après jour ! Ces dernières années n'auraient pas été aussi heureuses et amusantes sans toi !

Pour finir je souhaite remercier Brice, en sa qualité de directeur de thèse, pour ses compétences dans tous les domaines, pour la patience qu'il a eu à mon égard, et pour sa gentillesse. Tu m'as dit un jour : "La première année de thèse, tu écoutes docilement le directeur de thèse, la deuxième année tu commences à te rebeller contre lui et la troisième année tu ne le supportes plus !". Cette phrase est une des rares fois ou tu as eu tord pendant ces 3 ans. Je te suis extrêmement reconnaissant pour l'attention et le temps que tu m'as consacré, et pour toutes les choses que tu m'as appris. C'était un honneur et un immense plaisir de naviguer pendant 3 ans sous pavillon Bathellier !

Contents

To my parents...

Introduction

Preamble

The mammalian brain processes sensory information to create an inner representation of the environment, a phenomenon called perception, and then uses this representation to select and perform the appropriate behavior. To construct this inner image, the brain decomposes the scene into individual objects that can be identified as perceptual objects. How perception and behavior arise from the encoding and interpretation of information by neuronal networks are ones of the most fascinating and challenging questions of neuroscience. Indeed, each object in the real world can exist with different shapes, different colors, different sizes, different sounds... but the brain categorizes it into the same perceptual object despite this diversity. Therefore our brain extracts relevant pieces of information from the object, called features, that are combined together to interpret the object as a recognizable percept.

This example illustrates the power of perception and raises two main questions : How does the brain extract information from the environment? How does the brain interpret this information?

While the first question concerns how the sensory systems transduce physical signals, like light or sound waves, into neuroelectric signals, the second is more focused on how the neuroelectric signals are then transformed to encode perceptual objects. To better understand how perceptual objects are encoded, it is necessary to work on awake animals which is very challenging because of all the active perception that the animal is doing (eye movements, sniffing, whisker movements, etc). In order to control as much as possible the experimental procedure, this PhD thesis focuses on hearing and how auditory objects are encoded in the auditory system, a system with arguably the most stable sensor across all sensory modalities.

From a sound to a neural code representation

What is a sound ?

A sound is a pressure wave that propagates through a medium (usually air) at a speed of 340m.s^{-1}. It results from the vibration of an object in the environment, like a guitar string for instance. It carries information about the amplitude, the frequency content of the wave, and the localization of the object. The amplitude of pressure variations is called sound level and it is measured

on a logarithmic scale, in decibels (dB), which is the log-ratio between measured pressure variations p_1 and a reference pressure p_{ref} : $20 * \log(p_1 / p_{ref})$. The reference pressure p_{ref} was chosen as the limit of human hearing of 20 micropascals. Sound waveforms are very complex and diverse. One way to decompose the information contained in a sound waveform is to compute its Fourier transform to represent the wave as a sum of frequency components, termed frequency content of the sound. For a very simple sinusoidal pressure wave, the frequency content would be a single frequency. But generally, natural sounds waveforms in the environment are much more complex with a rich multi-frequency content that changes over time. Together with the frequency content, the intensity and the localization information form the identity of each sound. The mammalian brain extracts these elements from the pressure wave in order to understand its auditory environment.

How does a sound enter in the brain ?

Sounds need to travel though the outer and middle ear to reach the cochlea that transduces sounds into a neural code. First, they enter the ear canal, which is part of the outer ear, and they are changed from air pressure waves into a fluid pressure wave by a set of 3 small ossicles that vibrate against the eardrum membrane (Figure 1), thus creating a pressure wave in the fluid of the cochlea (perilymph) on the other side of the membrane. The cochlea is a coiled tubular structure, filled with fluids (perilymph and endolymph), composed of two elongated compartments separated by a stiff basilar membrane (Figure 1). The basilar membrane vibrates locally, then a pressure wave passes through and the location of the vibration depends on the frequency content of the sound. On the basilar membrane, there are cells with mechanoreceptors that detect deformation. These cells are called Inner Hair Cells (IHN) and Outer Hair cells (OHN) and they contain stereocilia that are sensitive to the vibration of the basilar membrane (Figure 1).

How does the cochlea transduce a pressure wave into a neural code ?

The basilar membrane of the cochlea vibrates locally according to a precise organization with respect to the incoming pressure waves from the middle ear. Close to the eardrum, at the base, the basilar membrane vibrates for high frequency sounds and close to the end of the cochlea, at the apex, it vibrates for low frequency sounds (Figure 1). The cochlea is sensitive to a frequency range that depends on the species. For example, humans are limited to a hearing range from 20Hz to 20kHz, whereas mice have a hearing range between 2kHz and 85kHz[2]. Between the base and the apex, there is an increasing sensitivity from high to low frequency, also called the tonotopic gradient. During a vibration of the basilar membrane, the stereocilia receptors depolarize the IHCs which release neurotransmitters that excite fibers in the auditory nerve. It is interesting to note that there are no action potentials

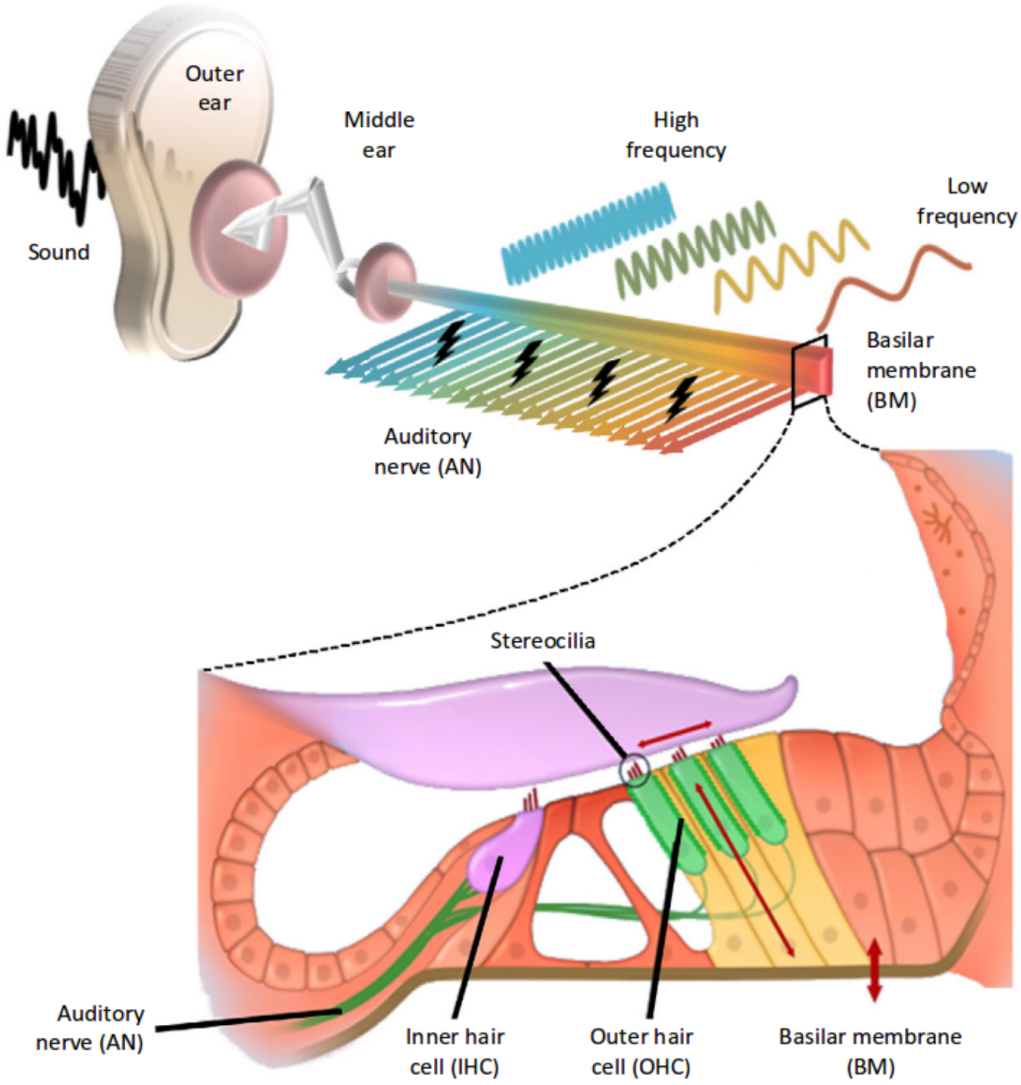

FIGURE 1: **The functional organization of the ear and the cochlea.**

Schematic description of the first processing that occurs in the auditory system. The sound enters by the outer ear canal and creates a vibration in the ossicles of the middle ear, that is transmitted to the cochlea. The basilar membrane is organized with tonotopy and it sends the frequency information to the rest of the auditory system through the auditory nerve. IHCs are mostly coding the information about the sound such as frequency and intensity whereas the OHCs are mostly modulating the vibration of the basilar membrane. (Illustration from Lesica 2018 [1])

in the hair cells, the first action potentials of the auditory system are generated in the auditory nerve. The pioneering work of Galambos and Davis in 1943 recording auditory nerve fibers from anesthetized cats provided the first functional understanding of the cochlea [3]. They understood that a local vibration of the basilar membrane creates a depolarization of a local population

of IHCs that in return excites only a small number of fibers in the auditory nerve. Therefore, as the basilar membrane is tonotopically organized, the identity of the fibers getting excited during a sound encodes the frequency content[3, 4]. The sound level is encoded by the number of action potentials created by the fibers because the IHCs produce a larger depolarization for loud sounds [3, 4, 5]. In contrast to IHCs, OHCs have mostly a modulatory role of the basilar membrane vibration [1, 6, 7]. They can produce an amplification of the vibration for the weakest sounds and a compression of the vibration for the loudest sounds [1, 7, 8]. The modulation of the vibration allows the auditory system to encode a very large range of sound intensities. OHCs are also involved in tone-suppression, as shown by Geisler and colleagues in 1990, a phenomenon described as a reduction of the firing rate induced in the auditory nerve by one sound in the presence of another sound[9, 10, 11]. Tone-suppression could be involved in complex hearing situations like a conversation in a noisy environment, a daily hearing challenge called the "cocktail-party problem" [12, 13].

What is the neural code in the auditory nerve ?

The auditory nerve transmits sound information now encoded with action potentials as in the central nervous system, and thus the rest of the auditory system. At this level, the sound information is encoded with two different but complementary neural codes: the precise timing of the action potential and the firing rate of the auditory nerve fibers[14, 5].
The precise timing is a code mostly used for binaural difference and to precisely encode the frequency of low frequency sounds [14]. It also preserves the high temporal resolution of the auditory system. The binaural difference is a method used by the brain to estimate the localization of the sound source based on the precise time at which the sound reaches the auditory nerve in each ear [15, 16]. The small time difference between the two ears, caused by the traveling time of the pressure wave in the air, is used to determine the location of the sound source. The precise timing of each spike could also be used to encode low frequency sounds. Palmer and Russels show that the IHCs have a voltage that follows the oscillation of the sound for low frequencies, a phenomenom called phase-locking [17, 18] . In others words, for low frequency sounds the fibers from the auditory nerve are spiking with the same frequency as the sound if the frequency is low enough [17, 18]. In general, precise timing of spike is mostly observed in the brainstem, and tends to be discarded in favor of a firing rate code in higher areas of the auditory system [14] (but not entirely, see [19]).
The firing rate of the fibers is mostly used to encode sound frequencies [20] and intensity (as shown before, the intensity of the sound: the louder the sound, the higher is the firing rate of the fibers). Even if this code is temporally less precise, it probably encodes the relevant sound features needed to construct or recognize perceptual objects [14]. The precise firing rate of each fiber depends on the intensity and on the frequency content of the sound. Indeed, a fiber typically fires action potentials for a very precise frequency

(called the characteristic frequency) at low intensity but it gets less specific with high intensity sounds[5]. As a consequence, when represented on a graph, the firing rate of a fiber as a function of the frequency (in the x axis) and sound level (in the y axis), forms a V-shaped pattern (Figure 2a). Each fiber can thus be summarized as a V-shaped frequency band filter; which means it is sensitive to a frequency band that is dependent on sound intensity, and the louder the sound is, the broader is this sensitivity[5] (Figure 2b). Therefore, all together these fibers decompose the sound into its frequency components thanks to V-shaped filters in a similar way to the Discrete Fourier Transform (DFT), which decomposes sounds into sinusoids [21].

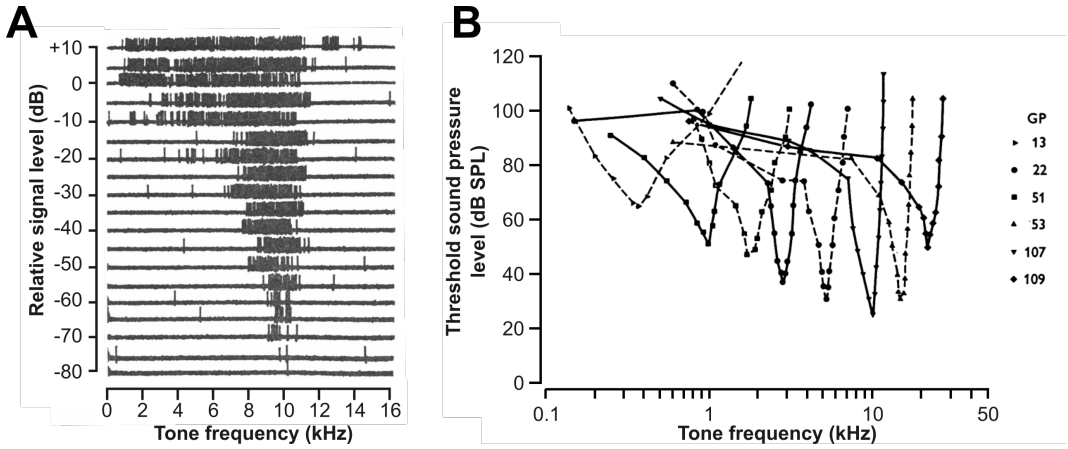

FIGURE 2: **Sound encoding in the auditory nerve.**

A. *Action potentials (verticals bars in the signal) recorded from a guinea pig auditory nerve during frequency modulated sounds (14kHz/s or -14kHz/s) of different loudness. The fibers often show a tuning to a particular frequency to low intensity sounds and increase their tuning width for high intensity sounds resulting in a V-shaped tuning curve. Adapted from Evans et al 1972 [22].* ***B.*** *Recordings of eight fibers from the auditory nerve of 6 guinea pigs (symbols) described eight V-shaped tuning curves that specify the auditory range of the animal. Each curve is obtained by testing for each frequency the loudness that elicits a response for the fiber. The horizontal axis is logarithmic, and the V-shape has the same general shape over a large range of frequencies, indicating a logarithmic perception of the frequency. Adapted from Evans et al 1972 [22].*

What are the inputs for the central auditory system ?

The inputs of the central auditory system are the information as it is encoded in the auditory nerve fibers by the cochlea. From a computational point of view, the cochlear operation on the sound can be seen as a DFT, a method to decompose a signal into a sum of sinusoids with different frequencies, different phases and different intensities (Figure 3). However, this comparison is a bit naive [21]. While the DFT decomposes the sound into sinusoids with linear frequency intervals (Figure 3a), the cochlea decomposes the sounds with logarithmic-like frequency intervals [21] (Figure 3b). Indeed, the difference between the characteristic frequency of adjacent fibers encoding low

frequency sounds is very small compared to the same difference measure for high frequency encoding fibers. Another difference is that the DFT separates the sounds in pure sinusoidal waves while the cochlea has broader filters centered on the characteristic frequency[21]. The decomposition of the sound into several bandpass frequency filters performed by the cochlea is called a cochleogram (Figure 3b). Because it contains the frequency content and the intensity information, the cochleogram contains the majority of the information carried by the firing rate, which is needed for the central auditory system to recognize auditory objects.

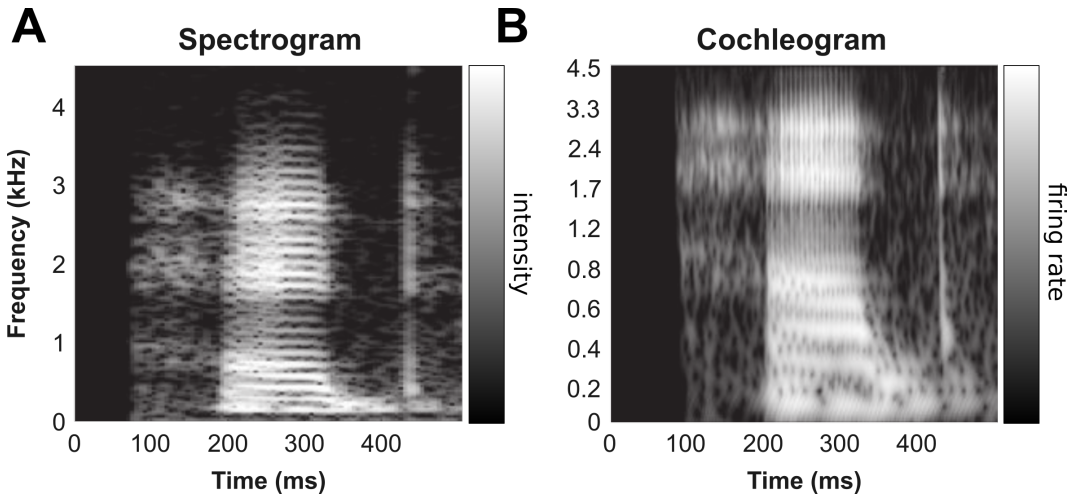

FIGURE 3: **The cochleogram is a decomposition of sounds into frequency bands**

A. *Spectrogram of a sound (the spoken word "head") with a Fourier decomposition in short temporal windows, to obtain the spectral content of the sound over time. Values range from black (0) to white.* ***B.*** *Cochleogram of the same word as in* ***A*** *but with a frequency decomposition similar to what is encoded in the auditory nerve fibers. The colormap ranges from white (0) to black and it represents the firing rate of the fiber. Note that the frequency bands on the vertical axis have logarithmic intervals. Adapted from Auditory Neuroscience, by Schnupp et al [21].*

From a sound to a neural code representation

To summarize, the role of the peripheral auditory system in object recognition is to transform the sound waves into a neural signal for the central auditory system. The sound arrives as a pressure wave at the cochlea where the frequency components of the sound are separated thanks to the basilar membrane. Each frequency component triggers a response from a local population of IHCs that elicit specific responses in the auditory nerve. These responses are spatially organized in the cochlea, following the tonotopy, and they form a cochleogram version of the sound that is transmitted as a neural code through the auditory nerve. The cochleogram information extracted

from the sound is then transmitted to auditory system to be encoded in high-level representations.

From the cochleogram to high-level representations

What is a high-level representation?

Representations in the brain can be very complex. In the visual system for example, while some neurons are sensitive to simple light dots in the visual field [23, 24], leading to a simple encoding of the stimulus, others are sensitive to faces or more complex objects [25, 26, 27, 28]. It has been shown in humans that some neurons encode distinct people. For example, a neuron fired every time someone mentioned or showed a picture of Jennifer Aniston [26]. A similar neuron was also recorded with tuning for Luke Skywalker, a fictive Star Wars character [27]. These very high-level neurons are activated by faces or persona, thus encoding very complex and abstract concepts. More generally, a high-level representation is a categorization of objects and the level of the representation depends of the abstraction and the complexity needed to categorize this type of object. With hearing, the encoding of a frequency or a specific intensity is relatively simple compared to the encoding of frequency-modulated sounds, natural sound categories, or words.

Hierarchical organization of sensory systems

In hierarchically organized sensory systems, the representation of external objects is less complex in the brainstem compared to high level areas like the sensory cortex [29]. The signals coming from the peripheral organs like the eyes or the ears get transformed in multiple areas in the brain. Every time the signal passes through a processing station or nucleus, there is a transformation, and these transformations pre-process the signal so that high-level areas can encode increasingly complex representations. The features extracted from the stimuli, which are initially simple (like the frequency content of a sound, the position of a point of light on the retina), are combined during these transformations to form more complex features (Figure 4a). The experimental and theoretical work of Hubel and Wiesel [30] sheds light on the hierarchical organization of the visual system and how simple feature detectors like neurons sensitive to a point of light in space (on-center cells)[23, 24], connect to a neuron that is responsive to a more complex feature, like a neuron sensitive to a line of light in space, like the so called “simple cells” [30] (Figure 4a). In the auditory system, there is also an increasing complexity in the features detected [31, 32, 33] suggesting a hierarchical organization from the cochlea to the auditory cortex [31, 34, 32, 35, 36, 37, 38]. In addition to the hierarchical architecture, the auditory system is organized with multiple parallel pathways (Figure 4b). From the cochlea where frequencies are separated in parallel fibers in the auditory nerve, passing through the cochlear nucleus where the sound information is split in three different nuclei, to the

cortex; the architecture of the auditory system is also distributed into specialized regions (Figure 4b)[39, 40].

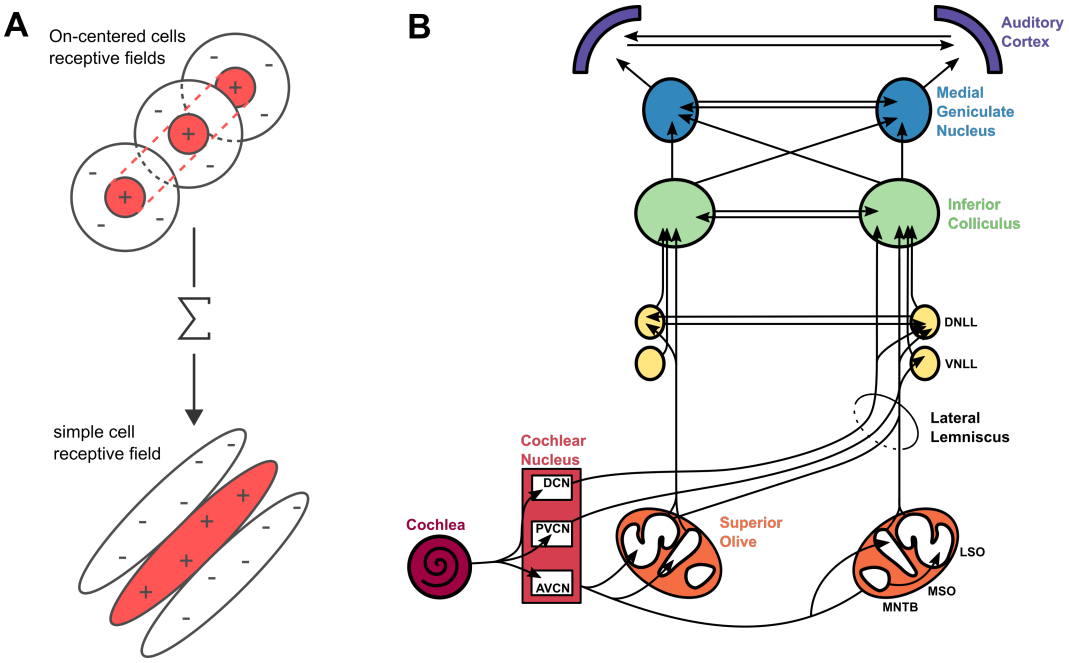

FIGURE 4: **Hierarchical feature extraction and architecture of the auditory system**

A. *The visual system is often used to explain how neurons detecting simple features are organized to extract more complex features from an image. On-centered cells are sensitive to dot of light in a particular spot of the field of view. On centered neurons with aligned sensitivity (represented here by the 3 receptive fields aligned in the field of view) connect to a cell called a simple cell that is sensitive to a line of light in the field of view. The sensitivity of neurons for aligned dots of light is directly linked to the receptive field of the neuron they are connected to. The red area is a region of the field of view the neurons are sensitive to, whereas white regions with minus signs represent inhibitory regions.* ***A.*** *The architecture of the auditory system in mammals is organized into multiple nuclei. The sounds are transduced to neural activity by the cochlea that projects to the cochlear nuclei (DCN: Dorsal Cochlear Nucleus, PVCN: PosteroVentral Cochlear Nucleus, AVCN: AnteroVentral Cochlear Nucleus). Then the information arrives to the several nuclei of the superior olive (5 in total, LSO: Lateral Superior Olive, MSO: Medial Superior Olivary Nucleus, MNTB: Medial Nucleus of the Trapezoid Body) and passes through the lateral lemniscus to reach the inferior colliculus and other small nuclei (DNLL: Dorsal Nucleus of the Lateral Lemniscus, VNLL: Ventral Nucleus of the Lateral Lemniscus). Finally, the sound information goes to the thalamus and then to the auditory cortex. This illustration only represents the feedforward connections in the auditory system and ignores the multiple feedback connections between high-level areas and the brainstem or the cochlea. Adapted from the book Auditory pathways: Anatomy and physiology from Pickles [41].*

Architecture of the central auditory system

The architecture of the mammalian auditory system is composed of multiple nuclei (Figure 4b). After the inner ear and outer ear, the sound is transformed into neural signals in the cochlea to be sent to the cochlear nuclei via the auditory nerve. The cochlear nuclei are a set of three nuclei: the dorsal cochlear nucleus, the posteroventral cochlear nucleus, and the anteroventral cochlear nucleus. The sound information is then sent to the superior olive, which is composed of 5 nuclei, with the three main ones being the lateral superior olive, the medial superior olivary nucleus, and the medial nucleus of the trapezoid body. It then reaches the inferior colliculus and other small nuclei (the dorsal nucleus of the lateral lemniscus and the ventral nucleus of the lateral lemniscus) through the lateral lemniscus. Finally, the sound information reaches to the thalamus, where it is routed to the auditory cortex. The auditory cortex itself is composed of subregions: a core, that includes the primary auditory cortex, a belt, and a parabelt. Because of the hierarchical organization of the auditory system, every time the sound information passes through the nuclei of the architecture, the features extracted from the sound get more abstract and more complex.

How to estimate a feature extracted by a neuron?

A receptive field is the subspace of the stimuli that trigger action potentials in a neuron. The complete list of features, or characteristics, extracted from the sounds by one neuron is therefore the receptive field of the auditory neuron. One method used to estimate the receptive field of a neuron consists of playing a set of very controlled sounds (like frequency modulated sounds, or intensity modulated sounds) and to find the common feature all the sounds eliciting response have in common. This technique was used to determine orientation selectivity in frequency modulated sounds[42, 43] or in intensity modulated sounds[44, 45]. This method is very powerful to determine the selectivity of a neuron for a given sound feature, but it is very dependent on the set of sounds. Another approach is to assume nothing about the sound characteristics a neuron is tuned to with a method calls reverse correlation [46]. It consists of playing a particular set of sounds to an animal while recording a neuron and then averaging all the spectrograms of the sounds that elicit spikes (spike-triggered averaging). The result obtained is called a spectro-temporal receptive field (STRF) (Figure 5) and it is the linear approximation of the spectrogram of the sound that triggers the most activity in the neuron. Therefore, the neuron's firing rate can be obtained very easily from the spectrogram and the STRF by a convolution of the two signals. STRFs are often used to describe qualitatively and quantitatively receptive fields of neurons in the auditory system (Figure 5). Parameters such as the characteristic frequency, the band width, the latency of the response of the neurons and even inhibitory sound features are deduced from STRF. However, if it is estimated with a given set of sounds, STRFs are very dependent on the set of sounds. To get rid of the bias introduced by sound selection, a few studies used white noise sounds that contain equal intensities at all frequencies [47,

48, 46]. Then, because white noise is not very efficient to drive activity in the late stages of the auditory system [49], it has been replaced by particular synthetic sounds. These sounds are for example modulated sinusoidally both in the temporal and spectral domains, like ripples[50, 51, 52], or they can be a randomly generated series of short pure tones[53, 54, 55]. A comparison between STRFs in the brainstem and in the cortex shows that the features detected in the cortex are more complex[54, 53] than the features extracted at the level of the cochlea[56]. Even if there are still neurons with tuning to a simple frequency in the cortex[57], multiple neurons have more complex receptive fields[54, 53] (Figure 5). For example, some neurons have specific interactions between frequency and time, and respond to sounds that are frequency modulated (Figure 5). Other neurons are sensitive to two different frequencies; thus, they respond to chords. Finally, some of them are tuned to very specific temporal patterns of excitation and inhibition (Figure 5).

Limits of the linear STRF approach

Despite the understanding it brings in the study of high-level representations, reverse correlation has strong limitations. First, the STRF does not generalize properly to other sets of sounds [58, 59, 60]. For example, if it is trained on responses to particular sound set A, the firing rate prediction for another set of sounds B will be poor (approximately 89% of error)[58]. One reason for this could be that the entire receptive field of a neuron cannot be represented in a single STRF [61], especially for high-level areas like the auditory cortex. As an example, if there is a neuron able to detect the word "house", it will be activated when this word is pronounced independently of the frequency of the speaker, and thus this neuron's STRF is impossible to represent. Indeed, if the neuron STRF is tuned to a certain speaker saying "house", the neuron will not respond for another speaker with a high pitch voice saying the same word, because the frequency content of the sounds will be to high for its receptive field. Therefor the STRF is not the best representation of the receptive field of a neuron for high-level representation. It is a simple description of the receptive field because it is a linear approximation of it. Moreover, it is known that the auditory system processes information with non-linear transformations, so a linear approximation of the receptive field is probably missing a lot of features extracted by each neuron [61, 62].

High-level representations are built with non-linear systems

What is a non-linear system?

A non-linear system is a system in which changes in the output are not proportional to changes in the input. As non-linear transformation is defined in opposition to linear transformation, it is important to first define what a linear transformation is. A linear transformation is defined as a function f

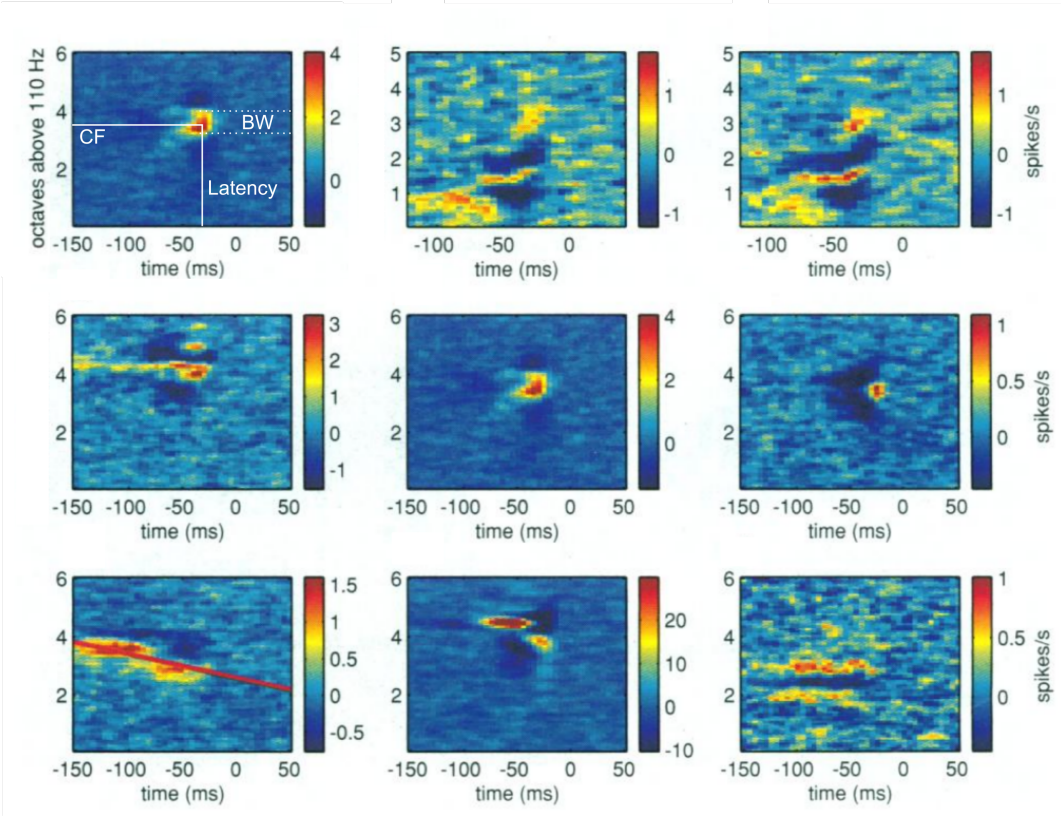

FIGURE 5: **Spectro-temporal receptive fields show complex patterns of sensitivity.**

*Spectro-temporal receptive fields define the frequency patterns and the latencies that neurons are tuned to. It is expressed as the average spectrogram of sounds that trigger activity in a neuron. (**top-left**) We can determine the characteristic frequency (CF), the band width of frequency the neuron responds to (BW) and the latency for the neuronal response. The red regions indicate the excitatory patterns whereas the blue regions are inhibitory patterns for the neuron. Some neurons are tuned to a complex pattern of frequencies (**top-middle and top-right**). Others have simple receptive fields with tuning to a particular frequency and a particular latency much more like the auditory nerve (**middle and middle-right**). Some neurons are tuned to organized features of the sounds like frequency modulated sounds (bottom-left) or chords (**bottom-right**). Adapted and completed from deCharms et al [54].*

such that $f(x) = A \cdot x + B$ for a set of parameters A and B, where $A \cdot x$ is the dot product between A and x. In the example of an auditory neuron and its STRF, $f(x)$ would be the firing rate of the neuron, A would be the STRF, B is a bias term that represents the basal firing rate, and x would be a part of the presented spectrogram with the same dimensionality as A. For each temporal window, we apply the transformation of the spectrogram by the STRF to obtain the firing rate of the neuron. Non-linear transformations are defined in opposition to linear transformations, which means every transformation that cannot be written in the form $f(x) = A \cdot x + B$ is non-linear. A simple example of a non-linear transformation is to apply a threshold θ so that if $f(x) < \theta$, then $f(x) = 0$, otherwise $f(x) = f(x)$. The threshold method

is widely used in machine learning and in neuroscience [63] to introduce a non-linearity, after a linear filter like an STRF for instance [64]. The threshold non-linearity is very convenient because it makes sure that, in the prediction, the firing rate of a neuron never reaches negative values because it is not physiologically possible. Yet there exists many other form of non-linearities.

Deep neural networks are extremely efficient to build high-level representations thanks to a succession of non-linearities

In order to understand the role of non-linearities in signal processing, the example of the deep learning revolution is very adequate[65]. Less than 10 years ago, signal processing and machine learning underwent huge improvements in object detection and categorization tasks for images and sounds, thanks to deep neural networks (DNN)[65]. In the recent years, several algorithms were designed to recognize faces, cars, houses, cats, dogs, and about 1 000 other categories of the ImageNet challenge[66] from pictures (Figure 6). Since 2012, with AlexNet[67], the first deep convolutionnal neural network, this type of model is showing extraordinary performance in this challenge especially since 2015 where the classification error dropped below human error levels of 5.1% (Figure 6) [68, 69]. The DNN architecture is designed with multiple layers each composed of computational neurons, called units, that perform a linear filter and a threshold non-linearity. In computational terms, DNNs encode images with a succession of linear transformations and non-linearities; and generally, the more layers there are, the better the categorization performance [70]. Generally, a DNN can be split in two parts: A first part of the network is extracting features from the image, and the second part uses the features extracted to take a decision on the categorization problem. Interestingly, the model learns the linear filters by itself during the training period with supervised learning, and the best encoding (i.e. the best representations) of the stimulus set to perform the task. At the end of the training, in the last layers of the network, units extract very complex features like faces, animals, street signs ... independently of the size, or the angle of the picture, showing a very high level of abstraction[65, 71]. Non-linearities in DCNNs are essential to encode these complex representations. Indeed, the parameters for the nonlinearities constitute a very small fraction of the total parameter set of the model, but still if they are removed from the layers, the computation of the all network is summarized as a linear transformation of the image which is not sufficient to perform the categorization task with great accuracy. In sound processing, deep learning algorithms have already achieved impressive results such as recognizing speech [72, 73], music genres [74], or even music composers [75]. The vast majority of these algorithms do not recognize sound features on the signal directly but they often first transform the sound into a spectrogram, and then use the spectrogram as an image to do the categorization task.

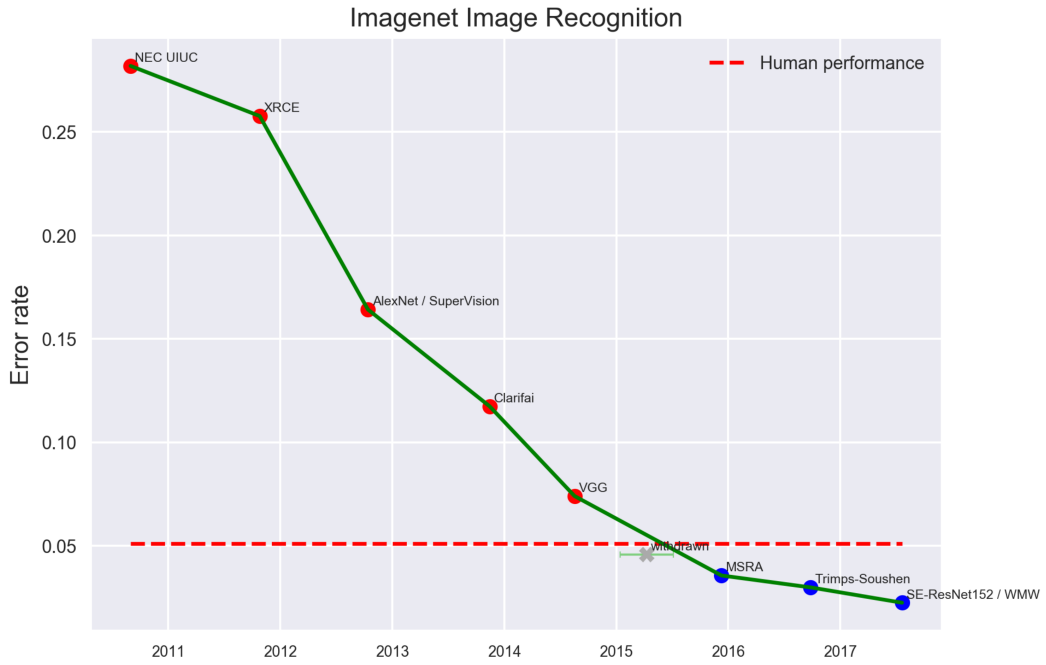

FIGURE 6: **Performance of DNNs for image recognition reaches human-level.**

The performance of the best algorithms each year measured as the error rate of misclassification from the ImageNet challenge. Deep convolutionnal neural networks are winning the challenge since 2012 with the introduction of the AlexNet[67]. The algorithms reached human performance (dot red line) in 2016. From Peter Eckersley, Yomna Nasser et al. 2017.

Deep neural networks are computationally similar to sensory systems

Interestingly DNNs are now challenging humans on some tasks as diverse as image recognition[68], speech recognition [73], or playing Go [76] (or even playing more complicated games such as DotaII). It is then legitimate to ask if there are similarities between DNN and sensory systems in computational terms. As both systems are trained (by an algorithm or by learning in the brain) with similar tasks, they could obey the same constraints and thus converge towards the same representations for objects. Recent studies follow this hypothesis and compare the performance of a DNN and the brain in both visual and auditory perceptual tasks [77, 78, 29]. In a serie of study, Yamins and colleagues compare how DNNs encode stimuli relative to how the brain does it [79, 29, 37]. In their work on the auditory system [37], they first train a DNN to perform two perceptual tasks, a word recognition task and a musical genre recognition task (Figure 7). Then, they recorded the cortical activity of human brains for a series of sounds with fMRI and they compare the representation of the stimuli found in these recordings to the representation of the DNN trained on the perceptual tasks only. To do so they predicted the fMRI voxel activity based on the activity of a certain layer in the network

with a linear model. In the middle of the network, where it is expected to have high-level representations, the DNN could accurately predict the fMRI activity, showing that the network representations of sounds are very similar to the human brain representations (Figure 7). Of course, the representations were not identical (one neuron for one unit of the network) but the encoding in the network was similar, up to a linear transformation, to the encoding found in the brain recordings. This study shows that even if the mechanisms are not exactly similar, there is some computational similarity between the auditory system and a DNN, suggesting that non-linearities are important in both systems to generate high level representations.

There are also differences between deep neural networks and sensory systems

Even if DNNs and sensory system share similarities, they also have great differences, especially in the learning process [65]. In terms of computational mechanisms involved in the brain for learning, there is no real evidence that the learning that is implemented in DNNs is similar. Some particular molecular mechanisms in the brain may be able to reproduce the learning process in DNNs[80], but this is very hypothetical so far. Another large difference between these two systems is the number of trials required to learn [65]. The amount of information, or trials, required to create a category out of family of objects is huge for DNNs, whereas it is very easy for humans to generalize from one or two samples. In other words, a human only needs to hear the word "banana" once to be able to recognize it later, even at a different pitch or with a different accent, whereas DNNs need multiple thousands of repetitions of the same word to be able to recognize it later. Also, humans can learn without specific supervision, whereas it is very difficult for DNNs to learn without supervision [65].

The auditory system is highly non-linear

The auditory system is a series of nuclei applying non-linear transformations

An intuition on the amount of non-linearities involved in the auditory system can be developed by considering the architecture of the system. If we consider the simplest computational model for a neuron, developed in 1943 by McCulloch and Pitts [81], where a neuron's output is the sum of the inputs weighted by the synaptic weights which is then passed into a threshold function; each neuron exhibits a threshold non-linearity. Interestingly, in biological neurons, there are other types of non-linearities than the threshold for the action potential, such as dendritic non-linear processing [82]. Even in the simplest configuration, with one non-linearity added every time the signal passes through a neuron, and with 5 steps of transformation in the shortest

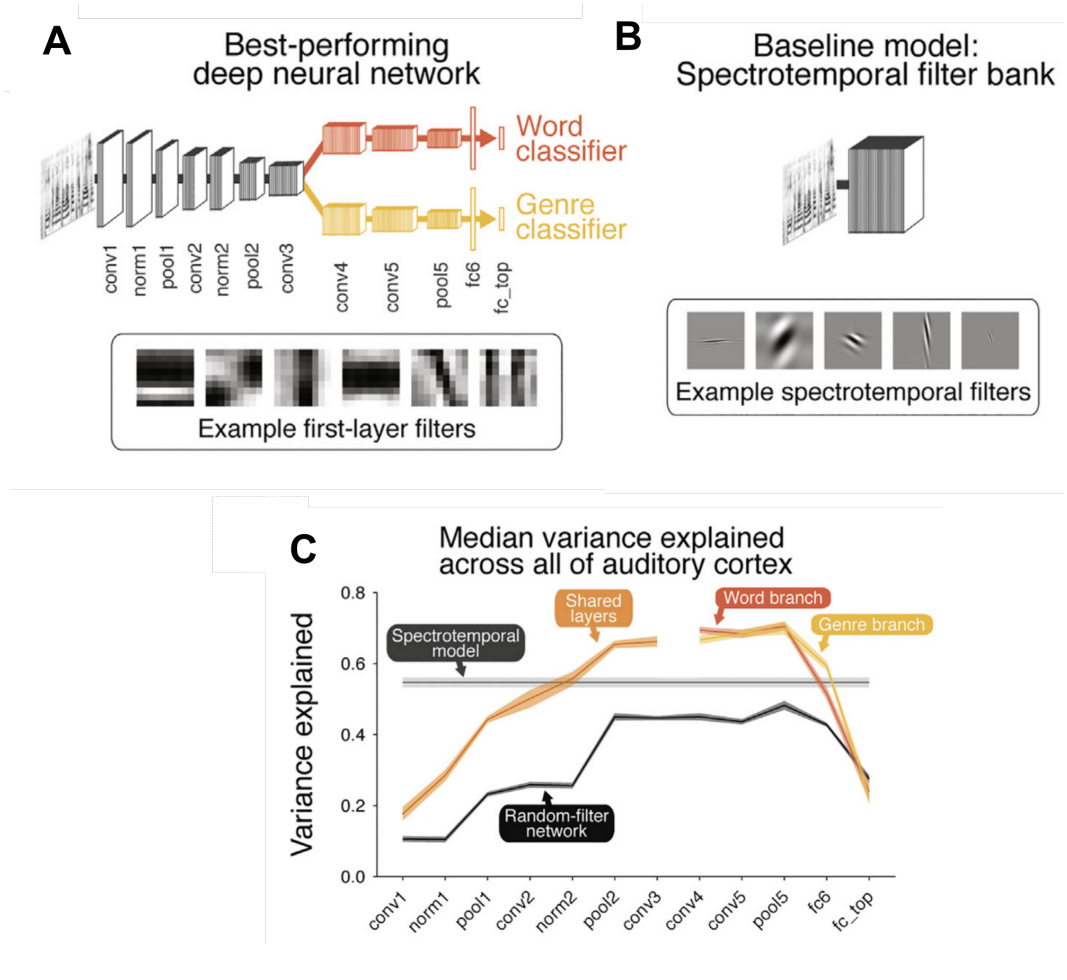

FIGURE 7: **Similar sound features are encoded in a DNN and the human brain.**

A. *Schematic of the architecture of the DNN used to solve a word classifier and a musical genre classifier. The network is composed of 5 convolutional layers followed by 2 fully connected layers. The network takes as input a cochleogram version of the sound. The first part of the network is where the features are extracted (black) and the second part is where the network takes the decision (colored). In the first part, there are also normalization layers to help for the learning process and pooling layers to increase the complexity of the features detected. Some examples of detected features on the cochleogram are shown below.* ***B.*** *Linear model used for control condition with examples of the features extracted below. As in the model in* ***A****, it works with the cochleogram as input.* ***C.*** *Variance of the fMRI voxels of the auditory cortex explained by a linear transformation of individual layers of the network. Two controls are presented: a linear model of the cochleogram (the spectro-temporal model) and a DNN with the same architecture of* ***A*** *with random filters (Random-filter network). The DNN presented in* ***A*** *shows increasing performance after each layer (colored curves) until a plateau where the representations of the sounds in the network explains 70% of the variance of the fMRI recordings. Figures from Kell and al. 2018 [37]*

path to reach the auditory cortex from the cochlea, the transformation applied to the sound information is highly non-linear in the auditory system [1]. This simple explanation also simplifies the architecture of the network

that is not only feed-forward, but that contains feedback able to modulate the responses in the brainstem and thus to introduce new non-linearities [83].

Feedback in the auditory system introduces non-linearities

The intuitive view to explain the complexity of the system is still very naive because it ignores a large part of the auditory pathway. The explanation is based on the feed-forward connections that are defined between the cochlea and the auditory cortex. However, there are neural connections going from the high-level areas to the low-level areas that are called top-down connections [41]. These connections are needed to redefine the representation of perceptual objects and convey information about attention and about the predictions from high-level areas to perform predictive coding as it was already describe in the auditory system [84, 85, 86]. In a complex environment, the top-down connections are very important to focus the attention on a sound source and to not get distracted by other sources, like in the "cocktail party" situation [12, 13]. Because of the top-down signals, the activity of a neuron does not depend only on the activity of the previous area (in the feed-forward organization) but it also depends on the activity of the high-level areas. In other words, the activity of a neuron also depends on its own activity a few seconds before and this is another source of non-linearity [83]. The top-down connections modify how the sounds are encoded by adding more contextual information to the neurons.

Non-linearities in the cochlea

The hierarchical organization of the auditory system predicts that the encoding in the brainstem is less complex and less non-linear than the encoding in the auditory cortex. Yet, the transformations in the auditory system are so complex that even at the level of the cochlear nucleus, the encoding is highly non-linear [1, 87, 88]. The cochlea role was described in the previous chapters as an organ that decomposes the sounds into frequency bands to produce a cochleogram version of the sounds. However, its role in the auditory system is more complex than that [7]. Indeed, the cochlea carries out non-linear transformations [89], mainly because the OHCs modulate the vibration of the basilar membrane. The vibrations for low level sounds are enhanced and they are reduced for high level sounds[1, 7, 8]. But the OHCs not only modulate the basilar membrane movement at their attachment location, they also can modulate the vibration at another location[1]. Therefor OHCs modulate the encoding of frequencies that are not encoded at their location. This process is similar in computational terms to a non-linearity already described in all the sensory systems [90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100]: divisive normalization. It was first described in the retina [94] and in the visual system [99, 100, 95, 96], and it is thought to be an active phenomenon in the auditory system [97, 98]. The idea is that neighboring neurons inhibit the neuron of interest in order to change its activity or its tuning [99, 100]. In the cochlea, the consequences can be the generation of an imaginary frequency

with the basilar membrane vibrating for a frequency which was not in the sound, or the suppression of a particular frequency in a chord by attenuation of the basilar membrane movement like in tone-suppression [9, 10, 11]. Because of these cross-frequency interactions, the cochlear response deviates from a simple decomposition in frequency bands and it is considered to apply a highly non-linear transformation to the sound [1]. This consideration is important especially in hearing loss where most of the hearing aids simply amplify the sounds or act like simple frequency decomposers[1]. With these hearing aids reproducing only the linear part of the cochlear computation, hearing impaired people complain that in a natural environment they struggle to understand words and sentences[1]. This example illustrates how important non-linear processing is to understand hearing and to find solutions for hearing impaired people.

Non-linearities in the auditory cortex

At the level of the auditory cortex, the sound information has already passed through multiple non-linear transformations. Cortical neurons thus encode sound information after a long series of transformations, and it is difficult to know precisely what non-linearities are generated specifically in the cortex if we ignore the representations of the previous areas. For that reason, it is important to record individually each area of the auditory system to understand precisely where the different non-linear transformations occur. As an example, some neurons in the cortex have a specific response to the beginning of the sound and they remain silent for the rest of the sound [101, 45]. This particularity that defines the On-neurons results from a non-linear operation. Indeed, it is impossible to obtain an similar On-response with a linear operation. Even if it is possible to get an increase of intensity at the beginning of a sound, this activity will be followed by a period of negative activity at the end of the sound if there is no threshold, and this is not physiologically plausible. Thus, in order to reproduce this response, a non-linear operation is needed. This could be adaptation or a threshold. These On-responses are also observed in the cochlear nucleus where they are first extracted from the sound probably with adaptation[102, 103] (Figure 8). With this example it is easy to understand that even if the cortex is maybe not the area where this type of feature is extracted, it gives a nice point of view on the system to observe all the consequences of non-linearities that transform the sound information. Therefore, this PhD thesis focuses on studying non-linearities in the auditory system by large-scale recording of the auditory cortex. In particular, this work aims at a better understanding of how non-linearities influence the representation of sounds, and thus how do these non-linearities influence perception.

This PhD thesis was motivated by four main questions: What are the consequences of non-linearities on the encoding found in the auditory cortex? How do non-linearities influence the representation and the perception of sounds? What are the consequences of the non-linearities on the encoding found in the auditory cortex? Are different types of non-linearities

interacting in the auditory system? In addition, we reasoned that another source of non-linearities could come from connections across sensory systems. Thus we also wondered how multisensory integration influences primary sensory encoding. To answer these questions, this work benefits from large population recordings in the mouse auditory cortex with calcium imaging, and it relies on a clustering analysis that preserves the non-linearity in the encoding.

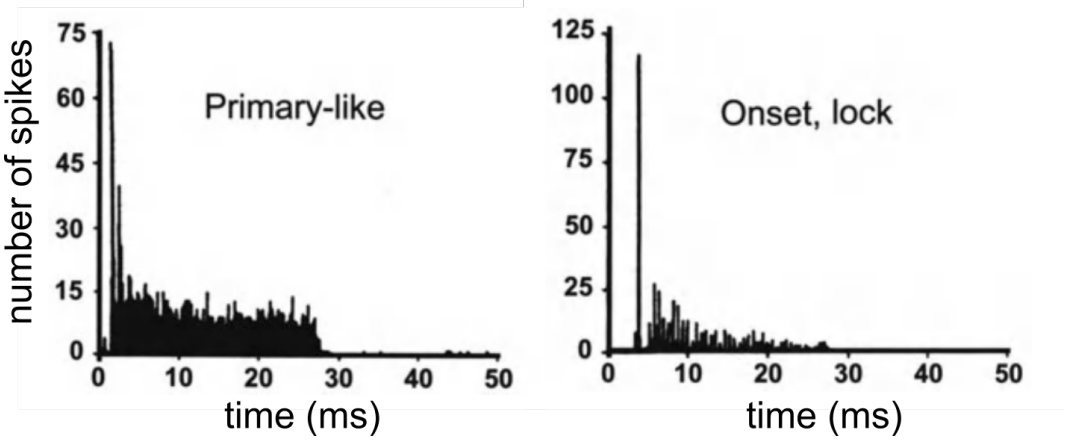

FIGURE 8: **On-responses are generated in the cochlear nucleus.**

*Post-stimulus time histogram of the response of two neurons recorded in the cochlear nuclei (ventral and dorsal nuclei) of the cat. Responses are calculated for pure tone burst at the characteristic frequency of each neuron (sound duration 25ms). (**left**) Some neurons have a response similar to the auditory nerve fibers with a sustain response during all the sound duration. (**right**) Other neurons exhibits a strong inhibition after the onset response and therefor only respond to the beginning of the sound.*

Techniques to study non-linear representations of sounds in the auditory cortex

The mouse model is a good model to study the auditory system

All the experiments performed in this PhD work were on mice because it is a powerful model for audition[55]. Mice use vocalizations for maternal care, communication, play and mating, thus their hearing is behaviorally important [104], and a lack of vocalizations lead to behavioral deficits like a reduction of female attraction[105]. In addition, they are a great biological model in general with vast numbers of transgenic lines, viral toolboxes, complex behavior, and the possibility to do physiological recordings in awake conditions. Genetic tools have been used to dissect neuronal circuits in the brain or to address specific populations of neurons. As an example, they are

very convenient for optogenetic stimulation because there are many transgenic lines that express channelrhodopsin in different neuronal populations. Because of the size of their brain, and all the transgenic lines expressing fluorescent markers, optical imaging is very well adapted to the mouse model. In this PhD thesis, all wild type and transgenic mice were coming from the most commonly used genetic background in neuroscience laboratories the C57/Bl6 background. Unfortunately for hearing research, this line has progressive hearing loss starting at the age of 1 or 2 months for very high frequencies and hitting the low frequencies in the next 12 months [106]. Even if a solution exists to correct the genetic disease of this line[107], all the mice used during this work were younger than 16 weeks old and the sounds were in the low frequency part of the hearing spectrum for mice that produce vocalizations up to 80 kHz.

Recording of large cortical populations with calcium imaging

To address the question of how complex representations form in high-level areas like the auditory cortex, it is important to consider the auditory cortex as a population of neurons encoding the stimuli. Therefore, in this study we could neither use methods with a spatial resolution larger than the neuron size (e.g. EEG or fMRI), nor a method that records only a few dozens of cells because high-level representations are possibly encoded by several hundred neurons (e.g. tetrodes, single cell patch). Calcium imaging is an optical method able to simultaneously record the activity of thousands of neurons with single cell resolution and to map the spatial location of the neurons in the auditory cortex. It works thanks to a calcium indicator: a protein that is fluorescent in the presence of a large concentrations of calcium. If the calcium indicator is inside the neurons, it creates a fluorescent event every time the neurons increase their firing rate because there is a calcium influx during the action potential generation. Therefore, the fluorescence of the neurons is a proxy for the neuronal activity. As of 2013, the calcium indicator most commonly used is GCaMP6 because of its low toxicity and high signal to noise ratio [108]. In order to express this protein in the cortical populations of neurons, the first method is to use endogenous genetic expression of the protein with genetically modified mice[109], the second is to inject a virus with the GCaMP6 gene into the auditory cortex so that it can infect neurons locally to express the protein. While the first method expresses the protein more homogeneously in the auditory cortex and is less lethal for cells after a long period of expression, it affects the brain development if uses in all the neurons and it modifies neuronal activity [110]. On the other hand, the viral injections are a bit more difficult to do for the experimentalist, but it creates large GCaMP6 expression sites where recordings have a nice signal to noise ratio. The cortex of mice is organized into layers with specific cell types, and as layer 1 is not very dense in terms of cell bodies, the recording depth is at least 15μm below the pia in order to measure the responses of large populations of neurons from the second and third layers of the cortex. Two-photon microscopy is a commonly used microscopy technique for calcium imaging

because it allows deep recordings in living tissues (up to 600μm) within a precise focal plane. The high precision in the focal plane is essential to avoid the fluorescence contamination from cells that are not in the focal plan, and it is the reason this microscopy technique has been widely used in many cortical areas. The main limitation of the technique is that the sampling rate is limited because images are recorded pixel by pixel. In this PhD thesis, the experiments were performed with a 31Hz frame rate, using the GCaMP6s ("s" stand for slow) version of the calcium indicator. The fluorescence events produced by this indicator decay as an exponential with time constant of about 2 seconds [108], which is long compared to the sampling rate of the microscope used in the study. Therefore, the number of points to describe a calcium event is large and the signal to noise is remains high. It was reported that calcium imaging spike-to-calcium ratio varies a lot and can be non-linear which makes it very difficult to estimate spikes for a single neuron [111, 108, 112, 113]. As in Seidermann et al [114] we assumed that the non-linearities of the GCaMP average out over the heterogeneous population responses in the auditory cortex, so it was important to have a very large number of neurons to reduce as much as possible this effect. We also controlled for these non-linearities when necessary [44].

Reducing dimensionality of large datasets with hierarchical clustering

The calcium imaging recordings generate a large amount of data that are sometimes too big to be analyzed, therefore we apply some dimensionality reduction to simplify the data size. The method to reduce dimensions should be used with care in order to preserve the non-linearities and to make as few assumptions as possible. There are several methods to reduce dimensions that are widely used in neuroscience such as principal component analysis (PCA), non-negative matrix factorization (NMF), or clustering. PCA and NMF are two methods that reduce a high-dimensional dataset into its principle axes of variation. For instance, if PCA was used to reduce the activity of two neurons into a single vector: imagine a neuron A that is excited every time neuron B is inhibited and vice versa. The PCA algorithm will find the axis that encodes the interaction between these two neurons. Low values on this axis will mean that neuron A is inhibited, and neuron B is firing, and high values will mean that neuron A is excited and neuron B is inhibited. While the PCA works best with positive and negative values, the NMF algorithm is optimized for positive values, which is very convenient for neuroscience since neurons cannot have negative firing rates. Unfortunately, these two methods are reducing the dimensions to a smaller set of dimensions with the constraint that the new dimensions should be orthogonal to each other, which is not necessarily the case in neuroscientific applications. PCA and NMF extract the more relevant stimuli representations from a population of neurons considering that these representations are orthogonal in the space of the neuron ensemble. However it is possible that the representations are not orthogonal, leading to a wrong interpretation of the representations.

Clustering is a technique based on the redundancy in the data to group together neurons with the same activity pattern. For instance, two neurons that are activated for a sound A and a sound B will be put in the same cluster, away from another cluster that contains neurons responsive for sound A but not for sound B. This technique takes advantage of the large redundancy in the response patterns in the auditory cortex to group neurons as functional populations. Clustering is also a way of isolating response patterns from a large diversity of cells responses. It preserves the response profiles, even if they are non-linear, and it is also a very elegant way of removing cells with no particular activity for the sounds tested because they are all grouped together in a single cluster. For these reasons, this is the method that is preferentially used for this PhD thesis.

Testing perceptual representations with behavioral tasks

Perception can be accessed in the mouse model with behavior tasks. These tasks are often associated with a reward, like a drop of water in water-deprived mice[115, 116], or a punishment, like an electric chock [117]. The most classical auditory task to test perception in mice is the detection task in which the animal needs to lick for a drop of water when a sound is presented [116, 118]. In mice, the task can be in freely moving conditions [115, 119] or with head fixation to facilitate physiological recordings and simplified animal monitoring [120]. A more complex version of this task, called the Go/NoGo adds another sound for which the animal should abstain from licking [115, 121]. If it licks, the animal gets punished by a puff of air, a loud white noise sound, or simply a time out. In this task, the water-deprived mice are first trained to lick for a drop of water in response to a sound (S+), and then, once they reach a good level of performance for this detection task, they are trained not to lick for the other sound (S-). The Go/NoGo task can be used as a readout to compare the perception of the two sounds. However, the results from these experiments are always difficult to interpret because the mouse is providing perceptual feedback on a single licking tube, creating ambiguities. For instance, is the mouse is not attentive to the task, it will not lick and this will be interpreted as it perceived a S- sound, and vice versa, if the mouse is too motivated it will lick for every stimulus. Another method to study the perception of two sounds is the two-alternative forced-choice paradigm where a water deprived mouse needs to choice the correct licking tube over two licking tubes based on the sound [122, 123]. This method correct for the attentional bias because the mice motivation is decorrelated from its choice. It is important to note that it is possible to avoid the motivation biais with a Go/NoGo task as in Bathellier et al.[115], if the task is split into two subsets of mice, one doing the correct task and the other doing the opposite task (the sound S+ for a subset is the S- sound of the other and vice versa). Using this method, it is possible to obtain unbiased categorization curves with the Go/NoGo behavioral task.

Context Chapter 1

This first study was published in Nature Communications in 2016[44] and it aims at understanding the consequences of non-linear processing of sound intensity in the auditory cortex. Starting from the observation that sounds ramping up in intensity (up-ramps) are perceived louder than sounds ramping down (down-ramps) in human, we took advantage of the mouse model to explore the neuronal mechanisms of this perceptual difference. We recorded cortical activity of the awake mice with calcium imaging to understand how these sounds are encoded in the auditory cortex. We observed that up-ramps trigger more activity than down-ramps and we identified the cortical subpopulations that encode these two sounds. We concluded that the brain use non-linear features to process these sounds and we created a multilayer model based on these features to reproduce the difference in cortical recruitment.

My contribution in this study was to analyze the cortical recordings which include the quantification of the cortical asymmetry, the implementation of the correlation matrices, and the hierarchical clustering. In addition to the data analysis I also tested the performance of the linear and the adaptation models, and I created the first version of the multilayer model. I also participated to the generation of the figures (1.1, 1.2, 1.3, 1.4, 1.6, 1.7, 1.8).

Chapter 1

Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities

Published in Nature Communication in 2016

Thomas Deneux[1,3], Alexandre Kempf[1,3], Aurelie Daret[1], Emmanuel Ponsot[2] and Brice Bathellier[1,4]

1. Unité de Neuroscience, Information et Complexité (UNIC), Centre National de la Recherche Scientifique, FRE 3693, Gif-sur-Yvette, 91198, France
2. Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), Centre National de la Recherche Scientifique, UMR 9912, Paris, 75004, France
3. Co-first authors

Abstract

Sound recognition relies not only on spectral but also on temporal cues, as demonstrated by the profound perceptual impact of the time-reversal of common sounds. To address the coding principles underlying such auditory asymmetries, we recorded a large sample of auditory cortex neurons using two-photon calcium imaging in awake mice while playing sounds ramping up or down in intensity. We observed clear asymmetries in cortical population responses, including stronger cortical activity for up-ramping sounds, which matches perceptual saliency assessments in mice and previous measures in humans. Analysis of cortical activity patterns revealed that auditory cortex implements a map of yet undescribed, spatially clustered neuronal ensembles detecting specific combinations of spectral and intensity modulation features. Comparing different models, we show that cortical responses must result from multilayered nonlinearities, which, unlike standard receptive field models of auditory cortex function, build divergent representations of sounds with similar spectral content but different temporal structure.

Introduction

Since the work of von Helmholtz[124], it is well recognized that sound perception involves frequency decomposition of the acoustic waves by the auditory system. The frequency spectrum is, however, not the only characteristic that influences perception and identification of sounds. Psychophysical experiments in audition have shown that temporal features, i.e. the sequence of intensity and frequency variations, are also crucial determinants of perception, not only for sound localization but also for identification[125, 126, 127, 128, 129, 130]. For example, the recognition of musical instruments by humans strongly depends on the time-intensity profile of the tones and is strongly impaired by time-reversal of the waveform[130, 131]. Even for simple percepts, such as loudness, temporal features play an important role[132]. Numerous psychophysical experiments have shown that sounds whose intensity is ramping up with time (up-ramps) are globally perceived as louder or changing more in loudness than their time-symmetric opposites (down-ramps). This perceptual asymmetry has been observed for a wide range of sound durations[133, 134, 135, 136] and is proposed to emphasize approaching sound sources relative to sources moving away[133] in order to favor threat detection. The physiological bases of this perceptual asymmetry are yet unknown, but several studies found activity correlates in later stages of the auditory system. In humans, fMRI studies have shown that up-ramps produce stronger BOLD signals than down-ramps already in the auditory cortex[137, 138, 139]. Similarly in monkeys, LFP and multi-unit recordings in auditory cortex have demonstrated a positive bias for up-ramps in the global cortical response, consistent with behavioral asymmetries[140, 141]. Recent single neuron recordings in cat auditory cortex have suggested the existence of a positive bias for up-ramps beyond primates, although this study was only focused on very short ramps and found a bias only for the duration of cortical responses[45].

While all these results suggest a coding asymmetry between up- and down-ramps, the representation principles of intensity modulated sounds in auditory cortex and the computational underpinnings of asymmetric responses to sounds are still largely unknown, despite their pivotal relevance to the understanding of natural sound perception. Moreover, it is unknown whether asymmetric perception of intensity modulated sounds is a shared property of the mammalian auditory system and could be studied with the powerful tools available for a simpler animal model such as the mouse. In this report, we combined two-photon calcium imaging experiments and behavioral assays to show that the positive bias for up-ramping sounds as compared to down-ramping sounds is also present in mice at both the cortical and perceptual level, indicating a remarkably general property of the mammalian auditory system. We demonstrate that this bias is the result of profound nonlinearities which go beyond sensory adaptation mechanisms. By analyzing the response properties of a large sample of supragranular cortical neurons, we show that the temporal modulations of sounds are encoded by spatially clustered ensembles of neurons that detect specific features about

the time-course and amplitude of the modulations. Using modeling, we also show that the mechanism of the observed perceptual asymmetry is probably rooted in the sequence of nonlinearities implemented in the multilayered architecture of the auditory system to extract divergent, high-level representations of intensity modulated sounds.

Results

Mean response asymmetry between up- and down-ramps in mouse auditory cortex.

We first asked whether, like in primates, mouse auditory cortex is more strongly driven by up-ramps than by the symmetric down-ramps although the two signals have equal cumulative physical intensity. To answer this question, we performed two-photon calcium imaging in large populations of supragranular neurons of the auditory cortex (imaging depth from 110 μm to 230 μm) expressing GCAMP6s through stereotaxic injection of an AAV-syn-GCAMP6s vector[108]. Mice were awake and held head-fixed under the microscope thanks to a chronic cranial window and head-post implantation (Figure 1.1a). This preparation allowed imaging multiple fields of view (550 x 540 μm) in the same animal across several days (note that a different neuronal population was imaged in each session). One or two horizontal locations at one to three vertical positions were sampled in five mice and horizontally remapped (translation and rotation) with respect to each other using blood vessel patterns. Moreover, a gross identification of auditory cortex subfields[142] was also obtained based on intrinsic imaging maps as previously reported[115] (Supplementary Figure 1.8). With this approach, we could verify that, across 15 imaging sessions in five mice, we densely sampled core subfields of auditory cortex including A1 (about 40% of the neurons, mice 2, 3 and 4) and the anterior auditory field (AAF, about 45% of the neurons, mice 1, 2 and 5), while we estimate the fraction of neurons from the belt regions (imaged at the ventral or dorsal borders of core fields) to be about 15% of the neurons. In total, we imaged 4088 auditory cortex neurons at a rate of 31.5 frames per second using a resonant scanner. Stimuli included a randomized presentation of white noise and 8 kHz harmonic sounds with durations ranging from 100 to 2000 ms and ramping up or down in intensity. The calcium signals (Figure 1.1 b) were corrected for neuropil contamination (Supplementary Figure1.9) and temporally deconvolved to more closely track the actual firing rate variations in each identified neuron[115, 143] than can be followed with raw calcium signals (Supplementary Figure 1.10, but note that deconvolved signals probably still contain residual time shifts on the order of tens of ms due to the slow rise time of GCAMP6s).

Averaging the estimated activity of all recorded neurons, we observed that population responses to up-ramps were in many cases larger than for the symmetric down-ramp (Figure 1.1c-f). This was particularly evident for the

longer white noise ramps (60 to 85 dB SPL) for which the activity was at almost all times larger for the up-ramp than for the down-ramp (Figure 1.1c), and this same trend was also clear for the 8kHz harmonic sound (Figure 1.1d). Typical responses to the ramps included onset and offset response peaks, which were merged into a single peak for the shortest ramps (Figure 1.1c-f, Supplementary Figure 1.8). Interestingly, the onset responses were attenuated with increasing ramp duration, probably due to superposition of on- and offset responses in shorter ramps but also potentially due to some inhibitory process (Figure 1.1f). To quantify the asymmetry over the entire time-course of the response, we measured the difference of response integrals for up- and down-ramps (Figure 1.1g,h). This difference was systematically positive and could be as high as 80% $\pm$ 22% (mean $\pm$ SEM, n = 15 populations) of the average ramp integral for 2s white noise ramps (60-85dB) (Figure 1.1g,h). Individual statistical analysis with correction for multiple testing (see Figure 1.1g,h), showed significant asymmetry for most ramp parameters although shorter ramps and 8kHz harmonic sounds displayed weaker asymmetry which did not reach significance thresholds (Figure 1.1g,h). Hence, asymmetries in the global cortical response between up- and down-ramps are clearly present in mouse auditory cortex as observed in other animal species such as monkeys[140, 141]. Also, the direction of the asymmetry is similar to the one observed in human sound level perception assays, with the difference that the perceptual asymmetry in humans is stronger for harmonic sounds than for white noise[133, 144, 135].

Linear and adaption models do not explain response asymmetry.

These observations raise the question whether current models of auditory cortex sound encoding can readily explain up- versus down-ramp asymmetries. The responses of auditory cortical neurons are often modeled as linear filters of the sound input, also called spectro-temporal receptive fields (STRFs)[145, 51, 48], which in fact correspond to a 2-dimensional linear filter acting on the sound spectrogram. In this study, we did not characterize the STRFs because we could show mathematically that the response of any STRF filter to our intensity ramps is equivalent to the convolution of the ramp envelope with a linear temporal filter (Supplementary Note 1).
Moreover, we demonstrated analytically in the Supplementary Note 1, that, as a general rule for any linear filter (or sum of linear filters), the integral of the output is independent of whether the input signal is played forward or backward. Hence, any linear filter model including STRF models, by construction, predicts equal response integrals for up- and down-ramps. The theorem is also true if a non-linear scaling function (e.g. logarithmic intensity scaling) is applied to the signal before passing it through the linear filter (Supplementary Note 1, Figure 1.2a). These analytical results can be illustrated by showing the responses (see Methods) to up- and down-ramps of a temporal linear filter optimized to fit the observed cortical data. Despite optimization, the agreement with measured responses is very poor (Figure 1.2

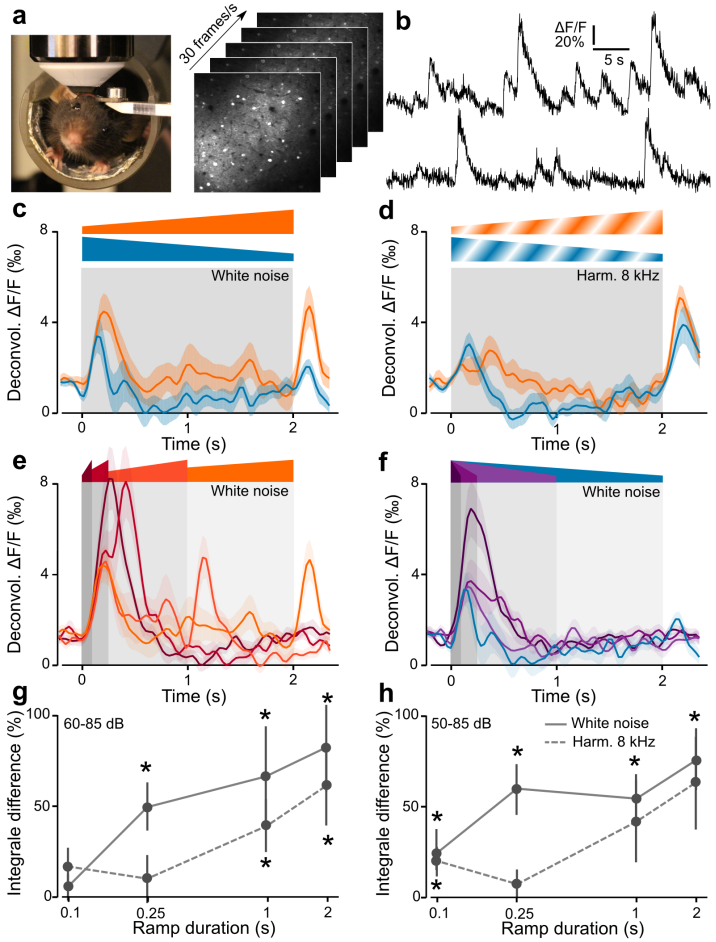

FIGURE 1.1: **Asymmetry of responses to intensity ramps in mouse auditory cortex.**

*(**a**) Awake head-fixed mouse under the two-photon microscope and an example of a recorded image time series of GCAMP6s labelled neurons in cortical layer 2/3 of the mouse auditory cortex. (**b**) Examples of raw GCAMP6s signals for one neuron (sampling rate: 31.5 Hz). Scale bars, vertical 20% $\Delta F/F$, horizontal 5 s. (**c**) Mean deconvolved calcium signals (that is, estimated firing rate) for 2 s white noise up- and down-ramps (range 60-85 dB SPL, shading indicates s.e.m. across n=15 imaging sessions). (**d**) Same as **c** for 2 s 8 kHz harmonic sound ramps (n=13 imaging sessions). (**e**) Responses to white noise up-ramps of 100 ms, 250 ms, 1 and 2 s. (**f**) Same as e for down-ramps. (**g**,**h**) Differences of the integrals of response signals between up and down-ramps (for example, integral of the difference of the two mean signals shown in **c**). The differences are normalized by the down-ramp integral. Error bars, s.e.m. When assessed globally (pooling durations together), the integral differences for each intensity range and spectral content was very significantly positive (Wilcoxon signed-rank test, white noise 60-85 dB: $P = 2.10^{-5}$, 50-85 dB: $P = 7.10^{-9}$ n=60 measurements; 8 kHz 60-85 dB: $P = 1.10^{-3}$, 50-85 dB: $P = 2.10^{-3}$, n=52 measurements). Statistical significance for individual stimuli is assessed across imaging sessions (white noise: n=15, 8 kHz: n=13) using the single-sided Wilcoxon rank-sum test and a Benjamini-Hochberg correction for multiple testing applied to the 16 tests (* P<0.05).*

b). This observation corroborates earlier demonstrations that STRFs fail in general to explain single neuron responses to a wide range of sounds[58, 61, 62]. We prove here, in addition, that STRFs are already inaccurate for simple intensity modulations, even at the population level.
Auditory cortical responses are known to show strong adaptation[146] and it was proposed that the STRF model can be combined with a synaptic adaptation model to better fit cortical responses[62]. However, we analytically showed that the adaptation + STRF model also preserves the equality of input integrals (see Supplementary Note 1), so that even the best fit of an adaptation model followed by a linear filter cannot explain the population data (Figure 1.2c,d). Therefore, the clear discrepancy between the data and the equality of response integrals predicted by simple or extended STRF models (Figure 1.2e) shows that the observed asymmetry between up- and down-ramps is the result of possibly unidentified nonlinearities at play in the auditory system.

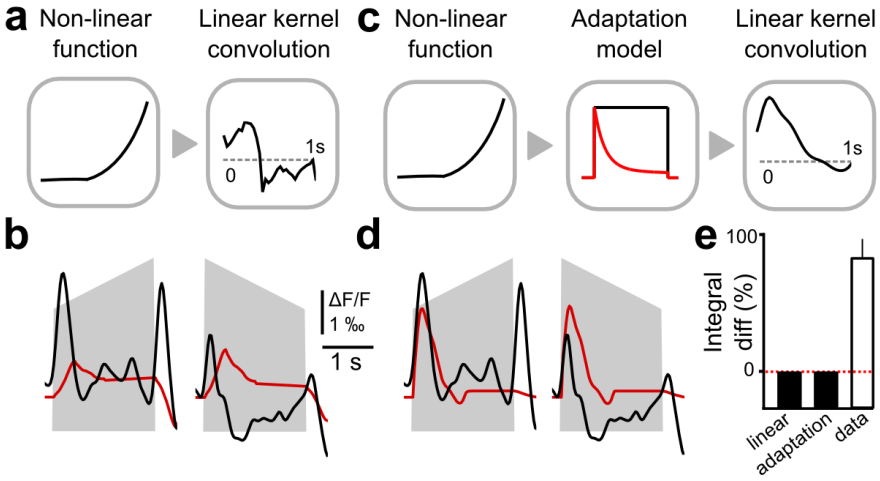

FIGURE 1.2: **Cortical response asymmetry is a non-linear effect.**

*(**a**) Sketch of the linear filter model. The input signal is scaled by a nonlinear function (left) and then goes through a linear kernel (right) to obtain the neuronal response. (**b**) Best fit by the linear model of the population responses to the 2 s white noise up- and down-ramps. (**c**) Sketch of the adaption model. The input signal is scaled by a nonlinear function (left), then undergoes adaptation (middle) and finally passes through a linear kernel (right). (**d**) Best fit by the adaptation model of the population responses to the 2s white noise up- and down-ramps. (**e**) Integral differences between up- and down-ramps for the linear and adaptation models for any choice of parameters and any ramp waveform (analytical result) versus experimental integral differences for the 2s white noise ramps.*

Asymmetric sequences of population patterns during up- and down-ramps.

To better understand the origin of the observed coding asymmetry, we analyzed the sequence of global cortical population activity patterns produced during up- and down-ramp presentations. To do so, we defined a measure of similarity between population activity patterns elicited at different time points of the stimulus presentation (Figure 1.3a, see Methods). To evaluate the similarity of population responses during the up- and down-ramps, we plotted similarity matrices for all time points of the sound presentations, including similarity across both ramps (Figure 1.3b,c). This allowed us to identify four different types of population patterns for each sound quality tested (white noise, Figure 1.3b; 8 kHz, Figure 1.3c). These included a response typical of the up-ramp onset, which was identical to the onset response to a 250ms constant intensity sound played at the ramp start level (filled green arrowheads). We termed this response type as "quiet ON". A reproducible response pattern was also seen at the up-ramp offset, which was almost identical to the offset response to a 250 ms constant intensity sound played at the ramp end level (empty magenta arrowheads). We termed this response type as "loud OFF". For the down-ramp, we observed a "loud ON" response pattern immediately after onset (empty green arrowheads), and a "quiet OFF" response pattern immediately after offset (filled magenta arrowheads), which was more evident for the 8kHz sounds. Although some residual similarity was observed between the "loud ON" and "quiet ON" response types, these patterns corresponded to a specific encoding of multiple sound features, including not only the direction (ON vs OFF), but also the intensity (quiet vs loud) of fast variations. Beside these salient responses to transients, specific but more variable activity patterns were produced during the slow ramping phase of the stimuli, most visibly for the white noise up-ramp (Figure 1.3b). This complex time-intensity code leads to very asymmetric response sequences for longer up- and down-ramps as seen in the matrices comparing up- and down-ramp responses (Figure 1.3b,c). Importantly, along with time-intensity coding, we also observed (as expected) sound quality coding, as the four ON and OFF response patterns for white noise were distinct from the ON and OFF response patterns for 8kHz sounds despite some similarities for loud ON and OFF patterns (Figure 1.3d, e.g. white arrowheads). Altogether, this analysis showed that the population encoding of multiple sound parameters (in particular the level and direction of intensity modulations) leads to an asymmetric cortical representation of up-ramping and down-ramping sounds.

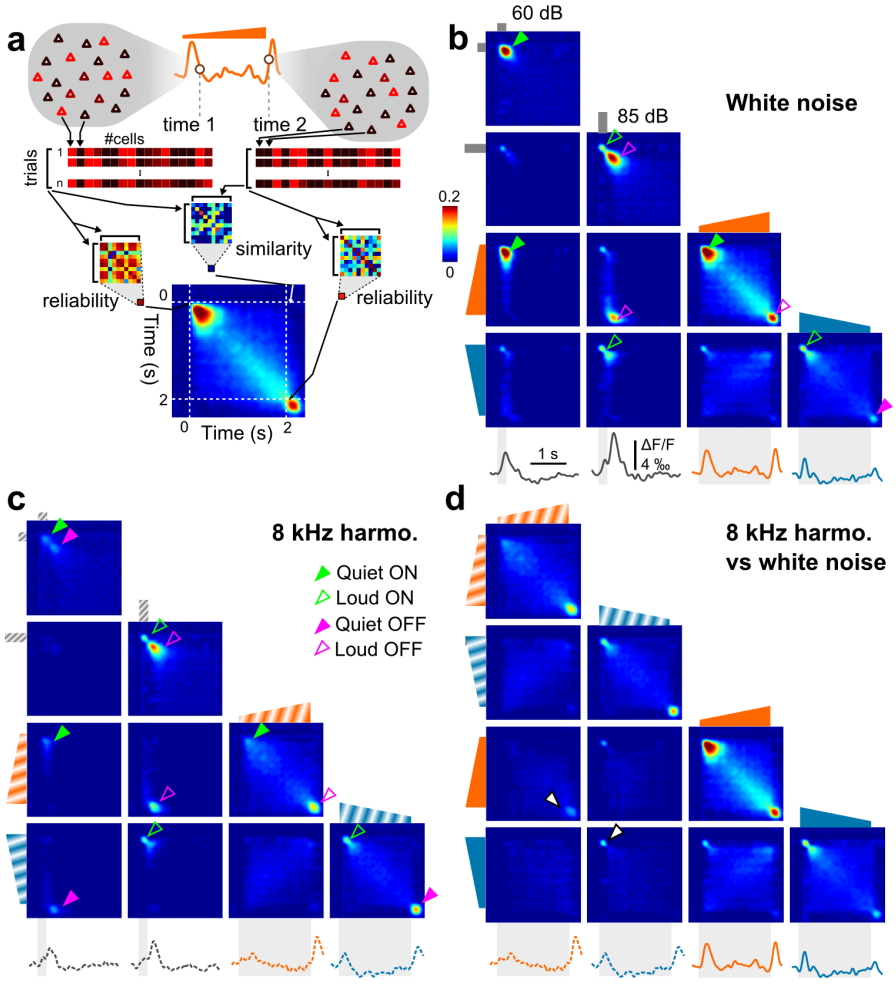

FIGURE 1.3: **Cortical population dynamics during up- and down-ramps.**

*(**a**) Schematic of the population vector similarity measure (see text). (**b**) Population similarity matrix across time bins and stimuli for four white noise sounds (60 dB and 85dB 0.25s duration and 60-85 dB 2s duration up- and don-ramps). Underneath the population firing rate waveforms are shown. The arrowheads on the diagonal indicate distinct activity patterns identified as "Quiet ON" (filled green), "Loud ON" (empty green), "Quiet OFF" (filled magenta) and "Loud OFF" (empty magenta). Arrowheads off the diagonal indicate strong similarities between different responses (e.g. empty magenta arrowhead indicates similarity between "Loud OFF" activity patterns observed after the 85dB 250ms sound and after the 60-85dB up-ramp). (**c**) Same as **b**. for the harmonic 8kHz tone. (**d**) Same as **b** and **c**. but the response to the white noise and 8kHz ramps are compared.*

Spatially organized multi-feature coding of intensity modulations.

To understand the functional properties of single neurons underlying the observed population code, we first aimed to determine the main types of individual responses present in the dataset and their distributions. We performed a hierarchical clustering of significant single neuron responses, using

the similarity of temporal response profiles across neurons as a metric (see Methods). First, about 63% of the neurons were classified as weakly or non-responsive with response profiles non-discriminable from noise, in line with the reported sparseness of auditory cortex response in awake rodents[147]. For the remaining 1341 neurons, we obtained 13 clusters displaying different average response profiles (Figure 1.4a,b). But note that, here, we used clustering mainly to organize the dataset and not to identify fully distinct clusters: while some clusters were clearly separated from each other, some others represented variations of one another in a continuum. Clusters were first distinguished by their selectivity to sound quality. Although several clusters responded both to white noise and 8 kHz sounds, seven of them showed preference (stronger responses) for white noise and six for 8 kHz sounds. Another important difference between clusters was their sensitivity to particular intensity modulation features. Eight clusters (70% of the clustered population) seemed to respond to a single precise feature of sounds. These included "loud" offsets as characterized by "OFF" responses to up-ramps but not down-ramps and to loud but not quiet constant sounds as observed in three clusters (Figure 1.4b, Loud OFF, 30% of the cells). We also observed two clusters of cells responding to "quiet" onsets as characterized by "ON" responses to up-ramps but not down-ramps and to quiet but not loud constant sounds (Figure 1.4b, Quiet ON, 18% of the cells) and a small population was found responding to "quiet" offsets (Figure 1.4b, Quiet OFF, 4% of the cells). In addition, 18% of all clustered neurons (two clusters) responded in a tonic manner to the loud (or intermediate loud) part of long ramps (Tonic). In contrast to these very specific clusters, we found five clusters signaling several intensity modulation features, including two clusters responding both to on- and offsets, (Figure 1.4b, ON + OFF, 15% the cells) and three clusters responding both to loud offsets and the central loud part of the ramp (Figure 1.4b, Loud OFF +Tonic, 15% of the cells). All these cluster subtypes were divided into one or two white noise or 8 kHz preferring cluster, except for Quiet OFF (8 kHz only) and Tonic (white noise only). Strikingly however, we did not find Loud ON clusters despite the presence of a specific Loud ON pattern at the population level (Figure 1.3). In fact, Loud ON patterns correspond to the response of ON+OFF neurons alone. Therefore, the four identified on- and offset population patterns as well as the population pattern observed during slow up-ramping (Figure 1.3) all reflected the combined activation of several neuronal types detecting different features of the intensity modulated sounds. Interestingly, loud and quiet offsets as well as quiet onset patterns contained cells very specific to the associated feature, while loud onset patterns were reflected by the activity of the less specific ON+OFF neurons. Importantly, almost all of these neuronal types showed asymmetric responses to up- versus down-ramps but only three clusters (239 out of 1341 neurons, quiet OFF and ON + OFF) preferred down-ramps for their preferred spectral signal (white noise or 8 kHz, Figure 1.4c). This sparser encoding of time-intensity features specific to down-ramps explains the response asymmetry at the population level.

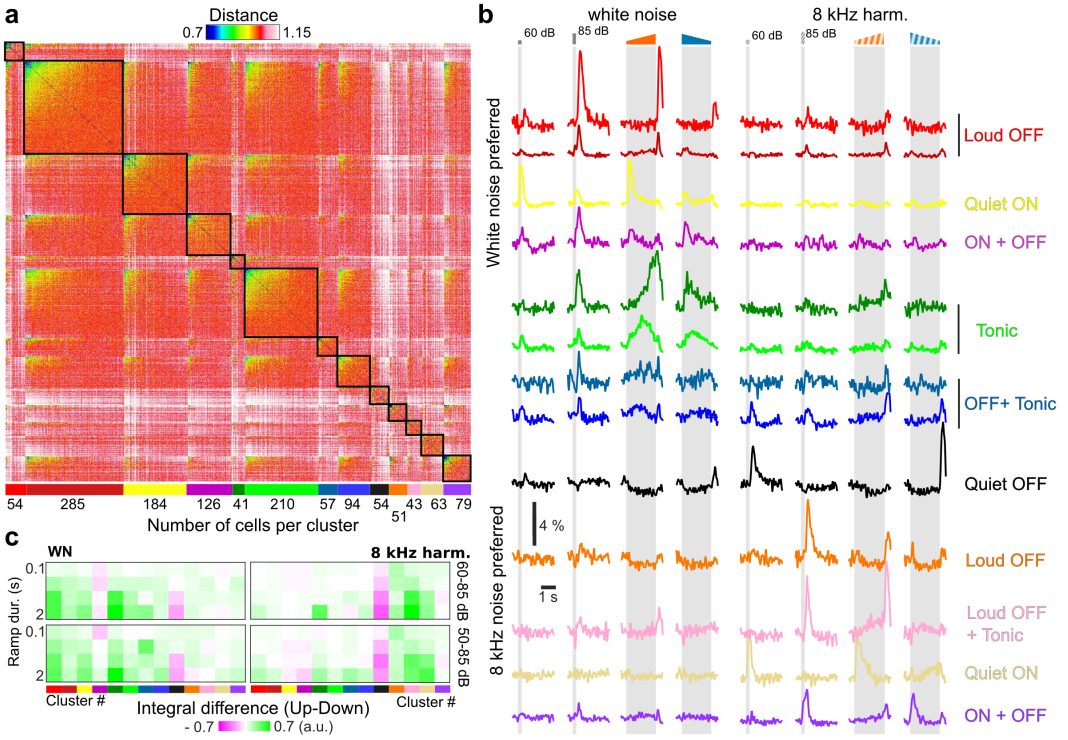

FIGURE 1.4: **Clustering of single neuron population responses.**

*(**a**) Distance matrix for the 1341 clustered neurons. The metric used is $d = 1 - cc$, where cc stands for the Pearson correlation coefficient between response traces. The identified clusters are delineated by a black square and labeled at the bottom of the matrix by a colored bar under which the number of cells in the cluster is indicated. Within each cluster the cells are sorted according to their mean distance with all other cells of the matrix. The gradient of distances within each cluster reflects the heterogeneity of the signal-to-noise ratio across cells. More reliable cells are on the left, less reliable cells on the right. (**b**) Mean response profiles of the twelve identified clusters to four white noise and 8kHz harmonic sounds (60 dB and 85dB 0.25s duration and 60-85 dB 2s duration up- and down-ramps). (**c**) Average absolute integral differences between up- and down-ramps for each cluster, ramp intensity ranges and durations.*

Next, we investigated the spatial organization of the clusters, by color-coding them in the imaging fields of view. We observed a relatively spread spatial distribution of the different cell types across imaging fields and mice (Figure 1.5a,b). However it clearly also appeared in some imaging fields that, despite some spatial intermingling, the clusters were unevenly distributed (e.g. mouse 1 Figure 1.5 a, Supplementary Figure 1.11a). In three out of five mice, we obtained multiple imaging sessions across different days in contiguous regions, either situated in nearby horizontal positions or at different cortical depths. In mouse 1 (Figure 1.5a) and 2 (Figure 1.5b) but less in the more sparsely responding region recorded in mouse 3 (Figure 1.5b), we also observed that the regions richer for one cluster were consistent across recording depths (e.g. 8kHz OFF + Tonic or Quiet OFF, Figure 1.5a) and were, in

some cases, horizontally continuous (white noise Loud OFF and Tonic clusters, Fig. 5a; Quiet ON, Supplementary Figure 1.11a). To quantify spatial clustering, we computed for each cluster a homogeneity index representing the average fraction of neighboring cells within a 30 μm radius that belonged to the same identified cluster (radius size effect are shown in Supplementary Figure 1.11) and compared it to maps in which cluster identity was randomly shuffled within each mouse (e.g. inset Figure 1.5c). For 12 out of 13 clusters, homogeneity was significantly higher than in shuffled maps ($p<0.05$, Benjamini-Hochberg correction for multiple testing, n=3 mice, Fig. 5c). Spatial clustering was very significant in mouse 1 and mouse 2 (Supplementary Figure1.11) but was just above the significance threshold in mouse 3 probably due to the sparser activity observed in this animal. Together, these analyses show that the coding of time-intensity features is spatially organized in the mouse auditory cortex.

Data-driven model of sound encoding nonlinearities.

To better understand what types of nonlinearity could be responsible for the asymmetric encoding of up- and down-ramps, we searched for models that could account for our observations, including asymmetric response integrals, on- and offset responses and specific sound intensity coding in certain neurons. The simpler non-linear models applied to both auditory[64, 148] and visual[149, 150, 151] systems combine a linear filter (receptive field) with a non-linear function that is expected to capture output nonlinearities such as the spike threshold. As a first guess, we tried to fit such a linear-nonlinear model (LN model, Figure 1.6a) to all white noise responses of the 13 neuronal clusters. Because the data shows clear intensity tuning split into groups of cells that either respond to low or high amplitude changes, we first assumed that the sound envelope is encoded through a "quiet" and a "loud" channel modeled with two different nonlinear scaling functions applied to the input intensity (Figure 1.6a) whose parameters were optimized for each tested model. The modeled response of each cluster was thus the sum of two linear filters applied to each of these channels followed by a nonlinearity (Figure 1.6a, LN model, see Methods). Using this approach, the best fit on our training stimulus set (see Methods) left unexplained 44.2% of the total variance of the responses to the test stimulus set (Figure 1.6c). More importantly, the LN model was unable to reproduce the magnitude of the asymmetry between up- and down-ramp responses (Figure 1.6d). We could thus conclude that the structure of LN-type models does not reflect the computations underlying the observed cortical responses. Importantly, this implies that neither intensity tuning nor output nonlinearities are sufficient to explain the observed cortical asymmetries.

The main reason for the failure of LN-type models is their inability to account for encoding the combination of certain features as observed in a large number of recorded neurons. LN-type models, for example, fail to explain the responsiveness of visual cortex complex cells to ON and OFF oriented edges, a property better modeled by the summation of at least two inputs

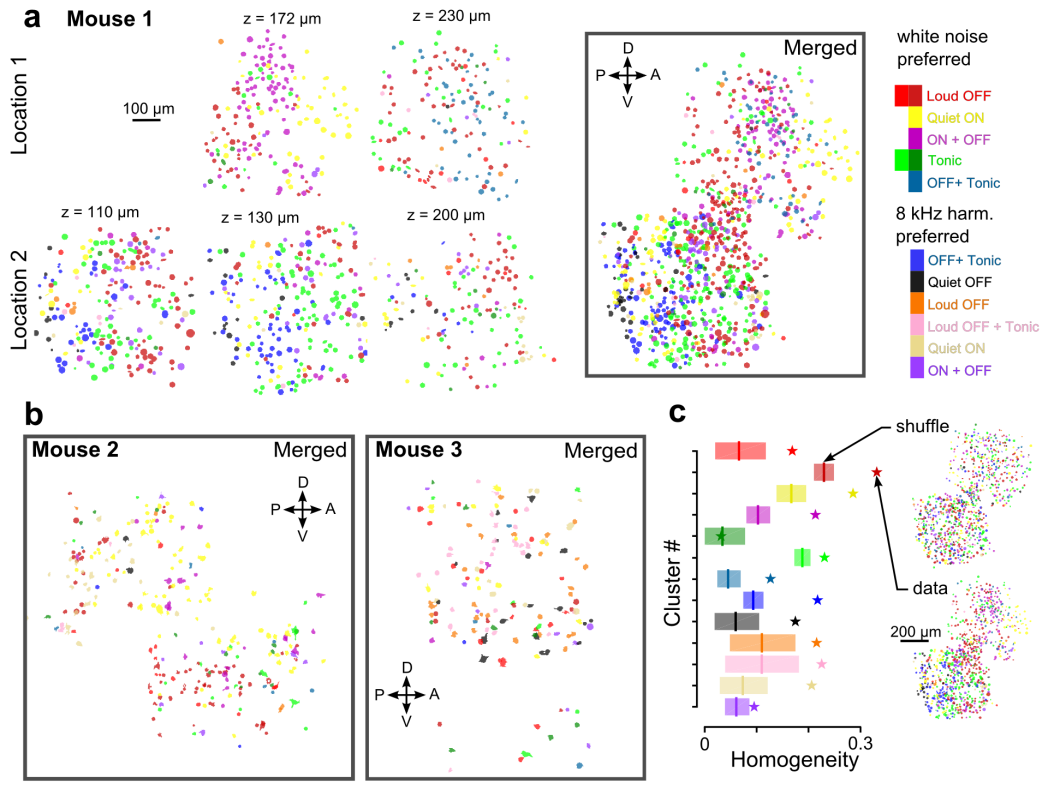

FIGURE 1.5: **Functional cell types correspond to cell assemblies clustered in space.**

*(**a**) Localizations of the cells belonging to the different identified clusters, color-coded as in Figure 1.4 (see color bar on the right), in five imaging sessions performed at two different horizontal localizations and different depths (z) across several days in mouse 1. On the right, the relative localization of all cells is shown in a horizontally-mapped z-projection. (**b**) Horizontally-mapped z-projection for mouse 2 and mouse 3 (4 imaging sessions each, see Supplementary Figure 1.11). (**c**) Each star represents the value of the homogeneity index calculated across 3 mice for each of the 13 clusters (same color code as in **a**). The vertical lines represent the value expected if each cluster was homogeneously spread in space obtained by shuffling and the shaded area is the 95% confidence interval.*

(ON + OFF)[151, 152]. Based on this idea, we added to our LN model a "middle layer" of units computing simple linear features and followed by a rectifying nonlinear function ("multilayer" model, Figure 1.6b). The features included tonic responses (transmission of their input) or first order-derivatives (either positive for ON-units or negative for OFF-units). In this multilayer model, cortical responses were then modeled as the sum of fitted linear kernels applied to these six nonlinear units (Figure 1.6 b; see Methods and Supplementary Figure 1.12 for the fitted kernels). In comparison with the LN model, this architecture left unexplained 28.8% of the response variance in the test set when the threshold of the rectifying nonlinearity was set to zero, and only 23.1% when it was fitted to an optimal positive value (Figure 1.6c, >0), without any further output non-linearity. When quantifying

only from the six clusters preferring white noise which have a larger signal to noise-ratio, the unexplained variance even dropped to 20.4% (Figure 1.6c, >0) with a visually striking fit to the data (Figure 1.6e). The multilayer model also closely reproduced the asymmetry of population integrals (Figure 1.6d). Moreover, the kernels obtained with the multilayer model were smooth positive or negative transient functions with decay time constants of 100 to 200 ms and thus compatible with long, polysynaptic, post-synaptic potentials (Supplementary Figure 1.12). This supports the functional plausibility of the chosen architecture despite its oversimplification with respect to the underlying biological network.
Together, this analysis shows that more than one nonlinear processing layer is required to explain the multifeature code observed in auditory cortical neurons. Interestingly, the result of this complex transformation of the auditory inputs is an asymmetric, divergent representation of temporally symmetric sounds as can be seen when comparing the population trajectories for the data and the multilayer model with trajectory results from the one-layer LN model (Figure 1.6f). Interestingly, this computational scheme and the overrepresentation of particular features also allows differentially boosting the overall saliency of up-ramps based on their temporal profile.

Up-ramps are behaviorally more salient than down-ramps in mice.

Are the nonlinearities observed in cortical sound encoding actually reflected in the perceived saliency of up- and down-ramps in mice? To answer this question, we first used the general observation that more salient stimuli lead to faster associative conditioning[153, 154]. We hence trained two groups of water-deprived, head-fixed mice to lick after ramping sounds to get a water reward (Figure 1.7a). While on the first training day both groups of mice licked almost irrespective of the sound cue, after seven training days, lick probability increased during and after sound presentation, following a similar profile when either up- or down-ramps were used as a cue (Figure 1.7b). However, the ratio between lick rate after sound offset and before sound onset increased faster for the group cued with a 2 sec white noise up-ramp (60 to 85 dB SPL) than for the group cued with the symmetric down-ramp (Figure 1.7c). This suggests that, in the context of this task, the up-ramp is more salient than the down-ramp, which results in faster learning. Given the duration of the ramps used in this task, it is unclear whether the up-ramp is overall more salient or only its terminal high intensity part which most closely signals the reward (mice had to do at least one lick following sound offset to get a reward). To test whether the earlier part of the up-ramp is also more salient, as suggested by cortical imaging, we used another associative task. In this task, freely moving mice first rapidly learned to lick at a tube after an S+ ramp to get a reward. Then, after 4 days, a non-rewarded S- ramp was introduced and mice learned to avoid licking following this new sound (Figure 1.7d). We observed that, in this task, the response to the S- ramp spontaneously occurred close to sound onset (Figure 1.7e). Moreover,

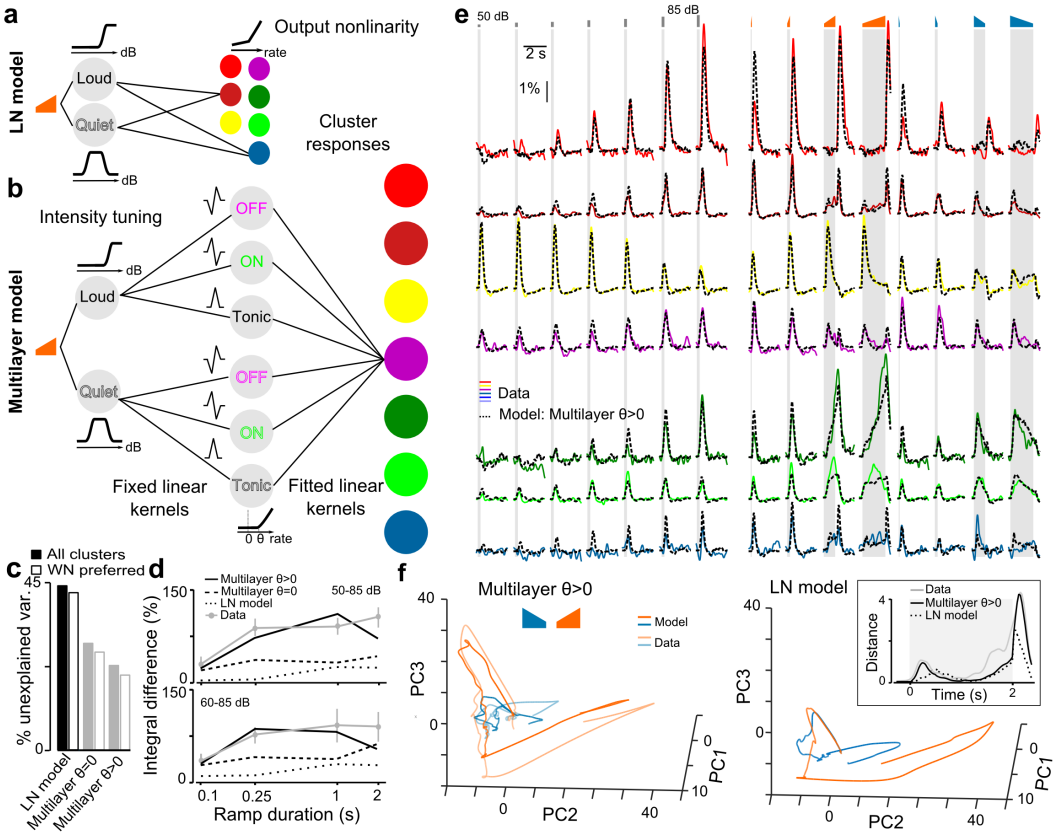

FIGURE 1.6: **Phenomenological model of intensity modulation coding.**

*(**a**) Linear-nonlinear (LN) model (linear filters + output non-linearity) applied to two different "intensity channels" obtained by nonlinear scaling of the input for intensity tuning. (**b**) The multilayer model with: (1) intensity channel as in **a**, (2) six fixed linear filters with a rectifying nonlinearity (threshold = θ), (3) linear sum of the feature detector outputs filtered by fitted kernels. (**c**) Fraction of unexplained variance for all clusters (filled bars) or for the seven clusters preferring white noise (empty bars) after fitting the LN model, and the full multilayer model with a fixed (multilayer $\theta = 0$) or a fitted (multilayer $\theta > 0$) rectifying threshold. (**d**) Normalized difference of the up- and down ramp responses for the clustered data (1341 neurons, n = 13 imaging session) and the different fitted models as in **b**. (**e**) Fit of the multilayer model ($\theta > 0$) to the responses of the six identified clusters which show preferred response to white noise (note that all twelve clusters where fitted by the model). Sounds are white noise: 250 ms constant (7 intensities) and 60-85dB up- and down-ramps. (**f**) Trajectories of the population responses to the 2 s white noise up- (orange) and down-ramps (blue) obtained for the fitted LN (right), the multilayer model and the data (left). The 13-dimensional data and model outputs are plotted in the space of the three first principal components of the data. The trajectories are more divergent for the multilayer model than for the LN model as corroborated by distance between the two trajectories at every time point (inset).*

because the S+ is already associated with licking, the association between S- and lick-avoidance is rate-limiting for the overall discrimination. Consequently, a learning speed analysis in this task makes it possible to assess

the respective perceptual saliency of ramp onsets when comparing groups avoiding the up- or the down-ramp. In accordance with the cortical population response, the rise of individual learning curves was significantly shorter (192 $\pm$ 28 trials, n=12 mice) when the 2 sec 60-85dB white noise up-ramp was the S- as compared to the opposite situation (310 $\pm$ 56 trials, n=12 mice, p=0.0046, Kolmogorov-Smirnov test, Figure 1.7g). Note that the individual learning curves for this task typically display a delay period during which performance stays at 50% (licking occurs both on S+ and S- sounds) followed by a learning phase in which mice start to avoid the S- sound[119] (Figure 1.7f) as observed in many learned behaviors[155]. Hence we measured rise duration on each individual curve from the end of the delay period to the end of the learning phase.
Taken together, this indicates that the 2 sec up-ramp is behaviorally more salient than the down-ramp both in its initial and terminal phase as predicted by our cortical activity measurements (Figure 1.1c). Hence, the asymmetric encoding of up- and down-ramps is reflected in the perceptual choices of the mouse.

Discussion

In this study, we demonstrate three important points. First, we show that to explain the asymmetric encoding of up- versus down-ramping sounds in auditory cortex, nonlinear processes are required (Figures 1.1 and 1.2). Several studies have already shown that auditory cortex is nonlinear through the limitations of linear models such as STRFs in approximating the responses of single neurons in the auditory cortex[58, 61]. However, neither the type nor the magnitude of the nonlinearities nor their consequences for perception are well characterized. With respect to magnitude and perceptual impact, our study provides a clear example that nonlinearities of the auditory system are so large that they produce population-scale differences in the cortical responses that correlate very well with the asymmetric perceptual saliency of up- and down-ramps observed both in mice (Figure 1.7) and humans[133]. With respect to the type of nonlinearities implemented by the auditory system, among the variety of models proposed in the past[62, 64, 148, 152], our study provides an important constraint. We show that simple nonlinear mechanisms operating at the output of a single layer such as adaptation or nonlinear output functions (LN models) can account neither for the combinations of different nonlinear features observed in the responses of single neurons, nor for the global asymmetry of cortical responses to up- and down-ramps.
Several studies have already shown the limits of LN-models, and have successfully proposed extensions including non-linear input scaling or frequency coding functions[156], adaptation[62] or gain control mechanisms[157]. Here we show that despite their importance, some of these nonlinearities (nonlinear input scaling, adaptation) are alone insufficient to explain up- versus down-ramp asymmetries (Figure 1.2). Instead, we propose that a sequence of nonlinearities embedded in multiple processing layers, as we exemplify in

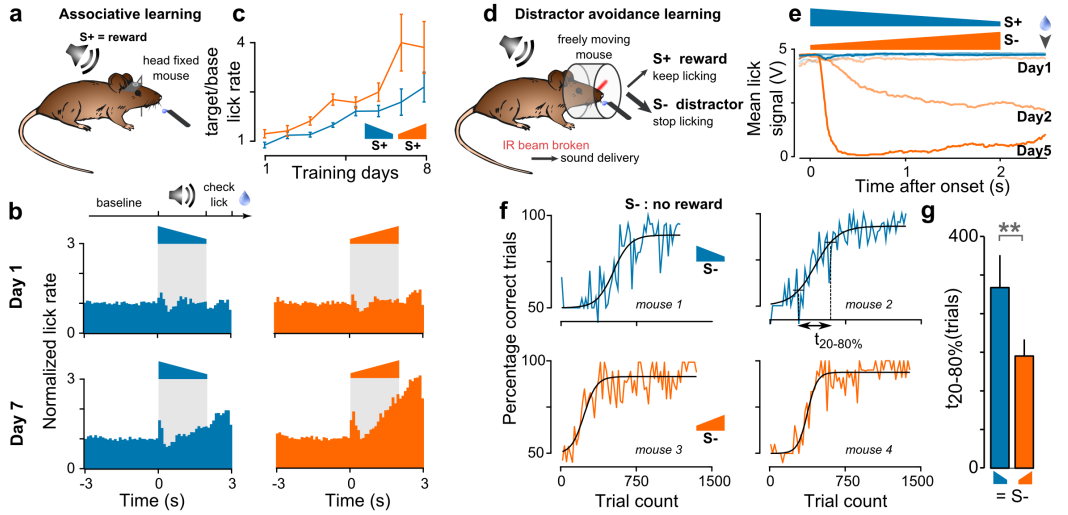

FIGURE 1.7: **Up-ramps are behaviorally more salient than down-ramps.**

*(**a**) Sketch of the head-fixed sound-reward association task. (**b**) Histograms of lick rates normalized to the baseline rate during the first and seventh days of training to the up- (right) and down-ramps (left). Average across all mice (n=6 per group). (**c**) Ratio of the post- and pre-stimulus lick rates over training days for the up- (blue) and down-ramps (orange) (mean ± SEM) showing increased sound-locked licking for up-ramps, Friedman test, $p=3.7 * 10^{-10}$, n=6 per group). (**d**) Schematic of the distractor avoidance learning task. Freely moving mice first learn to lick at a spout after an S+ sound to get a reward, then an S- sound is added and mice learn to stop licking after this sound. (**e**) Typical average IR beam break signal (5V = beam broken, 0V = beam intact) with respect to S+ and S- sound onsets for a mouse on the 1st, 2nd and 5th training days. (**f**) Examples of global performance learning curves (mean of S+ and S- performance) for the Go/NoGo distractor avoidance task. (**g**) Learning phase duration when either the down- (left) or the up-ramp (right) is the S- stimulus (mean ± SEM, n=12 per group, p=0.0046, Kolmogorov-Smirnov test). The learning phase duration if defined as the time necessary to go from 20% to 80% of maximum performance above chance level (i.e. above 50% correct) and is measured on the sigmoid fitted to the learning curve.*

Figure 1.6, is required to explain asymmetric coding of intensity modulated sounds as we observe in our data). This conclusion is further supported by the higher fitting accuracy, for various sets of sounds, obtained with other models simulating two layers[148] or expanding linear models to second order terms[64]. Indeed, these models try in essence to locally approximate complex multilayer architectures, potentially even more complex than the one we propose to account for our specific stimulus set.

While our conclusions constraint the generic type of computations underlying auditory processing, the biological implementations of such computations could be very diverse. Multilayer architectures computations could be, for example, implemented by a sequence of nonlinearities in the ascending auditory pathway through which increasingly complex features are produced. The observation of spatially segregated off-responses in the auditory

thalamus both in mice[158] and guinea pigs[159] is in support of this view. As these neurons could be a major contributor to cortical offset responses, it would be interesting to investigate if the dichotomy between quiet and loud offsets (as we observed for 8 kHz sounds in cortex) is also already present in the thalamus. But the multilayer structure of our model could in principle also reflect cellular nonlinearities (e.g. nonlinear dendritic integration can implement two processing layers[160]) or even recurrent connections within or across the networks of cells corresponding to the functional clusters we identified[161, 162].

The second point demonstrated in our study is that even stimuli as simple as up- and down-ramping sounds can reveal novel encoded features in the auditory system, challenging current models. In particular, we established that sound intensity modulations are encoded in the auditory cortex by multiple neuronal populations which respond to unexpectedly complex combinations of sound features that not only include frequency, but also direction of modulation (e.g. onset, offset) and sound level (Figure 1.4). A large fraction of cells code very precise feature combinations (e.g. offset of a loud white noise sound or onset of a more quiet sound, Figure 1.4), while larger groups of cells encode multiple feature combinations (e.g. offset and steady part of a loud sound, Figure 1.4). Previous reports have described tuning to sound intensity[148, 163] or to specific temporal features such as on- and off-responses[164] or the rate of click trains[165], but these were often described independently from other features[163] or under particular anesthesia[164], and could not be integrated in a general coding scheme. Here, we show that these features are actually combined with each other in the same neurons of the auditory cortex, potentially encoding higher level perceptual tokens. We also show that some specific feature combinations are favored, in particular the "quiet" onset and "loud" offset present in up-ramps (Quiet ON, Loud OFF and Loud OFF + Tonic, see Figure 1.4). This explains the observed asymmetry. It is noteworthy that, both in the data and the multilayer model, 2000 ms ramps globally generate larger asymmetry than 100 ms ramps (Figure 1.4 and Figure 1.6d) although the same on- and offsets are present for both durations. The reason for this discrepancy can be hypothesized based on the multilayer model simulations. In the multilayer model, the units detecting elementary on- and offsets (middle layer, Figure 1.6b) are thresholded (parameter θ). In longer ramps, because the central slope is shallow, its contribution to the activity is subthreshold when $\theta > 0$. In shorter ramps, the central slope is much steeper and this impacts the activity of on- and offset units which partially compensates for the asymmetry (note that when $\theta = 0$, there is no asymmetry discrepancy between shorter and longer ramps, Figure 1.6d).

As a striking side-result of our study, we also show that neurons encoding the same feature combination are non-randomly organized across the supragranular layer of the auditory cortex and tend to cluster spatially (Figure 1.5 and Supplementary Figure 1.11) forming a complex multifeature map. The structure of this map seems however more diffuse than expected for a purely

columnar architecture. This observation is in line with previous reports inferring[166] or describing[115] iso-functional intermingled subnetworks in mouse auditory cortex, as well as with reports indicating that strongly responsive, information-rich cells are sparse in auditory[147] and other cortical areas[167, 168]. Given the large number of feature combinations that could possibly exist but that we did not test in this study, further investigation will be required to reveal the exact organization of these subnetworks and their relationship to tonotopic organization. Nevertheless, our gross localization of imaging fields using intrinsic imaging indicates that these subnetworks exist within tonotopic fields (Supplementary Figure 1.8 and 1.11).
The third point of our study is that the multifeature code demonstrated in mouse auditory cortex and probably resulting from the multilayer nonlinear architecture of the auditory system allows the encoding of two distinct temporal modulations of the stimulus with divergent activity patterns, unlike models with a single output nonlinearity. This fact is evident when comparing the population activity trajectories produced by the fitted LN and nonlinear feature models respectively (Figure 1.6f). It suggests that the purpose of the nonlinearities implemented throughout the auditory system is to produce easily separable representations of distinct temporal sound intensity modulations. Remarkably, neural networks using multiple layers endowed with linearly filtered inputs and a rectifying nonlinearity are capable of generating complex representational features permitting impressive speech or object recognition performance55. It seems that such networks implement symmetry-breaking principles that are important for perception. Our results bring novel evidence suggesting that these principles are also at the source of the strikingly distinctive percepts generated by sounds bearing identical spectral content but different temporal dynamics2,3.

Methods

Animals.

All mice used for imaging and behavior were 6 to 16 weeks old male C57Bl6J mice. All animal procedures were approved by the French Ethical Committee (authorization 00275.01).

Two-photon calcium imaging in awake mice.

At least three weeks before imaging, mice were anaesthetized under ketamine medetomidine. The right masseter was removed and a large craniotomy (5 mm diameter) was performed above the auditory cortex. We then performed three injections of 150 nL (30 nL/min), AAV1.Syn.GCaMP6s.WPRE virus obtained from Vector Core (Philadelphia, PA) and diluted 10x. The craniotomy was sealed with a glass window and a metal post was implanted using cyanolite glue followed by dental cement. Two days before imaging, mice were trained to stand still, head-fixed under the microscope for 10 to 20 min per day receiving small sucrose rewards. Then mice were imaged one to

two hours per day. Imaging was performed using a two-photon microscope (Femtonics, Budapest, Hungary) equipped with an 8kHz resonant scanner combined with a pulsed laser (MaiTai-DS, SpectraPhysics, Santa Clara, CA) tuned at 920 nm. Images were acquired at 31.5Hz during blocks of 42s during which randomly chosen sounds were presented with 2.5s intervals. Blocks were interleaved by an 18 s pause repeated until all sounds were played 20 times. White noise and harmonic (8 kHz + the 5 first odd harmonics with a $1/(2n+1)^2$ spectrum) sounds were played in 23 different intensity modulations, including seven constant sounds of 250 ms at (50, 55, 60, 65, 70, 80, 85 dB SPL), eight up- and down-ramps between 50 and 85dB SPL with four durations (0.1, 0.25, 1 and 2 s) and eight up- and down-ramps between 60 and 85dB SPL. In two imaging sessions (431 neurons), only white noise sounds were tested. All sounds were delivered at 192 kHz with a NI-PCI-6221 card (National Instrument) driven by Elphy (G. Sadoc, UNIC, France) through an amplifier and high frequency loudspeakers (SA1 and MF1-S, Tucker-Davis Technologies, Alachua, FL). Sounds were calibrated in intensity at the location of the mouse ear using a probe microphone (Bruel&Kjaer).

Intrinsic optical imaging recordings.

To localize the calcium imaging recordings with respect to the global functional organization of auditory cortex, we performed intrinsic optical imaging experiments under isoflurane anesthesia (1%). The brain and blood vessels were illuminated through the cranial window by a red (intrinsic signal: wavelength = 780 nm) or a green (blood vessel pattern: wavelength = 525 nm) LED. Reflected light was collected at 20Hz by a CCD camera attached to a macroscope. The focal plane was placed 400 µm below superficial blood vessels. A custom-made Matlab program controlled image acquisition and sound delivery. We acquired a baseline and a response image ($164x123$ pixels, $\sim 3.7x2.8$ mm, image shown in Supplementary Figure 1.8 are cropped around the sound responsive area) corresponding to the average images recorded 3s before and 3s after sound onset respectively. The change in light reflectance ($\Delta R/R$) was computed then averaged over the 20 trials for each sound frequency. A 2D Gaussian-filter (σ = 45.6 μm) was used to build the response map (Supplementary Figure 1.8). Sounds were trains of 20 white noise bursts or pure tone pips (4,8,16 and 32 kHz) separated by 20 ms smooth gaps.

Data analysis.

Data analysis and modeling was performed with custom-made Matlab and Python scripts available upon request, as well as the datasets. Only recordings performed within auditory cortex as assessed with intrinsic imaging were included in the imaging. Sample size (about 4000 neurons) was chosen to obtain a representative sampling of auditory cortex as assessed in a previous study[115]. All images acquired during a session were registered by horizontal translation to a template image to correct for motion artifacts

(all sessions with visible z-motion were discarded). Regions of interest were then manually selected on the whole cell bodies of visually identifiable neurons and the mean fluorescence signal $F(t)$ was extracted for each region. We also estimated the local neuropil signal $F_{np}(t)$ for each neuron. Briefly, (Supplementary Figure 1.9a,b) we computed "filled-in" neuropil signal frames $Y(t)$ by spatially smoothing every data frame $X(t)$ with a Gaussian spatial kernel g (σ = 170μm) after excluding the neuron's region-of-interests represented by a masked binary image M. This is done using the formula:

$$Y(t) = (X(t) \cdot M) \circledast g / (M \circledast g) \tag{1.1}$$

in which "$\cdot$"and "$/$" denote the element-wise multiplication and division and $\circledast$ is the spatial 2D-convolution. Then for each neuron, we computed the neuropil corrected fluorescence signal $F_c(t) = F(t) - 0.7 * F_{np}(t)$ where $F_{np}(t)$ is the mean value of $Y(t)$ in the neurons' region of interest. The 0.7 correction factor was chosen according to calibration made in another study[108] for GCAMP6s in mouse visual cortex but we could visually verify that for our data neuropil contamination was removed with very little artefact while neuron-specific responses were preserved (Supplementary Figure 1.9c). Baseline fluorescence F_0 was calculated as the minimum of a Gaussian-filtered trace over the 42 s imaging blocks and fluorescence variations were computed as $f(t) = (F_c(t) - F_{c0})/F_{c0}$. The approximate time course of the firing rate was estimated using temporal deconvolution as $r(t) = f'(t) + f(t)/\tau$ in which $f'(t)$ is the first derivative of $f(t)$ and $\tau = 2s$ as estimated from the decays of the GCAMP6s fluorescent transients[108]. This method efficiently corrects the strong discrepancy between fluorescence and firing time courses due to the slow decay of spike-triggered calcium rises as we show in simulations based on GCAMP6s kinetic parameters (Supplementary Figure 1.10). However, as our simulation also shows, it does not correct for the relatively slow rise time of GCAMP6s, producing a time delay on the order of 70 ms between peak firing rate and peak deconvolved signal (see Supplementary Figure 1.10). Note that deconvolution is a linear operation and thus cannot be the cause of asymmetric integrals observed for up and down ramps (see Supplementary Note 1). Weak nonlinearities have been observed for the conversion of action potentials into the GCAMP6s signals, most particularly superadditivity of calcium transient amplitudes. As we show that a linear model followed by a non-linear function cannot explain the observed data (Figure 1.6), we can rule out the involvement of GCAMP6s superaddivity in the asymmetry observed for up- and down-ramps. However, in the absence of extensive characterization of GCAMP6s in our imaging conditions, we cannot fully exclude that another uncharacterized nonlinearity of GCAMP6s participates in the asymmetry. The integrals of population and single cluster responses were computed between the time of sound onset and the time of sound offset + 500 ms. The normalized difference of integrals (I_{up} and I_{down}) between up- and down-ramps was computed as $2 * (I_{up} - I_{down})/(I_{up} + I_{down})$.

Analysis of population pattern similarity.

Population activity at time t of the i^{th} repetition of sound j was represented by a 4088-dimension vector containing the firing rates of all imaged neurons. The similarity between population responses to sound j at time t and to sound j' at time t' was computed as the mean correlation between all single trial vectors pairs

$$s((t,j);(t',j')) =< \rho(V^{i}_{t,j}; V^{i'}_{t',j'}) >_{i,i'} \tag{1.2}$$

where ρ is the Pearson correlation coefficient between two vectors. For $(t,j) = (t',j')$, this measure evaluates the reproducibility of the response patterns across trials (in this case the pairs $i = i'$ are excluded from the averaging). In the plotted similarity matrices, similarity was evaluated frame by frame after smoothing the signals with a Gaussian filter (σ = 60 ms). Note that because trial to trial variability of single neuron responses is very large in mouse auditory cortex[115], the correlation between single trial population vectors is generally low ($\sim$ 0.2). Thus the average cross-trial correlation (i.e. our similarity measure) takes generally low values. However, the idea of this framework is to compare the similarity measure to the self-reproducibility measure. In particular, two population patterns can be considered indistinguishable, given the observed single trial variability, when their similarity is as high as the individual reproducibility measures (diagonal of the similarity matrix).

Single cell clustering analysis.

Clustering was used to organize the imaged neuronal responses. Due to the large variability observed in many neurons, this analysis is not exhaustive but rather aims at identifying principal classes of responses within our dataset. Clustering was performed across the 13 imaging sessions in which we played both white noise and 8 kHz harmonic sounds (3657 neurons). Response traces for all sounds were smoothed using a Gaussian filter (σ = 31 ms). Before clustering, we selected significantly responsive (assessed by testing for a difference of the pooled responses to all stimuli against their baseline using a paired Wilcoxon signed rank test p<0.05) and selective neurons (significant modulation by one of the stimuli, Kruskal-Wallis test, p<0.05) neurons. For the 2343 neurons which passed both tests, the signal-to-noise ratio (SNR) was calculated as:

$$(\int dt \langle |r(t)| \rangle_{trials}) / \sqrt{\int dt \langle |(r(t) - \langle r(t) \rangle_{trials})| \rangle_{trials}} \tag{1.3}$$

We observed that the SNR distribution was long-tailed with a small fraction of cells responding with high SNRs. To base the clustering on the clearest signals, we first selected the 30% of the cells with largest SNR. Using the Euclidean distance on z-scored response traces (i.e. normalized by their

standard deviation) as a similarity metric across cells and the "furthest distance" as a measure of distance between clusters, we established a hierarchical clustering tree. The tree was thresholded to yield 50 different clusters. This method yielded a large number of small clusters, which after visual inspection appeared to contain noisy responses (hence very dissimilar to other clusters). We therefore excluded clusters containing less than 10 cells. Applying this criterion, we obtained 13 clusters. Non-clustered cells were then assigned to one of the 13 clusters with which they had the highest correlation (Pearson correlation coefficient) provided that this correlation was higher than 0.1. After this procedure, 1341 neurons were assigned to a cluster while 1002 cells were not assigned. Inspection of their responses showed that the latter were weakly or non-responsive cells.

Behavior.

We measured sound salience indirectly by measuring learning speed during associative conditioning. In the first task, water-deprived mice (33 μL/g/day) were head-fixed and held in a plastic tube on aluminum foil. A two-second white noise ramp sound (range 60-85 dB) was presented every 6 to 16 s (uniform distribution) followed by a 1 s test period during which the mouse had to produce at least one lick on a stainless steel water spout to receive a 5 μL water drop. Licks were detected by changes in resistance between the aluminum foil and the water spout. By increasing random lick rates, mice received almost all available rewards within 2 to 3 days, but the time-locking of licks to the sound increased more slowly (Figure 1.7c). Learning speed through which we estimate relative salience of up- and down-ramps was calculated by quantifying the number of post-stimulus licks divided by the number of pre-stimulus licks. The second task was a stimulus avoidance (or Go-NoGo) task. Mice were freely moving in a transparent box equipped with a water spout flanked by an infrared detector. When a mouse approaching the spout was signaled by the detector, a two-second white noise ramp was played (range 60-85dB). During the first four days, only a rewarded S+ ramp was presented (either up or down, depending on the training group). Mice had to signal their licking by breaking the infrared beam for more than 1.125 s (licking threshold) during a 1.5s time window after sound offset to get a reward. All mice reached more than 80% correct performance on this task after four days of training. Then, a distractor S- ramp was introduced with direction opposite to the S+. If mice licked above the S+ licking threshold after the S- was presented, a time out of 8 s was issued in addition to the 5 s interval before the next trial. No time out was issued after incorrect S+ trials. Beam breaks were measured as a continuous voltage signal (Figure 1.7e) that was thresholded to compute lick duration. Salience of the sounds in this task could be compared through the time necessary for going from 20% to 80% of plateau performance. Behavioral analyses were all automated thus no animal randomization or experimenter blinding was used.

Cortical response models.

We tested different models to account for the mean cortical response $r(t)$ to the envelope $s(t)$ of the intensity-modulated sounds converted to dB SPL (non-linear function). The linear model (Figure 1.2a) corresponds to the convolution of $s(t)$ with a causal kernel $h(t)$, defined over $t \in [0;1]$ sec, that was fitted to the data using the Moore-Penrose pseudo-inverse method[58]. For the adaptation model[62] (Figure 1.2b), an adapted version of the input $s_d(t)$ was computed as $s_d(t) = s(t)(1 - d(t))$ with $d(t)$ solution of the differential equation:

$$\frac{\partial d}{\partial t} = u[1 - d(t)]s(t) - d(t)/\tau \tag{1.4}$$

Then a linear kernel was applied to $s_d(t)$ to fit the cortical response as for the linear model. The best fit of the model to the 2s white noise ramps was obtained by a brute force search for parameters u and τ. This fitting approach illustrates our analytical demonstration that both the linear and adaptation model cannot account for asymmetric responses to the up- and down-ramps described in Supplementary Note 1. Note also that the linear temporal filter used here is fully equivalent to a spectro-temporal filter when applied to sounds that have identical spectral content (see Supplementary Note 1).
We also tested different models to account for the 13 response clusters $r_i(t)$ observed in our dataset (Figure 1.6). For all models, the input was split into "loud" $s_L(t)$ and "quiet" $s_Q(t)$ channels computed as sigmoid functions of the input:

$$s_L(t) = \frac{1}{e^{-\frac{s(t)-\mu_L}{\sigma}} + 1} \tag{1.5}$$

$$s_Q(t) = \frac{1}{\left(e^{\frac{s(t)-\mu_L}{\sigma}} + 1\right)\left(e^{-\frac{-s(t)-\mu_Q}{\sigma}} + 1\right)} \tag{1.6}$$

In all models, the three parameters of these two functions were optimized using a brute force approach to best approximate the fit between the final output of the model and the data. For the linear-nonlinear (LN) model, we modeled the response of cluster i as

$$r_i(t) = F_i(s_Q \circledast h_{Q,i}(t) + s_L \circledast h_{L,i}(t)) \tag{1.7}$$

in which $h_{Q,i}$ and $h_{L,i}$ are two kernels defined on $t \in [0;2]$ seconds and $F_i(x) = a_i(x - x_{0,i}) + c_i$ if $x \leq x_{0,i}$ and $F_i(x) = b_i(x - x_{0,i}) + c_i$ if $x > x_{0,i}$ is a monotonous piecewise-linear function. Fitting was done by first determining the kernels that best fit the data before applying the nonlinearity, and then finding the parameters of $F_i(x)$ that minimize the discrepancy between the sum of the kernel outputs and the data.
For the full nonlinear feature model, a layer of six nonlinear feature detectors (three "loud", three "quiet") were constructed as $f_{Q|L,on}(t) = G(\frac{\delta s_{Q|L}}{\delta t} \circledast \chi(t))$, $f_{Q|L,off}(t) = G(-\frac{\delta s_{Q|L}}{\delta t} \circledast \chi(t))$, and $f_{Q|L,tonic}(t) = G(s_{Q|L} \circledast \chi(t))$ in

which $\chi(t) = \Theta(t)e^{-t/\tau}$, $\Theta(t)$ is the Heaviside step function,τ = 0.05 s and G is piecewise-linear function ($G(x) = x - \theta$ if $x > \theta$ and $G(x) = 0$ otherwise (note that a single threshold value θ is used for the six feature detectors). We then modeled the cortical responses as a weighted sum of the two "tonic" features (no-kernel) and of the four transient features convolved with linear kernels:

$$r_i(t) = a.f_{Q,tonic}(t) + b.f_{L,tonic}(t) + \sum_{\substack{p=\{Q,L\} \\ q=\{on,off\}}} f_{p,q} \circledast h_{p,q,i}(t) \qquad (1.8)$$

To fit the model, the scalars and as well as the kernels $h_{p,q,i}(t)$ were obtained using linear regression solved exactly (Moore-Penrose pseudo-inverse method) once the nonlinear features had been generated. We either set the threshold θ of $G$to zero or optimized it together with the parameters of the "loud" and "quiet" input channels. For evaluation of the fraction of variance accounted by the model, we first trained the model on a subset of the white noise stimuli (0.25 s constant sounds at 50, 60, 70 and 85dB + all up- and down-ramps between 50 to 85dB) and measured the unexplained variance on the response of the model to a test set of white noise stimuli (0.25 s constant sounds at 55, 65 and 80dB + all up- and down-ramps between 60 to 85dB).

Acknowledgements

We thank Y. Frégnac, S. Rumpel, S. Ostojic and E. Harrell for comments on the manuscript, S. Sikirić and L. François for help with behavioral and imaging experiments. This work was supported by the Agence Nationale pour la Recherche (ANR "SENSEMAKER"), the Fyssen foundation, the DIM "Region Ile de France", the Marie Curie Program (CIG 334581), and by The International Human Frontier Science Program Organization (CDA-0064-2015).

Author contributions

BB, TD, AK and EP designed the study; BB and TD performed imaging experiments; AD performed behavioral experiments; BB, TD and AK performed data analysis; BB and AK designed the model; BB wrote the manuscript.

Supplementary Note 1

1 General question: transformations which preserve of the integral of their output after time-reversal of their input

For any input signal $s(t)$, defined as an integrable function on $\mathbb{R}$, we are interested in transformations F from the space of integrable function to itself, for which the time integral of the output signal is invariant with respect to time-reversal, i.e. the transformation that satisfy the property P_0:

$$\int_{-\infty}^{+\infty} F\left[s(t)\right] dt = \int_{-\infty}^{+\infty} F\left[s(-t)\right] dt$$

We here describe analytical proofs of this property for specific transformations or classes of transformation. Note that in the following, the notation $\int$ is used for $\int_{-\infty}^{+\infty}$.

2 Effect of an arbitrary function applied to the input before the transformation

It is interesting to mention, that if F is a transformation that satisfy P_0, this applies to any integrable function on $\mathbb{R}$. So for any function $f : x \to f(x)$ from $\mathbb{R}$ to $\mathbb{R}$ such that $f(s(t))$ is still integrable, the transformation $F[f(s(t))]$ also satisfies P_0. In other words, any function (including non-linear functions) applied to the input signal before the transformation does not affect the invariance of the output integrals to a time reversal.

3 Invariance for a linear transformation

A general linear transformation of a function $s(t)$, invariant by translation (i.e. the transformation does not depend on the absolute time at which is occurs) can be written as a convolution with a filter $h(t)$.

$$F : s(t) \to \int h(t-u)s(u)du$$

For such a transformation the integral of the time-reversed signal is:

$$\int F[s(-t)] = \int \int h(t-u)s(-u)dtdu = \int s(-u)du \int h(t-u)dt$$

So by setting $t' = t - u$ and then $u' = -u$ one easily obtains the equality of the integrals:

$$\int F[s(-t)] = \left(\int s(u')du'\right)\left(\int h(t')dt'\right) = \int F[s(t)]$$

4 Case of STRF filters

In the particular case of a STRF filter, the input signal is the spectrogram $\hat{s}(t, f)$ of the signal $s(t)$. The response $r(t)$ of a neuron predicted by its associated spectro-temporal receptive field, is computed by first convolving the spectro-temporal kernel $STRF(t, f)$ with $\hat{s}(t, f)$

$$\hat{r}(t, f) = \int duSTRF(u, f)\hat{s}(t-u, f)$$

This transformation is linear and invariant by time translation, thus for all frequencies f the time integral of $\hat{r}(t,f)$ is not affected by time-reversal. $r(t) = \int \hat{r}(t,f)df$ corresponds to the sum of $\hat{r}(t,f)$ over all frequencies f. This integration step is independant of time and thus is also unaffected by time-reversal. Therefore the integral of STRF predictions of a neuron's response is in all cases unaffected by time reversal of the stimulus.

In addition in the particular case of the stimuli used in this study which have a frequency content that is invariant over time, the spectrogramm can be written as a product of a spectral and enveloppe component $\hat{s}(t,f) = g(f)S(t)$. In this case:

$$r(t) = \int duS(t-u) \int g(f)STRF(u,f)df = \int duS(t-u)\tilde{STRF}(u)$$

Thus the STRF framework simplifies for this particular case to convolution with a frequency independent effective kernel $\tilde{STRF}(t)$ valid for a particular frequency content. This justifies that the use of a frequency independant kernel to fit, within the STRF framework, the neuronal responses to enveloppe variations of white noise stimuli.

5 Invariance for the synaptic depression model

The model of synaptic depression is defined by David et al. (2009) as a discrete time equation for a depression variable d:

$$d(t+1) = d(t) + s(t)[1-d(t)]u - d(t)/\tau$$

from which the output signal is obtained as:

$$s_d(t) = s(t)(1-d(t))$$

The first equation yields in continuous time:

$$d'(t) + [1/\tau + us(t)]d(t) = us(t)$$

in which $d'(t)$ is the first derivative of $d(t)$. If we take that $s(t) = 0$ for $t \leq 0$, (i.e. the signal starts at $t = 0$) the solution of this first order linear equation can be written as:

$$d(t) = u \int s(x)e^{-\int_x^t (1/\tau + us(v))dv}\theta(t-x)dx$$

in which θ is the Heaviside step function.

Because $s_d(t) = s(t) - s(t)d(t)$ the invariance to time-reversal will be obtained if and only if $A_s = \int s(t)d(t)$ is invariant to time-reversal. For the forward signal A_s (normalized by u) writes as:

$$A_{s+} = \iint dxdts(t)s(x)e^{-\int_x^t (1/\tau + us(v))dv}\theta(t-x)$$

And for the time-reversed signal it writes as:

$$A_{s-} = \iint dtdxs(-t)s(-x)e^{-\int_x^t (1/\tau + us(-v))dv}\theta(t-x)$$

Setting $t' = -t$, $v' = -v$ and $x' = -x$ yields,

$$A_{s-} = \iint dt'dx's(t')s(x')e^{-\int_{t'}^{x'} (1/\tau + us(v'))dv'}\theta(x'-t')$$

In the expression above, the x' and t' are equivalent. Hence:

$$A_{s-} = \iint dxdts(x)s(t)e^{-\int_x^t (1/\tau + us(v))dv}\theta(t-x) = A_{s+}$$

proving that the output integral of the synaptic depression model is invariant to time-reversal of the input signal despite its nonlinearity.

6 Some sufficient conditions for a linear non-linear transformation (LN model)

We now suppose that the transformation F is a linear filter of kernel h followed by a non-linear function f, i.e.

$$F : s(t) \to f\left(\int s(t-u)h(u)du\right)$$

In this case two sufficient conditions for P_0 can be derived.

Sufficient condition 1 If h has a vertical symmetry (i.e. it exists x_0 such that for all x, $h(x-x_0) = h(x_0-x)$) then F satisfies P_0.

Proof Three changes of variable: $u \to x_0 - v$ followed by $v - x0 \to u'$ and $t \to t'$ yield the equality.

$$\int f\left(\int s(-t+u)h(u)du\right)dt = \int f\left(\int s(-t+x_0-v)h(x_0-v)dv\right)dt = \int f\left(\int s(t'-u')h(u')du'\right)dt'$$

Sufficient condition 2 If h has a central symmetry (i.e. it exists x_0 such that for all x, $h(x-x_0) = -h(x_0-x)$) and if $f(x) = 0$ for $x \leq 0$ and $f(x) = x$ for $x > 0$, then F satisfies P_0.

Proof If h has central symmetry then $\int h(u)du = 0$ and $\int\int s(u)h(t+u)dudt = 0$. So if we call $\mathbb{H}^+$ the sub-ensemble of $\mathbb{R}$ in which $\int s(u)h(t+u)du > 0$ and $\mathbb{H}^-$ its complementary in $\mathbb{R}$, we have $\int_{\mathbb{H}^+}\int s(u)h(t+u)dudt = -\int_{\mathbb{H}^-}\int s(u)h(t+u)dudt$. The proof then comes from the fact that the integral of the time-reversed signal can be re-written as:

$$\int f\left(\int s(-u)h(t-u)du\right)dt = \int f\left(\int s(u')h(t+u')du'\right)dt' = \int_{\mathbb{H}^+}\int s(u')h(t+u')du'dt'$$

an that using the central symmetry of h we get for the integral of the forward signal:

$$\int f\left(\int s(u)h(t-u)du\right)dt = \int f\left(\int s(u)h(-t'-u)du\right)dt' = -\int_{\mathbb{H}^-}\int s(u)h(t'+u)dudt'$$

which thanks to the above mentioned equality leads to the proof.

7 Conclusions

In our experiments, we have observed that ramping-up sounds produce cortical responses with a larger time-integral that ramping-down sounds, although the time-integral of the envelop of the two sounds are the same. The above proofs show that models with a non-linear intensity scaling function followed by a linear filter are mathematically unable to explain this property in the general case. Moreover, the addition of a previously described non-linear adaptation model is also mathematically unable to explain the data.

Lastly, we show that models constructed with a linear filter followed by a non-linearity (LN models) will not be able to reproduce the observed experimental property if the kernel of the filter is has vertical-symmetry or in the case of very simple rectifying non-linearity if the the kernel has a central symmetry.

Note that, in other conditions, LN-models actually can produce unequal output time-integrals although input integrals are equal. Nevertheless, LN-models are unable to reproduce the temporal profile of recorded neuronal responses (see Fig. 7), because these responses encode features that are incompatible in a LN-model. For example, a linear filter cannot respond positively both at the onset and at the offset of a positive signal (as many neurons in auditory cortex do, e.g. ON-OFF cluster), because linear on-response (resp. off-response) filters also respond negatively at an offset (resp. onset). Hence the addition of linear on- and off-response filters produces overall no output which no subsequent nonlinear function can compensate. To model neurons that respond positively both to onsets and offsets, it is necessary to insert a nonlinearity before summing the two features as we did it our multilayer non-linear features model (Fig. 7).

Supplementary Figures

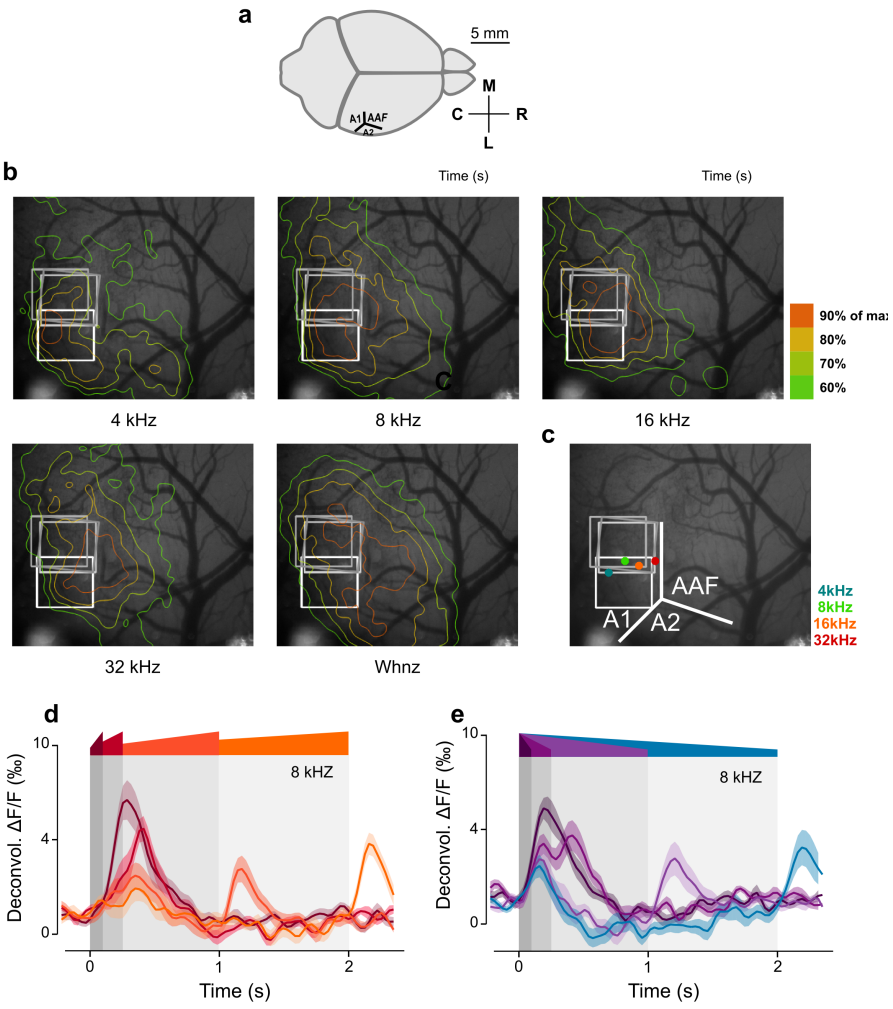

FIGURE 1.8: **Functional localization of the right auditory cortex with intrinsic optical imaging.**

*(**a**) Localization of the right auditory cortex relative to the mouse brain in mouse 3. Auditory core fields (composed of A1 and AAF) are located on the rostro-caudal axis while secondary fields of the belt region (e.g. A2) are more ventral or dorsal. (**b**) Identification of the auditory cortex and its subfields through intrinsic optical imaging of responses to pure tones. The contour maps superimposed to the blood vessels image represent the ratio of intrinsic signal before and during a 2 s auditory stimulation, here expressed as the percentage of the maximum response for each stimulus. For the animal shown in this example, four two-photon imaging sessions were performed at different location. They are represented as grey rectangles (overlapping locations indicate recordings at different depths) and cover a large part the A1 subfield coarsely identified from the tonotopic gradient observed in the intrinsic responses (see also c). (**c**) A tonotopic gradient perpendicular to the media-lateral axe can be deduced from intrinsic imaging signals. Each dot represents the centroid of the area in which the intrinsic signal is within 90% of the maximum response for each sound frequency. This rostro-caudal gradient from low-frequency (blue) to high frequency tuning (red) correspond to the A1 subfield21. The mirror symmetric gradient from AAF can be deduced from the anterior local response peak seen in the response to 4 kHz. These gradients were used to coarsely identify the location of auditory cortex subfields. (**d**) Mean deconvolved calcium signals (i.e. estimated firing rate) for 8kHz up-ramps of duration 100ms, 250ms, 1s and 2s (range 60-85 dB SPL, shading indicates SEM across imaging sessions, n=13). (**e**) Same as **d**. for 8kHz down-ramps of duration 100ms, 250ms, 1s and 2s.*

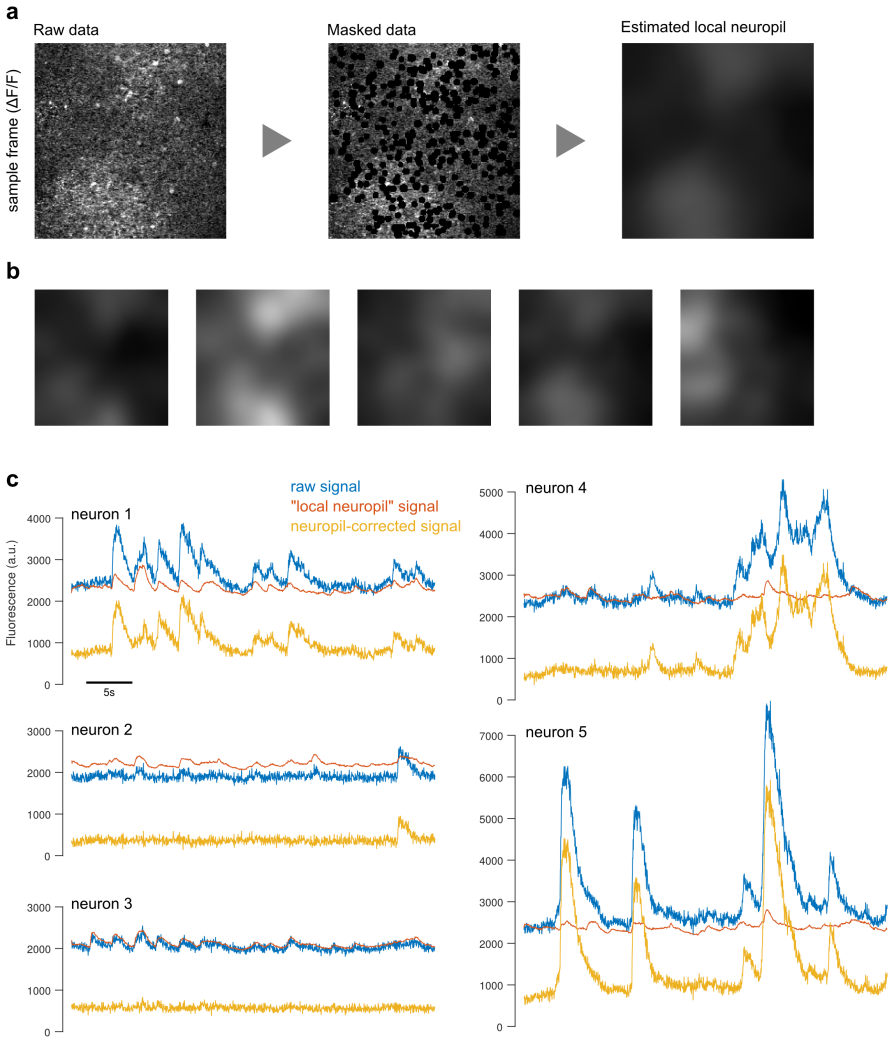

FIGURE 1.9: **Correction for neuropil contamination.**

*(**a**) Neuropil estimation method: each individual frame is multiplied with a mask that avoids all selected neurons, and appropriate smoothing (see Methods) is used to fill-in image parts that were masked out. This permits to estimate the average neuropil signal at the location of the neuron. (**b**) Sample images of local neuropil estimation from the same imaging session showing clear variations across different time points and spatial locations. (**c**) Raw signals from five individual neurons (in blue) and corresponding local neuropil signals (in red) extracted from the same ROIs. The neuropil-corrected signals (yellow) are obtained by subtracting from the raw signal a scaled version (by 0.7) of the local neuropil signals. For some neurons (e.g. bottom left) all the signal present in the raw data is removed, while for others simultaneously imaged neurons (e.g. bottom right) the signals are little affected by the correction.*

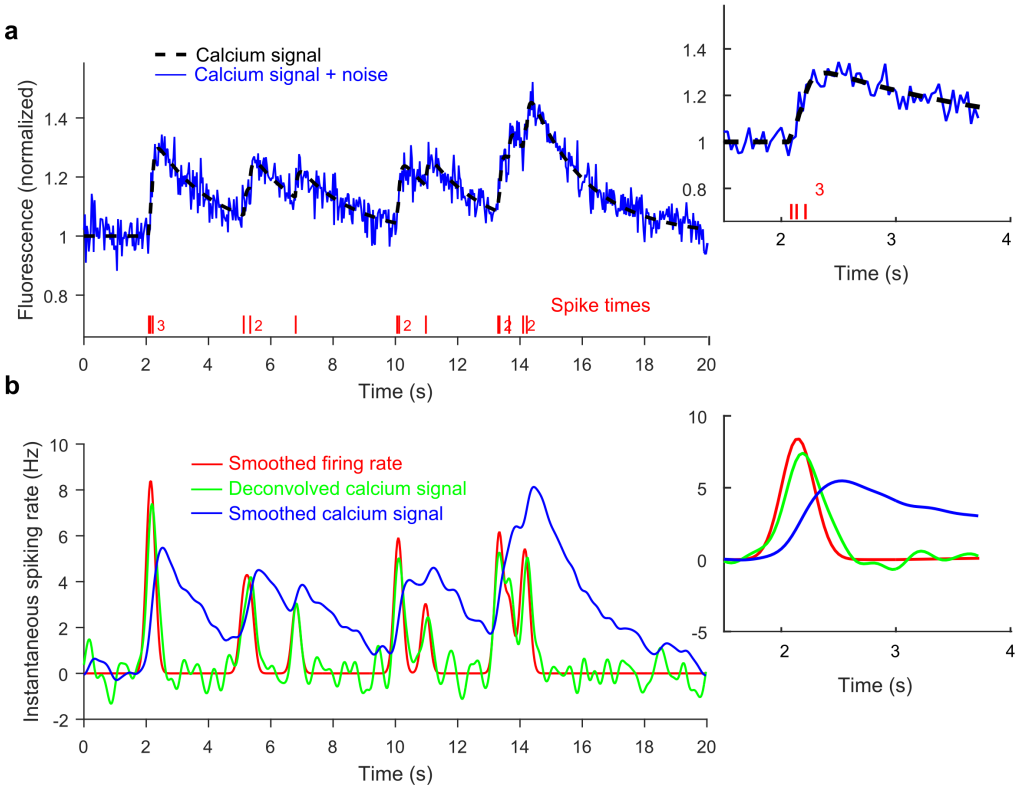

FIGURE 1.10: **Deconvolution of calcium signals: simulations.**

*(**a**) Simulated GCAMP6s fluorescence (black line) resulting from the train of spike shown below (red bars). The GCAMP6s signal resulting from a single spike is here modeled as double exponential with a unitary calcium increase a of 11.3%, a rise time τ_{on} of 70 ms and an exponential decay τ of 1.87s as described in mouse visual cortex (specifically, $F(t)/F_0 = \sum_{t_{spike}<t} a\left(1-\exp\left(1-\frac{t-t_{spike}}{\tau_{on}}\right)\right)\exp\left(\frac{t-t_{spike}}{\tau}\right)$). The blue line corresponds to the simulated signal superposed with white noise. Magnified signal in the inset highlights the temporal delay of the fluorescence peak compared to spikes due to the 70ms rise time. (**b**) Applying our linear deconvolution algorithm followed with Gaussian smoothing to the noisy fluorescence signal shown in a. yields an estimate of the time course of the instantaneous firing rate (green) which matches the smoothed instantaneous rate (red) much better than smoothed calcium signal (blue, the scale is hand-adjusted to match the rate signals). Correlation of the smoothed firing rate is much higher with the deconvolved calcium signals (0.91) than with the smoothed raw calcium signal (0.21), despite the fact that the deconvolution ignored the slow rise time of GCAMP6s, which results in a slight delay of the rate estimate, as can be seen in the inset.*

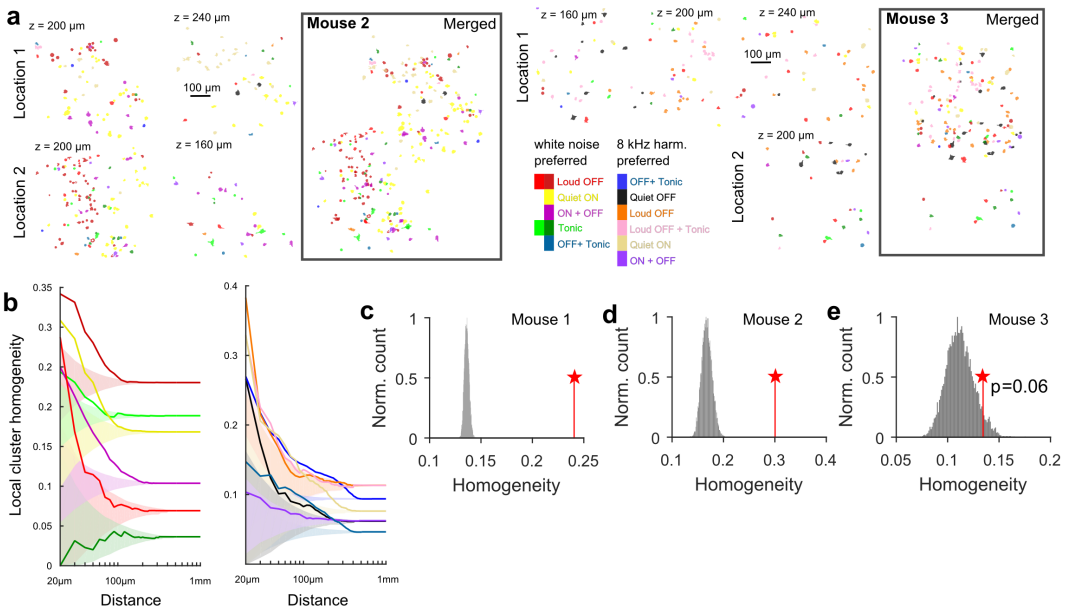

FIGURE 1.11: **Functional cell assembly organization in a third mouse.**

*(**a**) Localizations of the cells belonging to the different identified clusters in four imaging sessions performed at two different horizontal localizations and different depths (z) across several days in mice 2 and 3. The localizations of mouse 3 recordings within the auditory cortex can be found in Supplementary Fig. 1. The color-code used for the different clusters is consistent with Fig. 4 and Fig. 5 (see colorbar). On the right, the localization of all cells is shown in a horizontally-mapped z-projection. (**b**) Single cluster homogeneity for the 13 identified clusters (see color code in a) when the radius of analysis is varied. The shaded areas represent the range of values for the homogeneity index observed in 99% of the cell identity shufflings (bootstrap). (**c-e**) Spatial clustering is present across different mice and recordings. Distributions of the global homogeneity index (mean probability for the neighbors of any given neuron within an 80 m radius to belong to the same functional cluster) for 10,000 shuffling of the cell identities (bootstrap) compared to the experimental values for different mice.*

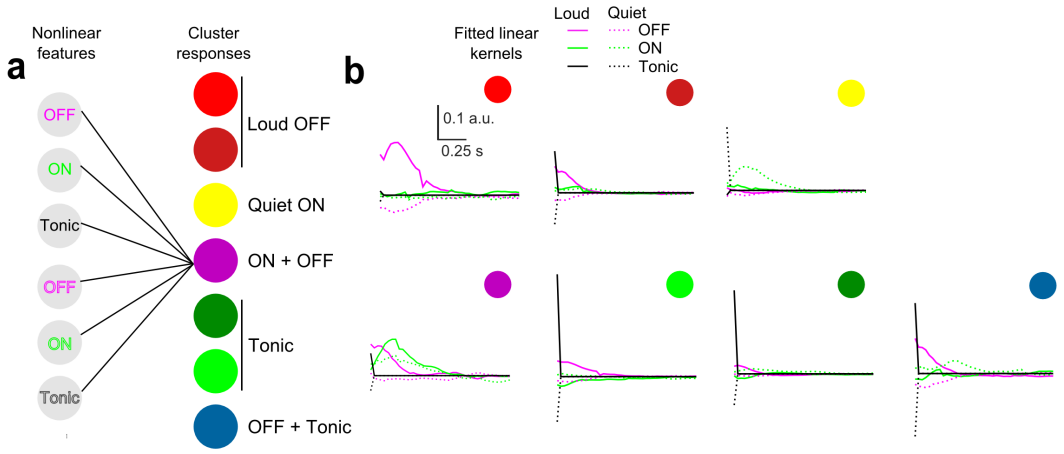

FIGURE 1.12: **Fitted kernels for the multilayer feature model.**

*(**a**) Schematics of the last two layers of the multilayer model. (**b**) Fitted kernels linking the input feature layer to the seven clusters showing preference for white noise. Green (resp. magenta) traces represent kernels from ON (resp. OFF) input features (plain line: loud; dashed line: quiet). Black peaks represent the weight of the constant input (Loud Tonic, continuous line, Quiet Tonic). These plots clearly show that certain clusters (e.g. red) mostly reflect a single input feature, while others are better modeled by mixed inputs. Moreover the time-course of the kernels reflect simple regular transient functions with temporal phasic temporal profile with a decay time constant of 200 to 300 ms resembling which are compatible with biological slow and polysynaptic post-synaptic potentials e.g. coming from an upstream neuronal population. This suggests that the model effectively summarizes the summation of different functional inputs (however complex the real presynaptic connectivity might be) in the cortical neurons and does not perform extensive overfitting.*

Context Chapter 2

This second study is currently under review in Nature Communications and it is related to the first study presented in the PhD thesis. The general scope of this study is to understand the link between cortical recruitment and learning speed. We observed in the first study that the up-ramps produce more cortical recruitment and that these sounds are learn faster compare to down-ramps. We wonder if a sound that produces more activity is generally learn faster in a behavioral task. With calcium imaging and a serie of Go/NoGo task, we showed a clear correlation between cortical recruitment and learning speed for a set of sounds. We used a reinforcement learning model that reproduces this correlation, and with a behavioral task drove by optogenetic we demonstrated the direct causality between cortical recruitment and learning speed.

My contribution in this study was to performed some of the surgery and calcium imaging experiments. I also analysed the recordings from calcium imaging and computed the sounds similarity matrices. I participated to the design of some figures and the generation of the figure 2.2.

Chapter 2

Cortical recruitment determines learning dynamics and strategy

Under review in Nature Communication in 2018

Sebastian Ceballo[1,3] , Jacques Bourg[1,3] , Alexandre Kempf[1,3] , Zuzanna Piwkowska[1,4] , Aurélie Daret[1] , Thomas Deneux[1] , Simon Rumpel[2] and Brice Bathellier[1]

1. Unité de Neuroscience, Information et Complexité (UNIC), Centre National de la Recherche Scientifique, FRE 3693, Gif-sur-Yvette, 91198, France
2. Institute of Physiology, Focus Program Translational Neuroscience, University Medical Center, Johannes Gutenberg University, Mainz, Germany.
3. Co-first authors
4. Present address: Institut Pasteur, Dynamic Neuronal Imaging Unit, Paris, France.

Abstract

Salience is a broad and widely used concept in neuroscience whose neuronal correlates, however, remain elusive. In behavioral conditioning, salience is used to explain various effects, such as stimulus overshadowing, and refers to how fast and strongly a stimulus can be associated with a conditioned event. Here, we identified sounds of equal intensity and perceptual detectability, which due to their spectro-temporal content recruit different levels of population activity in mouse auditory cortex. When using these sounds as cues in a Go/NoGo discrimination task, the degree of cortical recruitment matches the salience parameter of a reinforcement learning model used to analyze learning speed. We test an essential prediction of this model by training mice to discriminate light-sculpted optogenetic activity patterns in auditory cortex, and verify that cortical recruitment causally determines association or overshadowing of the stimulus components. This demonstrates that cortical recruitment underlies major aspects of stimulus salience during reinforcement learning.

Introduction

Sensory stimuli can vary substantially in their efficacy to serve as a conditioned stimulus during behavioral conditioning. In the context of classical conditioning, a well-known example is the so called "overshadowing" effect. When animals are trained to associate two simultaneously presented stimuli (historically a tone and a flash) to a specific unconditioned stimulus (e.g. foot- shock), it is often observed that, after training, the animal is conditioned more strongly to one stimulus than to the other[169, 170] . In their seminal theoretical work originally developed for classical conditioning, but later extended to operant conditioning (e.g. in reinforcement learning), Rescorla and Wagner[154] introduced the notion of salience to explain the overshadowing phenomenon. In their model, salience is a parameter affecting the speed at which a given stimulus is associated with the unconditioned stimulus. Thus, when behavior reaches maximal performance and learning stops, the more salient of the two stimulus representations has been associated more strongly with the unconditioned stimulus, leading to overshadowing. While this theory captures a number of phenomena and is the basis for important frameworks such as reinforcement learning[171, 172], the neural underpinnings of the salience parameter remain elusive. Salience in this context is usually seen as the global amount of neural activity representing the stimulus, like in "pop out" models of attentional salience[173, 174, 175, 176]. This intuitively follows from the idea that if more spikes are involved in representing a stimulus, they can produce more synaptic weight changes, as expected from the firing rate sensitivity of typical learning rules[177, 178, 179, 180, 181], and thus modulate more rapidly the relevant connections. However widespread, this idea lacks direct causal experimental verification in a learning task. Moreover, other theories propose that salience could also be encoded in other parameters such as neuronal synchrony levels[182, 183, 184, 185], which could influence learning via the temporal properties of biological learning rules[186, 187, 188, 189, 190]. Thus, the neuronal correlate of stimulus salience is a key question with broad implications for learning theories.
Using auditory discrimination tasks of sounds with different global cortical response strengths, we show that cortical recruitment impacts learning dynamics[155, 119] in a manner similar to the salience parameter of a reinforcement learning model. To explore this result in more precise experimental settings, we trained mice to discriminate optogenetically-driven response patterns that elicit different levels of cortical activity. Using this paradigm, we directly demonstrate that cortical recruitment determines which part of a compound stimulus drives a learned association while "overshadowing" other parts of the stimulus. This validates a generic prediction of reinforcement learning models and causally establishes the role of cortical recruitment as a neuronal correlate of stimulus salience.

Results

Sounds with identical physical levels and perceptual detectability can recruit different levels of cortical activity.

In order to investigate the relationship between stimulus salience and neuronal recruitment, we first aimed to identify sounds recruiting different amounts of cortical activity. A previous report has shown that complex sounds with different frequency content but equal duration and sound pressure level can recruit population responses of different sizes in cat auditory cortex[191]. To test if a similar result could be obtained in mice, which would then allow us to experimentally decouple recruitment from physical intensity, we chose three short, complex sounds (70ms duration) containing a large range of frequencies and temporal modulations, but normalized at equal mean pressure level (73dB SPL, Figure 2.1a). The three sounds displayed different power spectra in the 10-30kHz range (Figure 2.1a) where the mouse ear is most sensitive[192, 193, 194]. We thus wondered if this discrepancy was affecting their detectability. To do so, we trained mice to lick on a water port after presentation of each of the three sounds (all mice experienced the three sounds in the same task) to obtain a reward (Figure 2.1b) and then measured response probability to decreasing intensity levels. We observed that, for all sounds, response probability steadily decreased down to chance level as measured in the absence of sound (Figure 2.1c). Yet, no significant difference in response probability curves was observed across the three sounds (Friedman test – non-parametric anova, p = 0.43, n=6), indicating that the chosen 73dB SPL was at a comparable distance from the detection threshold for the three sounds.

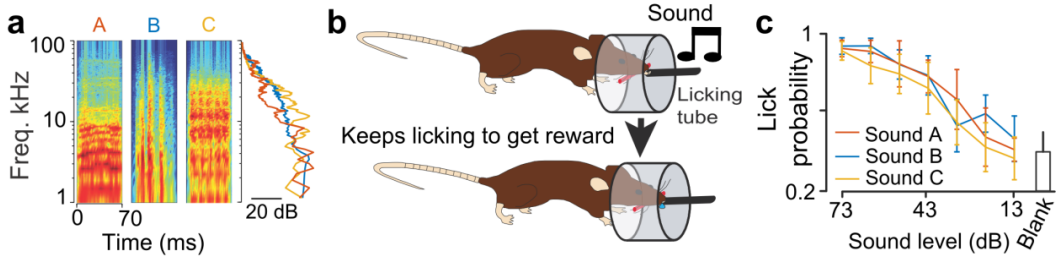

FIGURE 2.1: **Spectro-temporal differences in complex sounds do not affect near-threshold detectability.**

*(**a**) Spectrograms of three 70 ms long complex sounds, with power spectrum on right. (**b**) Schematics describing the auditory detection task. (**c**) Mean response probability for 6 mice trained to detect sounds A, B and C at 73 dB to get a reward and probed with lower sound intensities. While the effect of intensity was significant, there was no effect of sound identity (Friedman test, $p_{intensity} = 2.3 * 10^{-9}$, $p_{sound} = 0.43$, n=6 mice).*

We assessed recruitment of neural activity in the auditory cortex in response to these three sounds using two-photon calcium imaging in awake,

passively listening mice, a technique that offers access to large samples of neuronal activity. We imaged 6 mice that were injected with AAV1-GCAMP6s virus in auditory cortex (Figure 2.2a). Recordings were followed by an automated image registration and segmentation algorithm (Figure 2.2b) 30 that allowed the isolation of 15,511 neurons across 27 imaging sites, from which large fluorescence signals could be observed (Figure 2.2c). The fields-of-view were either $0.5 * 0.5$ or $1 * 1$ mm (Figure 2.2a, b), allowing a rapid tiling of the full extent of primary and secondary auditory cortex (Supplementary Figure 2.7). Cortical depths were randomly chosen ranging between 100 and 300 μm corresponding to layer 2/3. The mouse auditory cortex (primary + secondary) contains approximately 200,000 neurons in one hemisphere[195] and thus about 50,000 neurons in layer 2/3, so we expect our sample of $\sim$15,000 neurons to be representative for supragranular auditory cortex. Comparing the amplitude of the mean-deconvolved calcium signals recorded across the entire duration of the response (see example population response profiles in Figure 2.2d), we observed that at 73 dB intensity, sound A elicited at least two-fold less cortical activity than sounds B and C (0.05 against 0.10 and 0.12% $\Delta F/F.s^{-1}$, Figure 2.2e). This was consistently observed across mice (Supplementary Figure 2.7). Furthermore, sound A triggered a significant response (Wilcoxon sign test across 20 sound repetitions) in only about 35% of all neurons, while sounds B and C significantly excited 41% and 45% of all neurons (Figure 2.2e). Furthermore, this difference in population responsiveness was consistent with previous, independent measurements performed under anesthesia (Supplementary Figure 2.7) 32.
It is noteworthy that all three sounds elicited distinct response patterns as evaluated by correlation-based population similarity measures. Thus, sound identity could be decoded with high accuracy, on a single-trial basis, using linear classifiers (Figure 2.2f). This shows that sound discriminability was not affected by cortical recruitment.

Another discrepancy between cortical recruitment and the physical intensity of a stimulus can be observed using sounds with different temporal intensity profiles. Up-ramping sounds elicit larger cortical responses in mice[44] and other animals[140, 45] than their time-symmetric down–ramps, despite the equality of their cumulative physical energies. This effect correlates with asymmetries in subjectively perceived loudness in humans[133, 136]. We confirmed this result (Figure 2.2g) for 2s white noise sounds ramping between 60 and 85dB (mean deconvolved $\Delta F/F$ measured from 0 to 2.5 s after sound onset, up: 0.166 $\pm$ 0.58% vs down: 0.157 $\pm$ 0.57%, Wilcoxon rank-signed test, p = 0.034), with a clear effect even at the onset despite the lower start intensity level in up-ramps (mean between 0 and 0.5s for up: 0.80 $\pm$ 0.08 % vs down: 0.62 $\pm$ 0.11 %, Wilcoxon rank-signed test $p = 3.75 * 10^{-4}$, n = 29 sites, in 12 mice). Another striking example was obtained with rhythmic amplitude modulations. We observed across a large sample of cortical neurons that a white noise sound modulated at 1 Hz produces more activity than when modulated at 20Hz, although the two sounds have the same physical energy (Figure 2.2h).

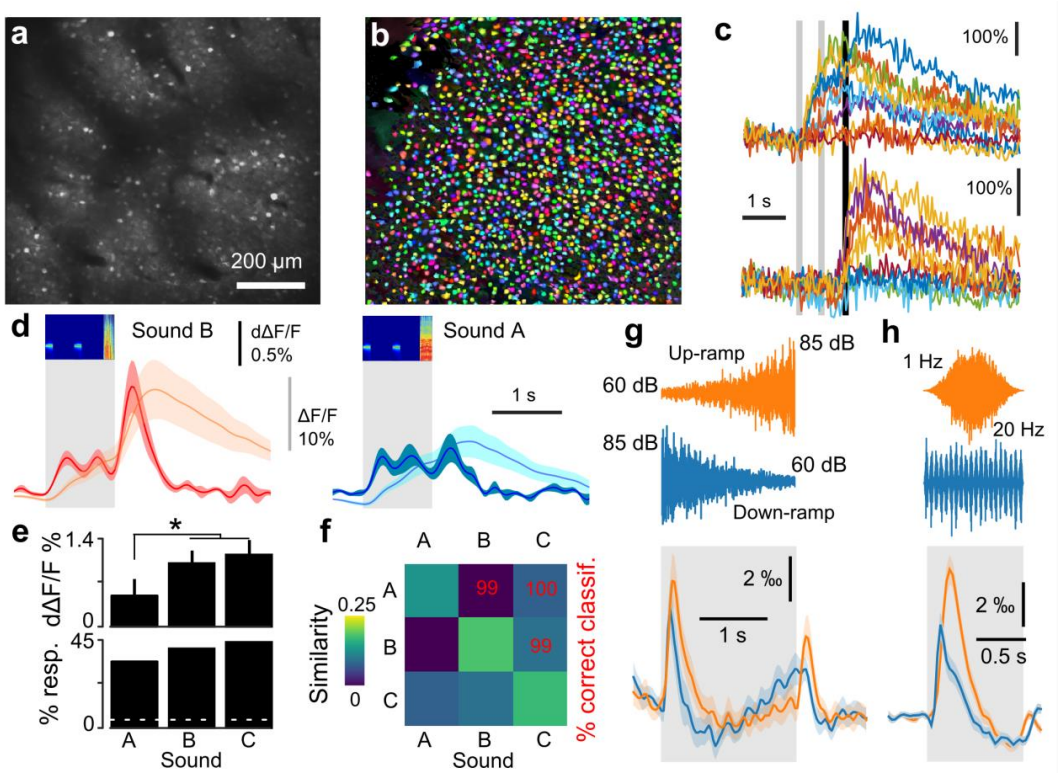

FIGURE 2.2: **Spectro-temporal differences impact on cortical recruitment.**

*(**a**) Example field of view illustrating GCAMP6s labeling of L2/3 auditory cortex neurons. (**b**) Result of automated cell segmentation run on the data acquired in the example shown in a. (**c**) Example single trial responses to sounds (different colors) for two neurons (top and bottom). Gray bars = sound duration. (**d**) Population responses (n = 26 sessions, 15511 neurons in 6 mice) to Sound B (red) and A (blue). Both normalized fluorescence (light colors) and deconvolved (dark colors) calcium signals are shown. (**e**) Mean deconvolved signal and fraction of significantly responding neurons (Wilcoxon signed test, p<0.05) to sounds A, B and C. Mean calcium responses to sound A (0.05 $\pm$ 0.03% $\Delta F/F.s^{-1}$) were significantly smaller than to B (0.10 $\pm$ 0.02% $\Delta F/F.s^{-1}$) and C (0.12 $\pm$ 0.02% $\Delta F/F.s^{-1}$; sign test, p = 0.0094 and p = 0.029, n = 26 sessions, 15511 neurons in 6 mice). The fractions of responding neurons (35, 41 and 45 %) were all significantly different from each other (χ^2 test, p = 10^{-12}, 10^{-27} and 3.10^{-72} , n = 15511). (**f**) Population response reliability (diagonal) and similarity (off-diagonal) matrix for sounds A, B and C. The pair-wise discriminability value, computed with a linear classifier is indicated in red. (**g**) Mean deconvolved calcium signals for 6757 auditory cortex neurons in 12 awake mice during 29 calcium imaging sessions for 2 s long white noise sounds modulated in intensity between 60dB to 85dB upwards and downwards. (**h**) Mean deconvolved calcium signals for 59590 auditory cortex neurons in 7 awake mice during 60 calcium imaging sessions for white noise sounds modulated in intensity at 1 Hz and at 20 Hz (0.357 $\pm$ 0. 032 % and 0.15 $\pm$ 0.034 % $\Delta F/F.s^{-1}$, signed test, p = 0.0009).*

In summary, when different sounds are played at an intensity above the detection threshold, the specific amount of recruited cortical activity in mouse auditory cortex strongly depends on factors other than intensity and can vary across different sounds. Based on this observation, we asked whether cortical

recruitment could be related to stimulus salience in a learning task.

Cortical recruitment influences learning speed.

Classically, relative salience measures are performed using an overshadowing paradigm in which two stimuli (e.g. a flash and a sound) are conditioned together, as a compound stimulus, to an unconditioned stimulus (e.g. footshock). Then, salience is derived from the level of the conditional response elicited by each stimulus component individually. While this approach is valid when the compound is made of stimuli from two different sensory modalities, two simultaneous sounds are likely to fuse perceptually, precluding measurement of their individual saliences with the classical overshadowing design[196]. Alternatively, the seminal model of Rescorla and Wagner 3 postulates that learning speed follows stimulus salience, and this has not yet been addressed experimentally. We thus decided to confirm if stimulus salience can be estimated through learning speed, using an auditory-cued Go/No-Go task. To do so, water-deprived mice were first trained to visit a lick-port and to receive a water reward if they licked after being presented with an S+ sound. This pre-training phase mainly aimed at raising motivation in all mice and was not used to measure learning speed. When mice collected rewards in at least 80% of their port visits, the Go/NoGo task was started by introducing a non- rewarded S- sound in half of the trials (Figure 2.3a). After a large number of trials, mice succeeded to both sustain licking to the S+ and withdraw from licking for the S- (Figure 2.3b), thereby demonstrating their ability to discriminate the two sounds. Importantly, as typically observed in such tasks[119], the S+ sound was rapidly associated with the lick response and the rate limiting factor in the acquisition of the task was to associate the suppression of licking with the S- sound (Figure 2.3b). Hence, learning speed depends more on the salience of the S- than of the S+ sound in this task. The relative saliences of two stimuli X and Y can thus be measured by comparing the learning speed of the X versus Y Go/NoGo discrimination when X is the S- against the speed observed when Y is the S-. If X is less salient than Y, we expect learning to be slower when X is the S-. We therefore trained six cohorts of mice to compare the salience of sounds pairs A-B, A-C and B-C, and we used data from an earlier study to compare salience of up and down-ramping sounds[44]. Plotting the population learning curves for the sound pairs with maximum cortical recruitment differences (A-B & A-C,), we qualitatively observed that the average learning speed was faster when cortical recruitment for the S- sound was larger than for the S+ sound (Figure 2.3c), suggesting a link between salience observed via learning speed and cortical recruitment.

However, looking at learning curves from individual mice, we noticed that the qualitative difference observed at the group level hides a more complex effect. As often observed in animal training[155] and as we previously reported for the particular task used in this study[119], most individual learning curves had a sigmoidal rather than exponential time course. Specifically,

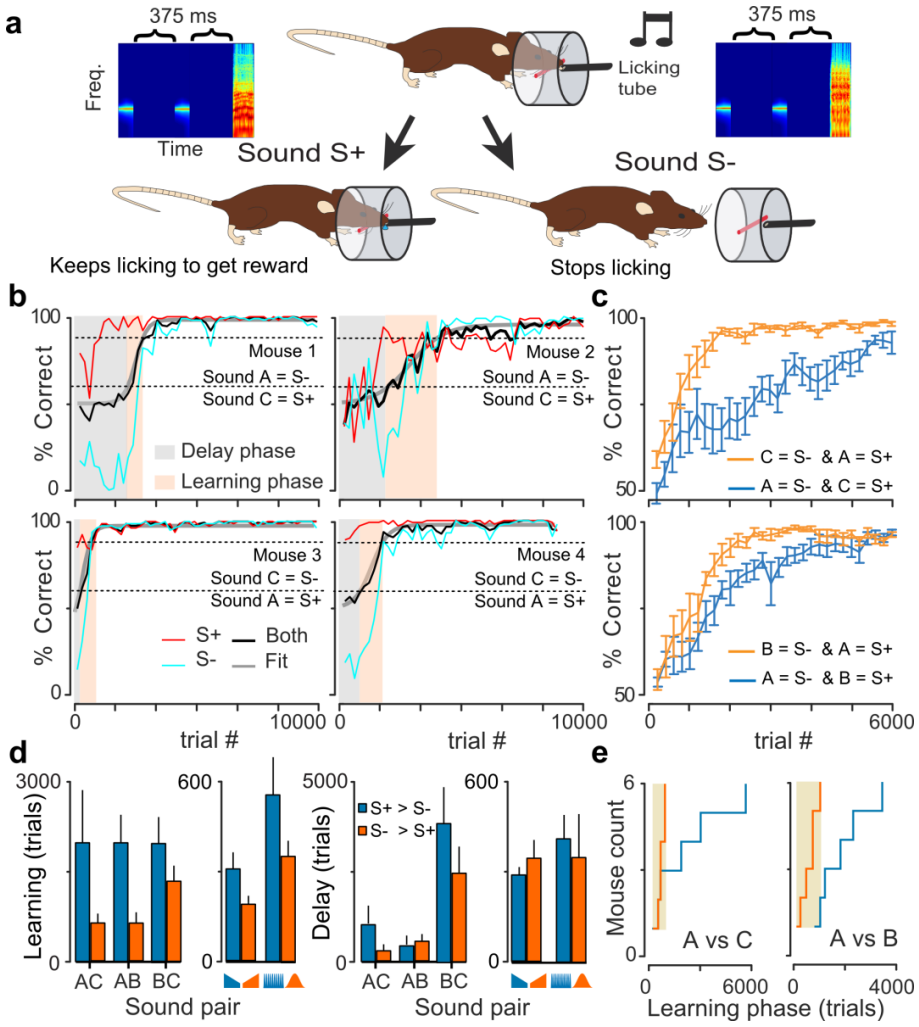

FIGURE 2.3: **Learning phase duration correlates with cortical recruitment differences.**

*(**a**) Schematics describing the auditory Go/NoGo discrimination task. (**b**) Individual learning curves for 4 mice discriminating sounds A and C. Performance for S+ (red), S- (light blue) and both (black) sounds are displayed. Mice from the top row have sound A as the S- stimulus while mice from the bottom row have sound C as the S- stimulus. Typical learning curves display a delay and learning phase as shown in light gray and orange colors. (**c**) Mean learning curves for different groups of mice (n = 6 for each curve) discriminating between sounds A and C (top) or A and B (bottom). Slower learning is observed when the S- sound recruits less cortical activity than the S+ sound (blue) as compared to when sound valence is swapped (orange). (**d**) Mean ± standard error for the learning and delay phase for the five discriminated sound pairs. The conditions "S- recruitment > S+ recruitment" (blue) and "S+ recruitment < S- recruitment" (orange) are significantly different for the learning phase but not the delay phase (Friedman test, p = 0.0005 and p = 0.72, n = 6 mice per group except for the up- and down ramps, n = 12). (**e**) Cumulative distributions of learning phase durations for sound pairs A-C, A-B.*

the curves displayed a delay phase with no increase in performance followed by a learning phase with an often very steep performance increase. Also, the

duration of each phase was highly variable across animals as exemplified in Figure 2.3b. We wondered whether cortical recruitment was affecting one particular phase or both. We performed a sigmoidal fit (Figure 2.3b) on each individual curve from which the delay phase duration was computed as the number of trials necessary to reach 20% of maximal performance level, and the learning phase duration as the number of trials necessary to go from 20% to 80% maximal performance. We observed across the five sound pairs tested that learning phase duration was systematically longer when the S- sound recruited less activity than the S+ sound (Figure 2.3d). A non-parametric analysis of variance showed this effect to be highly significant, while no systematic effect of cortical recruitment was observed for the delay phase (Figure 2.3d, Supplementary Figure 2.8). In addition, we noticed that cortical recruitment had an effect not only on the mean duration of the learning phase at the group level but also on the inter-individual variability. When the S- sound recruited more activity than the S+ sound, learning phase duration was more homogenous than for the opposite sound assignment, especially for the two sound pairs with a large difference of cortical recruitment (mean normalized standard deviation difference: 93% $\pm$ 18%, n = 5 sound pairs, p = 0.008 Wilcoxon rank-sum test, Figure 2.3e, Supplementary Figure 2.8).

A reinforcement learning model captures the effects of cortical recruitment on learning dynamics.

In order to better understand why cortical recruitment specifically impacts the learning phase and why variability is larger when S+ recruits more activity than S-, we employed a recently developed model of the discrimination task. This model is based on the Rescorla-Wagner reinforcement learning framework but extending it to a simple but more biologically interpretable model (Figure 2.4a)[119]. In short, the model postulates that associative learning occurs by adjusting the synaptic weights between "sensory" and "decision" neural populations described by population firing rate variables. At the input, two populations are specific for the S+ and S- sounds respectively, which we denote as $\hat{S}$+ and $\hat{S}$-, and one population, $\hat{C}$, which represents information common to S+ and S- trials (e.g. overlap between the S+ and S- representations or activity independent of sound, for example, related to visiting the lick port). Population $\hat{C}$ is an essential element of the model to reproduce high initial 'hit'-rates combined with delayed discrimination learning[119]. The "decision" population has two ensembles: one promoting and one inhibiting licking (Figure 2.4a). Adjustment of synaptic weights happens through a Hebbian learning rule modulated by Rescorla and Wagner's (1972)[197] δ-rule which gates weight updates by the reward expectation error. However, the employed δ-rule is asymmetric, meaning that the learning rate is larger by a factor v when an unexpected reward occurs, as compared to when an expected reward does not occur. This asymmetry is crucial for capturing the asymmetry of the learning process (i.e. the fast adjustment of the 'hit'-rate, and slower adjustment of the 'correct rejection'-rate; Figure 2.3b). In addition, synaptic updates are multiplicative, meaning that weight

updates are proportional to the current weight value[198, 199, 200, 201]. The key feature of multiplicative learning is that learning speed depends on the current strength of the synapses. Thus, the same model can have very slow learning (low weights) as in the delay phase and faster learning (high weights) as in the learning phase. Furthermore, this feature makes learning dynamics highly sensitive to the initial synaptic weights, which become important parameters that can even account for most of the inter-individual variability[119]. Most importantly here, the activity level of the Ŝ+ and Ŝ- populations, typically set to 1 (arbitrary units), can be varied in the model, allowing us to simulate the impact of cortical recruitment by setting either Ŝ- or Ŝ+ activity level to 2.
We therefore wondered whether this model reproduces the observed relationship between cortical recruitment and mean learning speed and its variability. The dynamics of the model depend on the choice of its three core parameters (noise level, learning rate and asymmetry, see Methods) and of its initial synaptic weights. As shown previously[119], individual learning curves could be fitted by adjusting these parameters, even without accounting for recruitment difference. Nevertheless, to test if the model captures the effect of recruitment, one can use a realistic set of parameters and test if asymmetric recruitment produces the effects seen during the behavior. We thus looked at the qualitative behavior of the model using a set of parameters obtained in a previous group of experiments[119] by fitting the individual learning curves from 15 mice trained in a task identical to the one used in this study. This parameter set included 15 different values of the initial weights, and core parameters were identical for all mice, which we showed is sufficient to account for inter-individual variability[119]. Based on these parameters, and systematically varying the recruitment values in simulations, we observed that recruitment differences were positively correlated to learning phase duration and variability in the model, similar to the experimental results (Figure 2.4b). Also, as illustrated when neuronal recruitment is doubled for one of the two stimuli, the model reproduced two other experimental observations. First, the delay phase was not significantly influenced by recruitment (Figure 2.4c). Second, the variability of the learning phase duration was much stronger when S- recruited less activity than S+ (Figure 2.4c,d). Thus, without any tuning, the model qualitatively reproduced the complexity of the experimental dynamics. This model thus provides a testing ground for important factors influencing the complex effects of stimulus-induced cortical recruitment on behavior in a precise theoretical framework.

Learning speed effects related to neuronal recruitment crucially depend on initial synaptic strengths in the model

To understand the origin of neuronal recruitment-based effects in our simulations, we plotted the learning phase duration against the values of the three initial synaptic weight parameters (w_{Ce} and w_{Ci} for the Ĉ population, and a single initial value w_S for the four weights of the Ŝ+ and Ŝ- population, as in Figure 2.4a). First, we observed that the initial weight between

sound-specific neural populations and the decision populations (w_S) had no correlation with learning speed duration. This was expected as w_S mostly impacts the delay phase and not the learning phase. Indeed, because of the multiplicative rule, small initial synaptic weights lead to slow initial learning, creating a period in which the performance does not improve above noise (delay phase). When sufficient learning has occurred (end of delay phase), the sound-specific synaptic weights become large enough to increase performance at a high learning speed. At this stage, the initial weight value is virtually "forgotten" and does not influence learning speed anymore. Our earlier fitting results[119] showed that w_S is the main determinant of the delay phase duration and is highly variable across mice. In our simulations, inter-individual w_S variations induced large variations of delay phase duration masking the smaller impact of neuronal recruitment on delay phase. Thus, our model suggests that the independence between cortical recruitment and delay phase duration in our experiments is due to inter-individual variability in initial connectivity.

A second observation was that the learning phase duration is large when the initial weights w_{Ce} and w_{Ci} (non-specific population $\hat{C}$) are small, but only when S- recruits less activity than S+ (Figure 2.4d). In contrast, for large weights, the recruitment differences between S+ and S- have no effect on learning phase (Figure 2.4d). Thus, our model suggests that the Go/NoGo task does not systematically reveal the influence of neuronal recruitment on learning speed, which the model predicts to appear only in some animals, depending on their initial synaptic state. This phenomenon could explain the large variability observed when cortical recruitment for S- is smaller than recruitment for S+ (Figure 2.3e).

To better understand why the initial conditions of w_{Ce} and w_{Ci} gate the influence of neuronal recruitment on learning speed, we plotted the time-course of both excitatory and inhibitory connection weights for four combinations of recruitment and initial weights values (Figure 2.5a-d). The performance plots illustrate that learning phase is strongly prolonged when neuronal recruitment is lower for S- than for S+ only when w_{Ce} and w_{Ci} are small (Figure 2.5d). The synaptic plots show that this large prolongation of the learning phase is due to the low initial weights from both the $\hat{S}$- and $\hat{C}$ to the NoGo decision population. These low weights are maintained during the delay phase and because of the multiplicative rule strongly slow down learning. The performance of correct rejection responses to S- takes longer and flattens the overall learning curve. In contrast, when synaptic weights from the $\hat{C}$ population to the NoGo population are initially high, more rapid S- rejection learning can be obtained solely based on the $\hat{C}$ common population (Figure 2.5c).

When S- recruits more activity, learning is always fast (Figure 2.5a-b) because, in all cases, the rate limiting process remains the abolition of licking to the NoGo stimulus (due to learning rule asymmetry, see Supplementary Figure 2.9). This process is boosted by strong $\hat{S}$- recruitment. Also, in these conditions, acquisition of the NoGo stimulus is independent of w_{Ce} and w_{Ci} , because the $\hat{C}$ population drives the Go response. Thus, the refined analysis

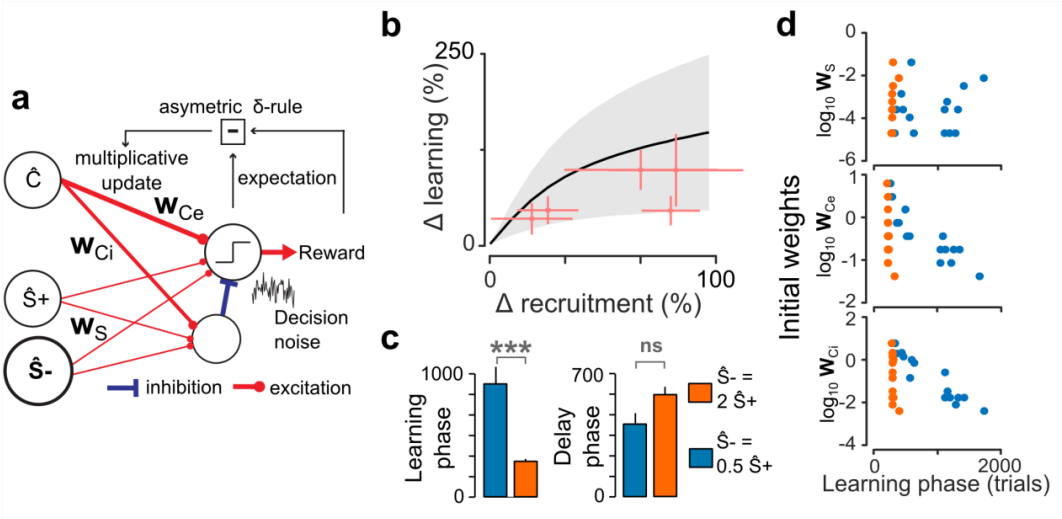

FIGURE 2.4: **A multiplicative reinforcement learning model reproduces modulations of learning speed by neuronal recruitment.**

*(**a**) Schematics describing the auditory Go/NoGo discrimination model. (**b**) Mean ± standard deviation of the difference of learning phase duration against difference in neural recruitment by the stimuli for the model initialized with 15 initial conditions obtained by fitting learning curves from a previous study[119]. The experimental observations of Figure 2.3d are superimposed in red. (**c**) Mean ± standard errors for the learning and delay phases obtained with the model for a two-fold difference in cortical recruitment between the two stimuli. A longer learning phase (Kolmogorov-Smirnov test, p=8.10^{-7} , n=15 initial conditions per group) but not delay phase (Kolmogorov-Smirnov test, p=0.13, n=15 initial conditions per group) is observed when S- recruits less activity (blue) as compared to when S- recruits more activity than S+ (orange). (**d**) Learning phase duration plotted against the value of the modeled initial synaptic weights. Same simulations and color code as in **c**. Significant correlations were observed only when S- recruits less activity (blue), and for w_{Ce} (ρ = -0.60, p = 0.016, n=15) and w_{Ci} (ρ = -0.68, p = 0.005, n=15).*

of the model indicates that the complex modulation of learning phase duration by neuronal recruitment is due to a non-trivial assignment of the three "sensory" populations to either Go or NoGo responses, based on neuronal recruitment distribution. Specifically, when the Ŝ- population recruits more activity, it is assigned to NoGo, while the Ĉ population drives the Go response. In contrast, when the Ŝ+ population recruits more activity, it is assigned to Go, and in this case, the Ĉ population drives the NoGo response. Thus, different neuronal recruitment distributions result in different solutions of the binary discrimination problem based on the three available cues. These solutions could be seen as different strategies chosen by the model or eventually the animal during the learning process.

Interestingly also, this provides a very general and testable prediction of reinforcement learning models using population activity as a salience parameter. The test would be to isolate and drive the neurons potentially corresponding to Ĉ in the brain. Activation of the Ĉ population alone should drive

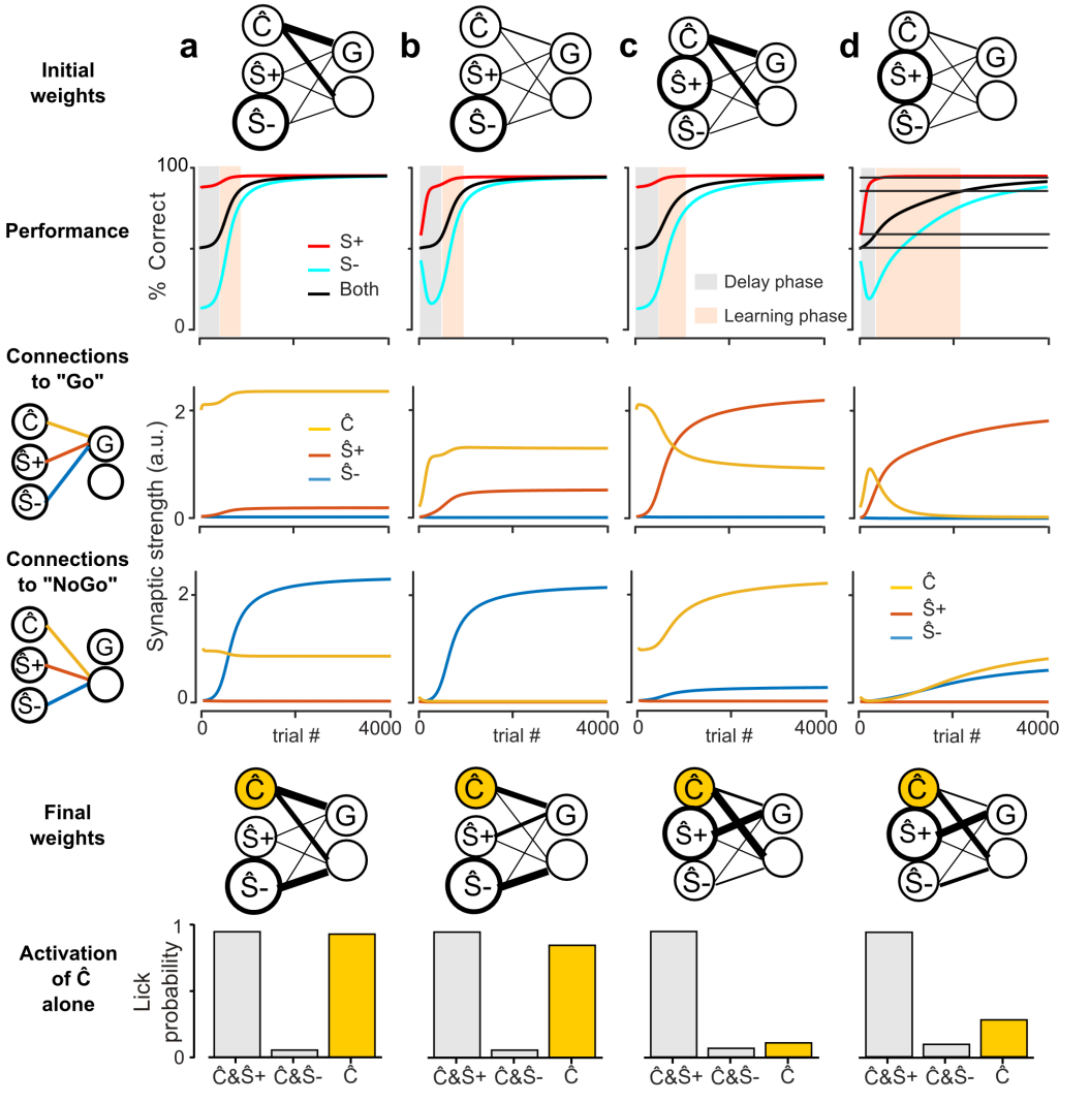

FIGURE 2.5: **The effects of neuronal recruitment in the model are explained by the differential adjustment of synaptic weights during learning.**

*(**a**) (top) Sketch of the initial synaptic weights and simulated model performance for S+ (red), S- (light blue) and both stimuli (black). (middle) Values of the connections to the excitatory (G) and inhibitory decision populations as indicated by the schematics on the left-hand-side. Yellow: connections from the "common" $\hat{C}$ population. Red: connections from the $\hat{S}$+ population. Blue: Connections from the $\hat{S}$- population. (bottom) Sketch of the connectivity pattern after and response probability after learning for the S+ (co-activation of $\hat{S}$+ & $\hat{C}$) and S- (co-activation of $\hat{S}$- and $\hat{C}$) reinforced stimuli as well as for the common stimulus component alone ($\hat{C}$, yellow). Simulation parameters: $x = [1; 0.66; 1.33]$, $\alpha = 0.01$, $\sigma = 0.6195$, $v = 6$, $w_{Ci} = 1$, $w_{Ce} = 2$, and $w_S = 0.01$. (**b**) Same as **a**, but with $w_{Ci} = 0.1$, $w_{Ce} = 0.2$. (**c**) Same as **a**, but with $x = [1; 1.33; 0.66]$. (**d**) Same as **c**, but with $w_{Ci} = 0.1$, $w_{Ce} = 0.2$.*

licking when S- recruits more activity, and should not drive licking when S- recruits less activity (Figure 2.5).

Biological learning associates neural populations to different responses based on activity recruitment

Testing this prediction was impractical with our sound-based Go/NoGo discrimination protocol, because the neurons encoding common information between S+ and S- trials ($\hat{C}$ population) likely code for multiple cues, including (i) the overlap of S+ and S- representations and (ii) all cues related to the decision to visit the lick port, and cannot be isolated. Therefore, we decided to test the model predictions in an artificial but better controlled experiment in which head- fixed mice had to discriminate optogenetically driven cortical ensembles. To do so we used a custom-made setup, based on a video projector[202] to project precise 2D light patterns through a cranial window placed above the auditory cortex in Emx1-Cre x Ai27 mice (Figure 2.6a and b, Supplementary Figure 2.10). Using intrinsic imaging, we identified the main tonotopic fields of mouse auditory cortex (Figure 2.6c)[115, 203] and thereby reliably positioned optogenetic stimulation spots in homologous regions across mice (Figure 2.6c). We defined three circular optogenetic stimulation spots out of which we constructed two stimuli. One of the three spots, the $\hat{C}$ spot, was common to the two S+ and S- stimuli and the two other spots corresponded to the stimulus-specific neuronal populations $\hat{S}$+ and $\hat{S}$- (e.g. Figure 2.6b). Thus, the cue S+ used for Go-trials consisted of simultaneous activation of $\hat{S}$+ and $\hat{C}$ and the cue S- consisted of simultaneous activation of $\hat{S}$- and $\hat{C}$. Furthermore, we were able to exclude cues common to S+ and S-, related to the visit of the port because, in the head-fixed task design, mice did not initiate the trials, which occurred at random time intervals. Thus we ensured that the only cue common to the S+ and S- stimuli was the $\hat{C}$ spot. We doubled the diameter of either the $\hat{S}$+ or the $\hat{S}$- spot to create a difference in cortical recruitment between the two input representations as in the model. Electrophysiological measurements of the population firing rate elicited by the small and large spots showed that the recruitment difference between the stronger and weaker stimuli was about 110% (Supplementary Figure 2.10), comparable with the population differences observed for sounds (Figure 2.4b). Mice were then initially trained to obtain a water reward by licking after the coincident activation of the $\hat{S}$+ and $\hat{C}$ spot. When 80% performance was reached in this detection task, the discrimination training started. During this stage, activation of the $\hat{C}$ population continued to occur in all trials: in half of the trials together with $\hat{S}$+ and in the other half together with $\hat{S}$-. Mice kept licking in the presence of the S+ spot and learned, within hundreds of trials to avoid licking in the presence of the S- spot (Figure 2.6d), reaching a steady state performance of 94.7% $\pm$ 4.5% correct trials (hit rate 93.9% $\pm$ 3.3%, false alarm rate 4.5% $\pm$ 1.3%, n=8 mice, see Figure 2.6e,f). Importantly, in this task setting, the stringent definition of the common $\hat{C}$ population, activated during the initial motivation training, likely resulted in the systematic establishment of strong initial connections for this population at the beginning of the discrimination training, leading to homogenous durations of the learning phase (212 $\pm$ 117 trials for the large $\hat{S}$- vs 260 $\pm$ 220 for the small $\hat{S}$-, p = 0.26, Wilcoxon rank-sum test, see also

Figure 2.6d) independent of recruitment (Figures 2.4 and 2.5).

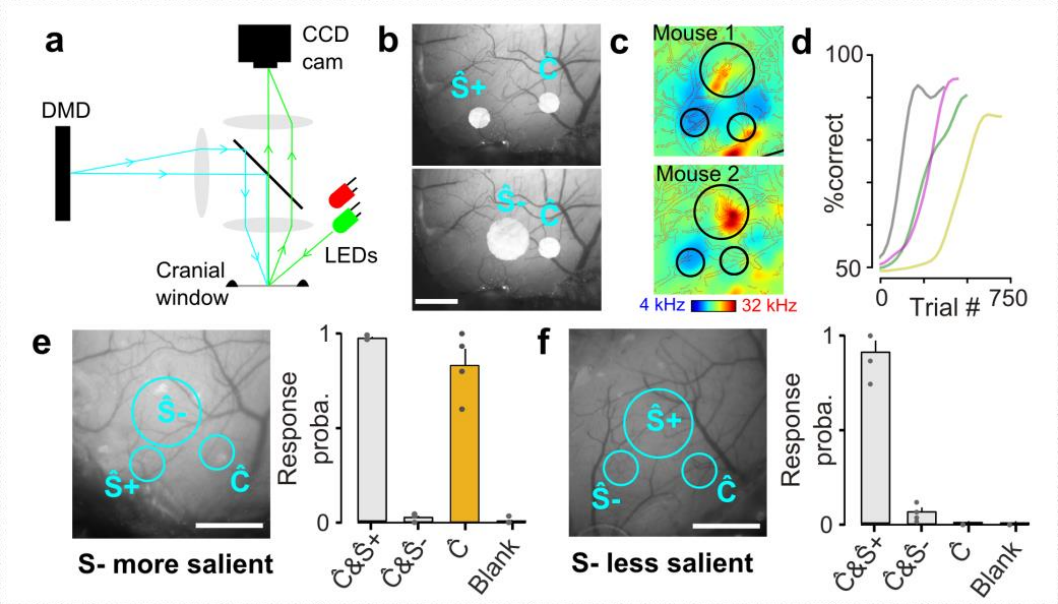

FIGURE 2.6: **Discrimination training of multi-spot optogenetic patterns reveals a choice of learning strategy depending on the level of cortical recruitment**

*(**a**) Schematics of the optical setup to project arbitrary 2D light patterns onto the surface of auditory cortex. Blue light patterns from a Digital Micromirror Device (DMD) are collimated and deflected through the objective lens by a beam splitter. The surface of the cranial window can be simultaneously imaged by a CCD camera using external LEDs (green for blood vessel, red for intrinsic imaging). (**b**) Examples of light patterns used for the discrimination task. Scale bars: 600 µm. (**c**) Tonotopic maps obtained with intrinsic imaging for two mice, and localization of the three optogenetic stimulation spots in the same high, low and mid frequency fields. (**d**) Four example learning curves for the optogenetic discrimination task. Grey: smaller Ŝ- ; color: larger Ŝ-. (**e**) (left) Light pattern in the task in which the S- stimulus has higher level of cortical recruitment (Ŝ- > Ŝ+). Ĉ = common component of S+ and S- stimuli. (right) Response probability on the two learnt target stimuli (Ĉ & Ŝ+ vs Ĉ & Ŝ-) and on presentation of the common part of the stimulus alone in 15 catch trials (yellow). Mean ± SEM, n = 4 mice. Gray dots: single animals. (**f**) Same as **e**, but with Ŝ+ larger than Ŝ-. Mean ± SEM, n=4 mice. Scale bars: 600 µm.*

However, once mice had learned the behavioral task, we measured their response to Ĉ activation alone in catch trials that were not rewarded (catch trial probability = 0.1; 15 catch trials per mouse). In the group of mice that had a larger Ŝ- spot, we observed that activation of Ĉ elicited strong licking responses (83% ± 10% response probability, n = 4 mice, Figure 2.6e). In contrast, in the group of mice that had a larger Ŝ+ spot, activation of Ĉ elicited no licking response (0% ± 0% licks per trials, n = 4 mice, Figure 2.6f). These results confirm the model prediction, and demonstrate, in a causal manner, that cortical recruitment affects the choice of which stimulus is associated to

a particular response. In this case, we show that the same stimulus ($\hat{C}$ activation) bearing neutral behavioral meaning is assigned by learning to be subthreshold when a larger population drives a suprathrehold response, and is assigned to be suprathreshold when a larger population drives a subthreshold response. Even if simple in essence, this result shows that cortical recruitment is a parameter influencing learning, in a manner compatible with the role of a salience parameter in reinforcement models.

Discussion

Combining behavioral measurements, large scale two-photon imaging, optogenetics and theoretical modeling, we have shown that sounds of different quality but equal mean pressure levels can recruit highly variable levels of neuronal activity in auditory cortex, measured as the mean amount of activity in a representative subsample of neurons. We showed that cortical recruitment levels correlate with learning speed effects in a Go/NoGo task as expected if neuronal recruitment corresponds to stimulus salience. Moreover, the details of these can be precisely reproduced by a reinforcement learning model of the task. Finally, training mice to discriminate optogenetically evoked cortical patterns, and manipulating these patterns, we showed that neuronal recruitment determines which elements of the cortical representation are selected to drive each conditioned action. This corroborates, in a causal manner, the idea that cortical recruitment is a neuronal correlate of stimulus salience.

Several studies indicate that cortical recruitment can vary across stimuli, even when played at the same sound pressure level [191, 193, 44]. These discrepancies may have multiple origins. First, it is well known that the mouse cochlea is more sensitive to some sound frequencies than others[194], which could explain the overrepresentations of sounds in the middle frequency range (10-30kHz) in cortex[193]. In this case, cortical recruitment is expected to reflect recruitment throughout the auditory system, making it a good proxy for sound salience independent of whether the discrimination task requires auditory cortex[204, 120, 205] or does not require it[206, 207, 208]. But a second source of recruitment differences may be the nonlinearities of cortical representations[44, 209, 83, 64]. For example, a recent study suggested that cortical response patterns can be invariant to changes in intensity[210]. In this case, cortical recruitment should also depend on the higher level features composing the representation of a given sound and on how broadly these features are represented in cortex.

The idea that the amount of neuronal activity recruited by a stimulus influences behavior has been proposed in different contexts. For example, several studies indicate that attention can boost the activity of the neurons representing behaviorally relevant stimuli[173, 211, 212] and thereby make it more discriminable from other stimuli[213]. Also, several theoretical studies have proposed that attention impacts learning[214, 215] and some reinforcement learning models can account for such effects by dynamically weighting stimuli according to their predictive relevance[216]. It will be an interesting

research avenue to analyze the relative contribution of bottom-up sound encoding and attentional top-down mechanisms to the level of cortical recruitment. Earlier reports, using direct microstimulation of the cortex, showed that low levels of neuronal recruitment can impact detection probability[217, 218]. Here, we show that neuronal recruitment for stimuli that are well beyond detection threshold still impact the learning process by which they are associated to particular responses. Even if such effects are predicted by the Rescorla and Wagner model[154], capturing their details requires a refinement of the original model. In particular, we had to introduce a more realistic multiplicative learning rule which renders learning speed not only dependent on neuronal recruitment, but also on the current synaptic strength. This property has important consequences. First, it introduces variability in the relationship between recruitment and learning speed, through large inter-individual variations of the synaptic weights present at the beginning of the task. Second, the fact that learning speed is proportional to the product of neuronal recruitment and connectivity, allows the system more flexibility, in particular by compensating weak neuronal recruitment with stronger initial connections. In our experiments, this phenomenon tends to stabilize learning speed, explaining why neuronal recruitment does not always impact the learning phase duration, except for particular initial conditions for which compensation occurs too slowly (Figures 2.3 and 2.5). More generally, strong pre-established connections can help implementing fast learning for specific stimuli with innate meaning.
These complex dynamical phenomena make learning speed measurements a more complicated proxy for stimulus salience than the overshadowing protocol which relies on steady state behavior, after the dynamical phase of the association. However, it allows the comparison of salience for stimuli from the same sensory modality. Because our extended multiplicative model only diverges from the Rescorla-Wagner model for the transient dynamical part of the association process (delay and learning phases), it also reproduces overshadowing effects[119], and more generally predicts which part of a sensory input representation is assigned to which conditioned response in a more complex task. Here, by conditioning mice to compound stimuli composed of multiple optogenetically activated neuronal ensembles of different sizes in auditory cortex (Figure 2.6), we show, in line with reinforcement learning models, that the brain can actually establish a stimulus discrimination strategy based on the amount of activity recruited by the different subpopulations representing the stimuli.

Methods

Animals.

All mice used for imaging and behavior were 8 to 16 weeks old C57Bl6J and GAD2-Cre (Jax #010802) x RCL-TdT (Jax #007909) mice. Mice used for optogenetics were 8 to 16 weeks old males and female obtained by crossing homozygous Emx1 IRES-cre (Jax #005628) mice with Ai27 (Jax #012567) mice

to obtain expression of Td-Tomato-tagged channelrhodopsin (ChR2) in excitatory neurons of the cortex. All animal were group housed. All procedures were approved by the Austrian laboratory animal law guidelines (Approval #: M58 / 02182 / 2007 /11; M58 / 02063 / 2008 /8) and the French Ethical Committee (authorization 00275.01).

Two-photon calcium imaging in awake mice.

At least three weeks before imaging, mice were anaesthetized under ketamine medetomidine. The right masseter was removed and a large craniotomy (5 mm diameter) was performed above the auditory cortex. We then performed three to five injections of 200nL (35-40nL/min), rAAV1.Syn.GCaMP6s.WPRE virus obtained from U. Penn Vector Core (Philadelphia, PA) and diluted 10 times. The craniotomy was sealed with a glass window and a metal post was implanted using cyanolite glue and dental cement. At least three days before imaging, mice were trained to stand still, head-fixed under the microscope for 20 to 60 min per day. Then mice were imaged one to two hours per day. Imaging was performed using a two-photon microscope (Femtonics, Budapest, Hungary) equipped with an 8kHz resonant scanner combined with a pulsed laser (MaiTai-DS, SpectraPhysics, Santa Clara, CA) tuned at 920 nm or 900nm depending on the experiments. Images were acquired at 31.5Hz. All sounds were delivered at 192 kHz with a NI-PCI-6221 card (National Instrument) driven by Elphy (G. Sadoc, UNIC, France) through an amplifier and high frequency loudspeakers (SA1 and MF1-S, Tucker-Davis Technologies, Alachua, FL). Sounds were calibrated in intensity at the location of the mouse ear using a probe microphone (Bruel&Kjaer). In a first experiment, we played three 70 ms complex sounds at 73dB SPL preceded by two 50 ms 4kHz pure tones (inter-tone interval: 375 ms) sounds as in the behavioral task. The three sounds were played in a random order and repeated 30 times. In a second experiment, we played white noise sounds ramping -up or -down in intensity between 60dB and 85dB SPL during 2s. The ramps were repeated 20 times. In a third experiment, we played 20 repetitions of white noise sounds modulated in intensity at 1 and 20 Hz.

Data analysis.

Data analysis was performed using Matlab and Python scripts. Motion artifacts were first corrected frame by frame, using a rigid body registration algorithm. Regions Of Interest were selected using a semi-automated hierarchical clustering algorithm based on pixel covariance over time as described in 30 (see detailed method below). Neuropil contamination was subtracted[219] by apply the following equation: $F_{true}(t) = F_{measured}(t)\check{}0.7F_{neuropil}(t)$, then the change in fluorescence ($\Delta F/F_0$) was calculated as $(F - F_0)/F_0$, where F_0 is estimated as the minimum of the low-pass filtered fluorescence over $\sim$40 s time windows period. To estimate the time course of firing rate, the calcium signal was temporally deconvolved using the following formula: $r(t) = f'(t) + f(t)/\tau$ in which f' is the first time derivative of f and τ the

decay constant set to 2 seconds for GCaMP6s. For the complex sounds, the population response was computed as the mean deconvolved signal across all neurons from sound onset to 500ms after sound offset. For the ramps, because the behavioural discrimination response typically occurs within few hundreds of milliseconds after the ramp onset[44], the mean population response was evaluated from 0 to 500ms after sound onset. To estimate the discriminability of two sounds based on cortical population responses, linear Support Vector Machine classifiers were trained to discriminate population activity vectors obtained from 20 presentations of each sound (training set), and were tested on activity vectors obtained for 10 independent presentations of the same sounds (test set).

Patterned optogenetics and intrinsic imaging.

To flexibly activate different activity patterns in the mouse auditory cortex, we used a computer driven (VGA input) video projector (DLP LightCrafter, Texas Instruments) which includes a strong blue LED light source (460 nm) and from which we have removed the objective. To project a two-dimensional image onto the auditory cortex surface (Figure 2.6a, b), the image of the micromirror chip is collimated through a 150 mm cylindrical lens (Thorlabs, diameter: 2 inches) and focused through a 50 mm objective (NIKKOR, Nikon). Imaging of the cortex at the focal plane is obtained by side illumination with a green (525 nm, blood vessels) or far red (780 nm, intrinsic imaging) LED. The light collected by the objective passes through a dichroic beamsplitter (long pass, >640nm, FF640- FDi01, Semrock) and is collected by a CCD camera (GC651MP, Smartek Vision) equipped with a 50 mm objective (Fujinon, HF50HA-1B, Fujifilm). Note that the image projected to the cortical surface corresponds to a narrow cone of light extending below the surface and potentially activating ChR2 expressing neurons throughout the cortical depth. Intrinsic imaging was performed in isoflurane anesthetized mice (1.1% delivered through SomnoSuite, Kent Scientific). To compute intrinsic signal maps we divided the red light image of the cortical surface after the onset of a stimulation (average over 2 s) with 2s long pure tones (4, 8, 16 and 32kHz) by the mean image immediately before stimulus onset.

Silicon probe recordings.

For calibration of optogenetic stimulation, a small aperture was drilled in the cranial window with a diamond-coated dental drill during isoflurane anesthesia. 30 minutes after surgery, 4x8 silicon probes (Neuronexus) were implanted at a $\sim 35^{\circ}$ angle in auditory cortex in the awake head-fixed mouse. Recordings were performed at three different depths (400, 600 and 800 μm) using a pre-amplifier and multiplexer coupled to a USB acquisition card (Intan Technologies). Sounds and light stimulations were randomly presented at 2.5 sec and each repeated 10 times. Single unit spikes were detected and sorted from multi-unit spikes using the Phy Suite (https://github.com/kwikteam/phy).

Go/NoGo discrimination behavior.

Mice were water-deprived and trained daily for 200 to 300 trials. Mice first performed 4 habituation sessions to learn to obtain a water rewards ($\sim$5 μl) by licking on a spout over a threshold after the positive stimulus S+. After habituation, the fraction of collected rewards was about 80%. The learning protocol then started in which mice also received a non-rewarded, negative sound S- for which they had to decreasing licking below threshold to avoid an 8s time-out. For the freely moving complex sound discrimination, S+ and S- sounds consisted of two 4 kHz pips (50 ms) followed by one of the three 70 ms complex click shown in Figure 2.1a. The interval between the offset and onset of the pips and click was 375 ms. Licking was assessed 0.58 sec after the specific sound cue in a 1 sec long window by an infrared beam at the spout. For the intensity ramp discrimination, licking was assessed in a 1.5s window after sound offset. In both cases, licking was considered above threshold if the infrared beam in front of the licking tube was broken during 75% of the measurement time-window. Positive and negative sounds were played in a pseudorandom order with the constraint that exactly 4 positive and 4 negative sounds must be played every 8 trials. For learning curves, performance was measured as the fraction of correct positive and correct negative trials over bins of 100 trials. For the optogenetically-driven, head-fixed discrimination task, the S+ and S- stimuli were each composed of two disks of blue light (465 nm) flashing at 20 Hz for 1 s. One of the two disks (noted $\hat{C}$) was common to S+ and S- stimuli, the other disk was condition- specific. The three disk locations were chosen in similar tonotopic locations across mice based on intrinsic imaging maps. They were precisely re-positioned for every training session using an automated registration procedure based on blood-vessel patterns. A strong masking light was used to prevent the animal from using visual cues in the task. The common disk was 360 μm in diameter. In one set of mice, the disk specific to S- was 720 μm in diameter, while the S+ specific disk was 360 μm in diameter. In the other set of mice, sizes of the specific disks were swapped. Head-fixed mice performed 200 to 300 trials per day with an inter-trial interval randomized between 3 s and 7 s. Individual licks were detected through an electric circuit connecting the mouse and the lick tube. Then, each trial was started only if the mouse was not spontaneously licking for at least 3 s (in addition to the inter-trial interval). Mice were first trained to respond to the S+ stimulus by producing at least 3 to 5 licks (depending on the mouse) to get the 5 μL water reward. When the mouse could collect more than 80% of the rewards, the S- stimulus was introduced. Licking above threshold after S- was punished with a 7 s timeout.

Reinforcement learning model.

The model has been described extensively in a previous publication[119]. In short, it is composed of three sensory units ($\hat{S}$+, $\hat{S}$- and $\hat{C}$, representing populations of neurons) whose activity described by a three-dimensional vector $\vec{x}$ and which are connected to a simple decision circuit (Figure 2.4a). Cortical recruitment is modeled by changing the firing value of the sound units.

When the S- stimulus recruits less activity than the S+ stimulus, the input vectors are: $\vec{x}_{S+}$ = [1 0 2] or $\vec{x}_{S-}$ = [1 1 0]. When S- recruits more activity than S+, the input vectors are: $\vec{x}_{S+}$ = [1 0 1] or $\vec{x}_{S-}$ = [1 2 0].
The decision circuit is composed of all-or-none response unit (y = 0 or 1) which linearly sums the three sensory inputs (representing synaptic populations) under the form of three direct excitatory connections and of a graded feed-forward inhibition from a virtual inhibitory unit in fact equivalent to three direct inhibitory connections. The output of model is described by a single equation for the decision unit:

$$y = \Theta(\vec{w}_E \cdot \vec{x} - \vec{w}_I \cdot \vec{x} - \xi) \tag{2.1}$$

in which Θ is the Heaviside step function. $\vec{w}_E$ and $\vec{w}_I$ are three-dimensional positive vectors describing the excitatory synaptic weights from the sensory units to the decision and inhibitory units respectively. The variable ξ is a Gaussian random noise process of unit variance which models the stochasticity of behavioral choices.
Based on the action outcome (R = 1 for a reward, R = -1 for no reward), the learning rule for the synaptic weights is implemented as:

$$\delta\vec{w}_E = \alpha\vec{w}_E \odot f[R - \sigma(\vec{w}_E - \vec{w}_I) \cdot \vec{x}]y\vec{x} \tag{2.2}$$

$$\delta\vec{w}_I = -\alpha\vec{w}_E \odot f[R - \sigma(\vec{w}_E - \vec{w}_I) \cdot \vec{x}]y\vec{x} \tag{2.3}$$

in which $\odot$ is the Hadamard (element-wise) product implementing the multiplicative rule, $y\vec{x}$ is a Hebbian term, α is the learning rate and σ is a parameter related to the noisiness of the model and setting its asymptotic performance. To account for the faster improvement of performance for rewarded as compared to non-rewarded trials, positive expectation errors are more strongly weighted than negative ones, thanks to the asymmetric function $f[u] = u$ if $u \geq 0$ and $f[u] = vu$ if $u > 0$. The parameter v is typically larger than 1, consistent with the activity of basal ganglia dopaminergic neurons in mice[220] and monkeys[221] coding for reward expectation error.

As described above, the equations of the model are stochastic due to the Gaussian random noise process ξ. To compute the response probability estimates plotted throughout the study, we used a previously established probability equation[119], valid for learning dynamics much slower than fluctuations (ergodic approximation). The probability to make a lick response given the input vector $\vec{x_{S+}}$ or $\vec{yx_{S-}}$ is:

$$P_{S+|S-} = p(y = 1|\vec{x} = \vec{x_{S+|S-}}) = \frac{1}{2}\left(1 + erf(\frac{\Delta\vec{w} \cdot \vec{x_{S+|S-}}}{\sqrt{2}})\right) \tag{2.4}$$

where $\Delta\vec{w} = \vec{w}_E - \vec{w}_I$ represents now the average observed values of difference between the excitatory and inhibitory connections. In addition, the plasticity equations become:

$$\delta\vec{w}_E = \frac{\alpha}{2}\vec{w}_E \odot \left(f[1-\sigma\Delta\vec{w}\cdot\vec{x_{S+}}]p_{s+}\vec{x_{S+}} + f[-1-\sigma\Delta\vec{w}\cdot\vec{x_{S-}}]p_{s-}\vec{x_{S-}}\right) \quad (2.5)$$

$$\delta\vec{w}_I = -\frac{\alpha}{2}\vec{w}_E \odot \left(f[1-\sigma\Delta\vec{w}\cdot\vec{x_{S+}}]p_{s+}\vec{x_{S+}} + f[-1-\sigma\Delta\vec{w}\cdot\vec{x_{S-}}]p_{s-}\vec{x_{S-}}\right) \quad (2.6)$$

Statistical tests

Unless otherwise specified, all quantifications are given as mean ± standard error (SEM). To statistically assess the differences between paired measurements (e.g. activity for two different sounds elicited in the same neuronal populations) we used the non-parametric Wilcoxon signed rank test. To compare two sets of measurements (e.g. delay and learning phase duration for two groups of mice) we used the non-parametric Wilcoxon rank sum test. Assessment of the differences in the fraction of responsive neurons for different sounds was done with the χ^2 test which evaluates differences in the distributions of two binary variables.

Data and software availability

Datasets, analysis software and codes for running the simulations of our model are available upon request to Brice Bathellier (brice.bathellier@unic.cnrs-gif.fr).

Acknowledgements

We thank K. Kuchibhotla, E. Harrell and M. Stüttgen for comments on the manuscript, P. Pindi, S. Sikirić and L. François for help with behavioral and imaging experiments. We thank the GENIE Project, Janelia Farm Research Campus, Howard Hughes Medical Institute, for GCAMP6s constructs. This work was supported by the Agence Nationale pour la Recherche (ANR "SENSE-MAKER"), the Fyssen foundation, the DIM "Region Ile de France", the Marie Curie Program (CIG 334581), the International Human Frontier Science Program Organization (CDA-0064-2015), by the Fondation pour l'Audition (Laboratory grant), the École Doctorale Frontières du Vivant (FdV) – Programme Bettencourt (support to AK), the DIM Cerveau et Pensée and Ecole des Neurosciences de Paris Ile-de-France (ENP, support to SC) and the Deutsche Forschungsgemeinschaft (DFG CRC1080/2).

Author contributions

BB and SR designed the study. AK, SC, and BB performed and analyzed the imaging experiments. BB and JB performed the modeling. AD and BB performed and analyzed behavioral experiments. ZP designed the patterned optogenetic setup and SC performed the optogenetic experiments. TD designed software for data analysis and behavior. BB and SR wrote the manuscript with comments from all authors.

Supplementary Figures

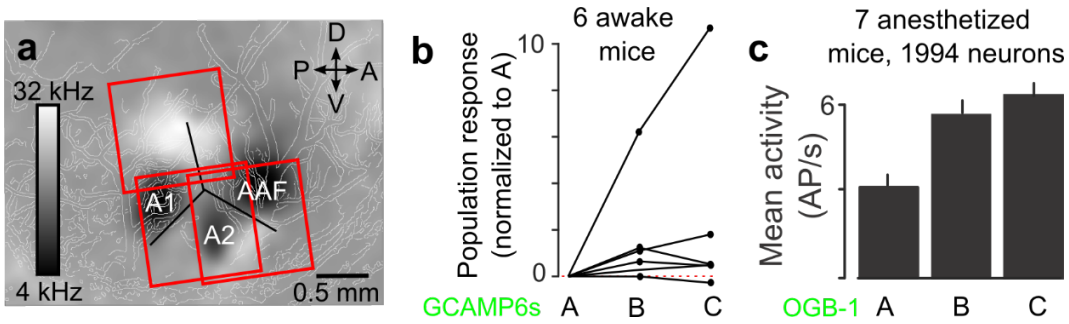

FIGURE 2.7: **Cortical recruitment differences are robust across mice and experiments.**

*(**a**) Localization and horizontal extent of the imaging fields-of-view (red squares) in one of the imaged mouse. The* $1 * 1$ *mm fields-of-view were localized with blood vessels patterns (white lines). The auditory cortex is identified by subtracting intrinsic imaging responses to 4kHz (black colors) and 32 kHz (white colors) tones revealing the main tonotopic gradients. (**b**) Responses of neural population recorded in each of the 6 awake mice included in the averages shown in presented in Figure 2.1 (activity is measured with GCAMP6s). Responses are measured as mean deconvolved calcium signals during the entire duration of the response (sound duration +0.5s). Normalization in obtained by subtracting and dividing with response to sound A. The distribution of responses to B and C are significantly above zero (n=6 mice, Wilcoxon signed rank test, p=0.031 for B and C) (**c**) Average population activity in OGB1 labelled neurons imaged with two-photon microscopy in 7 isoflurane anesthetized mice (28 population, 1994 neurons) for sounds A, B and C. Mean* $\pm$ *SEM, for A:* 3.1176 ± 0.3726 *spike/s, B:* 5.4527 ± 0.4982 *spike/s, C:* 6.3453 ± 0.4978 *spike/s. Responses to sound B and C are significantly different from A, but not from each other Wilcoxon rank sum test, p=0.0012, p=*$1.5 * 10^{-5}$*, p=0.31.*

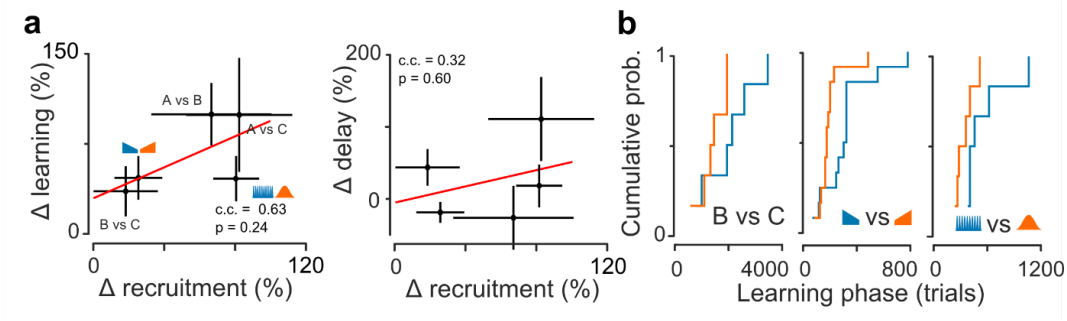

FIGURE 2.8: **Impact of cortical recruitment on learning and delay phase durations.**

*(**a**) Mean ± standard error of the difference of learning (left) and delay (right) phase durations plotted against the difference in neural recruitment by the stimuli. A correlation coefficient of 0.63 (p=0.24, n=5) is observed for the learning phase, while a weaker correlation of 0.32 is seen for the delay phase (p=0.60, n=5). (**b**) Cumulative distributions of learning phase durations for sound pairs B-C and for the up- and down-ramp.*

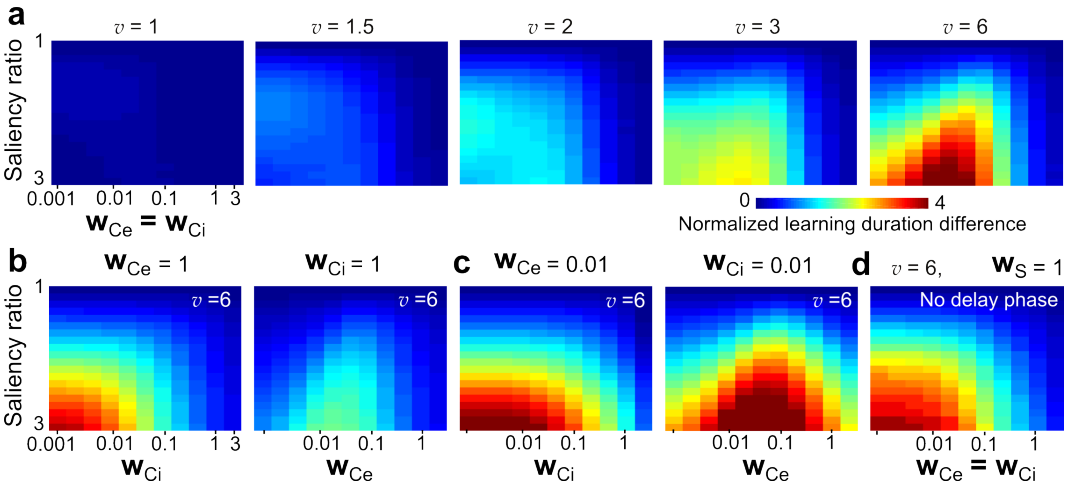

FIGURE 2.9: **The impact of neuronal recruitment on learning speed depends on the initial strengths of synaptic connections in the model.**

*(**a**) Normalized learning phase duration difference ((duration S- more salient) - (duration S- less salient))/(duration S- more salient) is color-coded and plotted against the cortical recruitment ratio of the discriminated stimuli and the strength of the initial inhibitory and excitatory connection to unspecific cues (w_{Ci} and w_{Ce}, same value for both). From left to right the value of the asymmetric learning rate parameter v is increased. Simulation parameters: $w_S = 0.03$, $\alpha = 0.01$, $\sigma = 0.6195$. (**b**) Same as **a**, but varying only the initial inhibitory (left) or excitatory connection (right) while the other connection is kept constant at value 1. (**c**) Same as (**b**), but with the non-varied initial connection kept at a constant value of 0.01. (**d**) Same as **a**, but now with initial connectivity from sound stimuli $w_S = 1$ such that there is no delay phase. In all plots, the recruitment ratio is equal to the larger of the two saliency values.*

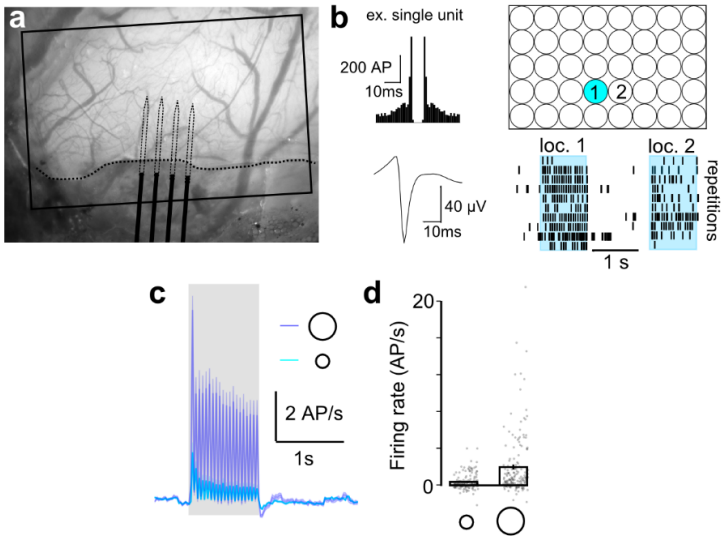

FIGURE 2.10: **Cortical responses to patterned optogenetic stimulations.**

*(**a**) Picture of the electrophysiolocal experiment setting. A 4-shanks silicon probe (sketched) is inserted with a shallow angle through a hole under the glass window (curved dashed line). The rectangle indicates the area covered by the videoprojector. (**b**) Example waveform (bottom left) and spike train autocorrelogram (top left) of a single unit. An array of* $5 * 8$ *stimulus locations was probed to reconstruct the spread of activity (top right). Example raster plots for two locations are shown at the bottom right. (**c**) Mean response time course for all probed locations and single units (188 single units, 2 mice, 5 recording locations), for the small (light blue) and large (dark blue) disk. (**d**) Mean population (bars) and single units (gray dots) firing rate estimated by averaging over the full grids of stimulus location for the small and large disk, in order to estimate the integral of the response elicited by one disk. At the population level the small disk elicits about 6 times less activity than the large disk (small* 0.36 ± 0.12 *Hz; large:* 1.7 ± 0.4 *Hz). Thus, the recruitment difference (as measured for sounds in Figure 2.4b and Supplementary Figure 2.8a) between the strong (1 small + 1 large disk) and weaker (2 small disks) stimulus is equal to* $2(S_{strong} - S_{weak})/(S_{strong} + S_{weak}) = 2(7-2)/(7+2) = 111\%$.

Context Chapter 3

This third study is not submitted yet, and it draws general conclusion on the auditory system based on the description of multiple non-linear features observed in the auditory cortex. From the first study we understood that the non-linear processing of sound intensity results in difference in cortical recruitment, and the second study emphasis the importance of this consequence on the behavior. We wanted to characterize the different non-linear processings in place in the auditory system, not only for sound intensity, but also in the frequency domain. We used calcium imaging to record the response of a very large population of cortical neurons to intensity- and frequency-varying sounds. We described 7 non-linear features encoded by subpopulations of neurons from the auditory cortex and we studied how multiple features can be extracted by the same subpopulation. We used this information to draw general conclusions on how non-linear transformations are organized in the auditory system. These conclusions can serve as constraints to design a complete model of the auditory system that takes non-linear processing into account.

My contribution to this article was to design the study, to generate the sounds, to perform the surgeries and some of the recordings. I also did the data analysis, the spatial functional maps, the hierarchical clustering, the description of the different non-linear features and their quantification, and the interactions between them. I designed and generated all the figures and I wrote the manuscript.

Chapter 3

Interactions between nonlinear features in the mouse auditory cortex

Alexandre Kempf[1], Thibaut Tarpin[1,2], and Brice Bathellier[1]

1. Unité de Neuroscience, Information et Complexité (UNIC), Centre National de la Recherche Scientifique, FRE 3693, Gif-sur-Yvette, 91198, France

2. Present Address : Institut de Biologie de l'Éccole Normale Supérieure, Inserm U1024, CNRS UMR8197, École Normale Supérieure, PSL Research University, France; Inserm UMR1120, University Paris 6, Institut Pasteur, France

Abstract

Human auditory perception involves a decomposition of the auditory scene into invariant auditory objects which requires complex non-linear operations throughout the auditory system. These non-linear transformations are still incompletely characterized. In this study, we used two-photon calcium imaging to record a total of 59 590 neurons from the mouse auditory cortex in response to 148 sounds, which is to our knowledge the largest auditory cortex activity sample at cellular resolution. From this large set of neural responses, we generated clusters of neurons according to the similarity of their temporal response patterns to the tested set of sounds. These clusters were spatially organized, and coded for diverse sound features. We identified within these features seven classes of nonlinearities ranging from non-monotonotic intensity tuning to non-additive response to chords or asymmetric responses to frequency modulated sounds. Interestingly, we observed that some pairs of non-linearities classes can be found in the same neurons while others exclude each other. Our results reveal a complex but non-random computational logic in the distribution of non-linear features in auditory cortex, setting new constrains for computational models of the auditory system.

Introduction

Hearing allows extracting information from the acoustic environment, eventually to identify relevant auditory objects[222, 223, 224]. The identification of auditory objects involve transformations of basic sound features extracted in the cochlea (frequency, amplitude) by high level brain structures like the auditory cortex [14]. The spectrotemporal structure and the amplitude modulation strongly influence object formation, and a gross encoding of these features might result in hearing impairment[223]. Recent evidences show that these transformations are highly nonlinear[58, 62, 1]. Even at the level of the cochlea, the nonlinear operations are so important[87, 88] that hearing aids that replace these operations by linear modification or amplification of the sound fail to restore hearing in a noisy environment[1]. At later stage, our understanding of the auditory system is poor in computational term in comparison to the visual system[37]. The limit is mainly set by our limited understanding of the nonlinearities involved in the auditory system. So far, several nonlinearities has been described at the level of the auditory cortex, like for example tuning to low intensity sounds[45, 225, 44] or direction selectivity in frequency modulated sounds (FMS)[226]. Unfortunately each nonlinear transformation of the sound is studied independently of the others and we still have a poor understanding on how theses transformations interact with each other. In a recent study focused on the asymmetric encoding of intensity ramps in the auditory cortex, Deneux and al[44] explains how two non-linear transformations, the encoding of quiet sounds and the encoding of onsets and offsets, are combine in cortical populations to generate an asymmetry in cortical activity. They show that populations encoding the quiet onset and the loud offset encode those features with more activity than the loud onset or the quiet offset, thus leading on a stronger cortical activity for up-ramps than for down-ramps. Here we aim at studying the interactions of multiple non-linear transformation and how the cortical population encode them. To do so, we record with calcium imaging the activity of 59 590 neurons from the auditory cortex in response to a set of 148 intensity and frequency modulated sounds. This dataset contains the recordings of 16% of the mouse auditory cortex with a cellular resolution[195] and to our knowledge it represents the largest dataset of cortical recording in the auditory system. We use hierarchical clustering to group neurons with similar activity patterns to the set of sounds, and to isolate distinct response patterns to sounds. Thanks to these patterns, we identify seven types of nonlinear transformations that emerge in the auditory system. We then compare functional populations of neurons to establish what non-linear transformations are encoded by the same populations, and what transformations are not. We describe four types of interactions between non-linear transformations and we interpret them to provide constraints for future models of the auditory system.

Results

Extensive recordings of the auditory cortex activity to sounds.

To access the diversity of responses in the auditory cortex, we recorded the activity of 59 590 cortical neurons in 7 mice in response to a set of 148 sounds (Figure 3.1a-b). The sound set was composed of 33 pure tones with frequencies ranging from 4kHz to 37kHz, sounds with the intensity modulated linearly or by a sine function from 60dB to 80dB, frequency modulated sounds, and chords (see Methods). (Figure 3.1b). We used chronic calcium imaging using GCAMP6s (Figure 3.1c,e) in awake mice and we recorded cortical neurons from -100μm to -600μm below pia (equivalent of layer II/III to layer V) (Figure 3.1d). In order to determine the location of our recordings site relative to the auditory cortex, and to verify that the sampling was covering all the auditory cortex, we identified the main auditory cortex regions with intrinsic optical imaging (IOI). We took advantage of the IOI response to a pure tone burst of 4kHz that produces a three spots pattern to obtain the location of the low frequency area of the three main regions of the auditory cortex (primary and secondary auditory cortex, A1 and A2, and the anterior auditory field, AAF)[115, 227] (Figure 3.1f). Once we localize these three regions of the auditory cortex, we rotate and align the IOI functional maps obtain across mice to obtain a frame of reference. Then we localized every field of view recorded with calcium imaging on the functional frame of reference thanks to the blood vessel patterns that are stable landmarks within the same animal. With this process, we could create a densely sampled, high resolution, map of the auditory cortex in three dimensions (Figure 3.1c,f). We confirmed with the coverage of recording sites (Figure 3.1d middle) that we recorded activity all over the auditory cortex. Interestingly we found that the general spatial pattern to a 4kHz sound obtain with calcium imaging at a cellular resolution reproduces the mesoscopic pattern of activation found with IOI for a 4kHz burst (Figure 3.1f). Although this observation support the tonotopy organization at the mesoscopic scale, the spatial pattern observe at the cellular resolution shows that the encoding of frequency is more salt and pepper at the microscopic scale, with some neurons responding for 4kHz outside of these three regions, and some neurons in these regions that do not respond for 4kHz sounds. We looked at the spatial organization of the auditory cortex for pure tones of different frequencies, intensities. As expected with a tonotopic organization, the spatial patterns of activation change with frequencies, but interestingly, the spatial organization is also modulated for sound intensity. We also observe a difference in the tonotopic maps for the sound onset compared to the sound offset (Supplementary Figure 3.6). The spatial patterns for responses at the offset of the sounds are less strong and they sometimes involve regions that are not activated during the onset of the sounds (Supplementary Figure 3.6). We did not observe any organization in the spatial patterns with depth, supporting the idea that the auditory cortex is functionally organized in a columnar fashion at least from layer 2/3 to layer 5.

The functional map obtained with cellular resolution from our recording confirm the tonotopic organization of the mouse auditory cortex and show that pure tones are encode by localized populations of neurons. We wondered if there is also a spatial organization in the response for more complex sounds.

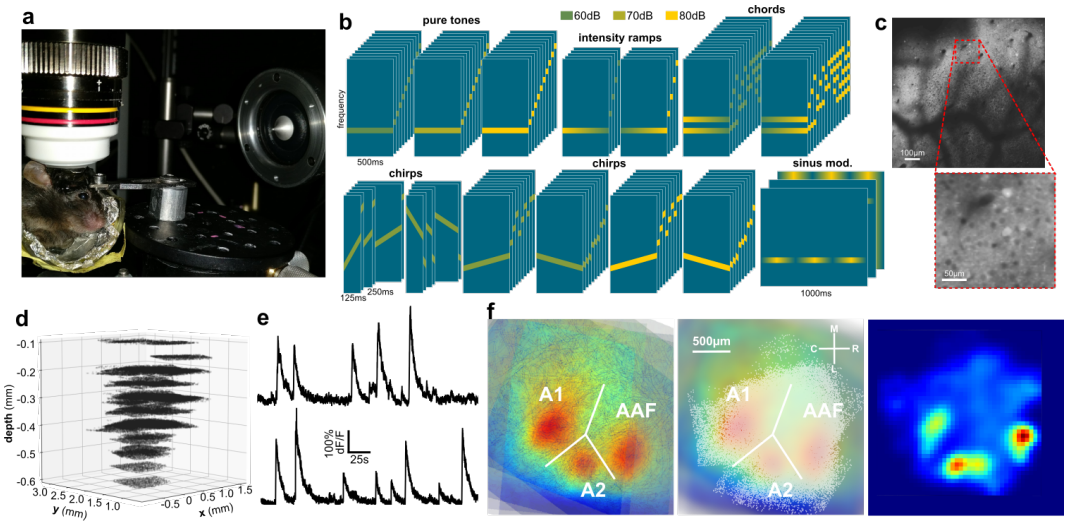

FIGURE 3.1: **Large population recordings with calcium imaging cover entirely the auditory cortex.**

*(**a**) Awake head fixed mouse under the two-photon microscope. A high frequency speaker is placed contralateral to the craniotomy on the auditory cortex. (**b**) Sketch of the set of sounds represented as spectrogram. The frequency axis has a logarithmic scale. All the sounds are 500 ms long except when specified. (**c**) Example of a field of view and a zoom of a recording in the auditory cortex. (**d**) Three dimensional location of our recordings sites. Depth ranges from -100 to -600 µm but most of the recordings (66%) have been performed in layers 2/3 (from -100µm to -400µm) (**e**) Example of calcium imaging traces obtained with GCAMP6s. (**f**) Superimposed intrinsic optical imaging from all mice for a 4kHz sound (left). The responses are aligned thanks to the three responsive regions that define the functional organization of the auditory cortex (A1: primary auditory cortex, A2: secondary auditory cortex, AAF: Anterior auditory field). The red regions correspond to low $\Delta R/R$ thus high response area. The field of view from the calcium imaging are aligned and superimposed to the intrinsic imaging response (middle). Average response to a 4 kHz pure tone of 60 dB from calcium imaging seeing from the top. All the recordings were aligned and the final image was binned into a $40 * 40$ grid, with each pixel size is $30 * 30$ µm. The red regions are regions with a high deconvolved $\Delta F/F$ response at sound onset, thus high activity regions.*

Spatially organized neuronal populations encode different features from sounds.

To have a better understanding of how the sounds are encode in the auditory cortex, we wanted to isolate the neuronal response patterns for the complete set of sounds. We used the deconvolved GCAMP6s traces as a proxy for the firing rate of the neurons[115], then we used hierarchical clustering to group neurons into functional populations with similar response patterns.

Calcium imaging data contains several sources of noise ranging from cortical responses variability to measurement noise from the imaging procedure that is amplified by the deconvolution process. To decrease this noise we first averaged responses across 15 repetitions of each sound. We thus used 148 mean responses profiles (Figure 3.2a top-left) to characterize each cell. We then decided to compare all cell using as a metric the correlation between these response patterns. Based on this metric we organized the set of recorded neurons by performing a hierarchical clustering (Figure 3.2a top-right). To define the clusters, we set a low distance threshold, that produced a large number of clusters with high homogeneity, in order to get as many different types of activity patterns as possible, eventually creating redundant clusters (overclustering) in order not to fuse together response profile that were clearly dissimilar (Figure 3.2a bottom). Then to avoid the bias of local noise correlation between small localized subset of neurons, we removed clusters with less than 10 cells (Figure 3.2a bottom). This technique allowed forming reasonably homogeneous clusters but discarded 87% of the cells from the analysis because other their noisy response profiles (Figure 3.2b). In order to take them into account in the analysis, we computed the average response of each identified cluster, and we assigned the discarded cells to the cluster with the most similar average activity. However if the similarity score (Pearson correlation) between a cell and the most similar cluster was below 0.2, the cell was not assigned to any cluster and removed from the analysis. We observed that the average response profiles of the clusters is not changed after this re-allocation procedure. After this step 61% of the cells are classified in 177 clusters. On average clusters contained 205 ± 289 cells (mean $\pm$ std) with the majority of the clusters containing between 60 and 200 cells (Figure 3.2c). We wondered if functional populations that encode similar features of the sounds are also spatially organized in the auditory cortex. We took advantage of the three dimensional map constructed earlier with cellular resolution to represents the localization of each cluster (Figure 3.2d). We observed that 47% of the clusters are spatially organized in three dimensions (permutation algorithm, see methods, pvalue < 0.05) (Figure 3.2d). The clustering algorithm isolates functional populations of neurons with a similar response pattern and half of those populations are located in space. It also allow the decomposition of our dataset into several canonical response patterns that can be used to better understand the auditory system.

Description of the response patterns observe in the auditory cortex.

After the clustering analysis, we analyzed in details the sound features encoded by the clusters (Figure 3.3). We used the pure tones of different frequency to determine the characteristic frequency of each population. A large portion of the clusters (93%) have a characteristic frequency between 4 kHz and 24 kHz with only a few clusters that respond for sound with higher frequencies. The tuning curve for frequency does not have the same broadness for each cluster (Figure 3.3). For a given sound level, some clusters are very

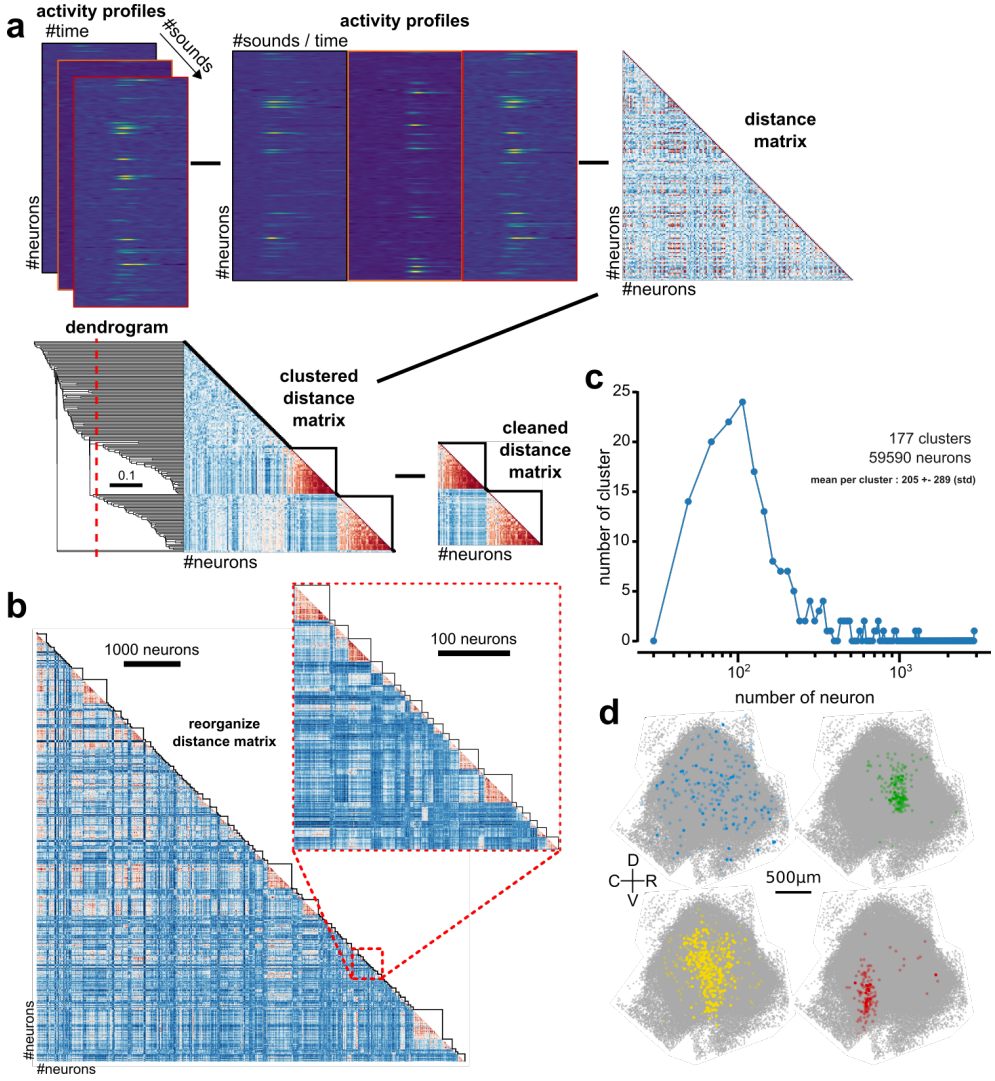

FIGURE 3.2: **Hierarchical clustering group neurons by functional populations with similar response pattern.**

*(**a**) The average neuron responses for each sound of the dataset is concatenate in time to form one time serie for each neuron (here ∼200 neurons are display). A distance matrix based on correlation is constructed and used from hierarchical clustering. A threshold of 0.7 was selected to isolate high correlation cluster. We remove smalls clusters of less than 10 cells to avoid the biais of high local correlation within a recording. (**b**) A total of 177 clusters were selected and we used their average time serie to merge all the cells that were discard from the clustering. Neurons were merge to their closest cluster (correlation score) except if the correlation score was below 0.2. (**c**) Distribution of the number of neurons per cluster. Most of the clusters contains 50 to 200 neurons. (**d**) Spatial distribution of the neurons of 4 clusters seen from the top view without spatial organization (top-left), and with spatial organization (top-right, and bottom). Colored dots without alpha correspond to the kernel of the cluster before agglomeration. Colored dots with alpha are neurons added during the agglomeration algorithm.*

specific for a frequency, responding for example at 5kHz but not at 4kHz nor at 6.2kHz, while others responds to almost all the frequencies tested

(responding with no particular tuning from 4kHz to 19kHz). Concerning the temporal profile of the responses, the majority of clusters responded to sound onset (49%, Paired student test on pure tones, N=34 sounds, p-value < 0.05 corrected for multiple testing with a Bonferroni correction) but 17% responded to sound offset (Paired student test on pure tones, N=34 sounds, p-value < 0.05 corrected for multiple testing with a Bonferroni correction) (Figure 3.3). There were only three clusters over 178 that respond to both onset and offset, and one of them responded specifically to the offset of low frequency sounds and to the onset of high frequency sounds. We then studied tuning to sound level, and we found around 45% of the clustered populations that responds significantly more to quiet ($\sim$60dB) than loud sounds ($\sim$80dB) (Wilcoxon test on pure tones, N=11 sounds, p-value < 0.05 not corrected for multiple testing) (Figure 3.3). By looking at others sounds than pure tones, we found clusters with frequency modulated sound (FMS) selectivity (21%, Student test N>34, p_value<0.05 not corrected for multiple testing), which means that they only responded more to FMS than to pure tones (Figure 3.3). However, because the FMS are continuous variation between two frequencies, it is hard to conclude if these populations are really FMS selective or if they are precisely tuned to a pure tone frequency not included in the sound set.

Some features encoded in the auditory cortex are non linear.

The auditory system is highly non-linear[58, 62, 1], which means that a variation in the sound intensity or frequency distribution does not lead to not proportional variations in the neuronal responses. For example neurons that are tuned to quiet sounds extract a non-linear feature because if the sound intensity increases, the activity of the neurons decreases. Multiple non-linearities have been described in the auditory system such as intensity ramps asymmetry, FMS asymmetry, non-additive response to chords, quiet tuning, and response tuned to the speed of variation in FMS. However, so far, they all have been study individually. We here took the opportunity offered by our dataset to study multiple non-linearities together to test if neurons encode only one non-linear feature or if they can encode a combination of them. First we identify manually the main classes of non-linear transformations that occurs in our recordings between the sounds and the cortical activity, and then we quantified them with a score centered around 0 and ranging from -1 to 1. We started to search for populations of neurons with a tuning to quiet sounds[45, 44] and we quantify it with a score defined by the response to 60, 70 and 80dB pure tones (see methods). A high value for the score represents a population of neurons that are tuned to quiet sound level. With this score we could perform a bootstrap analysis in order to know which population has a significantly tuning to quiet sounds. To do so, we calculate for each cluster the score for a subset of its neurons, and we repeat this process 10000 times (see methods) in order to get a distribution of score that characterize to which extend the cluster encode this non-linearity. If a cluster has a distribution with a average that is significantly different from zero, then this cluster has

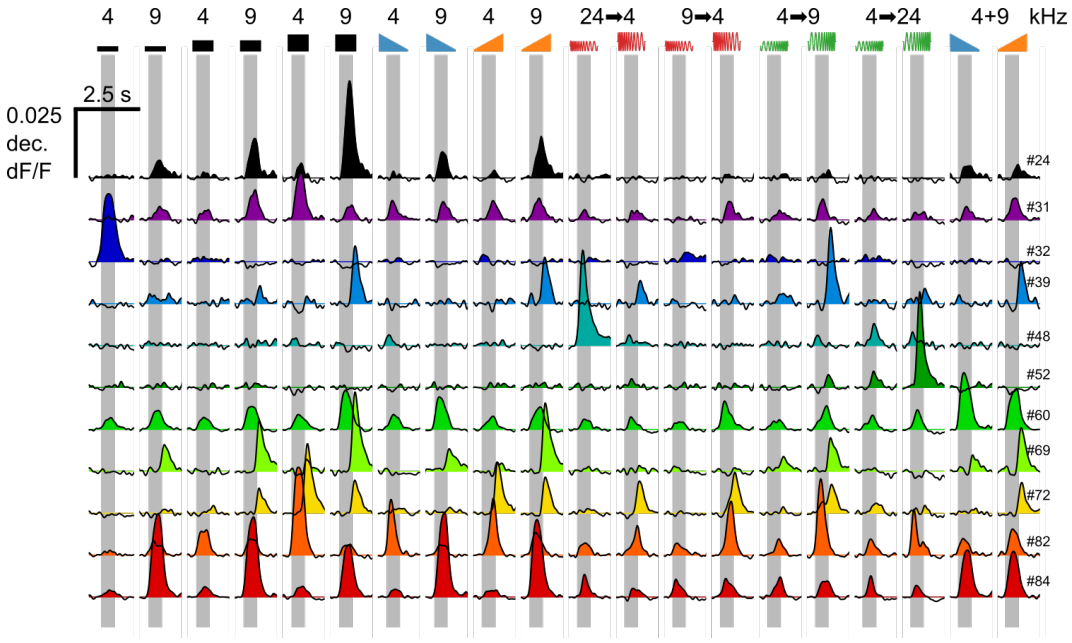

FIGURE 3.3: **Diversity of population responses in the auditory cortex.**

(***a***) *Examples of population responses after clustering for a set of sound focus around 4 and 9 kHz. Some populations respond only to 4kHz (#31, #32, #82, #72) where as others respond preferentially for 9kHz (#24, #39, #69, #84) or for both sounds (#60). Some population responds for onset (#32, #84 . . .) others to offset of sounds (#39, #69, #72). There are population that are tuned to the sound intensity. For example they prefer 60 dB sounds (#32, #84) or 80dB sounds (#24, #82). Some population seems to be FMS specific (#48, 52) but it is hard to say if they are FMS specific or narrowly tuned to a frequency different than 4 and 9 kHz as they respond to a sound varying between 24 and 4 kHz. However these populations are direction sensitive in the frequency variation. Some populations have a respond to chords that is the sum of the responses for the individual frequencies of the chord (here 4 and 9 kHz, #39). Others have a response which is inferior to the sum (sublinear response, #84, #72) or superior to the sum (supralinear response, #60).*

a homogeneous encoding of the non-linearity (i.e. a large proportion of its neurons encode the non-linearity). Over 24% of the clusters were tuned to quiet sounds (Figure 3.4a). We also expressed this percentage in term of cells by replacing the number for each cluster by the number of cells it represents, and we observe a similar percentage (22%) (Figure 3.4g). This non-linear encoding could be explained with a population of neurons tuned to all the sound intensities, that is inhibited by a population sensitive to loud sounds. Then we observed that some clusters were not producing an additive response to chords, in other words, their response for a chord was not the sum of the responses for the individual frequencies. Interestingly, we found three types of response to chords (Figure 3.4b): populations that are summing individuals frequencies thus creating linear response for the chords, and population that were responding more (supralinear response) or less (sublinear response) to the chords than what is expected by the sum of the responses of

individual frequencies. We define a score based on the difference between the response for individual frequencies and the response to the chord (see methods). A low value for this score indicates that the population is responding sublinearly to chords, where as a high value indicates that the population is encoding chords with a supralinear responses compare to the responses to individual frequencies. With a bootstrap method, we determine that 56% of the populations respond linearly, 10% respond significantly sublinearly and the remaining 34% respond significantly supralinearly (Figure 3.4g). This non-linear encoding can be explained by lateral inhibition across frequency channels for sublinear responses[228] or by multiple frequency sensitivity[229] with a threshold for supralinear responses.
Then we realized that the response of some populations for 100ms FMS sounds were stronger than the response to a 500ms sound with the same frequency boundaries (Figure 3.4c). Interestingly these populations also respond to 500ms FMS with a large frequency boundaries difference (like between 24kHz and 4kHz), so we deduced that they were tuned to a particularly rapid frequency modulation speed. We quantified this effect (see methods) with a score where high value means that the population is tuned to a rapid speed modulation and we found with bootstrap that around 30% of the clusters are significantly tuned to a high frequency modulation speed during FMS (Figure 3.4g). Multiple mechanisms could be responsible of this effect such as delayed high frequency inhibition, or two-tone facilitation as stated in Fuzessery et al[230].
We found another non-linear phenomenon by studying the intensity sine modulated sounds, because we found some population tuned to a specific frequency of modulation. As all the sounds have the same intensity, a tuning for one of these sounds implies that there is a non-linear transformation that occurs in the auditory system to encode that tuning. We found populations tuned to 1Hz modulation, and others to 20Hz modulation (Figure 3.4d). Interestingly some populations were also tuned to intermediate frequencies of modulation (3Hz or 7Hz) (Figure 3.4d). When we quantified this tuning and tested it with bootstrap, we realized that almost half of the populations we recorded were significantly tuned to a particular frequency of amplitude modulation (Figure 3.4g). A high value for this score indicates that the population has a tuning to low frequency of modulation and a low value indicates that the score has a tuning toward high frequency of modulation.
Finally, we observed that there was asymmetric responses between two symmetric sounds like sound ramping-up or -down in intensity [45, 44] (Figure 3.4e), or FMS with an increase or a decrease in frequency[231, 232, 233] (Figure 3.4f). We compute the response difference between the response for two symmetric sounds, and we used it as a score in a bootstrap to test for each population if it was encoding one of the two stimuli. A high value for these scores indicate that the population responds more to decreasing intensity or frequency, and a low value for these scores indicate that the population responds more to increasing intensity or frequency. We found that 50% of the clusters encode the intensity modulated sounds in an asymmetric manner and that 47% of the clusters have an asymmetric response to FMS (Figure

3.4g).
We together described 7 non-linear features that are encoded at the level of the auditory cortex and we quantified each of them with a score ranging from -1 to 1. We know from Deneux et al[44], that the tuning to quiet sounds is responsible for the intensity modulated sounds asymmetry because quiet onsets increase the response to the increasing intensity sounds. With this observation they could build and constraint a model to explain how the asymmetry to intensity modulated sounds emerge from the features encode in the auditory system. We wondered if we could find similar interactions that would help to build a more complete model of the auditory system.

The nonlinear features extracted by the auditory cortex interact together.

We wanted to investigate if several non-linearities interact with each other in the auditory system. Can a functional population of the auditory cortex encode multiple non-linear features? Are some populations specific to a single non-linear feature? If a population encode a non-linear feature A does it necessary encode another non-linearity feature B? To answer these questions, we compared how populations encode non-linear features two-by-two thanks to the score we determine previously for each non-linearity. In a two-by-two comparison there was only 4 types of interactions we could found : (i) a independent interaction where the probability to encode the feature A is not dependent of the encoding of the feature B. (ii) an exclusive interaction where the two non-linear features are encode by two distinct populations, and there is no population of neurons that encode both features; (iii) an inclusive interaction where if one population encode the feature A, it also encode the feature B, and this relation can be reciprocal; (iv) a particular inclusive interaction where if one population encode a non-linear feature B, it gives information about how it encodes the feature B (for example the value of its tuning, or which stimuli it encodes). By having a close look at the two-by-two comparisons, we could observe that the most frequent type of interaction (40%) is exclusion (Figure 3.5, green star). An example is the response to the speed of modulation in the FMS, that is not encoded by the same populations as the non-additive response to chords (Figure 3.5d). We realized that the population that encode the tuning for a frequency of sine intensity modulation interacts with exclusion with almost all the others non-linear features (Figure 3.5c). It means that these populations respond linearly to the others features, and for example, it respond linearly to the intensity modulated sounds. This last example illustrates two sound level related features that cannot be encoded by the same populations of neurons, indicating that the treatment of the intensity in the auditory system is performed with at least two parallel and independent pathways. In total, by considering how the populations encode the non-linearities, the information about the exclusion demonstrate that there should be at least 3 different pathways in the auditory system to encode the non-linear features (one for the sine intensity modulated sounds, one for the rapid speed of frequency modulation, and

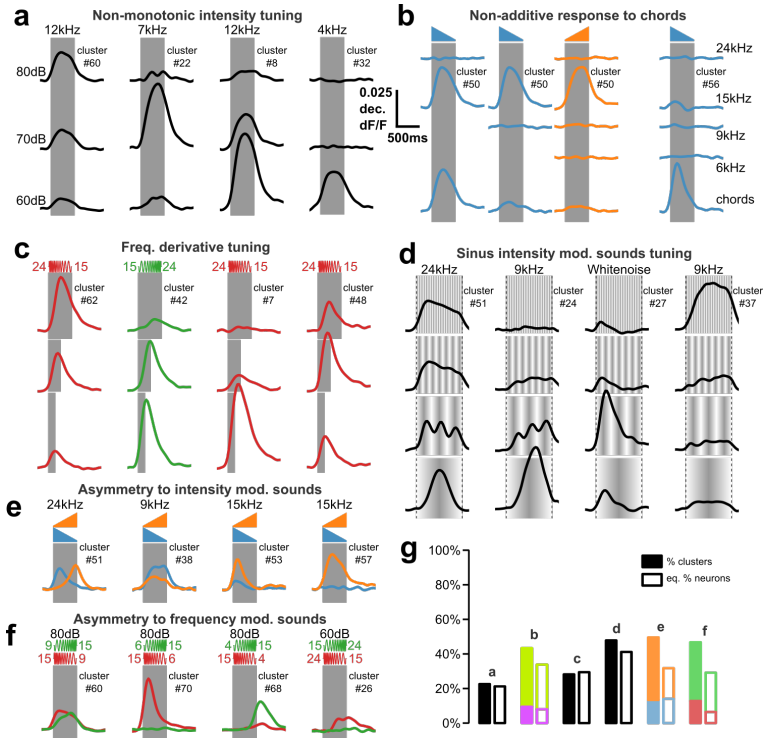

FIGURE 3.4: **Cortical recordings in the auditory cortex reveals several non linearities in sound processing.**

*Consequences of nonlinearities in the sound processing at the level of the auditory cortex. All the figures are express in deconvolved ΔF/F (equivalent firing rate) and the example on the left is a case of possible linear processing. Every column is one example of cortical population. (**a**) Examples of populations with a stronger responses for quiet than for loud sounds (except left). (**b**) Examples of population with sublinear or supralinear response to chords. The chords are also ramping up or down in intensity. The two example on the middle reveals a sublinear response to chords compare to the sum of the individual frequencies (green in **g**). The right example is an case of supralinear summation to chords (purple in **g**). (**c**) Examples of population with a tuning for FMS frequency speed modulation. The sounds with high speed modulation are also shorter. The second example and the last ones exhibits the same effect but for, respectively, increasing or decreasing FMS. (**d**) Example of population with a tuning to a frequency of sinus modulated sounds. The sounds have a duration of one second. (**e**) Example of population with an asymmetric response for increasing of decreasing intensity ramps, respectively, orange and blue curves. (**f**) Example of population with an asymmetric response for increasing or decreasing FMS, respectively, green and red curves. (**g**) Percentage of clusters significantly involved in the different nonlinear encoding. The percentage is express in number of significant clusters divided by the total of clusters (full bars) or express in number of cells in the significant clusters divided by the total number of cells (empty bars). The color indicate the preferred stimulus (green/purple for sub/supra linear responses to chords; orange/blue for up and down ramps; green/red for increasing or decreasing FMS).*

one for the rest of the non-linearities). A fraction of 37% of the interactions were asymmetric with populations that encode a non-linear feature also encoding another subtype of feature. For example the population encoding for

quiet sounds are more likely to respond to intensity ramps in a asymmetric manner because they respond for the onset of the ramps ramping up (with a quiet onset). This particular example was already described in Deneux et al[44]. And it suggests that the populations that encodes one non-linear feature (here quiet sounds) is responsible of one part of the encoding of the second feature (here a stronger response to increasing intensity ramps compare to decreasing intensity ramps) and therefore it reveals that one non-linear feature is extracted before the other one in the auditory system. This interaction is mostly encountered with populations that encode quiet sounds, a feature that is relatively simple compare to the others non-linear features, and thus that is probably encode very early in the auditory system, and that can be involve in the formation of the other non-linear features. It is easy to imagine how population tuned to quiet are responsible for intensity features such as the asymmetry between intensity ramps or the sine modulated sounds. However it is much more interesting to see that it is also responsible for features such as the non-additive response to chords or the FMS asymmetry. We observe another interaction which is the independent relation between two features (no symbols, Figure 3.5), and it concerns 17% of the interactions. For example the population that encode the asymmetry between intensity ramps can also be responsible of the encoding of the FMS asymmetry (Figure 3.5a). If we know that a population of neurons prefers increasing intensity ramps, it does not give information about how the population encodes FMS. This interaction constraint the architecture of a model because a population that encode one feature should also be able to encode another, or not. As a particular implementation, in the case of a multilayer model like in Deneux et al[44], or McFarland et al[152], this interactions shows that the two non-linearities should be extracted by at least two separated fully connected layers. Finally we also observe one inclusive interaction where almost all the populations that are tuned to a high speed of modulation for the FMS, also prefer a direction of modulation for FMS (Figure 3.5d). However all the population that are sensitive to a direction of modulation are not sensitive to high speed of modulation. This interactions concerns two non-linear features that are measured on the same sounds. This interaction suggest that the asymmetry for FMS is extracted first by the auditory system and that this information is split into two pathways, with one that extract the tuning for rapid frequency modulation. Together the description of the interactions between the non-linear features will helps us to understand the architecture of the auditory system. Indeed, some interactions imply separate pathways where as others should be extracted one after the other. The knowledge on these interactions will help to refining models of the auditory cortex by providing constraints to construct their non-linear architecture.

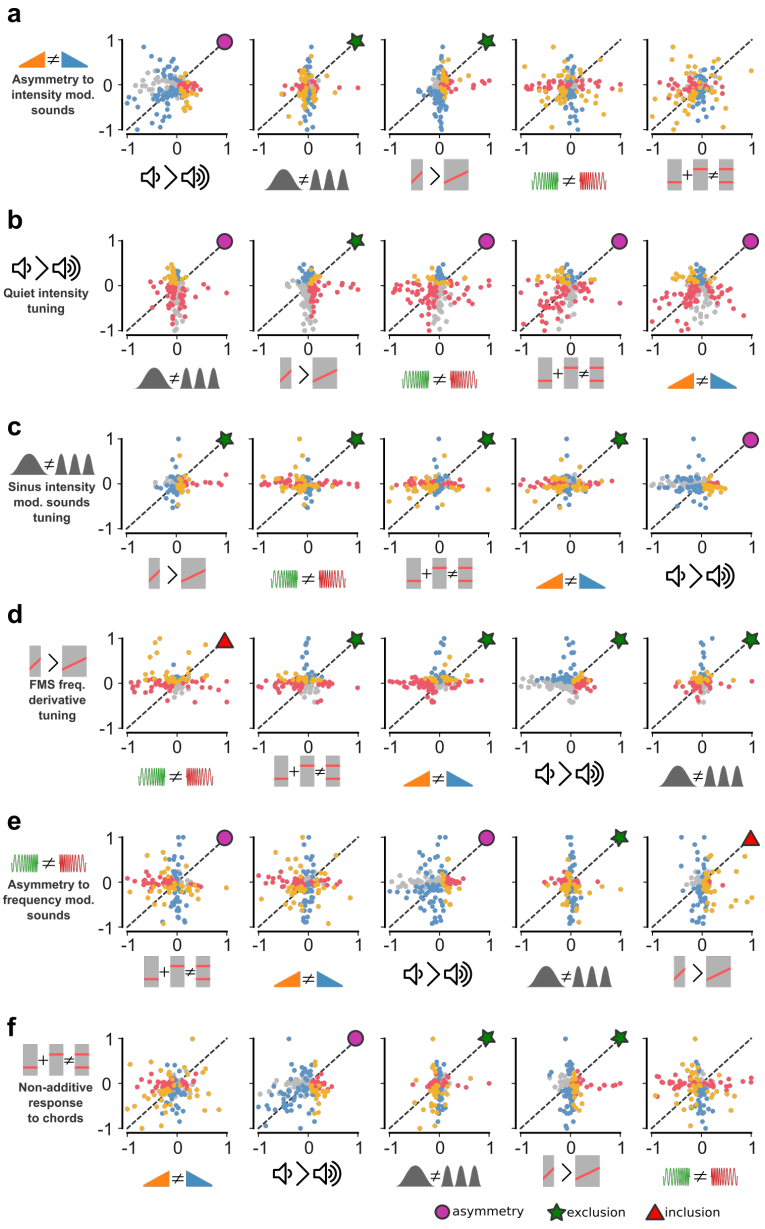

FIGURE 3.5: **Quantification of the features extracted by each population reveals different interactions between the nonlinearities features.**

*Comparison two by two of the score for each feature. Each point represents a cluster and can have a significant non linear response for the x-axis feature (red), for the y-axis feature (blue), or for both (orange). Interactions between the non linear features extracted reveals 4 types of interactions : An exclusion of the two features (green star), an inclusion of the two features (red triangle) an asymmetry (purple circle), and an independent relationship (no symbol). Each line describe the interaction between one feature and the others quantified by score normalized to set the maximum between -1 and 1. (**a**) Interactions between the tuning for a particular intensity ramps and the rest of the features described in Figure 3.4. Value for sounds ramping up or down in intensity correspond respectively to a score of 1 and -1. (**b**) Interactions between the tuning for quiet sounds and the rest of the features. If the score is*

*between 0 and 1, there is a tuning for quiet sounds, otherwise the population respond more for loud sounds. (**c**) Interactions between the tuning for a particular frequency of intensity modulation and the rest of the features. High value correspond to a tuning to low frequency of modulation. (**d**) Interactions between the tuning for the speed of the frequency variation in FMS. A score between 0 and 1 correspond to a population with a preference for fast frequency variation. (**e**) Interactions between the tuning for a direction frequency modulation in FMS and the rest of the features. The tuning for up and down ramps corresponds respectively to a score of -1 and 1. (**f**) Interactions between the non-additive response to chords and the rest of the features. If a cluster responds more for a chord than the sum of the responses for individual frequencies, then score is 1. If a cluster responds less for a chord than the sum of the responses for individual frequencies, then the score is -1.*

Discussion

In this study, we recorded the activity of 59 590 neurons from the auditory cortex of awake mice to a large set of synthetic sounds that vary in frequency and intensity. We found spatially localized populations of cortical neurons with the same response pattern and we describe the different features extracted by each population. We described 7 features that could not be extracted with linear processing of the sound information and we quantified the implication of each identified neuronal populations (cluster) in the encoding of each non-linear feature. Thanks to the quantification of the features we were able to study how each feature is extracted relative to the others. We defined 4 interactions: exclusion where two features are not encode by the same populations, inclusion were two features are encode by the same populations, asymmetry where one feature is encoded by populations that encode another feature with a bias, and independence where no systematic relationship is observed between the presence of one feature and of the other in neuronal responses.

Non-linearities may be generated before the auditory cortex. Indeed some non-linear phenomenon like tone-suppression [9] or adaptation [234] have been observed in the cochlea and can be responsible of the features such as the non-additive response to chords or the tuning to sine intensity modulated sounds. The approach in this study is to record from high level auditory areas and look at interactions between features to better understand the extraction of features in all the auditory system, and more generally its architecture. We could for example determine that the neurons tuned to quiet sounds appears probably very early in the auditory system since most of the others non-linearities relies on it.

We focused on seven non-linear features extracted by cortical neurons from sounds. However the set of non-linear features that can be observed at the level of the auditory cortex is probably larger than seven as we were limited here by the recording methods and by the set of sounds. Indeed, calcium imaging are suitable to record large populations of neurons in the awake animal and to study the encoding of sounds in the auditory cortex, but the technique lacks the temporal resolution that might be needed to study non-linear features encoded by single spikes or with temporal coding. However, the gap between features capture with imaging and electrophysiology in auditory cortex might not be to wide, as the precise temporal resolution seen in the brainstem tends to be replaced by a rate code for higher areas like the auditory cortex[235]. Another limitation to study non-linear features was set by our set of sounds. Indeed more non-linear features could have been tested if the sound set was larger.

We also did not present very complex sounds or natural sounds, which would have

led to a more complete description of the features extracted at the level of the auditory cortex such as neurons with specific tuning to vocalizations[236, 237, 238]. We here used synthetic sounds in our study to better control sound parameters and to provide more intuition on possible mechanisms behind the nonlinearities and their interactions.
Another limitation of the method is the use of GCaMP6s. This protein has some weak non-linearities especially between the number of action potentials and the amplitude of the fluorescence changes[44, 108]. We believe that this weak non-linear effect does not impact our results because the transfer function from the number of spike to the amplitude of the fluorescent events is monotonic with GCaMP6: if we observe a fluorescence event with a higher amplitude than a second event, its means that the first event produces more action potentials. As all the non-linear features we described is a comparison between two or more sounds, the effect we observe in this study cannot be explained by the non-linearity of the GCAMP6. What are the the neuronal mechanism by which non-linear features are extracted? Several of them were already described earlier than auditory cortex in auditory system, and this the mechanisms may be extra-cortical. Adaptation is a well-known non-linearity that can account for the tuning for sine intensity modulated sounds because it can decrease the response of neurons to high frequency intensity modulation. Lateral inhibition is another mechanism responsible for non-linearities and it can explain the tuning for quiet sounds, the non-additive response to chords, the asymmetry to FMS and the tuning to rapid frequency modulation of the FMS.
Each interaction between non-linear features characterized in this study can help designing models of the auditory systems because it brings constraints on the architecture of the model. Whereas a recent study emphasizes the hierarchical architecture of the auditory system [37], our results show that it also can be distributed with parallel pathways. Indeed, if two features are never encoded by the same population, we can deduce that two distinct set of populations encode those features. Our data indicate the existence of a minimal number of 3 parallel pathways because two features pairs are excluded from each other. In contrast, we noticed that the tuning to quiet sound is encode in all these three parallel pathways because this feature does not have an exclusion interaction with the rest of the non-linear features. The simplest hypothesis is that the tuning to quiet sounds is set before the separation of the auditory system into our three putative three l parallel pathways, maybe already at the level of the cochlear nucleus. As a matter of fact, non-monotonic tuning to intensity has been observed in the cochlear nucleus already which supports this idea [239, 240].
This study focused on the non-linear integration of sounds in the auditory cortex to draw more general conclusions on the architecture of the auditory system. We believe the non-linearities in sensory system are essential for perception and their understanding is key for next model generation. Here we provides information about how the non-linear features are extracted by populations and how these populations interacts to encode these features. We hope new models of the auditory system can emerge from these observations and be constrained by the conclusions of this study. Such more realistic models of auditory processing could help creating new algorithms for hearing aids for hearing loss patients.

Methods

Animals.

All mice used for imaging were 6 to 14 weeks old male C57Bl6J mice. All animal procedures were approved by the French Ethical Committee (authorization 00275.01).

Two-photon calcium imaging in awake mice.

Three weeks before the first imaging session, mice were anaesthetized under ketamine medetomidine. The right masseter was removed and a large craniotomy of diameter 5 mm was performed above the auditory cortex. We then inject 150nL (30nL/min) of the AAV1.Syn.GCaMP6s.WPRE virus from Vector Core (Philadelphia, PA, dillution 30x), at three different locations of the auditory cortex. The craniotomy was sealed with a glass window and a metal post was implanted using cyanolite glue followed by dental cement. Three days before imaging, mice were trained to head-fixation for 30 to 60 min per day. Then mice were imaged one to two hours per day. Imaging was performed using a two-photon microscope (Femtonics, Budapest, Hungary) equipped with an 8kHz resonant scanner combined with a pulsed laser (MaiTai-DS, SpectraPhysics, Santa Clara, CA) tuned at 920 nm. We used a 10x Olympus objectives (XLPLN10XSVMP), thus obtaining a field of view of $1 * 1$ mm. Images were acquired at 31.5Hz during blocks of 32 s during which randomly chosen sounds were presented with 1.5 s intervals. Blocks were interleaved by an 15 s pause and repeated until all sounds were played 15 times. All the sounds have a duration of 500 ms except when it is specified. The set of sound (Figure 3.1b) was composed of :

- 33 pure tones of different frequencies (4.0, 5.0, 6.2, 7.8, 9.7, 12.2, 15.2, 19.0, 23.7, 29.6, and 37.0 kHz) and intensities (60, 70, and 80 dB SPL).
- 10 pure tones intensity ramps (up-ramps going from 60 to 80 dB SPL and down-ramps going from 80 to 60 dB SPL) of frequency (4.0, 6.2, 9.7, 15.2, 23.7 kHz).
- 32 chords intensity ramps with all the possible combination of 2, 4, or 5 frequencies contained in the pure tones intensity ramps.
- 12 frequency modulated sounds (FMS) of different duration. The frequencies were increasing or decreasing on the range 24 kHz to 9 kHz or on the range 9 kHz to 4 kHz and with intensity of 60dB or 80dB. Each of these FMS were presented during 125 ms, 250 ms or 500 ms thus creating different slope of frequency variations.
- 40 FMS with frequency boundary in the list 4.0, 6.2, 9.7, 15.2, and 23.7 kHz. We created a new sound for every couple of values in this list (20 possible combination) and we played the sound at 60 dB or 80 dB.
- 12 amplitude modulated sounds, with a sinusoid wave as the modulation envelope. We used pure tones of 4.0, 9.74 kHz, and a whitenoise sound that we modulated with a sinusoid wave starting with phase $-\pi/2$ and that we normalized between 0 and 1. The frequency of the sinusoid wave was 1Hz, 3 Hz, 7 Hz, or 20 Hz. The duration of these sounds was 1 second.

- 8 sounds that were intensity and frequency modulated. The intensity modulation was ranging from 60 to 80dB and ramping up or down. The frequency modulation was ranging from 4.0kHz to 9.74kHz or from 9.7 kHz to 23.7 kHz and ramping up or down.
- 1 blank condition with no sound.

All sounds were delivered at 192 kHz with a NI-PCI-6221 card (National Instrument) driven by Elphy (G. Sadoc, UNIC, France) through an amplifier and high frequency loudspeakers (SA1 and MF1-S, Tucker-Davis Technologies, Alachua, FL). The head fixed mouse was isolated from other noise sources by a sound-proof box containing only the speaker and the animal. Sounds were calibrated in intensity at the location of the mouse ear using a probe microphone (Bruel&Kjaer).

Intrinsic optical imaging recordings.

In order to localize the calcium-imaging recordings with respect to the global functional organization of the cortex, we performed intrinsic optical imaging experiments under isoflurane anaesthesia (1%). In the mouse the blood vessels cannot be used to localized the functional parts of the brain because they form a non reproducible pattern across animals (Figure 3.1f). The brain and blood vessels were illuminated through the cranial window by a red (intrinsic signal: wavelength 780 nm) or a green (blood vessel pattern: wavelength 525 nm) light-emitting diode. The reflected light was collected at 20 Hz by a CCD camera (Smartek Vision, GC651M) attached to a custom-made macroscope. The focal plane was placed 400 mm below superficial blood vessels. A custom-made Matlab program controlled image acquisition and sound delivery. We acquired a baseline and a response image (164 x 123 pixels, $\sim$3.7 x 2.8 mm) corresponding to the average images recorded 3 s before and 3 s after sound onset, respectively. The change in light reflectance ($\Delta R/R$) was computed then averaged over the 20 trials for each sound frequency (4, 8, 16, 32 kHz, whitenoise). A 2D Gaussian filter (σ = 45.6 μm) was used to build the response map and determine the localization of the auditory cortex due to the stereotypical response map produced by the 4 kHz sound. Sounds were trains of 20 white noise bursts or pure tone pips separated by 20 ms smooth gaps. We benefited from the reproducible pattern of the 4 kHz response in order to define three part of the auditory cortex (the primary and secondary auditory cortex A1 and A2, and the anterior auditory field AAF). We also used this pattern to align (euclidean transformation) the recordings fields across mice based on the functional maps of their auditory cortex (Figure 3.1f).

Calcium imaging pre-processing.

We recorded 59,590 cortical neurons from 60 imaging area in 7 animals. The recording depth was between -600 and -100 μm corresponding to layer 2/3 and layer 5. Data analysis was performed with custom-made Matlab scripts available upon request. Each frame of the recording was corrected for horizontal motion to a template image using rigid body registration (all sessions with visible z motion were discarded). Regions of interest were then automatically selected and human checked as the cell bodies of neurons with visually identifiable activity. The mean fluorescence signal $F(t)$ and the local neuropil signal $F_{np}(t)$ was extracted for each region. For each region, we subtracted 70% of the the neuropil signal from the neuron signal to

avoid neuropil contamination. Baseline fluorescence F_0 was calculated with a sliding window computing the 3^{th} percentile of a Gaussian-filtered trace over the imaging blocks. Fluorescence variations were then computed as $f(t) = (F_c(t) - F_{c0})/F_{c0}$. Analyses were performed on the deconvolved version of the signal as an estimation of the firing rate. The deconvolution was performed in time as $r(t) = f'(t) + f(t)/\tau$, in which $f'(t)$ is the first derivative of $f(t)$ and τ = 2 s, as estimated from the decays of the GCAMP6s fluorescent transients. This simple method efficiently corrects the strong discrepancy between fluorescence and firing time courses due to the slow decay of spike-triggered calcium rises. However it does not correct for the relatively slow rise time of GCAMP6s, producing a time delay on the order of 70 ms between peak firing rate and peak deconvolved signal.

Hierarchical clustering.

Hierarchical clustering was performed with custom-made Python scripts (Figure 3.2). Calcium responses were deconvolved, baseline subtracted, blank subtracted, averaged over trials, and the responses to all considered stimuli were concatenated. Single cell average response were then smoothen with a Gaussian filter (σ = 90 ms). We performed the clustering with correlation distance (Pearson correlation) and the threshold was manually adjust to avoid cluster with redundant mean response profiles. We remove from the analysis, the clusters with less than 10 cells to avoid the bias of local correlation pattern. Once these clusters were identified, the 87% remaining cells were assigned to the closest average cluster response in term of correlation. If the best correlation value was below 0.2 we assign the cell in a specific cluster for weakly responsive cells. This procedure did not change the mean cluster response profiles.

Clusters localization.

The localization of the cluster was statistically tested with a permutation algorithms and the distance score to test each permutation. We permute the localization of every cells of the cluster with the location of random cells from our recordings and we compute the average distance between neurons of the cluster and their barycenter. We perform 1000 permutations and we tested if the distance score is in the 5% lowest value in order to know if the cluster is spatially grouped.

Quantification of nonlinear features.

We quantify each nonlinear features describe in Figure 3.4 with score for each cluster. Because we used bootstrap to test significance we force the score to be centered on zero if there is no activity in the cluster. To test whether or not a cluster respond stronger to a quiet sound compare to a loud sound, we used the score : $\sum_{t,f,i}(r_{tfi} - r_{tfj})$ where r is the average response of a cluster for a pure tone stimuli, t is time, f is pure tone frequency, i is the response to 60 dB or to 70dB, and j is the response to 80dB. The stronger the score, the more the cluster is tuned to quiet sounds compare to 80dB sounds

To test if the response to chords is the simple sum of the responses for the individual frequencies, we used the score : $\sum_{t,c}(r_{tc} - \sum_f r_{tf})$ where r is the average response of a cluster for a stimuli from the intensity ramps or the chords, t is a time point, c is one chord from the dataset, f is a frequency that compose the chord. If the score is

positive it reflects a supralinear response compare to the sum of the individual frequencies responses. If the score is negative, it reflects a sublinear response compare to the sum of individual frequencies responses.
To test if a cluster is tuned to the speed of frequency modulation is FMS, we used the score : $-\sum_{FMS} a_{FMS}$ where FMS is a frequency modulated sound identify by its boundary frequencies, a is the linear coefficient define after fitting a linear regression $f = a.x + b$ with $f = \sum_t r_{td}$(t being a time point) for each duration tested d, and x is the corresponding duration (and r is the response for a FMS). The higher the score, the more the cluster respond for short duration FMS, i.e. rapid frequency modulation in the FMS.
To detect tuning to the frequency of modulation in sinus modulated intensity sounds, we used the score : $\sum_{ft} \left(3r_{ft}^{1Hz} + r_{ft}^{3Hz} - r_{ft}^{7Hz} - 3r_{ft}^{20Hz}\right)$ where r is the average response of the cluster for a intensity sinus modulated sound, t is a time point, f is a frequency for the sound (4kHz, 9kHz, 24kHz or whitenoise), and the superscript indicate the frequency of modulation for the intensity of the sound. A high score corresponds to a tuning to low frequency of intensity modulation.
To determine if the response for each cluster was asymmetric for the intensity ramps we used the score : $\sum_{t,f}(r_{ft}^{down} - r_{ft}^{up})$ where r is the response for an intensity modulated sound going up or $down$ (superscript indices), t is a time point, and f in a particular frequency for the ramp. The higher the score is the more the cluster respond to ramps decreasing in intensity.
In order to know if a cluster was responding more to a direction for FMS, we used the score : $\sum_{t,FMS}(r_{tFMS}^{down} - r_{tFMS}^{up})$ were r is a response for a FMS sound, t is a time point, and FMS is a frequency modulated sound from the dataset. The superscript indicate if the frequency modulation is increasing (up) or decreasing ($down$). If this score is high, then the population respond more to the decreasing frequency sounds than for the equivalent sound with a increasing frequency.
For each score we divided it by the absolute of the maximum value so that all the value are between -1 and 1 and can be compare easily.

Bootstrap and significance of the score.

For each score we design a bootstrap method in order to find the clusters with a significant non linear encoding. From each cluster we select randomly a subpopulation and we average its activity to obtain the mean response of this subpopulation for each sounds. Then we compute the score design for each non-linearity on this population. We repeat the all process 10000 times to have a estimation of the variability of the score based on the variability of the activity in the cluster. Finally we test the distribution obtain with the subpopulation compare to a score of zero, which represent no non-linear encoding of the feature. The cluster is significant for a non-linear encoding if the distribution is statistically different from zero (p-value > 0.0001).

Acknowledgments

We thank E. Harrell for helpful discussions and comments on the manuscript. This work was supported by the Agence Nationale pour la Recherche (ANR "SENSE-MAKER"), the Marie Curie FP7 Program (CIG 334581), the Human Brain Project

(SP3 - WP5.2), the Fyssen foundation, the DIM "Region Ile de France", the International Human Frontier Science Program Organization (CDA-0064-2015), the Fondation pour l'Audition (Laboratory grant) and the Paris-Saclay University (Lidex NeuroSaclay, IRS iCode and IRS Brainscopes).

Author Contributions

AK and BB designed the study and wrote the paper. AK and TT performed the experiments. AK analyzed the data.

Supplementary Figures

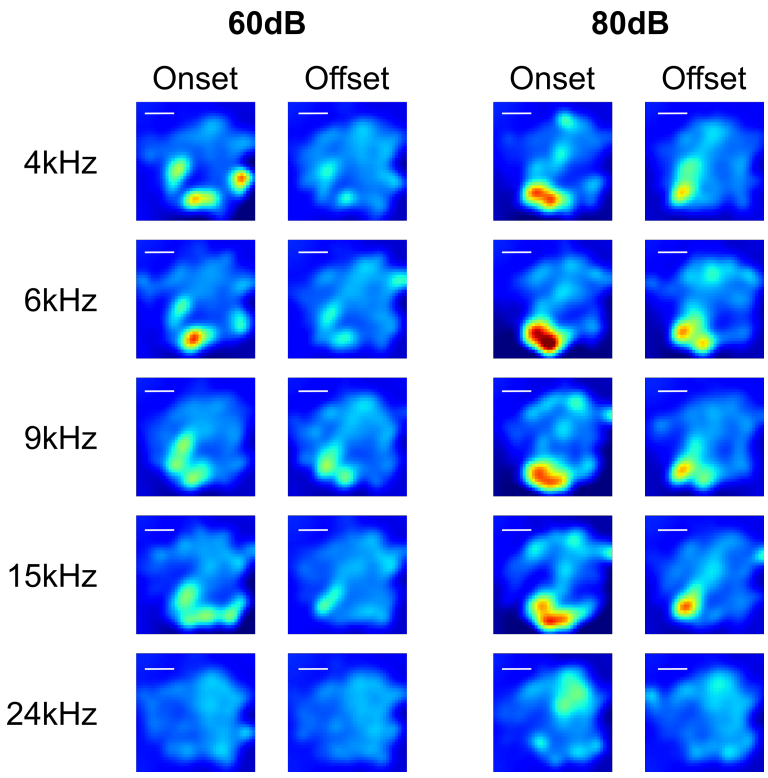

FIGURE 3.6: **Calcium imaging recordings in the auditory cortex reveals spatial patterns of activation.**

Top view of all the fields of view from calcium imaging recordings that were aligned based on the intrinsic optical imaging signals for each mouse. The signal comes from pure tone sounds recorded and binned in a grid with a pixel size of $30 * 30$ *µm. The activity of each pixel is represented as the deconvolved* $\Delta F/F$ *with a color code, blue for no activity and red for high activity. The scale bar represents 500µm. The images have the same colorbar and their orientation is similar as in Figure 3.1f. The tonotopic organization of the auditory cortex is reveals with some areas dedicated to low frequencies and others to high frequencies. It is interesting to note the difference in spatial activity pattern between quiet and loud sounds, and also between onset and offset especially for loud sounds.*

Context Chapter 4

The fourth study of this PhD work is currently under review in Nature Communications. After the three first studies we have a better idea on the importance of cortical representations and how they form in the auditory system. We wonderer if the information from other sensory modalities could also modify cortical representations. In particular, we focused on the visual and the auditory systems. We used calcium imaging in the mouse visual and auditory cortex, to understand how visual and auditory information combine in the brain. We didn't find any modification of sound processing in the auditory cortex from the visual system, however we found that sounds modify visual cortical representations. Especially we discovered that the onset of a loud whitenoise sound drives activity in the presence of visual information, which is surprising in the context of the first study where we observed more responses for quiet whitenoise sounds. We demonstrated that the neurons that specifically project toward the visual cortex are biased for the onset of loud sounds. In the dark, when there is no visual information, sounds elicit a strong inhibition of the visual neurons. We designed a model of the visual cortex that explain the light-context dependent modulation from the auditory cortex.

My contribution of that study was to perform surgeries and recordings for the activity of the auditory cortex and for the particular projection-neurons that connect the visual cortex. I also performed some of L1 recordings in the visual cortex. I performed a part of the data analysis, in particular the hierarchical clustering and the quantification of the responses in the auditory cortex. I contributed to the design of the model. I generated the figures 4.1 and 4.8 and I contributed to the redaction of the manuscript.

Chapter 4

Context-dependent signaling of coincident auditory and visual events in primary visual cortex

Under review in Nature Communication in 2018

Thomas Deneux[1], Alexandre Kempf[1], and Brice Bathellier[1]

1. Unité de Neuroscience, Information et Complexité (UNIC), Centre National de la Recherche Scientifique, FRE 3693, Gif-sur-Yvette, 91198, France

Abstract

Detecting rapid coincident changes across sensory modalities is essential to recognize sudden threats and events. Using two-photon calcium imaging in identified cell types in awake mice, we show that auditory cortex (AC) neurons projecting to primary visual cortex (V1) preferentially encode the abrupt onsets of sounds. In V1, a sub-population of layer 1 interneurons gates this selective cross-modal information by a suppression specific to the absence of visual inputs. However, when large auditory onsets coincide with visual stimuli, visual responses are strongly boosted in V1. Thus, a dynamic asymmetric circuit across AC and V1 specifically identifies visual events starting simultaneously to sudden sounds, potentially catalyzing localization of new sound sources in the visual field.

Introduction

Numerous multisensory illusions[241, 242, 243, 244] show that audition and vision have strong perceptual bonds. For example, in the double flash illusion[243] a brief sequence of two sounds played simultaneously to a single visual flash leads to the impression of two flashes. While the mechanisms of cross-modal perceptual interactions remain unclear, the anatomy of the mammalian brain shows multiple sites of auditory-visual convergence. One of the best studied examples happens in the superior colliculus, in which visual and auditory cues are combined in the networks computing gaze direction[245, 246]. In cortex, secondary associative areas, such as parietal cortex [247, 248], are not the sole cortical sites of auditory-visual convergence: increasing evidence shows that functional auditory-visual interactions exist

already in primary sensory cortex[249, 250, 251, 252, 253]. Moreover, axonal tracing studies indicate that significant direct connections exist between primary auditory and visual cortex [254, 255, 256, 257, 258, 259, 260], suggesting that important auditory-visual computations are implemented before the associative stage. The role of these direct connections has started to be addressed in the mouse model. Recent studies indicate that, in mice, auditory to visual connections are much stronger than their reciprocals, and that they provide very significant inputs to visual cortex that can modulate visually-driven activity[251, 252]. Nevertheless, the computational role of early auditory-visual connections remains unclear, in particular because information is lacking about the auditory features channeled from auditory to visual cortex and how they are combined with the visual processing stream.

Mouse auditory cortex encodes a wide variety of acoustic features [261] ranging from sound frequency [193, 262] to temporal features such as modulations of frequency[263] or intensity[165] including the salient intensity variations occurring at sound onsets and offsets[164, 44]. The question of how sound frequency information would map onto visual cortex is a difficult one, because of the lack of perceptual and ethological data on the particular frequency cues that could potentially be associated with particular visual stimuli. In contrast, temporal information is known to be used for perceptually assigning auditory and visual stimuli to the same object, based on temporal coincidence as in the double flash[264] and ventriloquist illusions [241, 265] and on covariations of particular features such as the size of the visual input and of the sound intensity envelope as classically evidenced with looming and receding auditory-visual stimuli[266]. Yet it is not known what time and intensity features are channeled through the direct connection between AC and V1 and how they impact visual processing.

To address this question, we here used two-photon calcium imaging and intersectional genetics in order to identify the particular intensity envelope features encoded by AC neurons that project to V1. We show that V1-projecting neurons preferentially encode a subset of the intensity variations extracted by auditory cortex, with a strong enrichment in cells sensitive to the onset of steeply rising sounds. Strikingly, we also show that this selective information impacts V1 in an illumination-dependent manner, with a net inhibitory effect in darkness and a net positive effect in the light, thus reconciling opposing observations reported previously[251, 252]. Analyzing the responses of V1 excitatory and inhibitory neurons in layer 1 and layer 2/3, we show evidence that inhibition in darkness originates from a specific subpopulation of L1 inhibitory cells which masks the direct excitatory drive provided by AC axons to pyramidal cells[252]. Interestingly, the activity of this L1 subpopulation is reduced in normal light condition releasing the excitatory drive from AC. Furthermore, we show that this mechanism also allows for a specific boosting of visual responses which occurs together with large and rapid auditory onsets, and that this boosting correlates with improved detection of visual stimuli in a behavioral task. We thus propose that one important role of auditory to visual cortex connections is to increase saliency of visual events that occur simultaneously to the rise of a sudden sound source, possibly to help identifying the particular visual events responsible for new sounds.

Results

V1-projecting neurons in auditory cortex preferentially encode salient sound onsets.

Intensity variations are precisely coded in auditory cortex, in particular the direction of variations which is clearly reflected in so called “On” (increasing intensity) or “Off” (decreasing intensity) responses[164, 44, 267, 268], but also in neurons that more tonically follow sound amplitude[44, 269]. We have recently shown that not only the direction but also the amplitude ofvariations is coded in different cell types. For example, “Loud On” neurons specifically respond to high amplitude sound onsets, while “Quiet On” only respond to low amplitude (and not high amplitude) onset[44]. The advantage of this coding scheme is to better tile the possible range of variations and thereby refine sound categories that reflect different types of events. For example, sounds coming from sudden physical events such as shocks tend to rise abruptly in intensity and rather activate “Loud On” neurons. On the contrary, sounds coming from more continuous events slowly ramp up in intensity and thus first activate “Quiet On” neurons. Ideal stimuli to capture the variety of On, Off and more tonic response types are long up- and down-ramping intensity profiles. We thus wondered whether such stimuli are similarly encoded by generic auditory cortex neurons and by neurons projecting to the visual cortex. To identify the latter, we injected in V1 a canine adenovirus (CAV) expressing Cre, which is retrogradely transported through axons. In the same mice, we injected in AC an adeno-associated virus (AAV) expressing GCAMP6s[108] in a Cre-dependent manner (Figure 4.1a). As a result, GCAMP6 expression was obtained exclusively in AC neurons that project to V1. Consistent with previous reports we observed that these neurons were located predominantly in layer 5 (Figure 4.1b)[252]. As a comparison, in another set of mice, we injected an AAV1-syn-GCAMP6s virus to broadly express the calcium reporter in generic auditory cortex neurons. Using two-photon microscopy, we imaged calcium responses at single cell resolution in these two set of mice (layer 2/3: 3771 neurons, 18 sessions, 7 mice; V1-projecting: 1593 neurons, 14 sessions, 3 mice) while playing white noise sounds ramping up or down in intensity (range 60 to 85dB - note that visual stimuli were also played in these experiments but had no impact on AC activity, Supplementary Figure 4.8). After automated ROIs extraction[270], calcium signals were deconvolved[143, 115] and smoothed to obtain a temporally more accurate estimate of actual neuronal firing rates. The resulting averaged response profiles were then submitted to a hierarchical clustering[44] (see Methods) to identify the different types of responses. In generic AC neurons, we found 7 distinct clusters, unsurprisingly corresponding to “On”, “Off” or tonic response types, each with a preference for either lower or higher sound amplitudes (Figure 4.1c). Also, as reported previously[44], a majority of neurons preferred up-ramps (62% of clustered neurons, Figure 4.1c). Interestingly, in V1 projecting neurons we identified the same response clusters, to the exception of the sparser “Quiet Off” responses (Figure 4.1d), however, the distribution of each response types was clearly different. First, in contrast to generic L2/3 neurons, a majority of V1-projecting neurons preferred down ramps (67% of clustered neurons, $p<0.0001$, bootstrap, see Methods Figure 4.1d). Second, some cell types that were rare in the generic L2/3 AC population were clearly enriched in V1 projecting neurons. Most prominently, a 2.5-fold increase was observed for "Loud On" responses (10% of generic clustered cells, and 26% of clustered V1 projecting neurons, $p<0.0001$, bootstrap, Figure 4.1c,d). Thus,

while V1-projecting neurons encode a broad range of intensity variation features, they clearly emphasize a subset of them, in particular sudden onsets.

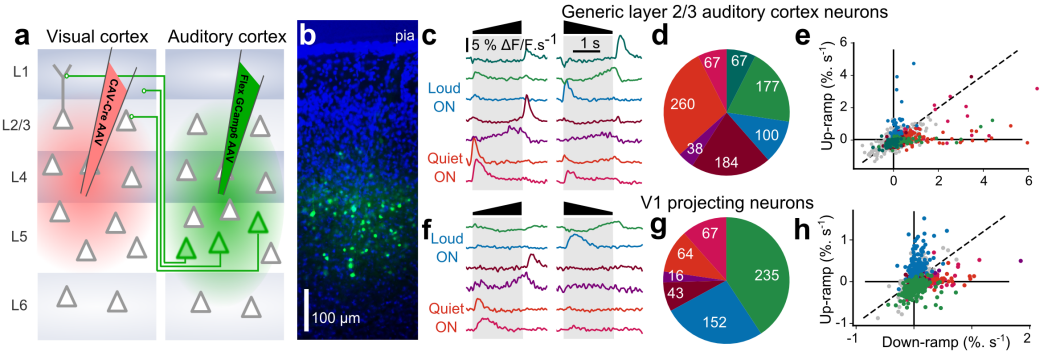

FIGURE 4.1: **V1-projecting neurons over-represent loud onsets of sounds.**

*(**a**) Viral expression strategy for GCAMP6s labelling of V1 projecting neurons in AC. (**b**) Epifluorescence image of an AC histological section (blue = DAPI, green = GCAMP6s) showing V1 projecting neurons (cortical layers are matched to **a**). (**c**) Response profiles and cell count distribution of the 7 different activity clusters fund in layer 2/3 AC neurons. Color code: Blue and greens = clusters preferring down ramps, warm colors = clusters preferring up-ramps. (**d**) Fraction of neurons corresponding to each cluster shown in c, with the absolute number neurons superimposed. (**e**) Distribution of the mean response to up- and down-ramps for the neurons included in the responses clusters shown in **c**. (**f-g**) Same as **c-e** but for the 6 clusters identified in V1 projecting neurons.*

The sign of auditory responses in V1 depends on the illumination-context.

How do these specific AC inputs impact V1? One study in which mice received no visual input (dark environment) suggests an inhibitory effect[251] while another study in which mice received visual stimuli describes excitatory effects[252]. Thus, the sign of AC inputs to V1 pyramidal cells could be illumination-dependent. To address this question, we performed two-photon calcium imaging in head-fixed awake mice (Figure 4.2a) using the calcium sensor GCAMP6s (Figure 4.2b) expressed in V1 through stereotaxic injection of an AAV-syn-GCAMP6s viral vector. Imaging of the same neurons was performed either in complete darkness or with a visual context, in the light, in front of a grey screen at low luminance (0.57 cd/m^2; note that when screen is on, the mouse also receives visual inputs from the surrounding of the screen). To monitor gaze stability, pupil position and diameter was tracked during the experiment (Figure 4.2c,d, Supplementary Figure 4.9). V1 was identified using Fourier intrinsic imaging[271] as the largest retinotopic field in visual areas (Figure 4.2e,f), and all two-photon imaging fields-of-view were mapped to the retinotopic field thanks to blood vessel landmarks (Figure 4.2e,f). All neurons imaged outside V1 according to this criterion (Figure 4.2f) were analyzed separately. As visual sampling can be influenced with auditory stimuli[245], we first measured whether sounds impacted triggered eye-movements. We found that sounds and in

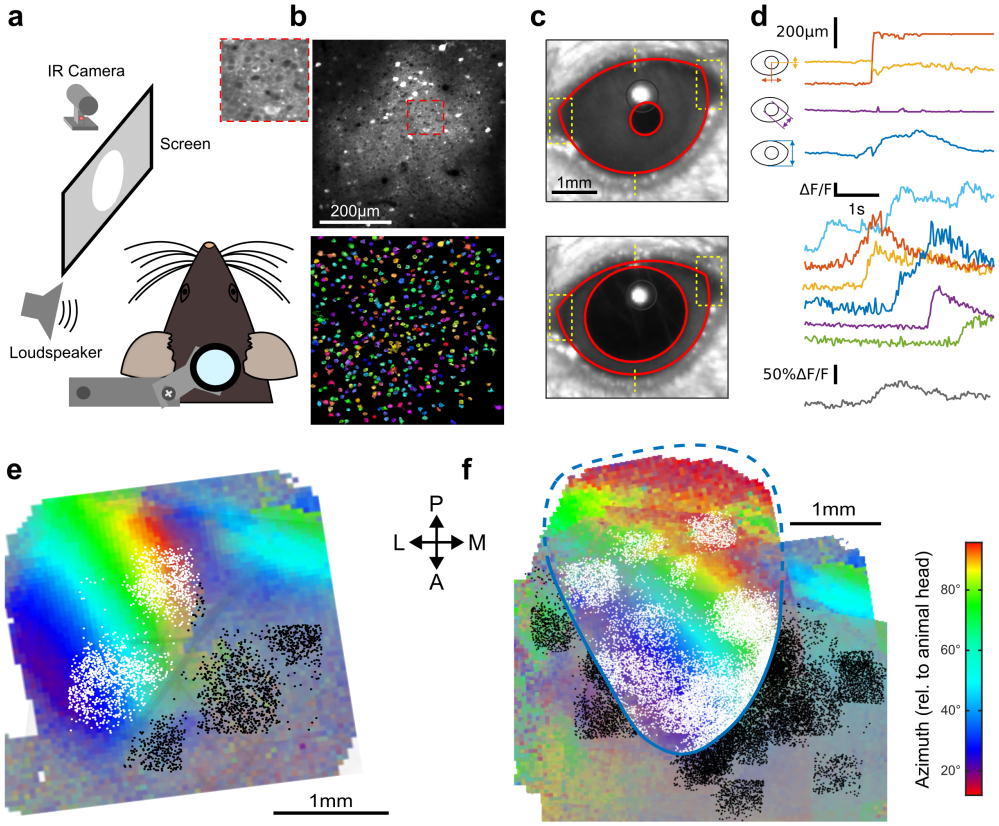

FIGURE 4.2: **Retinotopically-mapped two-photon imaging fields in V1.**

*(**a**) Sketch of the experimental setup. (**b**) (top) Example of a 0.5x0.5 mm imaging field of view showing GCAMP6s-expressing neurons in V1 (top with magnification in inset) and the ROIs that were automatically detected as putative neurons (bottom, see Methods). (**c**) Examples of eye tracking images. (**d**) (top) Eye tracking showing a large saccade during a blank trial. The blue, purple and yellow-red traces indicate apertures between eye lids, pupil diameter, x-y motion of pupil center is shown in red. (bottom) Examples of raw GCAMP6s traces from individual neurons recorded concomitantly in V1 (black, frame rate: 31Hz) and population average (grey) showing saccade-related neuronal activity. (**e**) (left) Example of a retinotopic map obtained with Fourier intrinsic imaging (see Methods) and (right) registered maps across 9 mice. The color code indicates the azimuth in the visual field.*

particular the sharp onsets of down-ramps triggered occasional and mostly horizontal eye saccades (Figure 4.3a). A sound induced change in pupil diameter was also observed (Figure 4.3b, Supplementary Figure 4.9). In the light, the saccades triggered responses in V1 likely reflecting visual inputs as these responses vanished in the dark (Supplementary Figure 4.9). Therefore, we excluded all trials with a saccade larger than mouse visual acuity (2° of visual angle) from our analyses (Figure 4.3a). After this correction and pulling together the activity of 18925 V1 neurons (35 sessions, 9 mice), we observed that sounds alone trigger responses in supragranular V1 neurons, and that, as expected from the activity of V1-projecting AC neurons, these responses were stronger for the down-ramp (Figure 4.3b,c).

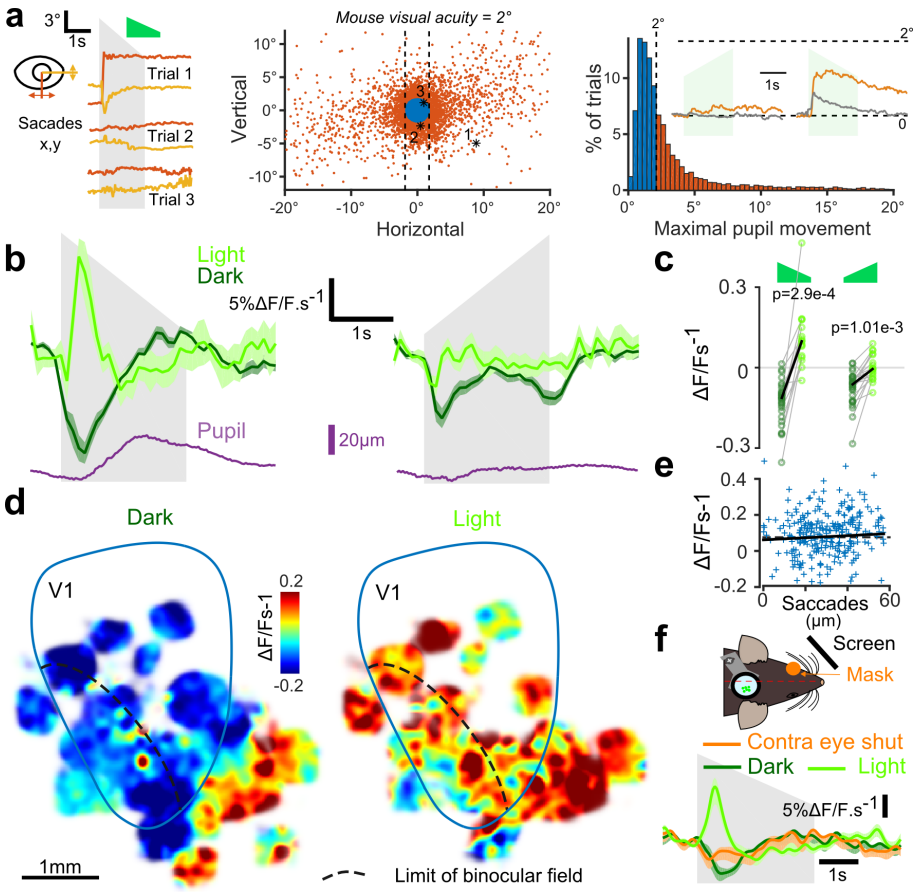

FIGURE 4.3: **Auditory responses in V1 switch sign with the presence of visual inputs.**

*(**a**) (left) Horizontal and vertical pupil movements recorded during three trials. (center and right) Distribution of saccades across all mice (n = 7) and trials. The blue portion indicates trials where eye pupil movement did not exceed visual acuity in the mouse (2° of visual angle: 84% of all trials). Inset: average saccade responses to up-ramps and down-ramps, before (orange) and after (grey) trial filtering. (**b**) Averaged deconvolved calcium traces of V1 neurons in the light (light green) and in the dark (dark green) (6207 neurons, n = 17 sessions in 7 mice). The purple line is the average pupil diameter. (**c**) Average V1 response are larger in the light than in the dark (n = 17 recording sessions in 7 mice, Wilcoxon signed rank test). (**d**) Smoothed maps of the local responses to the down-ramping sound, averaged across sessions and animals after registration with respect to the retinotopic map. (**e**) Single trial saccade amplitudes (below the 2° visual acuity threshold) do not correlate with the amplitude of V1 population responses to sounds (Pearson correlation coefficient 0.05, p = 0.42). (**f**) Mean V1 responses (2226 neurons, n = 13 sessions in 2 additional mice) in the dark, in the light and with the contralateral eye reversibly occluded as depicted in the sketch above.*

In addition, we clearly observed that the net population response was globally inhibitory when mice were in complete darkness and excitatory when mice were

in the light (Figure 4.3b,c), corroborating the idea that the sign of AC inputs depends on the visual context (darkness vs visual scene). To test whether this context-dependency was a general property of AC outputs or if it was specific to AC inputs impacting V1, we measured auditory responses in the secondary visual or associative areas next to V1 which also receive inputs from AC[259]. Interestingly, these areas displayed excitatory responses both in the light and in the dark (Figure 4.3d). Moreover, they showed richer auditory information, including selectivity to frequency (in FM sweeps) and to sound offsets, suggesting that they perform different cross-modal computations than V1 (Supplementary Figure 4.10). As a control, we verified that all excitatory effects in the light were not due to unfiltered micro-saccades as the amplitude of these extremely small saccades was not correlated to sound responses (Figure 4.3e). Moreover, excitatory responses to sounds could not be attributed to sound-induced pupil dilation as this effect occurred after V1 responses (Figure 4.3b). We then wondered whether the context dependence was related to a global change in arousal caused by complete darkness which would involve both hemispheres simultaneously, or to the absence of inputs to the visual system, which could be induced unilaterally. We therefore repeated the experiments with masking of the contra-lateral eye and observed that this condition reproduced the responses of the dark context (Figure 4.3f), ruling out the possibility of a global arousal effect. Thus, altogether we concluded that V1 implements a specific mechanism that reverses the impact of AC inputs depending on whether or not visual inputs are available.

Minimal model for the context-dependence of auditory responses.

To better apprehend the circuit mechanisms of this context-dependence, we looked for a minimal model of the effect, based on the facts that AC projections excite both pyramidal cells and interneurons in V1[252], and that V1-projecting AC neurons are not affected by light (Supplementary Figure 4.8). Considering a generic population of L2/3 V1 neurons (E-neurons), the minimal circuit which can provide sound-dependent inhibition to this population is a connection from an inhibitory population (I-neurons) driven by AC inputs (Figure 4.4a). In this case, as we exemplified in a simple simulation (Figure 4.4b,c), to render this inhibition context-dependent, the I-neurons musthave (i) a non-linear response curve (threshold function), and (ii) must be inhibited in the presence of visual inputs (light modulation, Figure 4.4a). Then, if I-neurons are close to activation threshold in the dark, and if inhibition by visual inputs brings them well below threshold in the light, I-neurons will be less active in the light than in the dark, leading for E-neurons to a dominance of excitation in the light and of inhibition in the dark (Figure 4.4b,c). This simplistic model could be extended into more complex models in which inhibition of I-neurons in the light is more explicitly implemented, but as we show in Supplementary Figure 4.11, whatever the complexity of the model, the context-dependent sign of auditory response in V1 requires the existence of neurons that have the typical response signature of I-neurons in light and dark (Figure 4.4d).

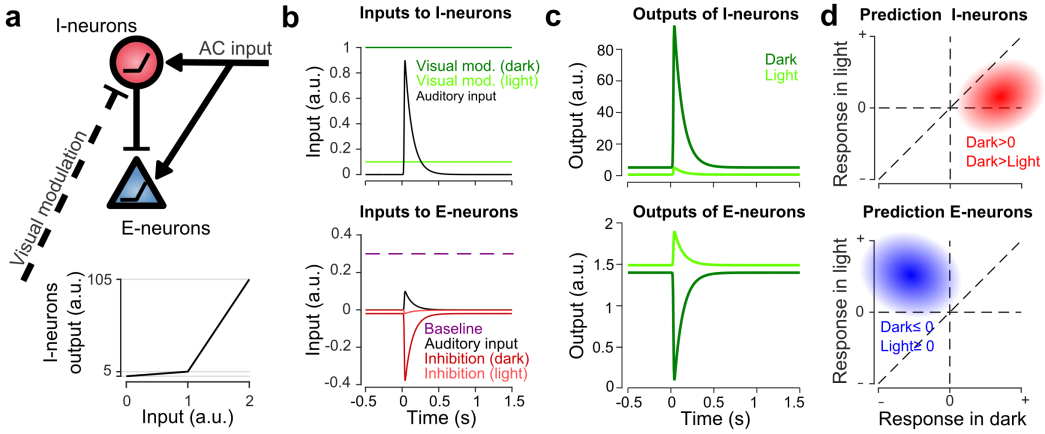

FIGURE 4.4: **A minimal model predicts the response signature of neurons mediating sound-induced inhibition.**

*(**a**) Sketch of the minimal model for the switch between negative and positive sound responses in dark vs in light. An inhibitory population (I-neurons) endowed with a non-linear response function receives positive auditory and negative visual inputs. If both inputs are present the drive is below threshold and no inhibition is delivered. Thus, only direct excitatory inputs impact the main neuronal population (E-neurons). (**b**) Input currents delivered to the I- and E-neurons during the simulations. (**c**) Outputs of the two populations in the presence (light) and absence (dark) of visual input. (**d**) Predicted responses signatures for E- and I-neurons as defined in the model.*

Layer 1 contains an interneuron subpopulation delivering context-dependent inhibition.

We thus aimed at identifying the predicted interneuron population that mediates context-dependent inhibition in V1. Because anatomy shows that AC to V1 projections are concentrated in L1 and L2/3[252], we imaged neurons in these two layers broadly labeled with GCAMP6s (synapsin promotor) in GAD2-Cre x flex-TdTomato mice which allow identifying inhibitory neurons based on red fluorescent protein expression (Figure 4.5a). We observed that in layer 2/3 both excitatory (n=4348) and inhibitory neurons (n=726) were globally inhibited by sounds in the dark and excited in the light, while the layer 1 interneuron population seemed globally unaffected in the dark and excited in the light (Figure 4.5b). However, the population trend concealed functionally distinct subpopulations. When we tested for significant positive or negative responses in the dark in single neurons (Wilcoxon rank-sum test, $p<0.01$), we found a large fraction of neurons significantly inhibited, but also, a subpopulation of 12.8% of all L1 inhibitory neurons that were significantly activated in the dark (Figure 4.5c). This positive response was not observed in L2/3 above significance threshold (Figure 4.5c). When plotting the response of L1 neurons in the dark against in the light (Figure 4.5d), it also became apparent that the L1 neurons that are excited by sounds in the dark tend to be less excited in the light (e.g. Figure 4.5e). In contrast, L1 neurons inhibited or unaffected by sounds in the dark became more activated in the light (Figure 4.5d). Consistently, statistical assessment showed that a significant fraction of 5.3% of L1 neurons (Wilcoxon rank-sum test, $p<0.01$) respond less in light than in dark (Figure 4.5f), and that almost all of

these neurons are significantly excited in the dark (4.9%, Figure 4.5g). In contrast, based on the same tests, there was no significant population of L2/3 neurons that responded less in light than in dark (Figure 4.5f, g). Thus, we concluded that layer 1 contains the only supra-granular subpopulation of GABAergic neurons that can provide a sound-induced inhibition gated by visual inputs. These neurons are thus good candidates to mediate the context-dependence of sound responses observed in the bulk of L2/3 V1 neurons (Figure 4.5h).

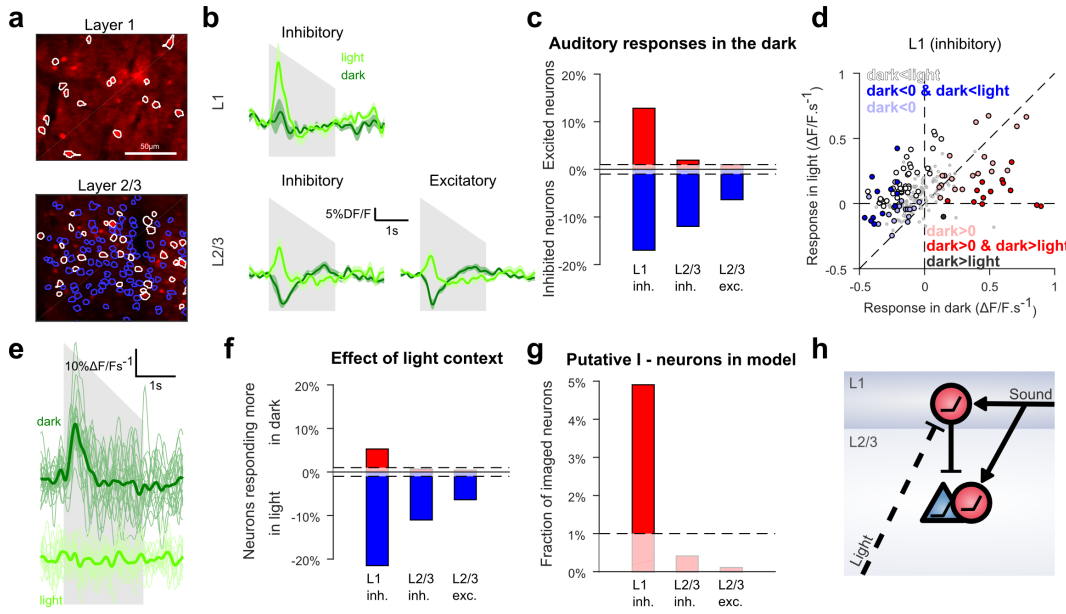

FIGURE 4.5: **Context-dependent inhibition by sound is mediated by a L1 subpopulation.**

*(**a**) In vivo 2-photon images of GAD2-positive V1 neurons in L1 and L2/3 expressing td-Tomato. Superimposed are the contours of the active regions of interest identified by GCAMP6s imaging. (**b**) Mean responses to the down-ramping sound for inhibitory and excitatory neurons in L2/3 and L1. (**c**) Fraction of neurons significantly excited (red) or inhibited (blue) by sounds for each layer and cell type (Wilcoxon rank sum test, p<0.01). (**d**) Scatter plot of the mean responses of all L1 inhibitory neurons to the down-ramping sound in the light against in the dark. Small grey dots indicates neurons that do not significantly respond in any condition. Larger, colored dots indicate significantly responding neurons. Dark blue indicates neurons significantly inhibited in the dark and responding less in dark than in light. Dark red dots indicates neurons significantly activated in the dark and responding more in dark than in light. Neurons responding equally in dark and light are marked with light blue (negative response) and light red (positive response) dots. Neurons that are non-responsive in the dark but respond more or less in the light are mark in white and dark respectively. (**e**) Single trial deconvolved traces for an L1 interneuron active in the dark and inactive in the light. (**f**) Fractions of neurons cell type responding significantly more in light than in dark (blue) or in more in dark than in light (red) (Wilcoxon rank sum test, p<0.01). (**g**) Fraction of neurons that can mediate context dependent inhibition in each layer i.e. interneurons significantly excited by sounds in the dark and responding significantly less in the light than in the dark. (**h**) Schematics of the core mechanism for context-dependent auditory responses in V1.*

Sounds with large onsets boost the representation and perception of coincident visual events.

We then assessed the impact of AC inputs to V1 on the representations of visual events. We recorded 9849 L2/3 neurons (23 sessions, 7 mice) expressing GCAMP6s under the synapsin promotor while delivering 2 sec long up- or down-ramping auditory stimuli together with a white disk looming or receding coincidently over the same duration as the sounds. The loudspeaker was placed such that auditory and visual stimuli came from a similar direction. Unimodal stimuli were also delivered to assess additivity for the bimodal conditions. We first observed that many V1 neurons displayed a supra-additive boosting of their visual response when coincident with the large onset of down-ramps (e.g. Figure 4.6a), an effect even visible at the population level (Figure 4.6b). Because looming and receding stimuli trigger different type of visual responses in V1, we performed a clustering in order to identify, in a model-free manner, the main types of visual responses and sound-induced modulations (see Methods, note that only the V1 neurons with high signal-to-noise and sufficient number of trials without eye movement, n=499, were retained by the clustering). This analysis identified seven distinct clusters (Figure 4.6c), among which three clusters (254 neurons) responded more to looming than receding disks, and three clusters (141 neurons) had the opposite preference. Four clusters were tightly direction-specific and responded only for their preferred stimulus, at its beginning or end, while the three others were less specific to direction, but also displayed temporal specificity. The temporal specificity was probably due to the location of the retinotopic receptive field of the neurons with respect to stimulus centre.

The direction specificity likely reflects the direction specificity of their receptive field, although we do not exclude that some of the specificity also relates to the global geometry of the stimulus as direction selectivity measured in the same neurons with drifting gratings was much less sharp than with receding/looming disks (Supplementary Figure 4.12). Also many cells responded to the disk but not to the gratings and conversely, suggesting complex non-linearities in receptive fields (Supplementary Figure 4.12). But more interestingly, the bimodal conditions revealed very clear supra-additive responses. Strikingly, all clusters responding at the onset of visual stimuli had their response strongly boosted by coincidence with the loud onset of down-ramps (Figure 4.6c, black arrows for clusters 1-5, bootstrap test, see Methods). No or much weaker boosting was observed with the quiet onset of up-ramps (Figure 4.6c), consistent with its reduced representation in AC to V1 projections (Figure 4.1). Responses occurring towardsthe end of the stimulus were minimally (although significantly, cluster 5) boosted, probably because they correspond to auditory features ("loud and quiet tonic") that are more weakly impacting V1 (see Figure 4.3b). Our data thus suggests that visual responses in V1 are boosted specifically by the coincidence of loud sound onsets. Note that a moderate response suppression was also observed, but only in cluster 7 (Figure 4.6c, blue arrow). Interestingly, the boosting of visual responses appeared to be a strong feature of V1, as the same analysis identified only one nonlinear cluster out of six clusters in the associative area medial to V1 (Supplementary Figure 4.13) .

We thus wondered whether V1 boosting could have a perceptual impact in mice. To do so, we trained head-fixed water-deprived mice to respond to looming or receding visual stimuli by liking at a water spout to get a reward (Figure 4.6d). After

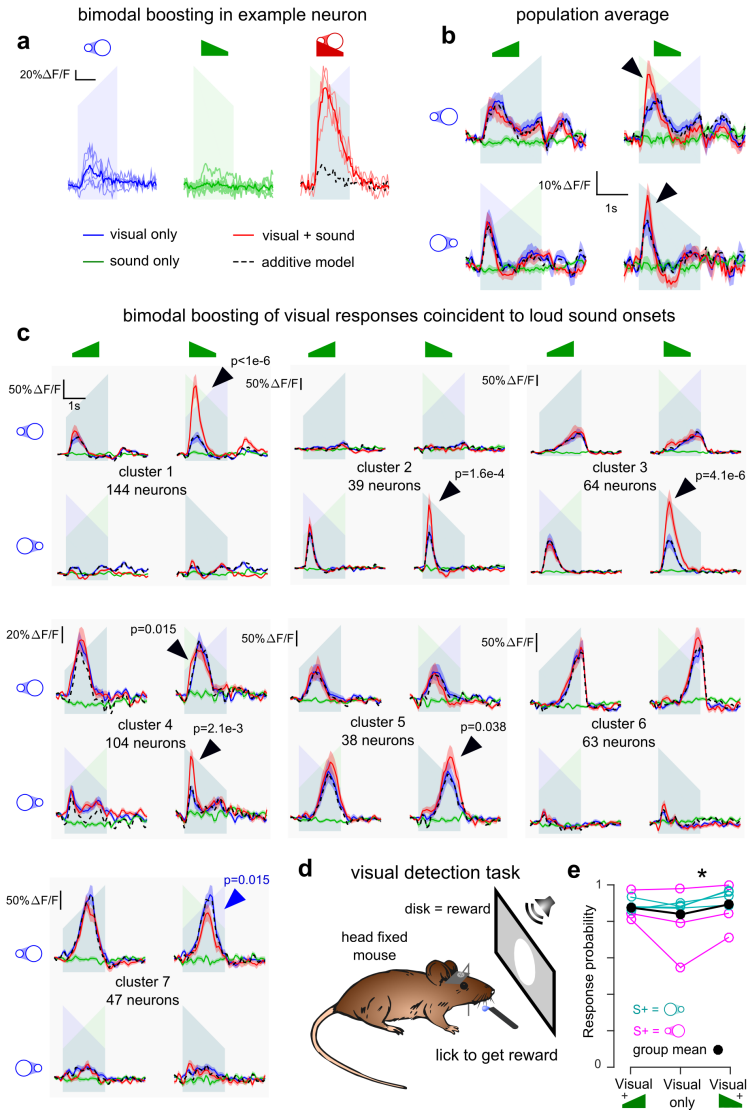

FIGURE 4.6: **V1 neurons are boosted by the coincidence of large auditory onsets with visual onsets.**

*(**a**) Examples of multisensory neurons in V1. Raw GCAMP6s traces for a visual stimulus (blue), a sound (green) and a combination of both (red) show a large neuronal response for bimodal condition compared to additive prediction (dashed black line) based on unimodal responses. Individual trials (thin lines) and their average (thick lines) show that the multisensory responses are robust for those neurons. (**b**) Average deconvolved calcium responses (mean ± sem) in V1 for up- and down-ramping sounds (green), looming and receding visual stimuli (blue) and their four bimodal combinations. Black dashed lines: linear prediction. (**c**) Uni- and bimodal responses, displayed as in **b**., of the six different functional cell types identified by hierarchical clustering. (**d**) Sketch of the behavioral experiments. (**e**) Increase of visual stimulus detection rate when a coincident down-ramping sound is played (Wilcoxon signed rank test, $p = 0.0156$); a non-significant increase is also observed with the up-ramping sound ($p = 0.375$).*

training, mice reliably responded to the visual stimulus with a success rate of 83.7 ±

5.7%. However, in trials in which the visual stimulus was paired with a down-ramp (loud onset) the response probability was significantly increased to 90.2 ± 4.1% (Figure 4.6e, Wilcoxon signed rank test, p = 0.0156). A weaker, non-significant effect was seen with up-ramps (Figure 4.6e, 88.4 ± 2.2%, Wilcoxon signed rank test, p= 0.375). Thus, loud sound onsets not only boost concomitant visual responses in V1, but also improve visual detection in a behavioral task.

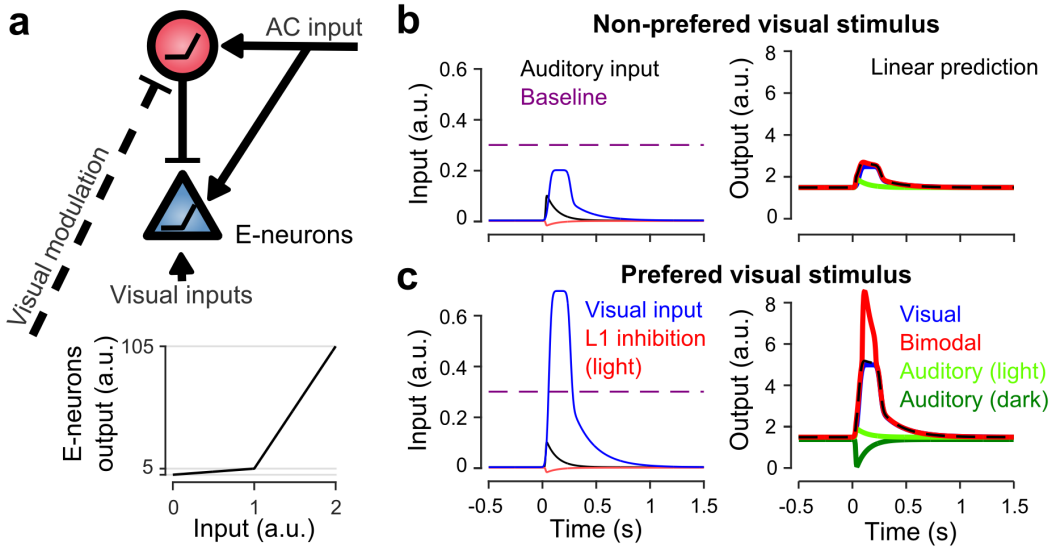

FIGURE 4.7: **Sound-dependent boosting is reproduced by the L1 inhibition model.**

*(**a**) Sketch of gated inhibition model as in Figure 4.4a. (**b**) Input currents (left) received by the E-neurons population and output response (right) for a non-preferred, "subthreshold" visual input. (**c**) Same as in **b**. but for a preferred, suprathreshold visual input.*

Audio-visual boosting can result from direct AC excitation if pyramidal cells are non-linear.

What could be the mechanism of sound-induced boosting? To answer this question, we attempted to extend our minimal model for sound responses in V1 (Figure 4.4), in order to reproduce also the boosting effect. We realized, that the same model can reproduce boosting by sounds (Figure 4.7a-c) provided that pyramidal cells have a simple non-linearity. We modeled this non-linearity with a thresholded activation function having a weak subthreshold output gain and a high supra-threshold gain (Figure 4.7a). This non-linearity is necessary to account for the fact that responses to sound alone are much weaker than the bimodal boosting effects. With this design, we could also reproduce the observation that boosting mainly occurs for the preferred visual stimulus (Figure 4.6c), provided that the auditory and non-preferred visual inputs are driving excitatory neurons in the subthreshold regime (Figure 4.7b) while the preferred input brings excitatory neurons close or above threshold (Figure 4.7c). In this case the auditory response sums with a high gain for the preferred stimulus and a low gain for the non-preferred. Thus altogether, our data and modeling work indicates that a minimal circuit including a direct excitation from auditory cortex onto L2/3 pyramidal cells and an indirect inhibition via a subpopulation of L1

interneuron is sufficient to provide an illumination-dependent auditory input that emphasizes visual events coincident to sound loud onsets.

Discussion

Using two-photon calcium imaging in identified excitatory and inhibitory neurons during auditory and visual stimulation, we demonstrated three important features of AC to V1 connections. First, we showed that they preferentially transfer a subset of the temporal auditory features encoded by AC, most prominently large amplitude onsets. Second, they are gated by a context-dependent inhibitory mechanism likely implemented by a subpopulation of L1 interneurons. Third, their most salient impact on V1 representations is to produce a strong boosting of responses to dynamical visual stimuli concomitant to sharp sound onsets. The preference of V1 projecting neurons as compared to bulk layer 2/3 neurons for the features of suddenly rising and slowly decaying sounds (Figure 4.1) is an interesting novel case of coding bias in a projection-defined neuronal subpopulation of a cortical area. Similar results have been found in somatosensory cortex for neurons projecting to secondary sensory or motor cortex[272]. In AC, it is well established that specific acoustic features are related to the horizontal localization of neurons in different subfields[193, 273, 274] and that, at least in higher mammals such as cats, these subfields can project to independent pathways propagating different acoustic information for behavioral decisions (e.g. localization and quality[275]). Our results show that projection specificity can also be a determinant of feature specificity, at least for intensity modulation features.

Strikingly, this select auditory information impacts V1 in a dynamic manner, with a net inhibitory drive in the absence of visual inputs and a net excitatory drive in layer 2/3 neurons in the presence of visual inputs (Figure 4.3b). Sound-induced inhibition of V1 has been described earlier[251]. We here show that this inhibition is specific to darkness, and thus could specifically serve to decrease V1 activity in the dark when visual information is irrelevant for sound source localization. The illumination-dependence of auditory responses in V1 is reminiscent of the recent observation that AC to V1 projections, activated by loud sounds and acting through L1, boost preferred orientation responses in V1 and inhibit non-preferred orientations[252]. The two effects share several properties including selectivity to high sound levels, the dominance of inhibition for weaker visual inputs and of excitation for stronger visual inputs, and the involvement of L1 neurons for mediating inhibition. The main discrepancy is the apparent lack of visual input specificity of the illumination-dependent excitation of V1 neurons by AC inputs alone. This could be due to differences related to the type of visual inputs used. One important result appearing in our data is that only a fraction of all L1 interneurons have response properties compatible with the function of releasing an illumination-dependent, sound-triggered inhibition (Figure 4.5c,f,g), while these properties are not found in L2/3 (Figure 4.5c,f,g). Thus the silencing of visual processing by sounds in the dark is not a generic function of L1 interneurons, compatible with the observation that only VIP negative interneurons receive AC inputs in V1[252] and the observation that L1 interneurons serve other function[276], such as disinhibition of L2/3 in various contexts[277, 278, 279, 280, 281]). An interesting question that cannot be addressed with the current observations is how the specific L1 neurons that mediate sound-induced inhibition are themselves inhibited in the presence of visual inputs. Several putative mechanisms are compatible with the data including disinhibition (see models

in Supplementary Figure 4.11).
Another important phenomenon observed in our data is the strong boosting of visual responses coincident to loud sound onsets (Figure 4.6). Our model suggests that this boosting results directly from the non-linear summation of visual and auditory inputs in V1 pyramidal cells, while sound-triggered inhibition is abolished by visual context (Figure 4.7). The large amplitude of the sound-induced boosting as compared to the weaker responses observed when sounds are played without coincident visual input (Figures 4.3 and 4.5 vs Figure 4.6, see also Figure 4.7) points towards a threshold mechanisms, allowing for amplification of the auditory input only if sufficient visual input concomitantly arrives. Such a gated amplification mechanism could be implemented directly in L2/3 pyramidal neurons using the direct inputs they receive from auditory cortex[252, 260]. For example, amplification of L1 input by coincident somatic inputs thanks to calcium spikes in the apical dendrite has been described both in L5[282] and L2/3[283] pyramidal cells. While this is a good candidate mechanism for the observed boosting effect, we cannot exclude alternative or complementary mechanisms such as a disinhibition mediated by a subpopulation of L1 neurons[252, 279, 280, 281], different from the subpopulation providing direct inhibition (Figure 4.5). Also, as down ramping sounds are salient stimuli, a cholinergic input to visual cortex, known to increase cortical responses via desinhibition [277, 278, 284] could partially contribute to the boosting effect. Independent on the mechanism, our data show that sound-induced boosting is a strong effect in V1 which tightly relates to coincidence with the onset of salient sounds. This effect provides a powerful way to highlight, in visual space, the visual events potentially responsible for the sound, and thus might be an essential element of the cortical computations related to the localization of sound sources[245].

Methods

Animals.

All animals used were 8-16 week-old male C57Bl6J and GAD2-Cre (Jax #010802) x RCL-TdT (Jax #007909) mice. All animal procedures were approved by the French Ethical Committee (authorization 00275.01).

Two-photon calcium imaging in awake mice.

Three to four weeks before imaging, mice were anaesthetized under ketamine medetomidine. A large craniotomy (5 mm diameter) was performed above the right primary visual cortex (V1) or the right auditory cortex. For the imaging of the auditory cortex, the right masseter was removed before the craniotomy. We then performed three to four injections of 150 nl (30 nl.min^{-1}), AAV1.Syn.GCaMP6s.WPRE virus obtained from Vector Core (Philadelphia, PA, USA) and diluted x10. The craniotomy was sealed with a glass window and a metal post was implanted using cyanolite glue followed by dental cement. A few days before imaging, mice were trained to stand still, head-fixed under the microscope for 30–60 min per day. Then mice were imaged 1-2h per day. Imaging was performed using a two-photon microscope (Femtonics, Budapest, Hungary) equipped with an 8 kHz resonant scanner combined with a pulsed laser (MaiTai-DS, SpectraPhysics, Santa Clara, CA, USA) tuned at 920 nm. We used 20x or 10x Olympus objectives (XLUMPLFLN and XLPLN10XSVMP) , obtaining a field of view of 500x500, or 1000x1000 microns, respectively. To prevent

light from the stimulation screen to enter the microscope objective, we designed a ring silicon mask that covered the gap between the metal chamber and the objective. Images were acquired at 31.5 Hz during blocks of 5s interleaved with 3s intervals. A single stimulus was played in each block and stimulus order was randomized. All stimuli were repeated 20 times except drifting gratings (10 repetitions).

Drifting square gratings (2Hz, 0.025 cyc/$^\circ$) of 8 different directions (0° from bottom to top, anti-clockwise , 45°, 90°, 135°, 180°, 225°, 270°, 315°), and a size-increasing (18° to 105° of visual angle) or decreasing white disk over a black background were presented on a screen placed 11 cm to the mouse left eye. The luminance of the black background and white disks was 0.57 and 200 cd/m^2 respectively. The screen (10VG BeeTronics, $22 * 13$ cm) was located 11cm from the contra-lateral eye, thus covering $90^\circ * 61^\circ$ of its visual field including some of the binocular segment. For auditory cortex recording, we used 250 ms constant white noise sounds at 4 different intensity modulations, and two up- and down-intensity ramping sound between 60 and 85 dB SPL. Note that microscope scanners emitted a constant background sound of about 45dB SPL. For visual cortex experiments, we used only the two intensity ramps and added two frequency modulated sounds, going linearly from 8 kHz to 16 kHz and vice versa. Up- and down- amplitude ramps were combined with the increasing and decreasing disks to form 4 multisensory stimuli. The loudspeaker was placed just next to the stimulation screen, facing the contralateral eye. All stimuli were 2 s long. Auditory and visual stimuli were driven by Elphy (G. Sadoc, UNIC, France). All sounds were delivered at 192 kHz with a NI-PCI-6221 card (National Instrument) an amplifier and high-frequency loudspeakers (SA1 and MF1-S, Tucker-Davis Technologies, Alachua, FL). Sounds were calibrated in intensity at the location of the mouse ear using a probe microphone (Bruel & Kjaer).

For V1 recordings, auditory only stimulations were performed in two different contexts: either in complete darkness (screen turned off in the sound and light isolated box enclosing the microscope) or in dim light (screen turned on with black background, luminance measured to 0.57 cd/m^2).

Intrinsic optical imaging recordings.

To localize the calcium-imaging recordings with respect to the global functional organization of the cortex, we performed intrinsic optical imaging experiments under isoflurane anaesthesia (1%). The brain and blood vessels were illuminated through the cranial window by a red (intrinsic signal: wavelength 780 nm) or a green (blood vessel pattern: wavelength 525 nm) light-emitting diode.

To localise the visual cortex, the reflected light was collected at 15 Hz by a charge-coupled device (CCD) camera (Foculus IEEE 1394) coupled to the epifluorescence light path of the Femtonics microscope (no emission or excitation filter). A slow drifting bar protocol was used: a white vertical bar was drifting horizontally over the screen width for 10 cycles at 0.1Hz, from left to right in half of the trials, and from right to left in the other half. After band-passing the measured signals around 0.1Hz, and determining their phase in each pixel and for each condition, both the preferred bar location and the hemodynamic delay could be determined for each pixel, yielding azimuth maps as in (Figure 4.2e). Similarly, elevation maps were obtained in a subset of the animals using a horizontal bar drifting vertically. These maps coincided

with those obtained in previous studies[271], so we could determine V1 border as the limit between pixels displaying and not displaying retinotopy tuning.

To localise the auditory cortex, the reflected light was collected at 20 Hz by a CCD camera (Smartek Vision, GC651M) attached to a custom-made macroscope. The focal plane was placed 400 mm below superficial blood vessels. A custom-made Matlab program controlled image acquisition and sound delivery. We acquired a baseline and a response image ($164 * 123$ pixels, $\sim 3.7 * 2.8$ mm) corresponding to the average images recorded 3 s before and 3 s after sound onset, respectively. The change in light reflectance ($\Delta R/R$) was computed then averaged over the 20 trials for each sound frequency (4, 8, 16, 32 kHz, whitenoise). A 2D Gaussian filter (σ = 45.6 μm) was used to build the response map and determine the localisation of the auditory cortex due to the stereotypical response map produced by the 4kHz sound. Sounds were trains of 20 white noise bursts or pure tone pips separated by 20 ms smooth gaps.

Eye tracking and trial filtering.

Left eye measurements (eye size, pupil diameter, pupil movement) were measured by tracking the eye using a CCD camera (Smartek Vision, GC651M). A Python software was used to capture images from the camera at 50Hz, synchronized with the cortical recordings.

These movies were analyzed off-line using custom automatic Matlab programs that traced the contours, first of the eyelid, second of the pupil. The eye lid shape was approximated by two arcs and involved the estimation of 6 parameters (4 for the coordinates of the two points where the arcs join, 2 for the y-coordinates of the crossings of the arcs with the vertical line at halfway between these two points; see Figure 4.2c, bounds for the parameters were set by hands and appear in yellow). The pupil shape was approximated by an ellipse, described by 4 parameters (center x and y, radius and eccentricity). Both estimations were performed by maximizing the difference between average luminance inside and outside the shape, as well as the luminance gradient normal to the shape boundary; in addition they were inspected manually and corrected occasionally inside a dedicated graphic interface. In the "dim light" and "bimodal" (but not the “dark”) contexts, it was observed that saccades correlated with population activity increase (e.g. Supplementary Figure 4.9), therefore we discarded all trials from these contexts displaying saccades. To do so we computed the maximal distance between pupil center positions at different instants during the 2 seconds of stimulation $\max_{0<t,t'<2s} \sqrt{(x_c(t') - x_c(t))^2 + (y_c(t') - y_c(t))^2}$ and discarded trials where the change in gaze exceeded the mouse visual acuity of 2°, i.e. where this distance exceeded 57.6μm (assuming an eye radius of 1.65mm).

Calcium imaging data summary.

The data obtained from the different imaging experiments consisted of the following.
A1 recordings: 3,771 cortical neurons from 18 L2/3 imaging areas in 7 animals, bimodal context only. And 1,593 neurons projecting to V1 from 14 L2/3 areas in 3 additional animals.
V1 recordings: 18,925 neurons from 35 L2/3 imaging areas in 9 mice, and included

the three "dark", "dim light" and "bimodal" blocks in random order. Eye tracking was performed in 23 of these sessions (9,849 neurons, 7 mice; only these sessions where used for the analyses of the "dim light" and "bimodal" context, which required filtering out trials with saccades).
V1 recordings with labelling of Gad-positive interneurons: an additional 4,348 neurons from 12 L2/3 imaging sessions (depth ranging from 140 to 300 microns), and 265 GAD2 positive neurons from 7 L1 imaging sessions (depth from 30 to 100 microns). All of these sessions had eye tracking, and included the "dark" and "dim light" blocks. 6 L2/3 sessions also included "bimodal" blocks, whereas the 13 remaining sessions included a second "dim light" block where the left (contralateral) eye was covered by a dark mask (Figure 4.3f).

Calcium imaging pre-processing.

Data analysis was performed with custom-made Matlab scripts available upon request. Every frame recorded was corrected for horizontal motion to a template image using rigid body registration (all sessions with visible z motion were discarded). Regions of interest were then automatically selected and human checked as the cell bodies of neurons with visually identifiable activity[270] and the mean fluorescence signal $F(t)$ was extracted for each region. We also estimated the local neuropil signal $F_{np}(t)$ for each neuron[44] and subtracted a fraction of it from the neuron signal. This fraction was set to 0.7, according to a previous calibration for GCAMP6s in mouse visual cortex[108] . Baseline fluorescence F_0 was calculated as the minimum of a Gaussian-filtered trace over the 5 neighbouring 5s imaging blocks and fluorescence variations were computed as $f(t) = (F_c(t) - F_{c0})/F_{c0}$. Analyses were performed either on these normalized fluorescence signals, or on estimations of the firing rate obtained by temporal deconvolution[143, 115] as $r(t) = f'(t) + f(t)/\tau$, in which $f'(t)$ is the first derivative of $f(t)$ and τ = 2 s, as estimated from the decays of the GCAMP6s fluorescent transients[108]. This simple method efficiently corrects the strong discrepancy between fluorescence and firing time courses due to the slow decay of spike-triggered calcium rises[115] , and was preferred to more advanced deconvolution methods because it does not bias deconvolution towards absence of activity (i.e. interpreting small signal as noise) in cells displaying poor signal to noise ratio. However it does not correct for the relatively slow rise time of GCAMP6s, producing a time delay on the order of 70 ms between peak firing rate and peak deconvolved signal[44].

Data analysis.

Data analysis was performed with custom-made Matlab and Python scripts available upon request.
Clustering of auditory responses in the auditory and visual cortices (Figure 4.1), and of bimodal conditions in the visual cortex (Figure 4.6) was performed using the following procedure. Calcium responses were deconvolved, baseline subtracted, blank subtracted, averaged over trials, and the responses to all considered stimuli were concatenated. Hierarchical clustering was performed using the Euclidean metric and Ward method for computing distance between clusters. To determine the number of clusters we moved down the clustering threshold of until clusters became redundant (overclustering) as assessed visually. This method clusters neurons

irrespective of whether they significantly responded to the stimuli. As a large number of neurons were not (or very weakly responsive) to the stimuli, a large fraction of the neurons were grouped into "noisy" clusters displaying no obvious sensory response. These "noisy" clusters were eliminated based on their noise levels measured as the mean distance between the raw average responses in the cluster and their time-filtered version (Gaussian kernel, σ = 95ms). All clusters with less than 5 neurons or with a noise level larger than 20% ($\Delta F/F.s^{-1}$) were discarded. These thresholds were chosen by visual inspection of the obtained clusters.

To make sure cell type distributions were not skewed by the fact that clustering outputs the most robust auditory responses (Figure 4.1), we re-aggregated neurons discarded as non-responsive if the mean correlation of their activity signature with any of the identified clusters was larger than 0.2. This procedure did not change the mean cluster response profiles and did not qualitatively impact the conclusions drawn on the distributions of the different clusters in V1-projecting and L2/3 neurons. In fact, the fraction of cells preferring down-ramps in V1-projecting non-re-aggregated clusters was even larger (71% in V1-project for 37% in L2/3, 4-fold difference for loud onset neurons), indicating that the observed preferences are even clearer in strongly driven neurons.

Statistics.

Data are displayed as mean $\pm$ S.E.M.

To assess the significance of the relative distribution of functional cell types in V1-projecting neurons and in layer 2/3 neurons, we performed a bootstrap analysis. The null-hypothesis is that both populations have the same distribution of clusters. To simulate this hypothesis we pooled together all neurons and performed 10,000 random partition of the pool population into the number of V1-projecting and L2/3 neurons. For each partition, we computed the difference of between the fractions of each clusters across the two partitions. This led to distributions of expected fraction differences for each cluster under the null hypothesis. The p-value was computed from the percentile of which the actual observed fraction difference was located.
Significant responses for individual neurons were detected using the non-parametric Wilcoxon rank-sum test. Raw calcium fluorescence responses were subtracted for pre-stimulus level, and averaged over a time window near the response peak. The vector of such responses for different trials was compared to the same computations performed on blank trials (unless responses to two different conditions were compared, such as auditory responses in the dark vs. in a dim light). Unless specified otherwise, we used a statistical threshold of 1% for detecting responding neurons; by definition this means that on average 1% of the neurons not responding to the condition would be detected as responding (false positive). Therefore in all the histogram displays of the fractions of responding neurons (e.g. Figure 4.5c, f) we masked the first 1% as being potentially only false positive detections.
To assess the significance of supra- or sub-linear responses to audio-visual combinations in individual clusters resulting from the clustering of bimodal responses, we used a bootstrap consisting in shuffling the different trial repetitions. The null hypothesis is that the cluster's average bimodal responses can be predicted as the sum of average visual and auditory responses, after baseline has been subtracted. To test this hypothesis, we first computed, for each neuron, individual "repetitions"

of the linear predictions by adding responses to one visual presentation and to one auditory presentation. Then we performed 10^6 shufflings of the labels of "linear prediction" and "bimodal response" trials, yielding a distribution of the expected difference between their averages under the null hypothesis. The p-values were computed from the percentile in which the actual nonlinear difference was located.

Model.

Simulations were performed using a rate model with 2 or 3 populations and no synaptic delay. The spiking activity r_i of population i followed the equation:

$$r_i = f\left(b^i_{context} + g^i_{auditory} s_{auditory}(t) + g^i_{visual} s_{visual}(t) + \sum_j g^i_j r_j(t-1)\right) \tag{4.1}$$

Where $s_{auditory}$ and s_{visual} are the auditory and visual input (displaying an exponentially decaying signal and a plateauing signal, respectively), $g^i_{auditory}$, g^i_{visual}, and g^i_j the connectivities to input and to the other populations (see model schematics in Figure 4.4a and Supplementary Figure 4.11 for which connections are non-zero). $b^i_{context}$ is a baseline accounting for other inputs, and whose value can depend on the context ("dark" or "light"). f is a nonlinear function consisting of two linear segments (Figure 4.4a) that accounts for spiking threshold or further nonlinear (e.g. dendritic) computations. The first segment is not a constant zero accounts for the fact that the other inputs summarized in $b^i_{context}$ are in fact stochastic, and can lead the cell to fire even when its average potential is below threshold.

Behavior.

We tested the effect of playing sounds during a visual detection task. Water-deprived mice (33 ml.g^{-1} per day) were head-fixed and held in a plastic tube on aluminium foil. The first day they were habituated to the fixation for one hour and manually rewarded. Then they were trained to lick when the rewarded stimulus (S+), the 2 second either looming or receding white disk on dark background, was presented. The mouse had to produce at least one lick on a stainless steel water spout to receive a 5 ml water drop. Licks were detected by changes in resistance between the aluminium foil and the water spout. To refrain mice from constantly licking, trials were started only following a period of 3 seconds without any lick. During the first sessions the reward was delivered immediately following the first lick; later the reward was delivered at least 1.7s after stimulation onset. After mice successfully responded to S+, achieving a steady state level of performance (6 sessions of $\sim$200 episodes), up- or down-ramping sounds were introduced and were played together with visual stimulation for half of the episodes; reward remained the same on these bimodal conditions.

Data availability.

The data that support the findings of this study and the analysis code are available from the corresponding author upon request.

Acknowledgments

We thank A. Fleischmann, Y. Frégnac and J. Letzkus for helpful discussions and comments on the manuscript. We thank A. Fleischmann for providing CAV2-Cre viruses. This work was supported by the Agence Nationale pour la Recherche (ANR "SENSEMAKER"), the Marie Curie FP7 Program (CIG 334581), the Human Brain Project (SP3 - WP5.2), the Fyssen foundation, the DIM "Region Ile de France", the International Human Frontier Science Program Organization (CDA-0064-2015), the Fondation pour l'Audition (Laboratory grant) and the Paris-Saclay University (Lidex NeuroSaclay, IRS iCode and IRS Brainscopes).

Author Contributions

TD, AK and BB designed the study, performed and analyzed the experiments, and wrote the paper.

Supplementary Figures

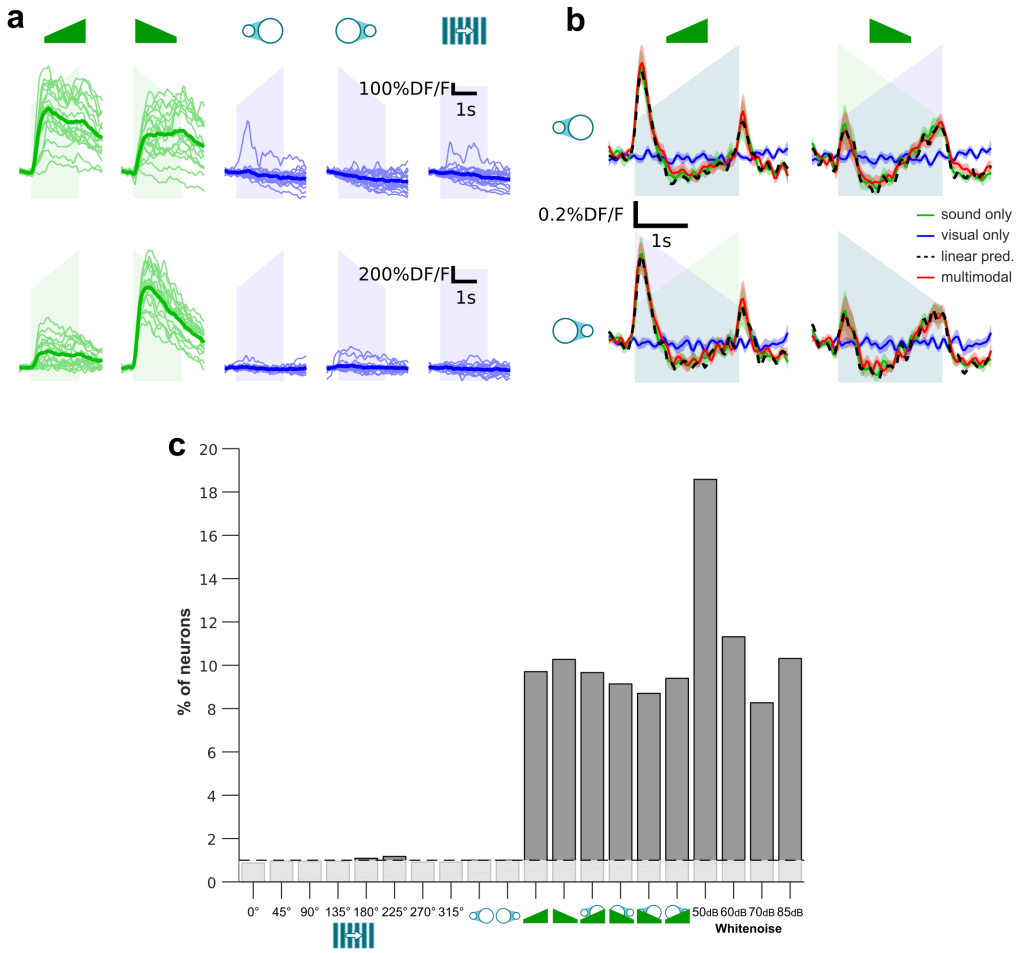

FIGURE 4.8: **Absence of visual responses or modulations in A1.**

*(**a**) Raw calcium activity trace of two auditory cortex neurons during auditory or visual stimulation (thick lines: mean and thin lines: single trials). (**b**) Mean layer 2/3 AC population response to sounds (green), visual stimuli (blue) and both (red) (n= 3771 neurons, across 18 sessions in 7 mice) show no impact of visual stimuli on AC responses to sounds. (**c**) Fraction of neurons responding to the different conditions (Wilcoxon rank sum test, p<0.01). Only those including an auditory stimulation elicited statistically significant responses.*

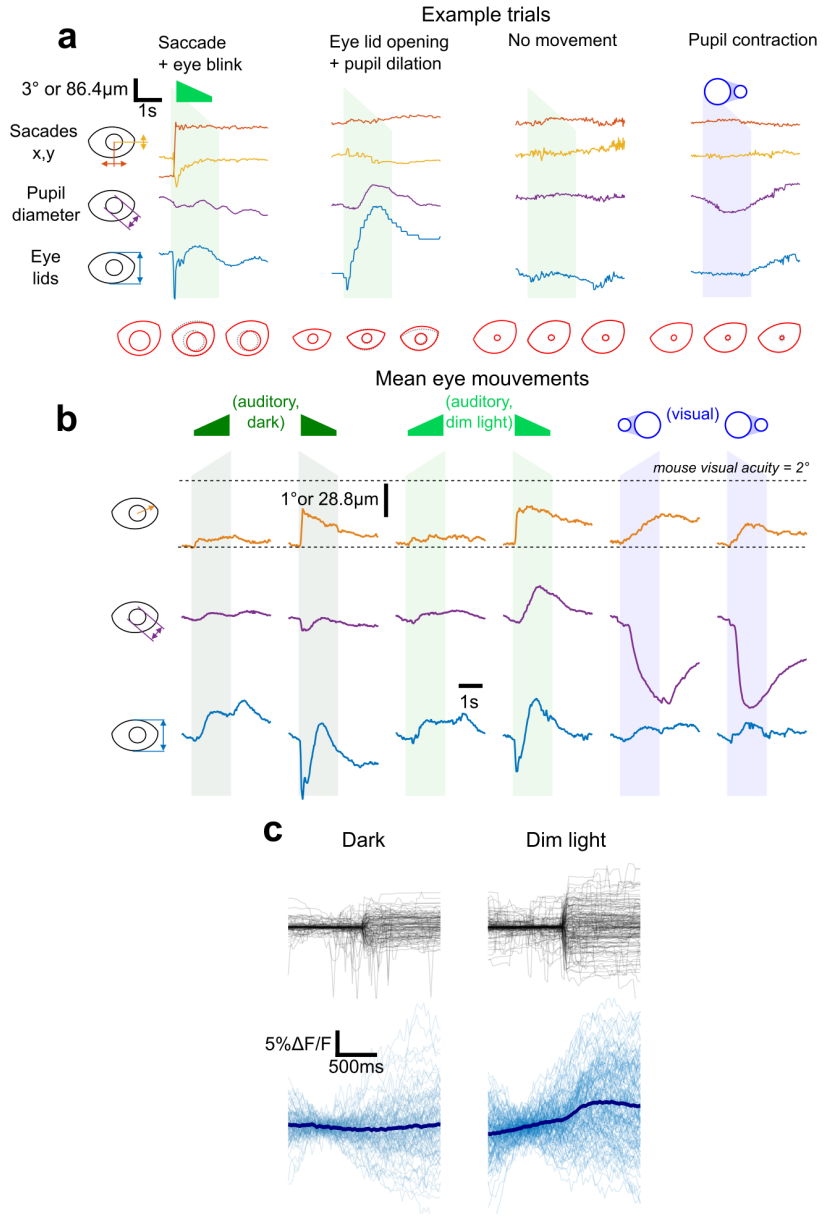

FIGURE 4.9: **Sounds generate occasional saccades, pupil dilation and eye lid movements.**

*(**a**) Examples of eye movements encountered during the recordings sessions. Auditory or visual stimuli can trigger eyelids movements (blue), pupil centre displacements (red and yellow), or pupil diameter variations (purple). The first three columns represent different single trial eye movement traces for a white noise sound decreasing in intensity (shaded green) and the last column represents a trial in which a only visual stimulus was played (receding disk, shaded blue). Below, schematics (red) representing the eyelids and pupil before, during and after the sound for each column. (**b**) Average eyelids and pupil measurements for 4 auditory and 2 visual conditions across 23 sessions in 7 mice. Dashed lines for the pupil angular position indicate mouse visual acuity (2° of visual angle). (**c**) Pupil movement traces in stimulus-free trials (black) at saccade onsets across 23 sessions in 7 mice. The corresponding single trial (thin lines) and average (thick line) calcium responses are shown below in complete darkness (left) and in dim light (i.e. low-level grey screen, right).*

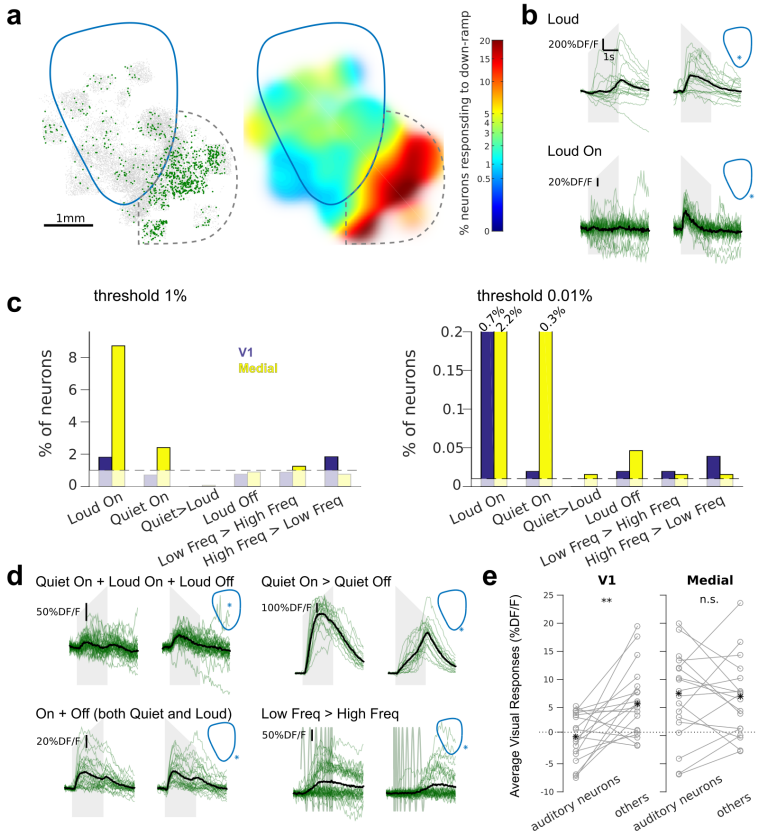

FIGURE 4.10: **Larger number of auditory-driven neurons in layer 2/3 for regions anteromedial to V1.**

*(**a**) Map of the neurons displaying significant positive responses to the down-ramping sound (Wilcoxon rank sum test, p<0.01). (Left) individual responding neurons are marked in green; (right) spatially smoothed ratio of positive responses. The area medial and frontal to V1 (secondary visual area and posterio-parietal cortex, delineated by a dashed grey line) displays a much higher number of positive responses than in V1. (**b**) Example raw calcium responses to the up- and down-ramps for two neurons, one in V1, the second in the medial area (insets indicate the neuron position on the map). (**c**) Ratio of neurons in V1 and in the anteromedial area displaying various auditory response properties assessed by a statistical test. From left to right: loud onsets (i.e. down-ramp onset), quiet onset (i.e. up-ramp onset), preference for quiet against loud onsets, loud offsets, preference for low against high frequencies based on frequency sweeps, preference for high against low frequencies. The ratios were calculated for two different statistical thresholds (left: p<0.01 right: p<0.0001, rank sum tests). Dashed line correspond to the statistical thresholds, thus indicating the expected false positive detection rate under null hypothesis. As also seen in **a**. the fraction of sound-responsive neurons is much higher in the anteromedial associative area than in V1, especially for the down-ramping sound. (**d**) Example raw calcium traces for neurons tuned to specific sound features including onset or offsets or frequency preference. (**e**) Average responses to visual stimuli across auditory responsive neurons and other neurons for V1 and the anteromedial area. Each dot corresponds to a recording session. Interestingly, neurons that are auditory-responsive in V1 (left) responded significantly less to visual stimuli than other V1 neurons (p=0.0048, Wilcoxon signed rank test). This discrepancy is not observed in the medial area (right, p=0.77). Altogether, these results show that many layer 2/3 neurons of secondary or associative areas anteromedial to V1 are directly driven by auditory inputs even in darkness whereas very few V1 neurons are (sound-driven neurons in V1 are actually mostly GABAergic neurons, with a very significant population in layer 1, see Figure 4.5). This suggests that auditory information has some driving role in associative areas while it has only a modulatory role in V1.*

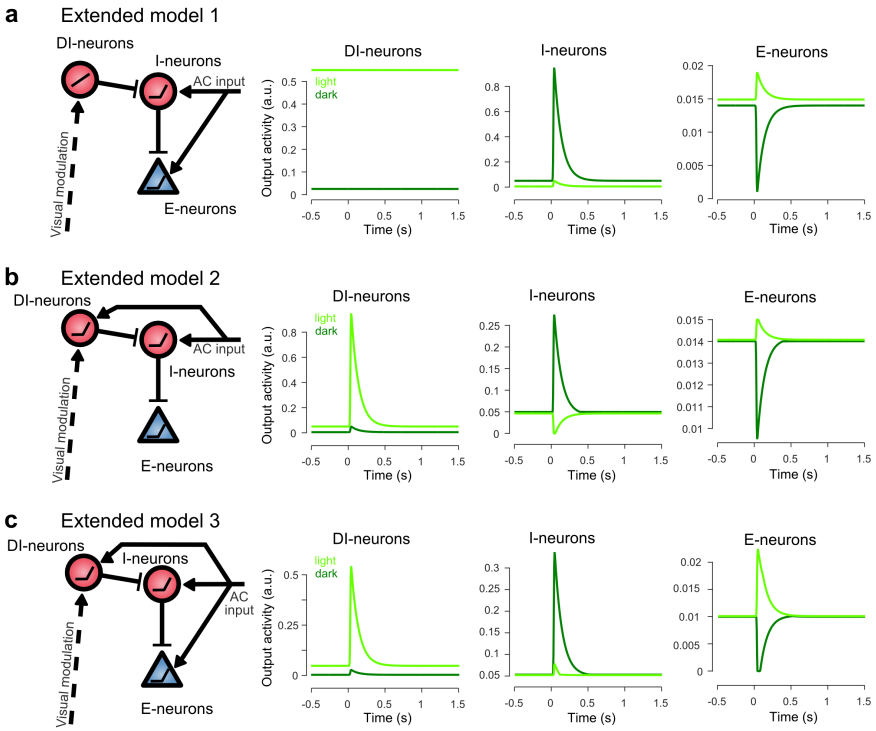

FIGURE 4.11: **Extended models of context-dependent responses to sounds in V1.**

In Figure 4.4, we have designed a minimal model of the context-dependent responses to sounds that we observe in layer 2/3 neurons of V1. This model includes an inhibitory population (I-neurons), driven by AC inputs, which is inhibited in the presence of visual input. We here show the results of simulations in which we have more explicitly modelled the inhibition by visual context.
*(**a**) A simple mechanism is to suppose that a population of desinhibitory interneurons (DI-neurons) linearly converts a tonic excitatory drive from visual context into inhibition of I-neurons. This model obviously recapitulates the output of the model shown in Figure 4.4. (**b**) Another possibility is that DI-neurons also receive AC inputs but are non-linear, such that unless they are driven by the visual context, they are subthreshold and do not inhibit I-neurons. In this case, the positive responses of pyramidal cells E-neurons to sounds in the light can be produced without AC inputs by disinhibition. Note however that this implies that I-neurons should be significantly inhibited by sounds in the light, and we found no population of neurons that is excited in the dark and inhibited in the light (Figure 4.5). Also connectivity data show that AC projection can drive layer 2/3 pyramidal cells in V1 (Ibrahim, 2016 #1909)[252]. However, numerous inhibitory neurons in layer 1 and 2/3 have the response signature of DI-neurons (Figure 4.5). (**c**) A more realistic design of the model is to implement AC inputs to all three neuronal populations. In this case the positive response of E-neurons results from both the direct AC drive and from the cancelation of I-neurons inhibition by the visual context-modulated DI-neurons. This model generates response signatures that are fully compatible with those observed in Figure 4.5.*

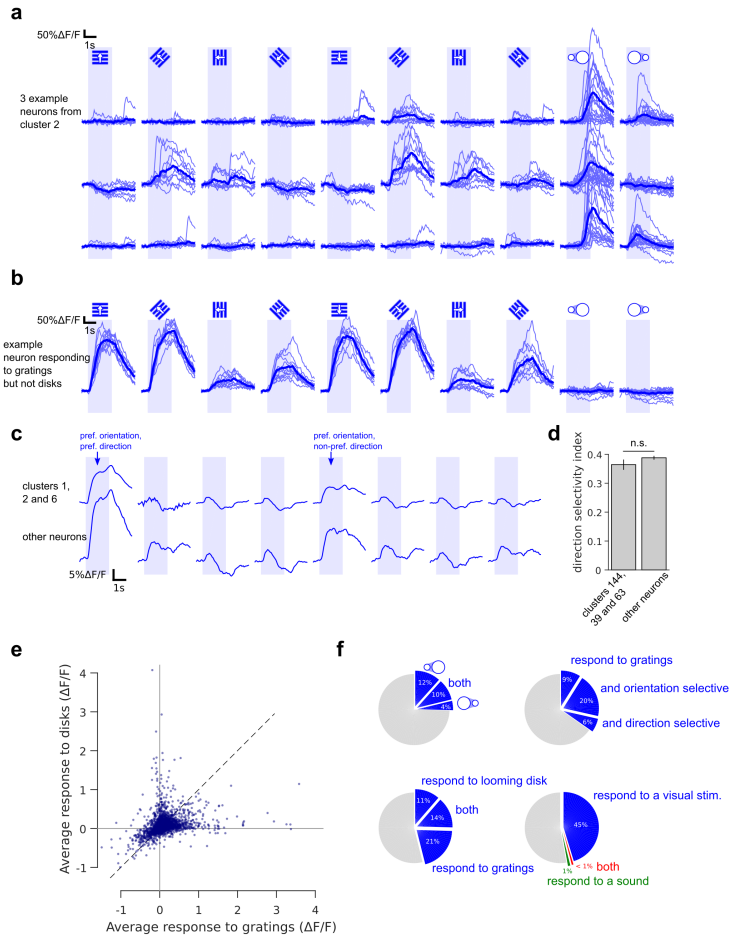

FIGURE 4.12: **Sharp tuning to looming or receding disks is not straightforwardly explained by tuning to gratings.**

*(**a**) Examples of visual responses in V1 (raw GCAMP6s signals), for gratings of different orientations and looming or receding disks as indicated above the traces. Individual traces (thin lines) and their average (thick line) are shown. Neurons that respond to disk do not systematically respond to gratings, although grating stimuli cover the entire visual field spanned by looming and receding disks. (**b**) Example neuron responding to gratings but not to disks. (**c**) Mean $\Delta F/F$ responses aligned to the preferred direction for the neurons contained in clusters 1, 2 and 6 of Figure 4.6c (which are sharply selective for looming or receding disks) (top), and for other clusters, less selective to disk direction (bottom). As quantified in **d**, the direction selectivity measured for gratings is identical for the two groups of clusters (sharply against weakly tuned to the motion direction of disk), indicating that direction tuning to disk cannot be deduced from the direction tuning properties measured with gratings. (**e**). Preference to disks or to gratings estimated by the average activity after stimulus onset for all V1 layer2/3 neurons. Equal preference is represented as the dash line. The distribution suggests that different neurons are strongly driven by disks and gratings. (**f**) Distribution of V1 neurons preferences for disks (top left), gratings (top right), both (bottom left), or multimodal stimuli (bottom right).*

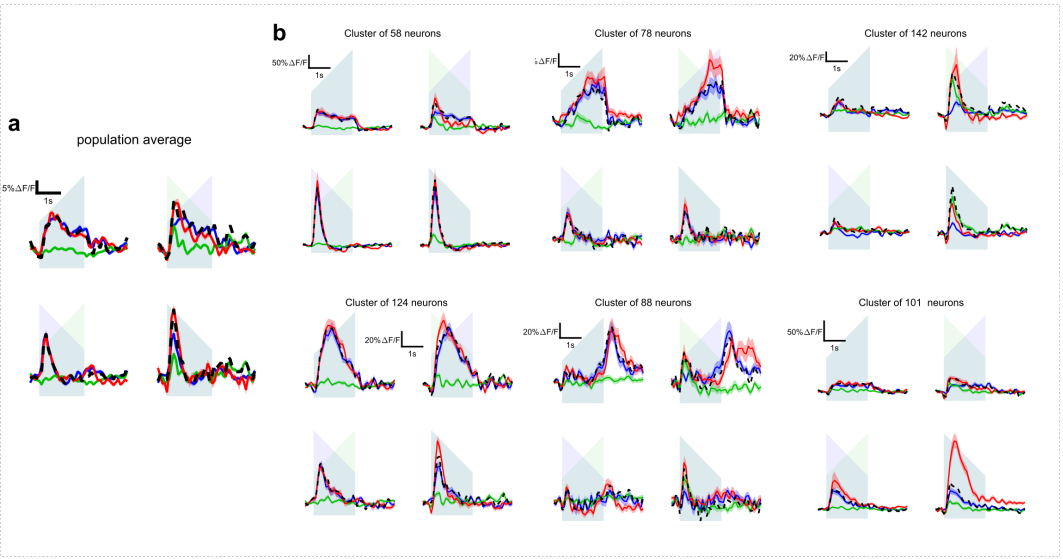

FIGURE 4.13: **Bimodal responses in the area anteromedial to V1.**

*(**a**) Population responses (3076 neurons, 9 sessions, 4 mice) to the 4 unimodal stimuli and 4 bimodal stimuli in the area anteromedial to V1 (see Supplementary Figure 4.10). Down-ramping sounds alone elicit strong onset responses which at the population level sum linearly with visual responses. (**b**) Result of the clustering of uni- and multimodal responses in the anteromedial area. Six clusters were identified, five of which integrate mostly linearly auditory and visual responses and only one (bottom right) shows non-linear auditory-visual responses. Interestingly also, one cluster shows auditory responses that are stronger than visual responses (top right), corroborating the idea that auditory input to the more associative region anteromedial to V1 have also a driving role and not only a modulatory role as we observe in V1.*

Discussion & Conclusion

This PhD thesis focuses on the contributions of non-linearities to sound perception and how non-linearities influence cortical representations of sounds. To address these two phenomena, this work takes benefits from large population recordings with calcium imaging in the awake mouse auditory cortex, and machine learning to investigate the neural representations. The first study shows that non-linear operations transform the sound intensity in the auditory system and that these non-linear transformations modify the cortical representations. Then the second study shows how cortical representations, and in particular their intensity, have an influence on behavior. The stronger is the cortical recruitment for one sound, the faster this sound is learned in a behavioral task. The third study aims for the description of several non-linear transformations. In particular it focuses on how cortical populations that encode these transformations interact. Finally, the last study explores how other sensory modalities, like sight, can influence the cortical representations.

This discussion adds context, and presents the limits and perspectives on the main results of the PhD thesis. It is composed of multiple independent sessions because these results are obtained from different questions with diverse approaches, therefore a general discussion would make little sense. The first part of the discussion concerns behavioral neurosciences and how to interpret the results in the behaving animals. Then it develops the limits and the advantages of the methods used in the PhD thesis. Finally it discusses about the general architecture of the auditory system.

Behavioral relevance of the cortical asymmetry between up- and down-ramps

The first study presented here aims to understand the perceptual differences between a sound that increases in intensity and a sound that decreases in intensity (respectively up- and down- ramps). Ponsot et al[135] studied the difference between these sounds in humans and showed that up-ramps were perceived louder than down-ramps. This auditory illusion is behaviorally relevant because up-ramps can be associated with threats that are approaching such as predators[133], and therefore there is a chance that this illusion works for others mammals. This first study recorded the activity of a large population of cortical neurons in response to these two sounds to understand how they are encoded in the mouse auditory system. Interestingly up-ramps produce more activity than down-ramps, and therefor they are learn faster in a behavioral Go/NoGo task (also see Chapter 2). Thus this asymmetry in the intensity of the cortical activity is probably relevant for the survival of mice to avoid predator such as cat or birds. In the third study, many populations of neurons responds more to increasing frequency modulated sounds than their decreasing equivalent. The behavioral reason for this asymmetry can also be linked to the detection of approaching objects because of the Doppler effect.

Hypothesis on the learning speed in the context of reversal learning

The second study of this PhD work generalizes an idea already tested with up- and down- ramps in the first study, that is: That are the behavioral consequences of the cortical recruitment for a sound? First it shows that there is a correlation between cortical recruitment and learning speed in a Go/NoGo task in naive animal. This finding shows first evidence for an hypothesis that is often implicit in neuroscience when the neuronal activity is link to behavior, and it provide a more general context for the cortical asymmetry for the up- and down-ramps (see Chapter1). The study uses naive mice for behavioral task, because expert mice could have enhance cortical representations for certain sounds and be very fast learners for others sounds due to different mechanisms. For example in the context of reversal learning, the learning rate for the task is fast but the intensity of cortical representations are unchanged. One interpretation of the fast learning speed in that particular case can be intuited from the model developed by Bathellier and al[119] also used in this article. If the synaptic weights change with a multiplicative rule, then when reverse learning occurs, the weights are strong and therefore the learning speed is fast.

Multisensory representation in the auditory cortex in naive and expert mice

The fourth study aims at understanding how information from multiple sensory modalities interacts in the brain to generate complete perceptual objects; in particular it focuses on the auditory and visual modalities. It is important to note here that in naïve mice these two sensory modalities interact asymmetrically, with sounds influencing cortical representations of images in the visual cortex, but images not modifying the representation of sounds in the superficial layers of the mouse auditory cortex (for infragranular layers see Morrill and Hasenstaub [285]). This statement might be false for mice trained on multisensory tasks. Even if Song et al.[247] shows that for a visual and auditory task, the decision was done by an associative area, it is possible that training produces a modification of the brain connectivity that would induce visual responses in the auditory cortex, especially considering the large plasticity of the brain in case of sensory deprivation.

Behavioral relevance of multisensory connections

The fourth study of this PhD also shows that the auditory cortex influences visual representations especially when the onset of the sounds are loud, where as the first study shows a larger activity in auditory cortex for quiet onset compare to loud onset. The reason is that direct connections from the auditory cortex to the visual cortex are biased towards neurons that encode loud onsets. One possible behavioral reason for this selectivity could be that loud sounds provide important information on how to redirect the visual attention, and therefore the sound information for loud sounds should be transmitted to the visual system.

Optogenetic stimulation does not reproduce simple sound perception

The second study not only shows a correlation between cortical recruitment and learning speed, it shows the causal link between the two. Indeed taking advantage of the optogenetic method to stimulate the activity of cortical populations, it shows that a light pattern of activation for a large population was learn faster than for a

small population. It is the first time that a study shows a discriminative task with optogenetic in the auditory cortex. The optogenetic patterns of activity, even if they target a tonotopical regions, are probably not perceived as simple sounds because of the salt and pepper organization of the tonotopy at cellular level (see Chapter 3). Indeed, if mice learn a discrimination task with optogenetic patterns on a tonotopic area such as it mimics a pure tones discrimination, a real pure tone does not replace the optogenetic pattern (data not show). With the development of hologram patterns for optogenetic it might be possible to identify functional populations with calcium imaging and then to modify their activity with optogenetic in a very precise manner in a few years.

Advantages and limits of the exploration of auditory system nonlinearities through a single, large scale dataset

The third study of this PhD thesis aims for the detail description of different types of non-linear features found in the auditory system. To do so, a very large population (59590 neurons) of cortical neurons was recorded with calcium imaging in response to 148 sounds. It is one of the biggest dataset available for the auditory cortex with cellular resolution. From this dataset, the study describes 7 features encoded in sub-populations that are the direct consequences of non-linear processing, such as the asymmetry between up- and down-ramps (see Chapter1). To this extent, this study can be thought of the generalization of the first study concerning the mechanisms responsible for non-linearities. All of the features were already identified previously in independent studies that used different sounds and techniques, and that are thus hard to compare. The study presented here uses a common method to describe and quantify these non-linear features. The drawback of this approach is that the problems from calcium imaging such as time resolution and the limited sampling to superficial layers (detailed below), as well as the problems of sounds sampling, cannot be corrected easily by another method. However this approach provides a dataset with comparable methods to study multiple non-linearities.

Limits and advantage of calcium imaging

One explanation as to why sound representations were not disturbed by visual information in the fourth study despite the large series of evidence that such interactions exist in other species[286, 287, 288] is that the methods were not adapted to distinguish subtle changes in cortical representations. Indeed, calcium imaging is limited by the temporal resolution (here 31Hz) and by the depth of recording. Even if Iurilli et al.[251] found no particular interactions between these two sensory modalities in the naive mouse using electrophysiology, the recent work of Morrill and Hasenstaub describes visual activity in the mouse auditory cortex in Layer 6. Microscopy techniques cannot reach such layers in the mouse cortex and therefore the exploration of the cortical representations was limited to superficial layers. This undersampling problem is a general problem in this PhD work. Therefore, it is important to consider the results of these studies as valid for supragranular layers (and when it is precise, the granular layers). It is also important to take into account that fast and subtle modifications of the cortical representations cannot be accessed with the calcium imaging technique. However for slower or larger variations the calcium imaging captures the difference thanks to the kinetics of the calcium indicator GCaMP6. Even if calcium imaging has some drawbacks, it is a very efficient method to access large-scale neuronal representations in sensory neuroscience. It gives access to the

precise activity and location of more than 1500 neurons at the same time, which is very convenient to study the brain at the population level.

Limits of the C57Bl6 strain

The experiments on the auditory system described in the work have all been performed on C57Bl6 mice. However, this strain is known for hearing loss after a few months, starting from very high frequencies and propagating to low frequencies with time. In order to avoid this problem, all the experiments were done with relatively young mice of 6 to 14 weeks old. Mice younger than 6 weeks old were too young and the skull was not fully grown which makes difficult the implantation of a chronic cranial window. However, some high frequency sounds were triggering only very small responses in the auditory cortex (above 30kHz), in a frequency band still in the audible spectrum for mice. This could be a problem due to C57Bl6, however it is interesting to note that some natural sounds with the main frequency above 30kHz still trigger activity in the auditory cortex (data not show) leading to the conclusion that it is not a hearing loss problem, but a characteristic of the auditory system to be more sensitive to high frequency natural sounds over pure tones. The choice of the strain has been done because it allows multiple genetics modifications and tools. This PhD work took full advantage of this strain in the second study with optogenetics and in the last study with labeling of different neuronal populations with genetic targetting.

Arguments for the parallel organization of the auditory system

The first study extracts the most relevant response patterns for up- and down-sounds to provide a more mechanistic description of the asymmetry. The two sounds are perceived differently, because they are encoded by a temporal pattern of distinct populations, "quiet on" and then "loud off" for the up-ramps and "loud on" and then "quiet off" for the down-ramps. None of these features can be extracted by the brain with linear processing and it needs at least 2 non-linear operations to encode one of them. Therefor non-linear transformations of sound intensity are responsible for the asymmetry in cortical activity between up- and down-ramps. The encoding of these two sounds provides a simple and understandable example of the role of non-linear processing because it is only based on the sound level and it has behavioral consequences. This study used two frequency contents (whitenoise and 8kHz) but the Chapter 3 describes the same temporal activity patterns for more frequency bands therefor this auditory illusion seems independent of the sound frequency content. Interestingly, the first study shows very little responses to "quiet offset", and this feature is also not extracted even on a larger set of sounds such as on Chapter 3. It seems that the auditory system works with parallel channels that extract the same features from the different frequency bands separated at the level of the cochlea.

This first study also proposes a multilayer model to explain the asymmetry between up- and down-ramps. The model is data-driven because it reproduces the activity patterns recorded in the auditory cortex. This model with linear transformations, interspersed with 2 non-linear operations can explain the cortical asymmetry for these two sounds.It is interesting because its architecture is composed of parallel channels for each features, that combine to form the different cortical representations. It is interesting to note however, that the operations between the channel "loud" and "quiet" are not exactly similar in the brain because it seems that the quiet channel has no offset response. On the model this problem is solved by setting low

weights on the last linear filters giving small importance to the "quiet offset" unit. This model could be extended to perform similar operations in multiple frequency channels, and thus it could generalize to a large diversity of sounds.

Lateral inhibition is required in the auditory system

From the interactions of the populations that encode the non-linearities, the third study draws conclusions on the general organization of the auditory system. Interestingly, some conclusions are in line with the model presented on Chapter 1 if it is extended to the frequency domain as discussed in the previous paragraph. For example, it suggests that "quiet" tuning is responsible for the asymmetry for up- versus down-ramps. It also suggests that "quiet" tuning is responsible for the tuning to sine modulated sounds, another feature that can be explained by such a model. However this model does not explain why there is a tuning to rapid frequency modulation or why some chords trigger non-additive responses. Indeed for the model to reproduce these two features it probably requires lateral inhibition. However the third study of this work does not give information on which layer of the model the lateral inhibition should be implemented. To have this information, additional recordings are require in sub cortical areas to understand where the features are formed in the auditory system (see perspectives below).

Hierarchical organization of the mouse auditory system

Stimuli are encoded in the brain in hierarchical systems such as the auditory system[37] with different levels of abstraction and complexity. High-level areas such as the auditory cortex have been hypothesized to encode perceptual objects[289], but this particularity means that the representations in the auditory cortex should be invariant in time and in frequency. For example, if a neuron of the auditory cortex encodes the word "phone", this neuron should be activated every time and for every pitch of pronunciation. The first and the third studies clearly identify populations of neurons that are tuned to particular frequencies, indicating that the auditory cortex still plays a role in the processing of the sound information, before the identification of the auditory object. However, in the large dataset from the third study, some cortical populations are encoding the frequency modulated sounds independently of the direction of modulation or the frequency range (data not show). Therefore, the mouse auditory cortex should have an intermediate position between processing signals and identifying auditory objects in the auditory scene.

Feedback in the auditory system

The information about the auditory objects and the auditory scene extracted at the auditory cortex are essential to focus attention on particular sounds. For example, if you are listening to an opera you can mentally isolate the first violin or the trumpet player, this process requires feedback to focus the auditory system into processing these particular sounds. Feedback in the auditory system goes from the auditory cortex and associative areas up to the cochlea[41]. It is probably playing a large role in sound processing especially when it comes to extracting non-linear features. In this PhD work, feedback has mostly been ignored because it is really difficult to take into account, especially with calcium imaging. However, the context of the sound was minimized in our protocol so that feedback connections did not play a major role in sound processing.

Perspective of this work

The third study in this PhD work aims at understanding the general architecture of the auditory system by looking at how non-linear features are encoded at the level of the auditory cortex. To design a computational model of the full auditory system is an enormous work and to address this problem by looking at the auditory cortex to draw general conclusions is a good starting point. However, it is difficult with this technique to get detail about the mechanism by which the non-linear features are extracted. It would be very interesting to study one by one the different nuclei that compose the auditory system to have a better understanding of how each nucleus encodes the information of the previous nucleus and how it extracts new features from the previous ones. This procedure reduces the general problem of understanding the complete auditory system into understanding each transfer function one by one. This approach raises two main concerns: (i) how to reach deep areas like the thalamus or the brainstem and (2) what to do with feedback projections? If deep areas do not have direct light access, it is possible to record their activity by recording their axonal projections in superficial areas[290, 291], and to use the activity of the axon terminals as a proxy to get the activity of the deep area. This particular technique can be used for the MGB of the thalamus that projects to the cortex and for the superior olive that projects to the inferior colliculus. Unfortunately, this approach limits the recordings of one nucleus to its projections to one particular area, and as the nuclei of the auditory system are interconnected, it can reduce a lot the diversity of responses for this nucleus. It is also necessary to block all the feedback from high-level areas so that they do not propagate high-level representations to lower areas. Unfortunately, blocking a complete area without lesions is difficult to perform, especially if calcium imaging is used to record neuronal activity because it makes difficult the use of optogenetic. Some techniques for chemical silencing can be good candidates such as muscimol or DREADs, but it is hard to precisely quantify and tune the amount of inhibition. Finally, this approach also record non-independent axon terminals because they come from the same neuron. This can be a problem for later data processing stage because the statistical tests are more complicated to perform on non independent individuals. However, the hierarchical clustering method is very efficient in this case to group together axons from the same neuron because they have the same activity patterns.

Last words

The full description of the non-linearities in the auditory system will require an enormous amount of work, and this is one of the goals of the Bathellier lab, but the relevance of this task is large considering the potential medical progress it represents for hearing loss patients that are mostly limited to linear hearing aids. I wish them good luck for this amazing endeavor.

Bibliography

[1] Nicholas A Lesica. "Why Do Hearing Aids Fail to Restore Normal Auditory Perception?" In: *Trends in neurosciences* (2018).

[2] Henry E Heffner and Rickye S Heffner. "Hearing ranges of laboratory animals". In: *Journal of the American Association for Laboratory Animal Science* 46.1 (2007), pp. 20–22.

[3] Robert Galambos and Hallowell Davis. "The response of single auditory-nerve fibers to acoustic stimulation". In: *Journal of neurophysiology* 6.1 (1943), pp. 39–57.

[4] Luis Robles and Mario A Ruggero. "Mechanics of the mammalian cochlea". In: *Physiological reviews* 81.3 (2001), pp. 1305–1352.

[5] Don H Johnson. "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones". In: *The Journal of the Acoustical Society of America* 68.4 (1980), pp. 1115–1122.

[6] William E Brownell et al. "Evoked mechanical responses of isolated cochlear outer hair cells". In: *Science* 227.4683 (1985), pp. 194–196.

[7] Peter Dallos. "The active cochlea". In: *Journal of Neuroscience* 12.12 (1992), pp. 4575–4585.

[8] Peter Dallos et al. "Prestin-based outer hair cell motility is necessary for mammalian cochlear amplification". In: *Neuron* 58.3 (2008), pp. 333–339.

[9] C Daniel Geisler et al. "Saturation of outer hair cell receptor currents causes two-tone suppression". In: *Hearing research* 44.2-3 (1990), pp. 241–256.

[10] Murray B Sachs and Nelson YS Kiang. "Two-tone inhibition in auditory-nerve fibers". In: *The Journal of the Acoustical Society of America* 43.5 (1968), pp. 1120–1128.

[11] PM Sellick and IJ Russell. "Two-tone suppression in cochlear hair cells". In: *Hearing research* 1.3 (1979), pp. 227–236.

[12] Simon Haykin and Zhe Chen. "The cocktail party problem". In: *Neural computation* 17.9 (2005), pp. 1875–1902.

[13] Josh H McDermott. "The cocktail party problem". In: *Current Biology* 19.22 (2009), R1024–R1027.

[14] Richard R Fay. *The mammalian auditory pathway: neurophysiology*. Vol. 2. Springer Science & Business Media, 2013.

[15] Richard Lyon. "A computational model of binaural localization and separation". In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'83.* Vol. 8. IEEE. 1983, pp. 1148–1151.

[16] Robert A Butler. "The bandwidth effect on monaural and binaural localization". In: *Hearing research* 21.1 (1986), pp. 67–73.

[17] AR Palmer and IJ Russell. "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells". In: *Hearing research* 24.1 (1986), pp. 1–15.

[18] Jerzy E Rose et al. "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey." In: *Journal of neurophysiology* 30.4 (1967), pp. 769–793.

[19] Thomas Lu, Li Liang, and Xiaoqin Wang. "Temporal and rate representations of time-varying signals in the auditory cortex of awake primates". In: *Nature neuroscience* 4.11 (2001), p. 1131.

[20] Robert D Frisina, Robert L Smith, and Steven C Chamberlain. "Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement". In: *Hearing research* 44.2-3 (1990), pp. 99–122.

[21] Jan Schnupp, Israel Nelken, and Andrew King. *Auditory neuroscience: Making sense of sound*. MIT press, 2011.

[22] EF Evans. "The frequency response and other properties of single fibres in the guinea-pig cochlear nerve". In: *The Journal of physiology* 226.1 (1972), pp. 263–287.

[23] Christina Enroth-Cugell and John G Robson. "The contrast sensitivity of retinal ganglion cells of the cat". In: *The Journal of physiology* 187.3 (1966), pp. 517–552.

[24] Stephen W Kuffler. "Discharge patterns and functional organization of mammalian retina". In: *Journal of neurophysiology* 16.1 (1953), pp. 37–68.

[25] Guy Wallis and Edmund T Rolls. "Invariant face and object recognition in the visual system". In: *Progress in neurobiology* 51.2 (1997), pp. 167–194.

[26] R Quian Quiroga et al. "Invariant visual representation by single neurons in the human brain". In: *Nature* 435.7045 (2005), p. 1102.

[27] Rodrigo Quian Quiroga. "Concept cells: the building blocks of declarative memory functions". In: *Nature Reviews Neuroscience* 13.8 (2012), p. 587.

[28] Rodrigo Quian Quiroga et al. "Explicit encoding of multimodal percepts by single neurons in the human brain". In: *Current Biology* 19.15 (2009), pp. 1308–1313.

[29] Daniel LK Yamins and James J DiCarlo. "Using goal-driven deep learning models to understand sensory cortex". In: *Nature neuroscience* 19.3 (2016), p. 356.

[30] David H Hubel and Torsten N Wiesel. “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. In: *The Journal of physiology* 160.1 (1962), pp. 106–154.

[31] Josef P Rauschecker. “Cortical processing of complex sounds”. In: *Current opinion in neurobiology* 8.4 (1998), pp. 516–521.

[32] Craig A Atencio, Tatyana O Sharpee, and Christoph E Schreiner. “Receptive field dimensionality increases from the auditory midbrain to cortex”. In: *Journal of neurophysiology* 107.10 (2012), pp. 2594–2603.

[33] Srivatsun Sadagopan and Xiaoqin Wang. “Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex”. In: *Journal of Neuroscience* 29.36 (2009), pp. 11192–11202.

[34] CM Wessinger et al. “Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging”. In: *Journal of cognitive neuroscience* 13.1 (2001), pp. 1–7.

[35] Gal Chechik et al. “Reduction of information redundancy in the ascending auditory pathway”. In: *Neuron* 51.3 (2006), pp. 359–368.

[36] Josef P Rauschecker, Biao Tian, and Marc Hauser. “Processing of complex sounds in the macaque nonprimary auditory cortex”. In: *Science* 268.5207 (1995), pp. 111–114.

[37] Alexander JE Kell et al. “A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy”. In: *Neuron* 98.3 (2018), pp. 630–644.

[38] Israel Nelken et al. “Primary auditory cortex of cats: feature detection or something else?” In: *Biological cybernetics* 89.5 (2003), pp. 397–406.

[39] Elia Formisano et al. “" Who" is saying" what"? Brain-based decoding of human voice and speech”. In: *Science* 322.5903 (2008), pp. 970–973.

[40] Noël Staeren et al. “Sound categories are represented as distributed patterns in the human auditory cortex”. In: *Current Biology* 19.6 (2009), pp. 498–502.

[41] James O Pickles. “Auditory pathways: anatomy and physiology”. In: *Handbook of clinical neurology*. Vol. 129. Elsevier, 2015, pp. 3–25.

[42] Nobuo Suga. “Classification of inferior collicular neurones of bats in terms of responses to pure tones, frequency-modulated sounds and noise bursts”. In: *The Journal of physiology* 200.2 (1969), pp. 555–574.

[43] BIAO Tian and Josef P Rauschecker. “Processing of frequency-modulated sounds in the cat’s anterior auditory field”. In: *Journal of Neurophysiology* 71.5 (1994), pp. 1959–1975.

[44] Thomas Deneux et al. “Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities”. In: *Nature communications* 7 (2016), p. 12682.

[45] J Wang et al. “Response characteristics of primary auditory cortex neurons underlying perceptual asymmetry of ramped and damped sounds”. In: *Neuroscience* 256 (2014), pp. 309–321.

[46] JJ Eggermont, PIM Johannesma, and AMHJ Aertsen. “Reverse-correlation methods in auditory research”. In: *Quarterly reviews of biophysics* 16.3 (1983), pp. 341–414.

[47] E De Boer and HR De Jongh. “On cochlear encoding: Potentialities and limitations of the reverse-correlation technique”. In: *The Journal of the Acoustical Society of America* 63.1 (1978), pp. 115–135.

[48] AMHJ Aertsen and PIM Johannesma. “The spectro-temporal receptive field”. In: *Biological cybernetics* 42.2 (1981), pp. 133–143.

[49] Jonathan Laudanski, Jean-Marc Edeline, and Chloé Huetz. “Differences between spectro-temporal receptive fields derived from artificial and natural stimuli in the auditory cortex”. In: *PloS one* 7.11 (2012), e50539.

[50] David J Klein et al. “Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design”. In: *Journal of computational neuroscience* 9.1 (2000), pp. 85–111.

[51] Nina Kowalski, Didier A Depireux, and Shihab A Shamma. “Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra”. In: *Journal of neurophysiology* 76.5 (1996), pp. 3503–3523.

[52] Jonathan Fritz et al. “Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex”. In: *Nature neuroscience* 6.11 (2003), p. 1216.

[53] Lee M Miller et al. “Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex”. In: *Journal of neurophysiology* 87.1 (2002), pp. 516–527.

[54] R Christopher deCharms, David T Blake, and Michael M Merzenich. “Optimizing sound features for cortical neurons”. In: *science* 280.5368 (1998), pp. 1439–1444.

[55] Jennifer F Linden et al. “Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex”. In: *Journal of neurophysiology* 90.4 (2003), pp. 2660–2675.

[56] Peter J Kim and Eric D Young. “Comparative analysis of spectro-temporal receptive fields, reverse correlation functions, and frequency tuning curves of auditory-nerve fibers”. In: *The Journal of the Acoustical Society of America* 95.1 (1994), pp. 410–422.

[57] Yael Bitterman et al. “Ultra-fine frequency tuning revealed in single neurons of human auditory cortex”. In: *Nature* 451.7175 (2008), p. 197.

[58] Christian K Machens, Michael S Wehr, and Anthony M Zador. “Linearity of cortical receptive fields measured with natural sounds”. In: *Journal of Neuroscience* 24.5 (2004), pp. 1089–1100.

[59] Yaron Rotman, Omer Bar-Yosef, and Israel Nelken. "Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex". In: *Hearing research* 152.1-2 (2001), pp. 110–127.

[60] Jos J Eggermont. "Context dependence of spectro-temporal receptive fields with implications for neural coding". In: *Hearing research* 271.1-2 (2011), pp. 123–132.

[61] G Björn Christianson, Maneesh Sahani, and Jennifer F Linden. "The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields". In: *Journal of Neuroscience* 28.2 (2008), pp. 446–455.

[62] Stephen V David et al. "Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli". In: *Journal of Neuroscience* 29.11 (2009), pp. 3374–3386.

[63] Nicholas J Priebe and David Ferster. "Inhibition, spike threshold, and stimulus selectivity in primary visual cortex". In: *Neuron* 57.4 (2008), pp. 482–497.

[64] Craig A Atencio, Tatyana O Sharpee, and Christoph E Schreiner. "Cooperative nonlinearities in auditory cortical neurons". In: *Neuron* 58.6 (2008), pp. 956–966.

[65] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), p. 436.

[66] J. Deng et al. "ImageNet: A Large-Scale Hierarchical Image Database". In: *CVPR09*. 2009.

[67] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[68] Kaiming He et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1026–1034.

[69] Olga Russakovsky et al. "Imagenet large scale visual recognition challenge". In: *International Journal of Computer Vision* 115.3 (2015), pp. 211–252.

[70] Christian Szegedy et al. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[71] Matthew D Zeiler and Rob Fergus. "Visualizing and understanding convolutional networks". In: *European conference on computer vision*. Springer. 2014, pp. 818–833.

[72] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks". In: *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*. IEEE. 2013, pp. 6645–6649.

[73] Aäron Van Den Oord et al. "WaveNet: A generative model for raw audio." In: *SSW*. 2016, p. 125.

[74] Keunwoo Choi, George Fazekas, and Mark Sandler. "Automatic tagging using deep convolutional neural networks". In: *arXiv preprint arXiv:1606.00298* (2016).

[75] Sergey Shuvaev, Hamza Giaffar, and Alexei A Koulakov. "Representations of Sound in Deep Learning of Audio Features from Music". In: *arXiv preprint arXiv:1712.02898* (2017).

[76] David Silver et al. "Mastering the game of Go with deep neural networks and tree search". In: *nature* 529.7587 (2016), p. 484.

[77] Michael Eickenberg et al. "Seeing it all: Convolutional network layers map the function of the human visual system". In: *NeuroImage* 152 (2017), pp. 184–194.

[78] Umut Güçlü and Marcel AJ van Gerven. "Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream". In: *Journal of Neuroscience* 35.27 (2015), pp. 10005–10014.

[79] Daniel LK Yamins et al. "Performance-optimized hierarchical models predict neural responses in higher visual cortex". In: *Proceedings of the National Academy of Sciences* 111.23 (2014), pp. 8619–8624.

[80] Pieter R Roelfsema and Anthony Holtmaat. "Control of synaptic plasticity in deep cortical networks". In: *Nature Reviews Neuroscience* 19.3 (2018), p. 166.

[81] Warren S McCulloch and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5.4 (1943), pp. 115–133.

[82] Ning-long Xu et al. "Nonlinear dendritic integration of sensory and motor input during an active sensing task". In: *Nature* 492.7428 (2012), p. 247.

[83] Kishore Kuchibhotla and Brice Bathellier. "Neural encoding of sensory and behavioral complexity in the auditory cortex". In: *Current opinion in neurobiology* 52 (2018), pp. 65–71.

[84] Torsten Baldeweg. "Repetition effects to sounds: evidence for predictive coding in the auditory system". In: *Trends in cognitive sciences* 10.3 (2006), pp. 93–94.

[85] Catherine Wacongne, Jean-Pierre Changeux, and Stanislas Dehaene. "A neuronal model of predictive coding accounting for the mismatch negativity". In: *Journal of Neuroscience* 32.11 (2012), pp. 3665–3678.

[86] Nachum Ulanovsky, Liora Las, and Israel Nelken. "Processing of low-probability sounds by cortical neurons". In: *Nature neuroscience* 6.4 (2003), p. 391.

[87] DO Kim, CE Molnar, and JW Matthews. “Cochlear mechanics: nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure”. In: *The Journal of the Acoustical Society of America* 67.5 (1980), pp. 1704–1721.

[88] David T Kemp. “Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea”. In: *Archives of oto-rhino-laryngology* 224.1-2 (1979), pp. 37–45.

[89] Graeme K Yates. “Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions”. In: *Hearing research* 50.1-2 (1990), pp. 145–162.

[90] Aurélie Vandenbeuch, Anne-Marie Pillias, and Annick Faurion. “Modulation of taste peripheral signal through interpapillar inhibition in hamsters”. In: *Neuroscience letters* 358.2 (2004), pp. 137–141.

[91] Rachel I Wilson and Zachary F Mainen. “Early events in olfactory processing”. In: *Annu. Rev. Neurosci.* 29 (2006), pp. 163–201.

[92] Christopher I Moore and Sacha B Nelson. “Spatio-temporal subthreshold receptive fields in the vibrissa representation of rat primary somatosensory cortex”. In: *Journal of Neurophysiology* 80.6 (1998), pp. 2882–2892.

[93] J Julius Zhu and Barry W Connors. “Intrinsic firing patterns and whisker-evoked synaptic responses of neurons in the rat barrel cortex”. In: *Journal of neurophysiology* 81.3 (1999), pp. 1171–1183.

[94] H KEFFER HARTLINE. “Inhibition of activity of visual receptors by illuminating nearby retinal areas in the Limulus eye”. In: *Studies on Excitation and Inhibition in the Retina: A Collection of Papers from the Laboratories of H. Keffer Hartline* (1974), p. 253.

[95] John M Crook, Zoltan F Kisvárday, and Ulf T Eysel. “Evidence for a contribution of lateral inhibition to orientation tuning and direction selectivity in cat visual cortex: reversible inactivation of functionally characterized sites combined with neuroanatomical tracing techniques”. In: *European Journal of Neuroscience* 10.6 (1998), pp. 2056–2075.

[96] UT Eysel, JM Crook, and HF Machemer. “GABA-induced remote inactivation reveals cross-orientation inhibition in the cat striate cortex”. In: *Experimental Brain Research* 80.3 (1990), pp. 626–630.

[97] Michael Brosch and Christoph E Schreiner. “Time course of forward masking tuning curves in cat primary auditory cortex”. In: *Journal of neurophysiology* 77.2 (1997), pp. 923–943.

[98] MB Calford and MN Semple. “Monaural inhibition in cat auditory cortex”. In: *Journal of Neurophysiology* 73.5 (1995), pp. 1876–1891.

[99] Matteo Carandini, David J Heeger, and Walter Senn. “A synaptic explanation of suppression in visual cortex”. In: *Journal of Neuroscience* 22.22 (2002), pp. 10053–10065.

[100] Jon Touryan, Brian Lau, and Yang Dan. "Isolation of relevant visual features from random stimuli for cortical complex cells". In: *Journal of Neuroscience* 22.24 (2002), pp. 10811–10818.

[101] EF Evans and IC Whitfield. "Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat". In: *The Journal of physiology* 171.3 (1964), pp. 476–493.

[102] WILLIAM S Rhode and PHILIP H Smith. "Encoding timing and intensity in the ventral cochlear nucleus of the cat". In: *Journal of neurophysiology* 56.2 (1986), pp. 261–286.

[103] William S Rhode and Steven Greenberg. "Physiology of the cochlear nuclei". In: *The mammalian auditory pathway: Neurophysiology*. Springer, 1992, pp. 94–152.

[104] Christine V Portfors and David J Perkel. "The role of ultrasonic vocalizations in mouse communication". In: *Current opinion in neurobiology* 28 (2014), pp. 115–120.

[105] Steven M Pomerantz, Antonio A Nunez, and N Jay Bean. "Female behavior is affected by male ultrasonic vocalizations in house mice". In: *Physiology & behavior* 31.1 (1983), pp. 91–96.

[106] Kenneth R Johnson et al. "A major gene affecting age-related hearing loss in C57BL/6J mice". In: *Hearing research* 114.1-2 (1997), pp. 83–92.

[107] Sebastian A Vasquez-Lopez et al. "Thalamic input to auditory cortex is locally heterogeneous but globally tonotopic". In: *eLife* 6 (Sept. 2017). ISSN: 2050-084X. DOI: 10.7554/elife.25141.

[108] Tsai-Wen Chen et al. "Ultrasensitive fluorescent proteins for imaging neuronal activity". In: *Nature* 499.7458 (2013), p. 295.

[109] Linda Madisen et al. "Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance". In: *Neuron* 85.5 (2015), pp. 942–958.

[110] Nicholas A Steinmetz et al. "Aberrant cortical activity in multiple GCaMP6-expressing transgenic mouse lines". In: *eNeuro* 4.5 (2017), ENEURO–0207.

[111] Jasper Akerboom et al. "Optimization of a GCaMP calcium indicator for neural activity imaging". In: *Journal of Neuroscience* 32.40 (2012), pp. 13819–13840.

[112] Martin Stetter et al. "Analysis of Calcium Imaging Signals from the Honeybee Brainby Nonlinear Models". In: *Neuroimage* 13.1 (2001), pp. 119–128.

[113] Vahid Rahmati et al. "Inferring neuronal dynamics from calcium imaging data using biophysical models and Bayesian inference". In: *PLoS computational biology* 12.2 (2016), e1004736.

[114] Eyal Seidemann et al. "Calcium imaging with genetically encoded indicators in behaving primates". In: *Elife* 5 (2016), e16178.

[115] Brice Bathellier, Lyubov Ushakova, and Simon Rumpel. "Discrete neocortical dynamics predict behavioral categorization of sounds". In: *Neuron* 76.2 (2012), pp. 435–449.

[116] Kelly E Radziwon et al. "Behaviorally measured audiograms and gap detection thresholds in CBA/CaJ mice". In: *Journal of Comparative Physiology A* 195.10 (2009), pp. 961–969.

[117] Cyril Herry et al. "Extinction of auditory fear conditioning requires MAPK/ERK activation in the basolateral amygdala". In: *European Journal of Neuroscience* 24.1 (2006), pp. 261–269.

[118] Shinsuke Ohshima et al. "Cortical depression in the mouse auditory cortex after sound discrimination learning". In: *Neuroscience research* 67.1 (2010), pp. 51–58.

[119] Brice Bathellier et al. "A multiplicative reinforcement learning model capturing learning dynamics and interindividual variability in mice". In: *Proceedings of the National Academy of Sciences* 110.49 (2013), pp. 19950–19955.

[120] Kishore V Kuchibhotla et al. "Parallel processing by cortical inhibition enables context-dependent behavior". In: *Nature neuroscience* 20.1 (2017), p. 62.

[121] Andreas L Schulz et al. "Selective increase of auditory cortico-striatal coherence during auditory-cued Go/NoGo discrimination learning". In: *Frontiers in behavioral neuroscience* 9 (2016), p. 368.

[122] Shogo Soma, Naofumi Suematsu, and Satoshi Shimegi. "Efficient training protocol for rapid learning of the two-alternative forced-choice visual stimulus detection task". In: *Physiological reports* 2.7 (2014).

[123] L Ian Schmitt et al. "Thalamic amplification of cortical connectivity sustains attentional control". In: *Nature* 545.7653 (2017), p. 219.

[124] Hermann Von Helmholtz. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Longmans, Green, 1912.

[125] James W Lewis et al. "Human brain regions involved in recognizing environmental sounds". In: *Cerebral cortex* 14.9 (2004), pp. 1008–1021.

[126] Michael K Mcbeath and John G Neuhoff. "The Doppler effect is not what you think it is: Dramatic pitch change due to dynamic intensity change". In: *Psychonomic bulletin & review* 9.2 (2002), pp. 306–313.

[127] Israel Nelken, Yaron Rotman, and Omer Bar Yosef. "Responses of auditory-cortex neurons to structural features of natural sounds". In: *Nature* 397.6715 (1999), p. 154.

[128] Frédéric E Theunissen and Julie E Elie. "Neural processing of natural sounds". In: *Nature Reviews Neuroscience* 15.6 (2014), p. 355.

[129] R Christopher deCharms, David T Blake, and Michael M Merzenich. "Optimizing sound features for cortical neurons". In: *science* 280.5368 (1998), pp. 1439–1444.

[130] Kenneth W Berger. "Some factors in the recognition of timbre". In: *The Journal of the Acoustical Society of America* 36.10 (1964), pp. 1888–1891.

[131] James E Cutting and Burton S Rosner. "Categories and boundaries in speech and music". In: *Perception & Psychophysics* 16.3 (1974), pp. 564–570.

[132] Sam Ferguson, Emery Schubert, and Roger T Dean. "Continuous subjective loudness responses to reversals and inversions of a sound recording of an orchestral excerpt". In: *Musicae Scientiae* 15.3 (2011), pp. 387–401.

[133] John G Neuhoff. "Perceptual bias for rising tones". In: *Nature* 395.6698 (1998), p. 123.

[134] G Christopher Stecker and Ervin R Hafter. "An effect of temporal asymmetry on loudness". In: *The Journal of the Acoustical Society of America* 107.6 (2000), pp. 3358–3368.

[135] Emmanuel Ponsot, Patrick Susini, and Sabine Meunier. "A robust asymmetry in loudness between rising-and falling-intensity tones". In: *Attention, Perception, & Psychophysics* 77.3 (2015), pp. 907–920.

[136] Patrick Susini, Stephen McAdams, and Bennett K Smith. "Loudness asymmetries for tones with increasing and decreasing levels using continuous and global ratings". In: *Acta acustica united with acustica* 93.4 (2007), pp. 623–631.

[137] Erich Seifritz et al. "Neural processing of auditory looming in the human brain". In: *Current Biology* 12.24 (2002), pp. 2147–2151.

[138] Deborah A Hall and David R Moore. "Auditory neuroscience: The salience of looming sounds". In: *Current Biology* 13.3 (2003), R91–R93.

[139] Dominik R Bach et al. "Rising sound intensity: an intrinsic warning cue activating the amygdala". In: *Cerebral Cortex* 18.1 (2007), pp. 145–150.

[140] Asif A Ghazanfar, John G Neuhoff, and Nikos K Logothetis. "Auditory looming perception in rhesus monkeys". In: *Proceedings of the National Academy of Sciences* 99.24 (2002), pp. 15755–15757.

[141] Joost X Maier and Asif A Ghazanfar. "Looming biases in monkey auditory cortex". In: *Journal of Neuroscience* 27.15 (2007), pp. 4093–4100.

[142] I Stiebler et al. "The auditory cortex of the house mouse: left-right differences, tonotopic organization and quantitative analysis of frequency representation". In: *Journal of Comparative Physiology A* 181.6 (1997), pp. 559–571.

[143] Emre Yaksi and Rainer W Friedrich. "Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca 2+ imaging". In: *Nature methods* 3.5 (2006), p. 377.

[144] Kirk N Olsen, Catherine J Stevens, and Julien Tardieu. "Loudness change in response to dynamic acoustic intensity." In: *Journal of Experimental Psychology: Human Perception and Performance* 36.6 (2010), p. 1631.

[145] Katherine I Nagel and Allison J Doupe. "Organizing principles of spectro-temporal encoding in the avian primary auditory area field L". In: *Neuron* 58.6 (2008), pp. 938–955.

[146] Michael Wehr and Anthony M Zador. "Synaptic mechanisms of forward suppression in rat auditory cortex". In: *Neuron* 47.3 (2005), pp. 437–445.

[147] Tomáš Hromádka, Michael R DeWeese, and Anthony M Zador. "Sparse representation of sounds in the unanesthetized auditory cortex". In: *PLoS biology* 6.1 (2008), e16.

[148] Nadja Schinkel-Bielefeld et al. "Inferring the role of inhibition in auditory processing of complex natural stimuli". In: *Journal of neurophysiology* 107.12 (2012), pp. 3296–3307.

[149] Markus Meister and Michael J Berry. "The neural code of the retina". In: *Neuron* 22.3 (1999), pp. 435–450.

[150] Matteo Carandini et al. "Do we know what the early visual system does?" In: *Journal of Neuroscience* 25.46 (2005), pp. 10577–10597.

[151] Nicole C Rust et al. "Spatiotemporal elements of macaque v1 receptive fields". In: *Neuron* 46.6 (2005), pp. 945–956.

[152] James M McFarland, Yuwei Cui, and Daniel A Butts. "Inferring nonlinear neuronal computation based on physiologically plausible inputs". In: *PLoS computational biology* 9.7 (2013), e1003143.

[153] Leon J Kamin. "Temporal and intensity characteristics of the conditioned stimulus". In: *Classical conditioning: A symposium*. 1965, pp. 118–147.

[154] A. R. Rescorla R. A. Wagner. *Classical Conditioning II: Current Research and Theory*. Black A.H. Prokasy W.F., 1972.

[155] Charles R Gallistel, Stephen Fairhurst, and Peter Balsam. "The learning curve: implications of a quantitative analysis". In: *Proceedings of the National Academy of Sciences* 101.36 (2004), pp. 13124–13131.

[156] Patrick Gill et al. "Sound representation methods for spectro-temporal receptive field estimation". In: *Journal of computational neuroscience* 21.1 (2006), p. 5.

[157] Neil C Rabinowitz et al. "Spectrotemporal contrast kernels for neurons in primary auditory cortex". In: *Journal of Neuroscience* 32.33 (2012), pp. 11271–11284.

[158] Lucy A Anderson and Jennifer F Linden. "Mind the gap: two dissociable mechanisms of temporal processing in the auditory system". In: *Journal of Neuroscience* 36.6 (2016), pp. 1977–1995.

[159] Jufang He. "OFF responses in the auditory thalamus of the guinea pig". In: *Journal of neurophysiology* 88.5 (2002), pp. 2377–2386.

[160] Michael London and Michael Häusser. "Dendritic computation". In: *Annu. Rev. Neurosci.* 28 (2005), pp. 503–532.

[161] Mark S Goldman. "Memory without feedback in a neural network". In: *Neuron* 61.4 (2009), pp. 621–634.

[162] Srdjan Ostojic. "Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons". In: *Nature neuroscience* 17.4 (2014), p. 594.

[163] Daniel B Polley et al. "Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex". In: *Proceedings of the National Academy of Sciences* 101.46 (2004), pp. 16351–16356.

[164] Ben Scholl, Xiang Gao, and Michael Wehr. "Nonoverlapping sets of synapses drive on responses and off responses in auditory cortex". In: *Neuron* 65.3 (2010), pp. 412–421.

[165] Xiang Gao and Michael Wehr. "A coding transformation for temporally structured sounds within auditory cortical neurons". In: *Neuron* 86.1 (2015), pp. 292–303.

[166] Gideon Rothschild, Israel Nelken, and Adi Mizrahi. "Functional organization and population dynamics in the mouse primary auditory cortex". In: *Nature neuroscience* 13.3 (2010), p. 353.

[167] Adrien Wohrer, Mark D Humphries, and Christian K Machens. "Population-wide distributions of neural activity during perceptual decision-making". In: *Progress in neurobiology* 103 (2013), pp. 156–193.

[168] György Buzsáki and Kenji Mizuseki. "The log-dynamic brain: how skewed distributions affect network operations". In: *Nature Reviews Neuroscience* 15.4 (2014), p. 264.

[169] L. J Kamin. *Classical conditioning*. Prokasy W.F., 1965.

[170] L. J Kamin. *Punishment and Aversive Behavior*. B.A. Campbell R.M. Church, 1969.

[171] A. G. Sutton R. S. Barto. *Reinforcement learning : an introduction*. MIT Press, 1998.

[172] Peter Dayan and Bernard W. Balleine. "Reward, motivation, and reinforcement learning". eng. In: *Neuron* 36.2 (Oct. 2002), pp. 285–298. ISSN: 0896-6273.

[173] R. Desimone and J. Duncan. "Neural mechanisms of selective visual attention". eng. In: *Annual Review of Neuroscience* 18 (1995), pp. 193–222. ISSN: 0147-006X. DOI: 10.1146/annurev.ne.18.030195.001205.

[174] L. Itti and C. Koch. "Computational modelling of visual attention". eng. In: *Nature Reviews. Neuroscience* 2.3 (Mar. 2001), pp. 194–203. ISSN: 1471-003X. DOI: 10.1038/35058500.

[175] Stefan Treue. "Visual attention: the where, what, how and why of saliency". eng. In: *Current Opinion in Neurobiology* 13.4 (Aug. 2003), pp. 428–432. ISSN: 0959-4388.

[176] Zhaoping Li. "A saliency map in primary visual cortex". eng. In: *Trends in Cognitive Sciences* 6.1 (Jan. 2002), pp. 9–16. ISSN: 1879-307X.

[177] L. F. Dayan P. Abbott. *Theoretical neuroscience : computational and mathematical modeling of neural systems.* MIT Press, 2001.

[178] S. J. Martin, P. D. Grimwood, and R. G. Morris. "Synaptic plasticity and memory: an evaluation of the hypothesis". eng. In: *Annual Review of Neuroscience* 23 (2000), pp. 649–711. ISSN: 0147-006X. DOI: 10.1146/annurev.neuro.23.1.649.

[179] Sophie Denève, Alireza Alemi, and Ralph Bourdoukan. "The Brain as an Efficient and Robust Adaptive Learner". eng. In: *Neuron* 94.5 (June 2017), pp. 969–977. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2017.05.016.

[180] Nicolas Frémaux and Wulfram Gerstner. "Neuromodulated Spike-Timing-Dependent Plasticity, and Theory of Three-Factor Learning Rules". eng. In: *Frontiers in Neural Circuits* 9 (2015), p. 85. ISSN: 1662-5110. DOI: 10.3389/fncir.2015.00085.

[181] E. L. Bienenstock, L. N. Cooper, and P. W. Munro. "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 2.1 (Jan. 1982), pp. 32–48. ISSN: 0270-6474.

[182] Rufin VanRullen. "Visual saliency and spike timing in the ventral visual pathway". eng. In: *Journal of Physiology, Paris* 97.2-3 (May 2003), pp. 365–377. ISSN: 0928-4257. DOI: 10.1016/j.jphysparis.2003.09.010.

[183] Pascal Fries. "Rhythms for Cognition: Communication through Coherence". eng. In: *Neuron* 88.1 (Oct. 2015), pp. 220–235. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2015.09.034.

[184] Thilo Womelsdorf and Pascal Fries. "The role of neuronal synchronization in selective attention". eng. In: *Current Opinion in Neurobiology* 17.2 (Apr. 2007), pp. 154–160. ISSN: 0959-4388. DOI: 10.1016/j.conb.2007.02.002.

[185] Shihab A. Shamma, Mounya Elhilali, and Christophe Micheyl. "Temporal coherence and attention in auditory scene analysis". eng. In: *Trends in Neurosciences* 34.3 (Mar. 2011), pp. 114–123. ISSN: 1878-108X. DOI: 10.1016/j.tins.2010.11.002.

[186] Stijn Cassenaer and Gilles Laurent. "Conditional modulation of spike-timing-dependent plasticity for olfactory learning". eng. In: *Nature* 482.7383 (Jan. 2012), pp. 47–52. ISSN: 1476-4687. DOI: 10.1038/nature10776.

[187] Robert Gütig and Haim Sompolinsky. "The tempotron: a neuron that learns spike timing-based decisions". eng. In: *Nature Neuroscience* 9.3 (Mar. 2006), pp. 420–428. ISSN: 1097-6256. DOI: 10.1038/nn1643.

[188] Timothy J. Buschman et al. "Synchronous oscillatory neural ensembles for rules in the prefrontal cortex". eng. In: *Neuron* 76.4 (Nov. 2012), pp. 838–846. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2012.09.029.

[189] Nancy J. Kopell et al. "Beyond the connectome: the dynome". eng. In: *Neuron* 83.6 (Sept. 2014), pp. 1319–1328. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2014.08.016.

[190] Shane Lee, Kamal Sen, and Nancy Kopell. "Cortical gamma rhythms modulate NMDAR-mediated spike timing dependent plasticity in a biophysical model". eng. In: *PLoS computational biology* 5.12 (Dec. 2009), e1000602. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000602.

[191] Sarah W Wong and Christoph E Schreiner. "Representation of CV-sounds in cat primary auditory cortex: intensity dependence". In: *Speech Communication*. The Nature of Speech Perception 41.1 (Aug. 2003), pp. 93–106. ISSN: 0167-6393. DOI: 10.1016/S0167-6393(02)00096-1. URL: http://www.sciencedirect.com/science/article/pii/S0167639302000961 (visited on 08/12/2018).

[192] Troy A. Hackett et al. "Linking topography to tonotopy in the mouse auditory thalamocortical circuit". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31.8 (Feb. 2011), pp. 2983–2995. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.5333-10.2011.

[193] Patrick O. Kanold, Israel Nelken, and Daniel B. Polley. "Local versus global scales of organization in auditory cortex". eng. In: *Trends in Neurosciences* 37.9 (Sept. 2014), pp. 502–510. ISSN: 1878-108X. DOI: 10.1016/j.tins.2014.06.003.

[194] Q. Y. Zheng, K. R. Johnson, and L. C. Erway. "Assessment of hearing in 80 inbred strains of mice by ABR threshold analyses". eng. In: *Hearing Research* 130.1-2 (Apr. 1999), pp. 94–107. ISSN: 0378-5955.

[195] Suzana Herculano-Houzel, Charles Watson, and George Paxinos. "Distribution of neurons in functional areas of the mouse cerebral cortex reveals quantitatively different cortical zones". eng. In: *Frontiers in Neuroanatomy* 7 (2013), p. 35. ISSN: 1662-5129. DOI: 10.3389/fnana.2013.00035.

[196] S. E. Brandon, E. H. Vogel, and A. R. Wagner. "A componential view of configural cues in generalization and discrimination in Pavlovian conditioning". eng. In: *Behavioural Brain Research* 110.1-2 (June 2000), pp. 67–72. ISSN: 0166-4328.

[197] Robert A Rescorla, Allan R Wagner, et al. "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement". In: *Classical conditioning II: Current research and theory* 2 (1972), pp. 64–99.

[198] Yonatan Loewenstein, Annerose Kuras, and Simon Rumpel. "Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 31.26 (June 2011), pp. 9481–9488. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.6130-10.2011.

[199] R. Gütig et al. "Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 23.9 (May 2003), pp. 3697–3714. ISSN: 1529-2401.

[200] Alexei A. Koulakov, Tomás Hromádka, and Anthony M. Zador. "Correlated connectivity and the distribution of firing rates in the neocortex". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 29.12 (Mar. 2009), pp. 3685–3694. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.4500-08.2009.

[201] Abigail Morrison, Ad Aertsen, and Markus Diesmann. "Spike-timing-dependent plasticity in balanced random networks". eng. In: *Neural Computation* 19.6 (June 2007), pp. 1437–1467. ISSN: 0899-7667. DOI: 10.1162/neco.2007.19.6.1437.

[202] Ashesh K. Dhawale et al. "Non-redundant odor coding by sister mitral cells revealed by light addressable glomeruli in the mouse". eng. In: *Nature Neuroscience* 13.11 (Nov. 2010), pp. 1404–1412. ISSN: 1546-1726. DOI: 10.1038/nn.2673.

[203] Kaja Ewa Moczulska et al. "Dynamics of dendritic spines in the mouse auditory cortex during memory formation and memory recall". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.45 (Nov. 2013), pp. 18315–18320. ISSN: 1091-6490. DOI: 10.1073/pnas.1312508110.

[204] Petr Znamenskiy and Anthony M. Zador. "Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination". eng. In: *Nature* 497.7450 (May 2013), pp. 482–485. ISSN: 1476-4687. DOI: 10.1038/nature12077.

[205] Hiroyuki K. Kato, Shea N. Gillet, and Jeffry S. Isaacson. "Flexible Sensory Representations in Auditory Cortex Driven by Behavioral Relevance". eng. In: *Neuron* 88.5 (Dec. 2015), pp. 1027–1039. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2015.10.024.

[206] Tyler L. Gimenez, Maja Lorenc, and Santiago Jaramillo. "Adaptive categorization of sound frequency does not require the auditory cortex in rats". eng. In: *Journal of Neurophysiology* 114.2 (Aug. 2015), pp. 1137–1145. ISSN: 1522-1598. DOI: 10.1152/jn.00124.2015.

[207] Shraddha Pai et al. "Minimal impairment in a rat model of duration discrimination following excitotoxic lesions of primary auditory and prefrontal cortices". eng. In: *Frontiers in Systems Neuroscience* 5 (2011), p. 74. ISSN: 1662-5137. DOI: 10.3389/fnsys.2011.00074.

[208] Natalia Rybalko et al. "Effect of auditory cortex lesions on the discrimination of frequency-modulated tones in rats". eng. In: *The European Journal of Neuroscience* 23.6 (Mar. 2006), pp. 1614–1622. ISSN: 0953-816X. DOI: 10.1111/j.1460-9568.2006.04688.x.

[209] Israel Nelken. "Processing of complex stimuli and natural scenes in the auditory cortex". eng. In: *Current Opinion in Neurobiology* 14.4 (Aug. 2004), pp. 474–480. ISSN: 0959-4388. DOI: 10.1016/j.conb.2004.06.005.

[210] Srivatsun Sadagopan and Xiaoqin Wang. "Level invariant representation of sounds by populations of neurons in primary auditory cortex". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 28.13 (Mar. 2008), pp. 3415–3426. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.2743-07.2008.

[211] Jonathan B. Fritz et al. "Auditory attention–focusing the searchlight on sound". eng. In: *Current Opinion in Neurobiology* 17.4 (Aug. 2007), pp. 437–455. ISSN: 0959-4388. DOI: 10.1016/j.conb.2007.07.011.

[212] Michael Goard and Yang Dan. "Basal forebrain activation enhances cortical coding of natural scenes". eng. In: *Nature Neuroscience* 12.11 (Nov. 2009), pp. 1444–1449. ISSN: 1546-1726. DOI: 10.1038/nn.2402.

[213] Siyu Zhang et al. "Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing". eng. In: *Science (New York, N.Y.)* 345.6197 (Aug. 2014), pp. 660–665. ISSN: 1095-9203. DOI: 10.1126/science.1254126.

[214] J. M. Pearce and G. Hall. "A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli". eng. In: *Psychological Review* 87.6 (Nov. 1980), pp. 532–552. ISSN: 0033-295X.

[215] N. J. Mackintosh. "A theory of attention: Variations in the associability of stimuli with reinforcement." en. In: *Psychological Review* 82.4 (1975), pp. 276–298. ISSN: 0033-295X. DOI: 10.1037/h0076778. URL: http://content.apa.org/journals/rev/82/4/276 (visited on 08/12/2018).

[216] Yael Niv et al. "Reinforcement learning in multidimensional environments relies on attention mechanisms". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 35.21 (May 2015), pp. 8145–8157. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.2978-14.2015.

[217] Daniel Huber et al. "Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice". eng. In: *Nature* 451.7174 (Jan. 2008), pp. 61–64. ISSN: 1476-4687. DOI: 10.1038/nature06445.

[218] Arthur R. Houweling and Michael Brecht. "Behavioural report of single neuron stimulation in somatosensory cortex". eng. In: *Nature* 451.7174 (Jan. 2008), pp. 65–68. ISSN: 1476-4687. DOI: 10.1038/nature06447.

[219] Aaron M. Kerlin et al. “Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex”. eng. In: *Neuron* 67.5 (Sept. 2010), pp. 858–871. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2010.08.002.

[220] Jeremiah Y. Cohen et al. “Neuron-type-specific signals for reward and punishment in the ventral tegmental area”. eng. In: *Nature* 482.7383 (Jan. 2012), pp. 85–88. ISSN: 1476-4687. DOI: 10.1038/nature10754.

[221] W. Schultz, P. Dayan, and P. R. Montague. “A neural substrate of prediction and reward”. eng. In: *Science (New York, N.Y.)* 275.5306 (Mar. 1997), pp. 1593–1599. ISSN: 0036-8075.

[222] Albert S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. en. Google-Books-ID: jI8muSpAC5AC. MIT Press, 1994. ISBN: 978-0-262-52195-6.

[223] Barbara G. Shinn-Cunningham and Virginia Best. “Selective Attention in Normal and Impaired Hearing”. en. In: *Trends in Amplification* 12.4 (Dec. 2008), pp. 283–299. ISSN: 1084-7138. DOI: 10.1177/1084713808325306. URL: https://doi.org/10.1177/1084713808325306 (visited on 08/14/2018).

[224] Jennifer K Bizley and Yale E Cohen. “The what, where and how of auditory-object perception”. In: *Nature Reviews Neuroscience* 14.10 (2013), p. 693.

[225] Paul V Watkins and Dennis L Barbour. “Specialized neuronal adaptation for preserving input sensitivity”. In: *Nature neuroscience* 11.11 (2008), p. 1259.

[226] Cal F. Rabang, Jeff Lin, and Guangying K. Wu. “Balance or imbalance: inhibitory circuits for direction selectivity in the auditory system”. en. In: *Cellular and Molecular Life Sciences* 72.10 (May 2015), pp. 1893–1906. ISSN: 1420-682X, 1420-9071. DOI: 10.1007/s00018-015-1841-2. URL: https://link.springer.com/article/10.1007/s00018-015-1841-2 (visited on 08/14/2018).

[227] Valery A. Kalatsky et al. “Fine functional organization of auditory cortex revealed by Fourier optical imaging”. en. In: *Proceedings of the National Academy of Sciences* 102.37 (Sept. 2005), pp. 13325–13330. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0505592102. URL: http://www.pnas.org/content/102/37/13325 (visited on 08/14/2018).

[228] Hiroyuki K. Kato, Samuel K. Asinof, and Jeffry S. Isaacson. “Network-Level Control of Frequency Tuning in Auditory Cortex”. eng. In: *Neuron* 95.2 (July 2017), 412–423.e4. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2017.06.019.

[229] M. L. Sutter and C. E. Schreiner. “Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex”. eng. In: *Journal of Neurophysiology* 65.5 (May 1991), pp. 1207–1226. ISSN: 0022-3077. DOI: 10.1152/jn.1991.65.5.1207.

[230] Zoltan M. Fuzessery, Khaleel A. Razak, and Anthony J. Williams. "Multiple mechanisms shape selectivity for FM sweep rate and direction in the pallid bat inferior colliculus and auditory cortex". en. In: *Journal of Comparative Physiology A* 197.5 (May 2011), pp. 615–623. ISSN: 0340-7594, 1432-1351. DOI: 10.1007/s00359-010-0554-0. URL: https://link.springer.com/article/10.1007/s00359-010-0554-0 (visited on 08/14/2018).

[231] Steffen R. Hage and Gunter Ehret. "Mapping responses to frequency sweeps and tones in the inferior colliculus of house mice". en. In: *European Journal of Neuroscience* 18.8 (Oct. 2003), pp. 2301–2312. ISSN: 0953-816X, 1460-9568. DOI: 10.1046/j.1460-9568.2003.02945.x. URL: http://doi.wiley.com/10.1046/j.1460-9568.2003.02945.x (visited on 08/14/2018).

[232] Richard I. Kuo and Guangying K. Wu. "The Generation of Direction Selectivity in the Auditory System". In: *Neuron* 73.5 (Mar. 2012), pp. 1016–1027. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2011.11.035. URL: http://www.sciencedirect.com/science/article/pii/S089662731200044X (visited on 08/14/2018).

[233] Chang-quan Ye et al. "Synaptic mechanisms of direction selectivity in primary auditory cortex". eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30.5 (Feb. 2010), pp. 1861–1868. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.3088-09.2010.

[234] Helen J Kennedy et al. "Fast adaptation of mechanoelectrical transducer channels in mammalian cochlear hair cells". In: *Nature neuroscience* 6.8 (2003), p. 832.

[235] Wei Guo et al. "Hearing the light: neural and perceptual encoding of optogenetic stimulation in the central auditory pathway". en. In: *Scientific Reports* 5 (May 2015), p. 10319. ISSN: 2045-2322. DOI: 10.1038/srep10319. URL: https://www.nature.com/articles/srep10319 (visited on 08/14/2018).

[236] X. Wang et al. "Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics". eng. In: *Journal of Neurophysiology* 74.6 (Dec. 1995), pp. 2685–2706. ISSN: 0022-3077. DOI: 10.1152/jn.1995.74.6.2685.

[237] X. Wang and S. C. Kadia. "Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat". eng. In: *Journal of Neurophysiology* 86.5 (Nov. 2001), pp. 2616–2620. ISSN: 0022-3077. DOI: 10.1152/jn.2001.86.5.2616.

[238] Isaac M. Carruthers et al. "Emergence of invariant representation of vocalizations in the auditory cortex". eng. In: *Journal of Neurophysiology* 114.5 (Nov. 2015), pp. 2726–2740. ISSN: 1522-1598. DOI: 10.1152/jn.00095.2015.

[239] JE Rose, R Galambos, and JR Hughes. "Microelectrode studies of the cochlear nuclei of the cat." In: *Bulletin of the Johns Hopkins Hospital* 104.5 (1959), p. 211.

[240] Eric D Young and WILLIAM E Brownell. "Responses to tones and noise of single cells in dorsal cochlear nucleus of unanesthetized cats". In: *Journal of Neurophysiology* 39.2 (1976), pp. 282–300.

[241] Bjoern Bonath et al. "Neural basis of the ventriloquist illusion". eng. In: *Current biology: CB* 17.19 (Oct. 2007), pp. 1697–1703. ISSN: 0960-9822. DOI: 10.1016/j.cub.2007.08.050.

[242] Fumiko Maeda, Ryota Kanai, and Shinsuke Shimojo. "Changing pitch induced visual motion illusion". eng. In: *Current biology: CB* 14.23 (Dec. 2004), R990–991. ISSN: 0960-9822. DOI: 10.1016/j.cub.2004.11.018.

[243] L. Shams, Y. Kamitani, and S. Shimojo. "Illusions. What you see is what you hear". eng. In: *Nature* 408.6814 (Dec. 2000), p. 788. ISSN: 0028-0836. DOI: 10.1038/35048669.

[244] R. Sekuler, A. B. Sekuler, and R. Lau. "Sound alters visual motion perception". eng. In: *Nature* 385.6614 (Jan. 1997), p. 308. ISSN: 0028-0836. DOI: 10.1038/385308a0.

[245] Barry E. Stein and Terrence R. Stanford. "Multisensory integration: current issues from the perspective of the single neuron". eng. In: *Nature Reviews. Neuroscience* 9.4 (Apr. 2008), pp. 255–266. ISSN: 1471-0048. DOI: 10.1038/nrn2331.

[246] Ilana B. Witten and Eric I. Knudsen. "Why seeing is believing: merging auditory and visual worlds". eng. In: *Neuron* 48.3 (Nov. 2005), pp. 489–496. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2005.10.020.

[247] You-Hyang Song et al. "A Neural Circuit for Auditory Dominance over Visual Perception". eng. In: *Neuron* 93.4 (Feb. 2017), 940–954.e6. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2017.01.006.

[248] David Raposo, Matthew T. Kaufman, and Anne K. Churchland. "A category-free neural population supports evolving demands during decision-making". eng. In: *Nature Neuroscience* 17.12 (Dec. 2014), pp. 1784–1792. ISSN: 1546-1726. DOI: 10.1038/nn.3865.

[249] Christoph Kayser, Nikos K. Logothetis, and Stefano Panzeri. "Visual enhancement of the information representation in auditory cortex". eng. In: *Current biology: CB* 20.1 (Jan. 2010), pp. 19–24. ISSN: 1879-0445. DOI: 10.1016/j.cub.2009.10.068.

[250] Christoph Kayser, Christopher I. Petkov, and Nikos K. Logothetis. "Visual modulation of neurons in auditory cortex". eng. In: *Cerebral Cortex (New York, N.Y.: 1991)* 18.7 (July 2008), pp. 1560–1574. ISSN: 1460-2199. DOI: 10.1093/cercor/bhm187.

[251] Giuliano Iurilli et al. "Sound-driven synaptic inhibition in primary visual cortex". eng. In: *Neuron* 73.4 (Feb. 2012), pp. 814–828. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2011.12.026.

[252] Leena A. Ibrahim et al. “Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition”. eng. In: *Neuron* 89.5 (Mar. 2016), pp. 1031–1045. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2016.01.027.

[253] Guido T. Meijer et al. “Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 37.36 (Sept. 2017), pp. 8783–8796. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.0468-17.2017.

[254] Jon Driver and Toemme Noesselt. “Multisensory interplay reveals cross-modal influences on ’sensory-specific’ brain regions, neural responses, and judgments”. eng. In: *Neuron* 57.1 (Jan. 2008), pp. 11–23. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2007.12.013.

[255] Asif A. Ghazanfar and Charles E. Schroeder. “Is neocortex essentially multisensory?” eng. In: *Trends in Cognitive Sciences* 10.6 (June 2006), pp. 278–285. ISSN: 1364-6613. DOI: 10.1016/j.tics.2006.04.008.

[256] G. M. Innocenti, P. Berbel, and S. Clarke. “Development of projections from auditory to visual areas in the cat”. eng. In: *The Journal of Comparative Neurology* 272.2 (June 1988), pp. 242–259. ISSN: 0021-9967. DOI: 10.1002/cne.902720207.

[257] Jennifer K. Bizley et al. “Physiological and anatomical evidence for multisensory interactions in auditory cortex”. eng. In: *Cerebral Cortex (New York, N.Y.: 1991)* 17.9 (Sept. 2007), pp. 2172–2189. ISSN: 1047-3211. DOI: 10.1093/cercor/bhl128.

[258] Arnaud Falchier et al. “Anatomical evidence of multimodal integration in primate striate cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 22.13 (July 2002), pp. 5749–5759. ISSN: 1529-2401. DOI: 20026562.

[259] Brian Zingg et al. “Neural networks of the mouse neocortex”. eng. In: *Cell* 156.5 (Feb. 2014), pp. 1096–1111. ISSN: 1097-4172. DOI: 10.1016/j.cell.2014.02.023.

[260] Marcus Leinweber et al. “A Sensorimotor Circuit in Mouse Cortex for Visual Flow Predictions”. eng. In: *Neuron* 96.5 (Dec. 2017), p. 1204. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2017.11.009.

[261] Adi Mizrahi, Amos Shalev, and Israel Nelken. “Single neuron and population coding of natural sounds in auditory cortex”. In: *Current Opinion in Neurobiology*. Neural maps 24 (Feb. 2014), pp. 103–110. ISSN: 0959-4388. DOI: 10.1016/j.conb.2013.09.007. URL: http://www.sciencedirect.com/science/article/pii/S0959438813001864 (visited on 08/13/2018).

[262] John B. Issa et al. “Multiscale optical Ca2+ imaging of tonal organization in mouse auditory cortex”. eng. In: *Neuron* 83.4 (Aug. 2014), pp. 944–959. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2014.07.009.

[263] Michael Trujillo et al. “Selectivity for the rate of frequency-modulated sweeps in the mouse auditory cortex”. eng. In: *Journal of Neurophysiology* 106.6 (Dec. 2011), pp. 2825–2837. ISSN: 1522-1598. DOI: 10.1152/jn.00480.2011.

[264] D. A. Slutsky and G. H. Recanzone. “Temporal and spatial dependency of the ventriloquism effect”. eng. In: *Neuroreport* 12.1 (Jan. 2001), pp. 7–10. ISSN: 0959-4965.

[265] Gregg H. Recanzone. “Interactions of auditory and visual stimuli in space and time”. eng. In: *Hearing Research* 258.1-2 (Dec. 2009), pp. 89–99. ISSN: 1878-5891. DOI: 10.1016/j.heares.2009.04.009.

[266] Sascha Tyll et al. “Neural basis of multisensory looming signals”. eng. In: *NeuroImage* 65 (Jan. 2013), pp. 13–22. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2012.09.056.

[267] Ling Qin et al. “Comparison between offset and onset responses of primary auditory cortex ON-OFF neurons in awake cats”. eng. In: *Journal of Neurophysiology* 97.5 (May 2007), pp. 3421–3431. ISSN: 0022-3077. DOI: 10.1152/jn.00184.2007.

[268] G. H. Recanzone. “Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys”. eng. In: *Hearing Research* 150.1-2 (Dec. 2000), pp. 104–118. ISSN: 0378-5955.

[269] Xiaoqin Wang et al. “Sustained firing in auditory cortex evoked by preferred stimuli”. eng. In: *Nature* 435.7040 (May 2005), pp. 341–346. ISSN: 1476-4687. DOI: 10.1038/nature03565.

[270] Benjamin Roland et al. “Odor identity coding by distributed ensembles of neurons in the mouse olfactory cortex”. eng. In: *eLife* 6 (2017). ISSN: 2050-084X. DOI: 10.7554/eLife.26337.

[271] James H. Marshel et al. “Functional specialization of seven mouse visual cortical areas”. eng. In: *Neuron* 72.6 (Dec. 2011), pp. 1040–1054. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2011.12.004.

[272] Jerry L. Chen et al. “Behaviour-dependent recruitment of long-range projection neurons in somatosensory cortex”. eng. In: *Nature* 499.7458 (July 2013), pp. 336–340. ISSN: 1476-4687. DOI: 10.1038/nature12236.

[273] C. E. Schreiner. “Order and disorder in auditory cortical maps”. eng. In: *Current Opinion in Neurobiology* 5.4 (Aug. 1995), pp. 489–496. ISSN: 0959-4388.

[274] Christoph E. Schreiner and Jeffery A. Winer. “Auditory cortex mapmaking: principles, projections, and plasticity”. eng. In: *Neuron* 56.2 (Oct. 2007), pp. 356–365. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2007.10.013.

[275] Stephen G. Lomber and Shveta Malhotra. “Double dissociation of ’what’ and ’where’ processing in auditory cortex”. eng. In: *Nature Neuroscience* 11.5 (May 2008), pp. 609–616. ISSN: 1097-6256. DOI: 10.1038/nn.2108.

[276] Ryoma Hattori et al. “Functions and dysfunctions of neocortical inhibitory neuron subtypes”. eng. In: *Nature Neuroscience* 20.9 (Aug. 2017), pp. 1199–1208. ISSN: 1546-1726. DOI: 10.1038/nn.4619.

[277] Johannes J. Letzkus, Steffen B. E. Wolff, and Andreas Lüthi. “Disinhibition, a Circuit Mechanism for Associative Learning and Memory”. eng. In: *Neuron* 88.2 (Oct. 2015), pp. 264–276. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2015.09.024.

[278] Yu Fu et al. “A cortical circuit for gain control by behavioral state”. eng. In: *Cell* 156.6 (Mar. 2014), pp. 1139–1152. ISSN: 1097-4172. DOI: 10.1016/j.cell.2014.01.050.

[279] Hyun-Jae Pi et al. “Cortical interneurons that specialize in disinhibitory control”. eng. In: *Nature* 503.7477 (Nov. 2013), pp. 521–524. ISSN: 1476-4687. DOI: 10.1038/nature12676.

[280] Soohyun Lee et al. “A disinhibitory circuit mediates motor integration in the somatosensory cortex”. eng. In: *Nature Neuroscience* 16.11 (Nov. 2013), pp. 1662–1670. ISSN: 1546-1726. DOI: 10.1038/nn.3544.

[281] Carsten K. Pfeffer et al. “Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons”. eng. In: *Nature Neuroscience* 16.8 (Aug. 2013), pp. 1068–1076. ISSN: 1546-1726. DOI: 10.1038/nn.3446.

[282] M. E. Larkum, J. J. Zhu, and B. Sakmann. “A new cellular mechanism for coupling inputs arriving at different cortical layers”. eng. In: *Nature* 398.6725 (Mar. 1999), pp. 338–341. ISSN: 0028-0836. DOI: 10.1038/18686.

[283] Matthew Evan Larkum et al. “Dendritic spikes in apical dendrites of neocortical layer 2/3 pyramidal neurons”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 27.34 (Aug. 2007), pp. 8999–9008. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.1717-07.2007.

[284] Johannes J. Letzkus et al. “A disinhibitory microcircuit for associative fear learning in the auditory cortex”. eng. In: *Nature* 480.7377 (Dec. 2011), pp. 331–335. ISSN: 1476-4687. DOI: 10.1038/nature10674.

[285] Ryan J. Morrill and Andrea R. Hasenstaub. “Visual Information Present in Infragranular Layers of Mouse Auditory Cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 38.11 (Mar. 2018), pp. 2854–2862. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.3102-17.2018.

[286] Jennifer K. Bizley and Andrew J. King. “Visual influences on ferret auditory cortex”. eng. In: *Hearing Research* 258.1-2 (Dec. 2009), pp. 55–63. ISSN: 1878-5891. DOI: 10.1016/j.heares.2009.06.017.

[287] Huriye Atilgan et al. “Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding”. en. In: *bioRxiv* (Jan. 2017), p. 098798. DOI: 10.1101/098798. URL: https://www.biorxiv.org/content/early/2017/01/06/098798 (visited on 08/14/2018).

[288] Christoph Kayser et al. “Functional imaging reveals visual modulation of specific fields in auditory cortex”. eng. In: *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 27.8 (Feb. 2007), pp. 1824–1835. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.4737-06.2007.

[289] Israel Nelken. “Processing of complex sounds in the auditory system”. eng. In: *Current Opinion in Neurobiology* 18.4 (Aug. 2008), pp. 413–417. ISSN: 0959-4388. DOI: 10.1016/j.conb.2008.08.014.

[290] Wenzhi Sun et al. “Thalamus provides layer 4 of primary visual cortex with orientation- and direction-tuned inputs”. In: *Nature neuroscience* 19.2 (Feb. 2016), pp. 308–315. ISSN: 1097-6256. DOI: 10.1038/nn.4196. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4731241/ (visited on 08/14/2018).

[291] Morgane M. Roth et al. “Thalamic nuclei convey diverse contextual information to layer 1 of visual cortex”. eng. In: *Nature Neuroscience* 19.2 (Feb. 2016), pp. 299–307. ISSN: 1546-1726. DOI: 10.1038/nn.4197.