IMT School for Advanced Studies, Lucca

Lucca, Italy

Naturalistic stimulation in sensory-deprived individuals reveals overlapping large-scale brain organization with differential cross-modal mechanisms

PhD Program in Cognitive, Computational and Social Neuroscience

XXXII Cycle

Francesca Setti
2020

The dissertation of Francesca Setti is approved.

Program Coordinator: Pietro Pietrini, IMT School for Advanced Studies Lucca

Advisor: Emiliano Ricciardi IMT School for Advanced Studies Lucca
Co-advisor: Giacomo Handjaras, IMT School for Advanced Studies Lucca

The dissertation of Francesca Setti has been reviewed by:

Prof Maurice Ptito,
Université de Montréal, (Qc), Canada;
University of Southern Denmark, Odense, Denmark

Prof Lotfi Merabet,
Harvard Medical School, Boston MA-USA

IMT School for Advanced Studies, Lucca

2020

Contents

Acknowledgements

The material accessible in this manuscript comprises much of the scientific work I devoted myself during the last three years.
The project presented in the following pages has been carried out in collaboration with the Department of Psychology of the University of Turin that provided a substantial contribution in the experiment setup, data acquisition and subjects recruitment.

I would like to express my sincere gratitude to the Molecular Mind Laboratory group, the academic crew of the IMT School for Advanced Studies Lucca and to my advisor, Prof. Emiliano Ricciardi, for the opportunity to carry out this challenging project and for the support I found throughout my PhD path.
A special thank goes to all the people that helped me during this journey and from which I received a great deal of encouragement and assistance.

First of all, to Dr. Andrea Leo and Dr. Giacomo Handjaras that patiently supervised me in different periods and whose expertise was of invaluable worth in drafting the distinct phases of the project. Andrea helped me in formulating the research questions, building up the experimental paradigm and acquiring the data while Giacomo closely mentored me through the devising and development of the methodological tools for data analysis and reviewed the Thesis draft.

I would also like to express my gratitude to Francesca Garbarini for her constant support throughout all the research phases and for her valuable suggestions and feedbacks; to Valentina Bruno and Carla Tinti for their precious dedication and enthusiasm in the recruitment of the different samples and to Matteo Diano for having introduced me to the MRI environment and assisted during data acquisition.

In addition, I whish to pay my special regards to all the people who dedicated their time in participating to the study and to the Unione Italiana Ciechi e Ipovedenti (UICI) and the Ente Nazionale Sordi (ENS) of the city of Turin that welcomed our ideas with excitement and, literally, made this research possible.

Finally, all my gratitude goes to my family that always encourages and sustains me.

Vita and Publications

Francesca Setti was born in Italy on 13/11/1992. From 2014 to 2016 she enrolled the Master's program in Cognitive Neuroscience at CIMeC (Centre for Mind/Brain Science) at the University of Trento, joining the *Visual Cognitive Neuroscience* group under the supervision of Prof. Marius V. Peelen and Prof. Daniel Kaiser. In that period she worked in the field of naturalistic vision exploring the perception and neural processing of objects regularities in complex real-world scenes. She acquired knowledge of MEG technique and MEG data analysis. In 2016 she took the Master's degree in Cognitive Neuroscience and she became a PhD student in Cognitive, Computational and Social Neuroscience at IMT School for Advanced Studies, Lucca under the supervision of Prof. Emiliano Ricciardi. During the last years, she worked on real-world perception through the study of congenital sensory deprivation. She developed her PhD project in first person from project plannig and subjects recruitment, to fMRI data acquisition and analysis.

- **Setti F.,** Handjaras G., Diano M., Bruno V., Tinti C., Pietrini P., Garbarini F., Leo A., Ricciardi E. Naturalistic stimulation in sensory-deprived individuals. Oral session presented at the XXVIII Congresso Nazionale SIPF – 'Real brains in the virtual SIPF Annual Meeting'; 2020 Nov 20-21, Nov 27-28; virtual edition.

- Avvenuti G., Handjaras G., **Setti F.,** Leo A., Cecchetti L., Betta M., Ricciardi E, Pietrini P, Bernardi G. Self-control performance, but not subjective vigilance level, decreases after brief periods of task practice and increases following caffeine intake. Poster session presented at: ESRS. 25th Congress of the European Sleep Research Society; 2020 Sept 22-24; virtual edition.

- **Setti F.,** Handjaras G., Diano M., Bruno V., Tinti C., Pietrini P., Garbarini F., Leo A., Ricciardi E. Naturalistic stimulation in sensory-deprived individuals reveals different reorganization mechanisms. Poster session presented at: OHBM. 26th Annual Meeting of the Organization for Human Brain Mapping; 2020 Jun 23- Jul 3; virtual edition.

- Papale, P., Betta, M., **Setti, F**., Malfatti, G., Pietrini, P., Ricciardi, E., ... & Leo, A. Fast Concurrent Processing of Object Shape and Category in Posterior MEG Sensors. 41st European Conference on Visual Perception; 2018 Aug 26-30; Trieste, Italy.

- **Setti F.**, Papale P., Betta M., Malfatti G., Pietrini P., Ricciardi E., Turella L., & Leo A. Spatiotemporal dynamics of object processing: a MEG study. Poster session presented at: IOP. 19th World Congress of Psychophysiology; 2018 Sept 4-8; Lucca, Italy.

- **Setti F,** Kaiser D, Peelen MV. (2016) Objects in commonly experienced configurations are less distracting: evidence from MEG. *CAOs Workshop on Concepts, Actions and Objects, Rovereto, Italy.*

Abstract

Animal and human sensory-deprived models offer the possibility to study the causal mechanisms implicated in perceptual processing and knowledge organization and investigate how the brain copes with the absence of modality dependent information. Congenital blindness and deafness represent unique tools to comprehend to what extent (the lack of) a specific sensory input is a necessary condition for the morpho-functional development of both early sensory and higher-level brain regions. Accumulated evidence from animal studies and comparative works in humans with congenital/early sensory loss has reliably described major general tenets: preservation of modality-independent large-scale organization with topological/regional task selectivity, and cross-modal, modality-dependent plasticity phenomena. Indeed, whilst on one hand the large-scale architecture of the brain conserves task specificity, at the local level, early sensory areas deprived since birth consistently show cross-modal engagement for the information coming from spared senses (e.g. auditory and tactile stimulation for the visual cortex in congenital/early blind individuals), whereas higher level cortical regions in both the auditory and visual hierarchy are still able to represent stimulus characteristics regardless from the sensory modality conveying the information to the brain. Although robust and replicable, these findings leave open questions about the functional mechanisms underlying the observed brain reorganizations.

Recently, a growing number of studies have used naturalistic stimulation in fMRI to convey complex, real-life-like perceptual and semantic information thus fulfilling the need of ecological validity and generalizability of results to daily-life perception and cognition. Indeed, contrary to traditional paradigms, that make use of simplified, artificial stimuli (static images, visual geometrical patterns or isolated sounds), the setup of naturalistic experiments consists in the usage of prolonged, complex stimulation with the aim to reflect more faithfully the dynamic structure of natural environments humans experience on a daily basis.

To this regard, movies and books audiodescriptions have been widely used to convey a continuous and rich stream of information whose processing requires the concerted deployment

and combination of different brain mechanisms, going from the online integration of the incoming multisensory perceptual input to more complex cognitive operations that impinge on attention, information retrieval/update and semantic knowledge. Thus, the richness of such paradigms offers the unique possibility to study brain functioning in complex settings that closely mimic everyday life experiences.

My PhD research meant to provide new insights on the role of experience-dependent plasticity in shaping brain functioning in everyday life. The project was carried out through a 3T functional Magnetic Resonance Imaging (fMRI) study with the aim to evaluate and compare patterns of brain response to a prolonged naturalistic stimulus (≅ 50 minutes long movie) in the early sensory areas of two models of sensory-deprivation, congenital blindness and deafness. The film One Hundred and One Dalmatians (Walt Disney, 1996) was shortened and edited (with the addition of subtitles and audio descriptions) in order to create three different versions: two unimodal conditions (i.e. auditory-movie with audiodescriptions or visual-movie complemented by subtitles) and one multimodal setting (audio-visual). Specifically, we measured brain responses in five experimental groups of participants: congenitally blind (n=9; 44±14 years), congenitally deaf (n=9, 24±4 years), and three control groups that attended the auditory (n=10, 39±17 years), visual (n=10, 37±15 years), or audio-visual (n=10, 35±13 years) variants of the movie.

After standard preprocessing, we took advantage of an Inter-Subject Correlation (ISC) analysis to measure to what extent sensory deprivation affected whole-brain participants synchronization in each group separately. We then compared congenitally deaf and blind with their control groups, demonstrating wide and overlapping modality-independent responses across groups, accompanied by a less lateralized recruitment (e.g., higher ISC in the right hemisphere) in both blind and deaf individuals. Afterwards, we focused our analysis on sensory areas, showing how V1 and A1 and late regions were differently affected in the congenitally deaf and blind groups. Finally, through computational modeling, we further described in each brain region to what extent the ISC was dependent to specific stimulus characteristics, in terms of both low- (i.e., visual or auditory) and high-level (i.e., semantic) features. Results clearly

indicated that V1 functional activity in blind individuals was driven by acoustic features (i.e., sound envelope), whereas the role of A1 in deaf individuals was not related to any of the low-level nor high-level stimulus descriptions explored. In this thesis, I presented the methodology and the result of the study and discussed the implications of the findings as compared to the current evidence in the literature.

Keywords: naturalistic stimulation, blind, deaf, ISC, cross-modal plasticity

Abbreviations

AG: Angular gyrus
ATP: Anterior Temporal Pole
BOLD: Blood-Oxigen Level Dependent
DMN: Default Mode Network
ECoG: Electrocorticography
EEG: Electroencephalogram
FA: Flip Angle
FDR: False Discovery Rate
fMRI: functional Magnetic Resonance Imaging
fNIRS: functional Near-Infrared Spectroscopy
FOV: Field Of View
GIMP: GNU Image Manipulation Program
GRE-EPI: Gradient Recall Echo Planar Imaging
HCP: Human Connectome Project
IAAFT: Iterative Amplitude Adjusted Fourier Transformed
IFG: Inferior Frontal Gyrus
ISC: Inter-Subject Correlation
LOC: Lateral Occipital Cortex
MEG: Magnetic Elettroencephalography
MT/MST: Middle Temporal area
PCA: Principal Component Analysis
PET: Positron Emission Tomography
PHG: ParaHippocampal Gyrus
PSD: Power Spectral Density
PSL: Superior Parietal Lobule
ROIs: Regions of Interest
SD: Standard Deviation
STS: Superior Temporal Sulcus
STG: Superior Temporal Gyrus
PCC: Posterior Cingulate Cortex
mPFC: middle Prefrontal Gyrus
TE: Echo time
TMS: Transcranial Magnetic Stimulation
TOM: Theory of Mind
TPOJ: Temporo-Parietal-Occipital Junction
TR: Repetition Time
VTC: Ventrotemporal Cortex
VVIQ: Vividness Imagery Questionnaire
VWFA: Visual Word Form Area

List of Figures

Introduction

The sensory deprived brain

Understanding the principles that guide human cognition and behaviour has been a matter of curiosity, speculation and extensive investigation since a long time in human history. The advent of new techniques for exploring brain structural and functional properties allowed researchers not only to describe its architecture and the relationships between different areas, but also to unveil the mechanisms that enable living beings to interact with the surrounding environment and make sense of the external world. Indeed, understanding and characterizing which are the general principles that drive structural and functional organization in the brain is a topic of main interest in neuroscientific research.

To this aim, animal and human sensory deprivation models offer the possibility to study the causal mechanism implicated in perception and investigate how the brain copes with the absence of modality dependent information. Congenital blindness and deafness represent unique tools to comprehend to what extent (the lack of) a specific sensory input is a necessary, sine qua non condition for the morpho-functional development and efficiency of both early sensory and high-level brain regions.

Is specific sensory experience a mandatory prerequisite for the brain to develop functional organization? What happens when a sense is ‘switched off’? How do the deafferented cortices work and which function do they serve?

Half a century ago, seminal studies both in animals and humans first showed that early sensory loss prompts structural and functional reorganizations in related sensory areas (Wiesel and Hubel, 1965; Wanet-Defalque et al, 1988; Buchel et al., 1998). Over the past two decades, accumulating evidence in sensory deprived individuals has been proving that the deafferented sensory cortices are consistently engaged by stimuli conveyed through the spared modalities (Sadato et al., 1996; Ptito et al., 2005; Amedi et al., 2007; Merabet and Pasqual-Leone, 2010; Renier et al., 2010). In this view, in the absence of typical sensory input, the deprived brain areas are taken over by the other senses, through cross-modal processes (Bavalier and Neville, 2002, Sur 2004; Frassinelli et al., 2011; Cecchetti et al., 2016).

Indeed, in blind subjects, the so called "visual" areas are activated by a variety of non-visual tasks that involve auditory (e.g., Röeder et al., 2002; Gougoux et al., 2005), tactile (e.g., Kujala et al., 1992; Sadato et al., 1996, 1998; Merabet et al., 2004) or olfactory stimulation (Kupers et al., 2011; Renier et al., 2013; Araneda et al., 2016; but see Gagnon et al., 2015b; Sorokowska et al., 2019). Interestingly, gustatory stimulation does not engage the deprived occipital cortex in congenital blindness (Gagnon et al., 2015a), suggesting that, unlike the other senses, taste processing does not exploit the vacant neural resources void of visual inputs. Importantly, cross-modal responses in those regions have been documented even for more complex tasks that rely on high-level functions such as spatial navigation (Kupers et al., 2010; Chebat et al., 2020), language processing (Röder et al., 2000; 2002; Bedny et al., 2011; 2015) and verbal memories retrieval (Amedi et al., 2003; 2004). These evidences supported the hypothesis of a reorganization of deprived sensory cortices and, in parallel, started questioning whether such principles could form the scaffolding of the entire brain architecture (Pascual-Leone et al., 2005).

Additionally, such observations paved the way for broader speculations about the definition of age-dependent neuroplasticity and devise effective intervention strategies for people suffering from early sensory loss (Collignon et al., 2011).

In the following sections we will review the state of the art on blindness and deafness research as traditional models of sensory deprivation. Then, we will address some open questions in the field with the aim of a better understanding of the brain mechanisms that enable humans to represent the external world.

Evidence from blindness

Over the past decades, neuroscientific research in humans has been focusing mostly on blindness as a reliable test bench to study how modified experience interacts with hard-coded physiology in driving functionality. Indeed, research on sensory deprivation allows to question the role of specific sensory input in shaping brain structural and functional properties (for a review Ricciardi & Pietrini, 2011). How much does experience matter in modelling brain architecture? Which are the stimulus features that sensory deprived brain areas care about and process? Is there such a thing as a biologically determined "visual" (i.e. modality specific) cortex or is rather the specificity of computations at hand that matters most? Answering such questions will inform the way we conceptualize brain organization, how brain reorganization works in the absence of a specific sensory experience and will open up new frontiers in rehabilitation programs.

Since vision is the more studied and developed sense in humans, initial investigations in the field looked at blindness as the optimal model to unravel the mechanisms guiding sensory perception in real-life contexts. Additionally, the differentiation in sensory loss onset time (i.e. congenital, early in life, late in life) allow researchers to study the mechanisms behind brain plasticity, estimating degree of reorganization the brain is still capable of at different moment in life.

In particular, congenital blindness represents an optimal tool to investigate how the brain "visual" circuitry develops in the absence of any visual information from birth. Such condition allows to "weight" the relative contribution of innate, biological factors from one hand and experience dependent neuroplasticity on the other, in driving perception. Hence, aside from inheritable factors, imposed by the principles of the evolution and thus, broadly speaking, pertaining to "nature", the environment (traditionally described with the term "nurture") provides living beings with the necessary resources to complement a balanced growth and maturation of the biological systems. The role of external world contingencies for typical development is well pictured by the conceptualization of the brain as an "experience expectant" system (Greenough, 1987): an already fully equipped machine that needs the proper fuel to optimally function.

Indeed, both postnatal experiences and innate, genetical factors play a pivotal role in constraining brain development and determining the nervous system phenotype. Hence, exposure to an atypical (e.g., impoverished/enriched) sensory environment from birth, along with changes in peripheral morphology due to injury, anomalies in gene expression or particular evolutionary trajectories may result in both physiological and behavioural deviants. The effects of external conditions on the individuals internal milieu give us the measure of the plastic capabilities of the biological matter (body and brain) following specific experiences. To this regard, research on the role of sensory pathways activity in shaping functional cortical networks, paved the way towards the understanding of the general organization of brain architecture.

Indeed, studies of the visual system in different species of mammals demonstrated that manipulations of the rearing environment, both in terms of augmentation (e.g., enriched cages) (Bealieu & Cynader, 1990; Coq and Xerri, 1998; Rampon & Tsien, 2000) or deprivation (e.g., enucleation, lid suturing, eye patching, stripe and dark rearing) (Hubel and Wiesel, 1965; Berardi et al., 2003, Toldi et al., 1996; Bavelier and Neville, 2002, Blakemore & Cooper, 1971; Sengpiel & Kind, 2002) give rise to phenomena of experience-dependent plasticity that have a strong impact in regulating the pruning and consolidation of connections among cortical and subcortical brain structures (Sur & Rubenstein, 2005; Henschke et al., 2018). Hence, evidence of neuroplasticity as a result of reduced or impaired visual sensory experience in early life, have been reported in rodents (Gordon & Stryker, 1996), cats, raccoon, ferrets and monkeys (Merabet and Pascual-Leone, 2010; Espinosa & Stryker, 2012). These findings were crucial for an accurate conceptualization of ocular dominance (Wiesel & Hubel, 1965; Blakemore & Sluyters, 1974) and orientation preference maps (Sengpiel et al., 1999) of the mammals visual system, while providing important insights for the characterization of the so called "critical periods", temporal windows during development in which the brain structures and networks are still permeable to the quality of peripheral stimulation and capable of plastic rearrangements. Changes in environmental and internal (i.e., molecular and chemical) factors could lengthen (both in terms of precocious onset or delayed closure) the duration of critical periods (for a review Hensch et al., 2005) with direct consequences on brain fictional circuitry maturation.

Direct proofs about the role of early visual experience in informing the development of the structural and functional properties of the brain come also from surgical investigations. Indeed, deafferentation (Angelucci et al., 1998) and re-routing of fibers (Roe et al., 1990; 1992; Sharma et al., 2000; Von Melchner et al., 2000) at birth provided clearly evidence of the capability of the deprived/rewired cortices to modify neural connections and/or physiologically re-adapt, at least to some extent, to the features of the incoming sensory information. However, studies showed that although pliable, the brain morpho-functional properties are somehow constrained by intrinsic, biological properties that are expressed despite dramatic changes in the external world.

The presence of both experienced and non-experienced orientation tuning preferences in the striate cortex of stripe-reared animals (Hubel & Wiesel, 1963; Gödecke & Bonhoffer, 1996) is just one compelling example of how brain functioning proceeds for adjustments of predefined lines accordingly to the available external resources.

To summarize, evidence from animal studies demonstrated that: 1) early visual deprivation results in impaired behaviour (improved abilities with the spared auditory and tactile inputs) due to compensatory plastic changes at multiple brain levels (Rauschecker, 1995), 2) susceptibility to the effects of deprivation is strongly related to the age of sensory loss (Hubel and Wiesel, 1964), 3) structural and functional changes are associated to specific underlying molecular mechanisms (Berardi et al., 2003), 4) the deprived sensory cortex undergoes cross-modal plastic mechanisms that enable the usage of freed neural resources to process inputs coming from the spared senses. Taken together, evidence shows the interplay between innate factors from one hand and experience-dependent modulation on the other, to shape the final phenotype.

Therefore, the knowledge derived from animal models (Wiesel and Hubel, 1965; Blakmore et al., 1970; Hyvärinen et al., 1981; Yaka et al., 1999; Sharma et al., 2000; Frost et al., 2000; Kahn & Krubitzer, 2002; Baldwin & Krubitzer, 2018) revealed the anatomical substrates of the multisensory plasticity observed in humans suffering of congenital and early sensory loss.

Primary and early "visual" areas in congenitally blind individuals

First hints about the recruitment of the "visual" cortex in blind individual came from Positron Emission Tomography (PET) studies that consistently reported high levels of glucose metabolism in striate and prestriate regions during rest (i.e., increased metabolism of occipital areas as compared to other brain regions) and task conditions (i.e., increased metabolism in occipital cortex during auditory or tactile tasks as compared to rest baseline) (Wanet-Defalque et al, 1988; Sadato et al., 1996; Buchel et al., 1998; Weeks et al., 2000; de Volder et al., 2001; Pito et al., 2005). These seminal observations demonstrated that, although deafferented from typical input, early visual cortices were still "active" and responsive to external stimuli. Such discovery raised questions about the functional role of the so called "visual" areas in blindness and determined a growing interest in the field. This phenomenon has been defined as cross modal plasticity (Bavalier et al., 2002). The term cross-modal plasticity refers to the adaptive, compensatory reorganization of neural resources in order to overcome the loss of a sensory modality through the integration of spared sensory information. In the last two decades, a plethora of studies investigated whether or to what extent this recruitment in early visual cortex in blind subjects was specific to stimulus conveyed by the spared sensory modalities (e.g., tactile, auditory), stimulus features (e.g., spatial position of the external sources, low-level spectral properties) or simply related to mental processes (e.g., perception, working-memory, language) (Ricciardi et al., 2014; Amedi et al., 2017).

Evidence studies has been consistently showing phenomena of cross-modal plasticity expressed as the recruitment of the primary and early visual cortices for tactile and auditory perception (Dormal et al., 2016; Striem-Amit et al., 2012). Indeed, in blind subjects the "visual" areas are engaged in a number of different non-visual perceptual and cognitive tasks such as sound processing (Vetter et al., 2020), auditory sentence comprehension (Röder et al., 2002), Braille-reading (Sadato et al., 1996; Cohen et al., 1997), and verbal memory performance (Amedi et al., 2003). The functional relevance of crossmodal plasticity in blind individuals has been confirmed by several TMS studies (Cohen et al., 1997; Ptito et al., 2008; Merabet & Pasqual-Leone, 2008; Kupers et al., 2006; Collignon et al., 2009) that proved the causal role of visually-deprived cortices in acoustic and somatosensory stimulus processing.

Hence, even in the absence of any visual experience or early vision loss, the architecture of what would classically be described as visual cortex adapts to process atypical sensory signals (i.e., input from spared modalities). Of note, these functional changes in early visual cortex may not alter the coarse retinotopic organization of the occipital cortex (Striem-Amit et al., 2015). Thus, these processes of cross-modal plasticity may be ascribed also to alterations in cortico-cortical connections among sensory areas (Collignon et al., 2013; Klinge et al., 2010; Müller et al., 2019).

However, other studies challenged this idea suggesting that the enhanced responses in early visual areas to non-visual stimuli could be guided by other factors aside task-specificity or, to some extent, by the interplay between them. Indeed, this argument is supported by the logical observation that a too varied range of experimental conditions activate overlapping regions of the occipital cortex in blind subjects (Amedi et al., 2017; Singh et a., 2018). One possible explanation for that, has been found in the modulatory effect of common, non-specific, top-down mechanisms involved in task performance such as attention, memory and executive functions (Lewis et al., 2010; Burton et al., 2010). Thus, potentially both bottom-up signals (e.g., inputs from the visual cortex) and feedback projections (e.g., from the associative frontal-temporal areas) could contribute to the observed cross-modal activity.

Although thoroughly showed in blind subjects, the recruitment of striate cortex for auditory processing has been recently documented also in sighted individuals with a variety of experimental setting comprising fMRI (Vetter et al., 2014; Bueti et al., 2010), TMS (Romei et al., 2009), EEG (McDonald et al., 2013, Feng et al., 2014), and ECoG (Brang et al., 2015) measurements. A recent study by Campus et al., (2019) demonstrated that sighted individuals engage their primary visual cortex significantly more than blind individuals when processing sounds related to space representation during a spatial bisection task. This finding shed new lights on role of vision in mediating the representation of the external space and of the importance of multisensory integration in typical learning and perception. Hence, the crossmodal recruitment of primary sensory areas might be a general principle of typical brain functioning that is merely unmasked in a primary sensory, such as primary visual areas, when the 'preferred' modality is unavailable (Ricciardi et al., 2020). These evidences question the nature of the cross-modal engagement of deprived areas found in visual sensory loss and highlight the role of multisensory integration in real-world perception even in primary sensory areas (Murray et al., 2016; Pattamadilok et al., 2019).

Functional preservation in high-order "visual" cortex

Is sensory experience necessary to shape the functional architecture of brain regions involved in processing high-level visual features (e.g., object shape, faces, motion)?
Sensory deprivation models allow to specifically address this question, evaluating the contribution of innate, and environmental factor in determining brain organization. To this regard, a work by Pietrini et al., (2004) showed that congenitally blind subjects exhibit topographically organized categorial responses in ventral temporal, "visual" regions during tactile object recognition, similarly the well-known object knowledge organization already described in sighted people (Haxby et al., 2001). This observation demonstrated that the absence of a sensory modality (vision) does not impair the development of brain functional specificity, which indeed is maintained through cross-modal processes.
Similar results have been consistently found for other ventrotemporal extrastriate areas (Amedi et al., 2007; Mahon et al., 2009; Striem-Amit and Amedi, 2014) and extended to dorsal "visual" areas as well (Ricciardi et al., 2007; Renier et

al., 2010; Collignon et al., 2011; Matteau et al., 2010; Ptito et al., 2012). Thus, the congentally blind brain does present a preserved functional organization into a ventral and dorsal pathways for processing non-visual stimuli.

These observations led to the idea that the traditionally conceived purely unisensory regions (e.g. "visual" cortex) may also process additional task-relevant information coming from different modalities in an integrated fashion (Rezk et al., 2020; Mattioni et al., 2020). Interestingly, this conceptual framework was extended to include high-level semantic processing, since several studies showed that the cortical organization of conceptual knowledge can develop independently from information modality (Mahon et al., 2009; Handjaras et al., 2016, 2017; Mattioni et al., 2020).

Overall these findings challenged the classical interpretation of functional selectivity as carved by sensory segregation, opening up a brand-new perspective on brain properties and functioning.

Thereafter, the idea of a sensory-independent, task-selective brain regional organization started to spread in the field. This paradigm shift promoted the flourishing of an entire line of research to investigate to what extent the cortical modular organization found in sighted is preserved also in blindness. A great amount of evidence in the last years has been reporting, in blind subjects, task-specific brain activations in well characterized anatomical regions, traditionally believed to be selective for *visual* properties of the presented stimuli. (Ricciardi et al., 2014; 2014; 2020; Cecchetti et al., 2016)

Taken together, all these results support the notion of local functional specificity as an innate brain organization feature developing independently from the modality of sensory experience. Moreover, those data highlight the preservation of high-order regions and other task-specific areas. For these reasons some researchers conceptualized the visual system (and sensory cortices in general), as *metamodal* (Pasqual-

Leone and Hamilton, 2001), *supramodal* (for a review Ricciardi et al., 2014) or *amodal* (Caramazza et al., 2003; Fairhall et al., 2013) in nature.

However, controversial results come from studies in sighted subjects enquiring the concept of supramodality as a general principle of brain functioning. Indeed, in sighted individuals both a-modal responses as well as deactivations in visual areas to cross-modal stimulation have been reported (Mozolic et al., 2008; Laurienti et al., 2002; Merabet et al., 2007). To note, the interpretation of the observed data in sighted "visual" areas during the presentation of non-visual stimuli in sighted can be driven by concurrent mental imagery (Ricciardi et al., 2011, Vetter et al., 2014) or by inhibition processes (Laurienti et al., 2002; Iurilli et al., 2012). Indeed, in such scenario, blind individuals offer the opportunity to control for visual mental imagery confounds. For what concerns the extent of observed neuroplasticity, some authors asked whether the large-scale reorganization principle is a distinctive feature of sensory deprived brain or it could be found, under specific circumstances, in sighted individuals as well. In tackling this question, Siuda-Kryzywicka et al., (2016) demonstrated that during tactile Braille reading, sighted individuals displays enhanced activity in the visual word form area (VWFA) and proved its causality in behavioural performance (reading accuracy) through Transcranial Magnetic Stimulation (TMS). Such results, along with other studies, revealed that cross-modal cortical reorganization can be observed also in non-deprived adults during complex skills learning, after intensive training (Saito et al., 2006; Kim et al., 2011; Powers et al., 2012) and in cases of peripheral injuries/damage (Kaas 2000; Dempsey-Jones et al., 2019). Although robust, such forms of short-term plasticity could be driven by other neurophysiological mechanisms that have nothing to do with the formation of novel connections among areas but that rather rely on the "unmasking" of pre-existent, but unexploited pathways.

To summarize, a large amount of evidence supports the notion that a sensory-independent processing may be the underlying principle of the structural and functional reorganization of visual areas in blind subjects.

Evidence from Deafness

The excitement for the discovery of cross-modal plasticity in blind individuals generated a growing interest in understanding whether the same principles could be extended outside the visual domain and applied to other cortical areas as well. How does early deafness affect the structural properties of auditory cortex? Does the auditory cortex functionally reorganize to process other stimulus modalities?

Auditory sensory loss has been extensively studied in animal models (Stanton et al., 2000; Wong et al., 2015; Kok et al., 2014) with an in-dept characterization of subcortical auditory structures and pathways (Butler et al., 2013; Kral 2013). Although the hierarchical organization of the auditory system appears to be preserved across mammals, thus offering broad opportunities to study typical and atypical sensory development, very much of the current knowledge comes from experiments with cats. Indeed, congenitally deaf or postnatally deafened cats have been providing a reliable model for the study of both neural and behavioral outcomes following hearing loss. Electrophysiological recordings acknowledged cross-modal processing of visual and somatosensory information in the auditory cortex of cats (Meredith & Lomber, 2011), mice (Hunt et al., 2006) and ferrets (Meredith & Allman, 2012). This crossmodal reorganization subserves the behavioral advantages (i.e., enhancement of perceptual performance) in the spared modalities. Although sensory loss should impact, at least in principle, all the deprived cortices to a similar extent, crossmodal colonization of vacant territories by ectopic fibers in early deaf animals seems rather sparse and heterogenous. Indeed, deprived primary auditory cortex appears to be marginally influenced by plastic reorganization induced by visual or somatosensory inputs (Kral et al., 2003, Meredith et al., 2011) while higher-level auditory areas do show cross-modal recruitment (Lomber et al., 2010). Thus, congenital deafness determines a cross-modal reorganization of secondary auditory areas without significant changes in the primary cortex that still, could potentially process cross-modal information, as proved by experimental surgical rewiring of fibers (Ptito et al., 2002, Mao & Pallas, 2011) .

To this regard, recordings in the auditory cortex of early-deafened cats showed functionally-preserved cross-modal responses to visual input. Indeed, likely to hearing conspecifics, (deprived) secondary, high-level (posterior and dorsal) auditory regions code for stimuli presented in the contralateral periphery, whilst in another sensory domain (i.e., visual). Further evidence about the causal role of secondary auditory cortex for processing cross-modal (i.e., visual) peripheral information is provided by the significant behavioural impairments in the perception of contralateral space caused by transient deactivation of high-order auditory fields by cooling techniques in deaf cats (Meredith et al., 2010). However, although appreciable responses to visual stimulation take place in absence of acoustic input from birth, large-scale microelectrode mapping revealed that the extent of cross-modal reorganization of secondary auditory cortices in cats is actually not large enough to prevent or degrade responsiveness of these areas to restored input by means of cochlear implants (Land et al., 2016). Indeed, evidence reported only a partial and scattered colonization of the deprived areas by crossmodal (e.g., visual) fibers suggesting a preservation, at least macroscopically speaking, of the genetically predisposed anatomical substrate of both cortico-cortical and thalamocortical connections (Barone et al., 2013; Meredith et al., 2016; Butler et al., 2016, 2017, 2018; Kral et al., 2017). Despite the gross-level anatomical similarities among congenitally deaf and hearing animals, striking differences have been reported at the cellular and synaptic (Klinke et al., 1999) stages affecting, in turn, the functional connectivity within and beyond the auditory cortex.

The fact that these auditory regions contain neural populations responsive for somatosensory and visual stimulations as well as bimodal neurons (visual-auditory), supports the idea of the presence of an already-set architecture that sustains subthreshold processes of multisensory processing (Meredith et al., 2006; Meredith and Allman, 2009). Although studies on animal models were crucial in understanding the role of environmental and genetical factors in the development of auditory cortex, comparative work on the human auditory system is not straightforward and it is constrained by phylogenetic differences across species mainly related to the acquisition of language via sign language and speechreading (Fullerton et al., 2007; Wirthlin et al., 2019). One common finding in the literature was the activation of secondary auditory cortices, especially the antero-medial portions of the Superior Temporal Sulcus (STS), in deaf subjects for sign-language processing (Bevalier et al., 2001; Butler et al., 2013; Moreno et al., 2018) and speech-related lipreading (Capek et al., 2008). Crucially, responses in these areas were not elicited by other visual, non-linguistic stimulation, that instead, along with tactile stimulation, activated more posterior regions of the STS (Sadato et al., 2005; Cardin et al., 2013; Finney et al., 2001; Pavani & Röder , 2012; Scott et al., 2014, Vachon et al., 2013; Bottari et al., 2014). This dissociation suggested an interplay among deprivation-related and linguistic factors in determining the observed cross-modal responses in deafness. However, while compelling evidence of the cross-modal recruitment of primary visual areas by non-visual stimuli in blind people has arisen in the last decades, research in deafness has been failing in providing similarly strong and unequivocal proofs both in animals and humans (Leonard et al., 2012). The following sections will discuss the current knowledge on how congenital deafness impacts primary/early and high-order auditory cortices in humans.

Primary and early "auditory" areas in congenitally deaf individuals

Despite numerous studies have been recorded in secondary auditory areas, evidence about crossmodal reorganization of primary regions in humans is scarce. This is mainly due to the difficulties to agree upon the precise localization of the core region and resolve the controversies about its functional organization. In spite of the effort to devise anatomical, cytoarchitectonic and functional maps of the auditory cortex (Morosan et al., 2001; Formisano et al., 2003; Striem-Amit et al., 2011), the comparability and correspondence between those areas across species is still unclear and matter of intense debate. Indeed, the anatomical landmark of the primary auditory cortex in humans, the so-called Heschl's gyrus, is not present in other mammals and its shape is highly variable across individuals (Penhune et al., 2003). Hence, all these factors make the study of cross-modal plasticity in primary auditory regions difficult and subordinate it to an *a priori*, detailed anatomical definition of its cortical extent.

In tackling these concerns through accurate parcellation of individual brains auditory cortex, recent evidence described cross-modal responses in Heschl's gyrus to visual (Scott et al., 2014), somatosensory and bimodal (combined visual plus somatosensory) stimulation (Karns et al. 2012) in congenitally deaf signers.

Although proving cross-modal plasticity in primary auditory cortex, such evidence was not sufficient to rule out a possible modulation of top-down projections related to visuo-spatial language processing (sign-language and speech-reading). Indeed, congenital and early deafness represent a particular and complex kind of sensory deprivation, in that it implies not only the absence of auditory input, but also little or no access to spoken language. Hence, communication for deaf individuals rely mostly on visual cues such as face and mouth movement interpretation (lip-reading) and the adoption of a language based on gestures. Therefore, a finer characterization of the relative contribution of language-related computations and auditory deprivation *per se* is fundamental to understand which mechanisms drive the observed cross-modal plasticity. Hints to answer this question, came from a study by Cardin et al. (2016) in which the variable "sign language knowledge" were used to disentagle the differential contribution of top-down (linguistic/task demands) and bottom-up (sensory) processes in determining responses in Heschl's gyrus. Hence, the authors compared the BOLD activity elicited by the presentation of videos containing sign language-related material or invented signs to congenitally and early deaf with and without sign language knowledge. Results failed in finding significant differences among the two samples suggesting that cross-modal plasticity in Heschl's gyrus is driven by general, visuo-spatial computations rather than by the linguistic processing of signs.

Another interesting piece of evidence was the work of Striem-Amit et al. (2016) which demonstrated, similarly to another research in congenitally blind individuals (Striem-Amit et al, 2015), that the auditory-based tonotopic organization traditionally found in normally hearing people is spared in adult deaf subjects. This finding showed that the formation of cochleotopic gradients does not require early auditory experience and that it is not detrimentally affected by the plastic changes subserving cross-modal recruitment.

These results, together with the evidence in blind people, speak in favor of cross-modal plasticity as a general principle of brain reorganization after sensory deprivation.

Functional reorganization in high-order "auditory" cortex

Although already proven in cats (Meredith et al., 2011), these data leaved unsolved the question whether a re-organization can be found beyond early auditory cortices in humans. The literature upon the functional selectivity of auditory cortices in congenital/early deafness is much more incomplete compared to the studies on blindness. Evidence of preserved functional specificity is mostly limited to sign language production and comprehension, a task that typically activate a well-known network of left-lateralized fronto-temporal areas involved in the computation of spoken language features (Marshall et al., 2004; Macsweeney et al., 2002, 2008). Further indications of differential functional tuning in the temporal regions of deaf people came from a study by Cardin et al. (2013) showing functional segregation between sensory perception (visual, non-linguistic hands-shape configurations) and cognitive (sign language) processing.

More recently, Bola et al. (2017) investigated whether deprived auditory cortices in deaf exhibit the typical subdivision of labor as hearing people show while listening to sound. Indeed, comparing the functional selectivity of specific areas between auditory deprived and non-deprived individuals allow to verify whether the concept of functional reorganization works beyond primary sensory areas and thus constitute a general, ubiquitous principle of brain functioning. The results of the study revealed overlapping activity in posterior and lateral auditory cortex for deaf and hearing individuals during the presentation of temporally complex patterns of visual and auditory rhythms respectively. Noteworthy, the observed activity did not peak in primary auditory areas but in high-level cortices, traditionally associated to complex sound and music processing (Zatorre et al., 2002; Saari et al., 2018).

Overall, data from research on deaf expanded the idea of a preserved large-scale patterns organization (found in

blindness) to the auditory domain; proving that some aspects of brain organization are, at least to some extent, predisposed before birth. At the same time, sensory experience during development is undoubtedly essential for refining (strengthening/weakening) pre-existing inter-regional pathways and ensuring the development of a completely functional system.
Thus, both retained organization and plasticity are key mechanisms of brain architecture.

To summarize, findings in sensory deprivation highlighted the role of both innate, genetic blueprints and external, environmental factors in determining, through plastic adaptations, the resulting cortical phenotype.
The study of how the brain cope with sensory impairment give us important hints to understand the general rules that govern brain functioning across species. To this regard, congenital sensory deprivation represents the ideal scenario to evaluate robust, long-term and large-scale neuroplasticity.
Ultimately, knowledge of the processes that arise in deafness and blindness is a crucial aspect when dealing with sensory loss restoration because it has major implications for the success of targeted interventions. Indeed, independently of the time of occurrence, losing one sense affects the brain structural and functional architecture properties. Hence, sensory recovery later in life poses several challenges due to the presence of an already substantially reorganized system. Thus, a deep understanding of the brain mechanisms beyond sensory plasticity is vital for planning accurate ad effective interventions and rehabilitation programs.

Naturalistic Stimulation and ISC analysis

In the last decades, the study of brain responses to real-life-like settings has drawn growing interest in the neuroscientific research. The need of understanding whether the discoveries in the lab apply to ecologic contexts brought researchers all over the world to devise new experimental paradigms to mimic the complexity of daily life experiences. Traditionally, neuroimaging experiments have made use of artificial, simplified stimuli, such as visual geometrical patterns, still images or isolated beep sounds to define the neurobiological substrates of sensory perception and cognition. Although needed for an in-depth characterization of the brain processes underlying human behavior, such oversimplified and highly controlled laboratory settings appear detached from reality and fail, for obvious reasons, to grasp the complex, dynamic structure of the natural environment humans live in. Moreover, the study of mental states taken in isolation necessarily produce an incomplete portray of a way more sophisticated picture in which cognitive, sensory and motor processes co-occur and integrate.

Naturalistic paradigms have been adopted to answer those needs and investigate human brain mechanism with greater ecological validity (Hasson & Honey 2012; Wang et al., 2016). Hence, the term *naturalistic* explicitly refers to the usage of rich, multimodal (Bishop & Miller, 2009; Maguire et al., 2012; Lahnakoski et al., 2012) complex stimuli (Kauttonen et al., 2015) that more closely reflect real-life cognition and perception. Indeed, through the exploit of stimuli that require dynamical, continuous processing of an ever-changing stream of information (e.g. movies, video games, books audiodescriptions and virtual environments) is it possible to study the brain functioning as an integrate system where top-down (e.g. attention-mediated) and bottom-up (e.g. multisensory integration) mechanisms act together in determining complex responses. Moreover, the presentation of realistic, interactive stimuli turned out to be an ideal tool for the study social cognition and emotions (Nummenmaa et al., 2018; Lettieri et al., 2019; Nastase et al., 2019).

However, naturalistic paradigms come with theoretical and practical challenges mainly related to data analysis and results interpretation. Indeed, although more engaging that traditional paradigms (Vanderwal et al., 2015), during task-free stimulation, is not straightforward to verify subjects' compliance and attentiveness throughout the experimental session. This is especially true for certain populations such as children (Richardson et al., 2018; Vanderwal et al., 2019) or patients that may show troubles in keeping sustained attention for prolonged period of time. Moreover, naturalistic stimulation hardly fits the classical task-based analysis pipelines, since the high-dimensionality (i.e., large amount of stimulus descriptors) and the impossibility to control for features collinearities thwart the a-priori definition of independent stimulation models. Thus, the necessity to manage stimulus complexity brought scientists to the development of novel and innovative data-driven analysis methods in order to describe the statistics of the brain responses elicited by real-time processing of multidimensional, structured information. One major factor contributing to the growing interest toward naturalistic paradigms has been the demonstration that exposure to prolonged, complex natural visual and acoustic stimulation (e.g. movie, audiodescriptions) synchronizes people brain activity over time, as reported by fMRI (Hasson et al., 2004, 2010; for a review Bottenhorn et al., 2018), MEG/EEG (Betti et al., 2013; Chang et al., 2015), fNIRS (Mukamel et al., 2005) and ECoG (Haufe et al., 2018) studies. This approach, known as Inter-Subject Correlation (ISC) analysis (Hasson et al., 2004), allows to identify shared, stimulus-driven responses by computing the averaged voxel-wise temporal correlations across subjects' fMRI timeseries. In other words, ISC analysis computes a measure of similarity among different subjects' brain responses when exposed to the same dynamic, complex stimulation. However, other than providing a measure of stimulus-locked neural responses consistency, ISC analysis can also highlight shared information across different experimental conditions and populations (Hasson et al., 2009).

In the last years, the usage of naturalistic stimulation has been widely embraced in neuroscientific research driving, at the same time, a substantial rethinking on the potentiality of data sharing. Indeed, in the last decade, the possibility to collect and handle massive quantities of data has led to the creation of open sources platforms such as *studyforrest,* a data repository (http://studyforrest.org) that collects fMRI and eye-tracking data registered while subjects were presented with the movie *Forrest Gump*, along with extensive annotations of the stimulus features (Hanke et al., 2014, 2016).

Recent findings and open questions

Naturalistic stimulation and ISC analysis can conjointly foster a better description of the functional implications of brain damage and sensory deprivation-related plastic remodeling on daily-life perception and cognition.

Recently, Loiotile et al., (2019) published a work in which they tested the ecological validity and reliability of the effects of sensory deprivation on cortical function in a group of congenitally blind subjects presented with naturalistic audio-movies. Results showed significant inter-subject correlation values in the occipital, “visual” cortex of blind individuals but not of sighted controls during stimulus listening.

This finding proved that the well-known structural and functional plastic modifications historically observed in congenital blindness arise to manage the complexity of real-world tasks. Indeed, complex and dynamic stimuli, as those encountered in real-life situations, synchronized large portions of the deafferented “visual” regions including both primary and more high-level areas of lateral and inferior occipito-temporal cortex. However, several questions about the significance of experience-dependent crossmodal plasticity in ecological settings remain open.

First, do different models of sensory deprivation (i.e. blindness and deafness) provide us converging information about the general principles that govern brain architecture and functional properties in real-life scenarios?

Second, which are the commonalities and the functional differences between deprived and non-deprived cortices during real-world interactions?

Third, what kind of perceptual and cognitive processes drive the observed synchronization of brain activity across subjects? Since naturalistic stimulation entails many cognitive operations at once, it is likely that the observed activations in the "visual" regions would be driven by different concurrent (i.e. bottom-up and top-down) processes that still need to be characterized. To this regard, an accurate modelling of the experimental stimuli would be necessary to understand which modality-(in)dependent stimulus features mediate synchronized responses within- and across- groups.

Lastly, do overlapping fMRI activity indicate same mental representations across groups?

Naturalistic stimulation in congenitally sensory deprivation

My PhD research project aimed to provide new evidence to address these critical, unsolved questions and offer new insights on the role of experience-dependent plasticity in shaping brain function in real-life contexts.
The research we performed is a 3T fMRI study carried out on both congenitally blind (n=10) and congenitally deaf

individuals (n=9), while presented with edited version of a 50 minutes long movie (101 Dalmatians, 1996, Walt Disney©). The two groups were paired with sighted (n=10) and hearing (n=10) control subjects that underwent the same unimodal experimental stimulation (auditory-movie and visual-movie respectively). Moreover, we included a fifth condition, administered on a different control group (n=10) and consisting in the complete audio-visual version of the movie (i.e., multimodal setting). Brain activity was analyzed using Inter-Subject Correlation (ISC) (Hasson et al., 2004) to measure the synchronicity across subjects within each experimental group and condition. To have a thorough characterization of the mechanisms driving ISC, we extracted featural descriptions of the movie using computational modeling. Specifically, we built multiple concurrent descriptors of both low-level and high-level stimulus features: two were based on static (i.e., power spectrum, de Heer et al., 2017) and dynamic (i.e., envelope modulation, Martinelli et al., 2020) acoustic properties; two on static (i.e., spatial frequencies, Oliva & Torralba, 2006) and dynamic (i.e., motion energy, Nishimoto et al., 2011) visual characteristics; two based on semantic categorial descriptors in auditory and visual domains separately (Huth et al., 2016; Mattioni et al., 2020) and, one based on computational linguistics (i.e., single concepts expressed as word embeddings of the movie dialogues Huth et al., 2016; Pereira et al., 2018; Vodrahalli et al., 2018). Moreover, since synchronization of brain activity in fMRI mainly arose from slow-frequency fluctuations (Hasson et al., 2004), we further controlled our computational models for these temporal components. To this aim, we first estimated a descriptor for the film editing process. Indeed, a movie is the result of the editing of quick shots (e.g., close-up) into scenes, artificial sequences which generally retained the unity of time and location and whose sound, music and dialogues were modified and assembled to be entrained with the visual stream. This technical process determined the pace of the audiovisual stimulation and the final rhythm of the narrative. Crucially, the description provided by film editing shared a consistent portion of variance with all the above mentioned computational models and represented an ideal covariate to control for slow temporal changes which cannot be ascribed to a single specific model and which can be modulated by other brain processes (e.g., working memory, attentional mechanisms) apart from sensory processing (Hasson et al., 2015). Indeed, real-life perception relies on dynamic processes that entail both storage of past information and continuous, on-line update of the current experience/knowledge basing on the occurrence of new events. In this way, the stream of information is integrated across different time scales in a hierarchical fashion, where primary sensory regions were modulated by transient inputs and high-order cortical areas (e.g., parietal and frontal areas) were sensitive to events persisting for minutes (Hasson et al., 2015; Baldassano et al., 2017). Therefore, we orthogonalize the film editing descriptor and the computational models to remove the slow-frequency oscillations which hardly represent purely sensory (i.e., acoustic and visual) or semantic (i.e., categorial or conceptual) short-term information.

The orthogonalized computational models served as mediator variables in the synchronization process across subjects (MacKinnon et al., 2007). Measuring the drop of ISC given a specific stimulus descriptor, we indirectly estimated how much of the inter-subject synchronization was driven by visual, auditory or semantic features. Similarly to Representational Similarity Analysis (Nili et al., 2014), this single-voxel approach was resilient to different model dimensionalities and offered a unique perspective where the ideal computational model led ISC to zero. Ultimately, we aimed to highlight both common, shared responses and differences between non-deprived and deprived (blind and deaf) individuals, weighting the relative contribution of each computational model into the synchronization of brain regions.

Materials and methods

Subjects

Fifty right-handed subjects took part in the study. Recruited participants were divided in five independent samples: congenitally blind (n=9, mean age 44±14 yrs, 3 females), congenitally deaf without cochlear implants (n=9, age 24±4, 5 females) and three control groups that attended the auditory (n=10, 39±17 yrs, 7 females), visual (n=10, 37±15 yrs, 5 females), or audio-visual (n=10, 35±13 yrs, 8 females) variants of the movie. This portioning allows us to have three different experimental settings: an auditory condition for the blind and the matched group, a visual condition (supplemented by subtitles) for the deaf and their paired controls, and lastly the complete, audio-visual version of the movie. The recruitment of the subjects was constrained by specific requirements: aside from being compatible for the MRI setting, blind and deaf participants were selected basing on the onset of their sensory deficit (all congenitally affected), no effect of rehabilitation programs (i.e. in the deaf impairment in hearing environmental sounds) and no history of neurological disorders. All congenitally deaf individuals were proficient in Italian sign language and did not use hearing aids at the moment of the study (the majority of the subjects used them for a limited period of time during childhood). All the subjects belonging to the control samples reported no hearing impairment and normal or corrected-to-normal vision. In addition, no-one of the hearing individuals has knowledge of Italian sign language. Additional information about the deaf and blind samples were provided in the Supplementary Material section. Only native Italian speakers were selected to be enrolled in the study. Each volunteer was instructed about the nature of the research and gave written informed consent for the participation in accordance with the guidelines of the institutional board of Turin University Imaging Center for brain research. The study was approved by the Ethics Committee of the University of Turin and conforms to the Declaration of Helsinki. Two blind subjects were removed from the analysis for poor behavioral performance and excessive head movement.

Stimulus

To provide a naturalistic stimulation to our five sets of participants, we selected the action movie “101 Dalmatians” (S. Herek, Great Oaks Entertainment & Walt Disney, 1996, based on Walt Disney's animated 1961 movie and adapted from the Dodie Smith's 1956 novel, “The Hundred and One Dalmatians”). The stimulus selection was done after considering the movie content, the participant characteristics and the fMRI setting. Indeed, we wanted our movie to include the classes of stimuli traditionally used in vision research (e.g., faces, houses/buildings, body-parts, man-made tools, words, vehicles, natural and urban landscapes), to last a duration viable for a relatively short fMRI acquisition session (about 1h) and to ideally favor subjects compliance and engagement (plot linearity and simplicity, familiarity with presented settings and characters). Moreover, considering the perceptual deficits of blind and deaf participants and our aim to measure the commonalities and differences among sensory deprived individuals, we considered the presence of unfamiliar characters, actions and scenes (referring with this term to those life scenarios that were somehow far from the cultural environment which our participants experience in daily life) as an exclusion principle, preferring instead plots that depict common, everyday life contexts. As concern the plainness of its narrative, we believed that plot linearity was crucial to facilitate comprehension and support subjects’ compliance especially for what concerns the auditory-only version that, otherwise, will be extremely difficult to follow.

The movie was shortened to make it suitable for a single scanning session. For this purpose, we discarded the scenes which exclusion do not alter the main narrative thread and edit the remaining parts together to ensure smooth transitions among cuts and preserve the continuity of narration. We edited the movie to a final version of about 54 minutes that was split into six runs (average time 8 minutes) accordingly to the main streams of the story. A six-second fade-in and fade-out period was added at the beginning and the end of each run. We added to the video stream a small red fixation cross at the center of the visual display (1280 x 720 pixels). Although movie watching intrinsically implies free viewing, leaving the person free to gaze any location in the space, we though having a central fixation cross would be an implicit cue, important in keeping people attentive for a long time. Indeed, participants in the fMRI session did not receive any specific instructions about the need to maintain the focus at the fixation, meaning they were free to move their eyes across the entire visual display accordingly to the saliency of the tale unfolding. However, given the need of preserve sustained attention throughout the entire length of the movie, we reasoned that the presence of a noticeable central fixation cross will help participants to maintain the gaze on the center of the screen, especially during scene transitions. Moreover, we included in each video frame subtitles related to the movie dialogues, text embedded in the video stream (e.g., newspaper), onomatopoeic sounds and voice-over audio descriptions (see below). Subtitles were written in different styles and colors according on the speaking voice (white for characters, blue for background sounds and yellow for the storyteller descriptions) to facilitate speech segmentation and aid understanding.

In addition to the six runs comprising the whole storyline, we built a scramble run as control condition. Specifically, the discarded segments during the editing of the main storyline were randomly sampled in accordance with the movie’s camera cuts duration distribution and concatenated to build a control stimulus that, although made by scrambled pieces, reflects the same structure of the original film. In this way, we made up a condition that, while sharing the same low- and middle-level visual and acoustic features with the actual stimulus, breaks the coherence of the storyline, crucially losing the meaning of the narration.

Video editing has been carried out using iMovie software (10.1.10) on a Apple Macbook Air, whereas for the creation of subtitles, we rely on the open-source cross-platform Aegisub 3.2.2 (http://www.aegisub.org/).

As concerns the audio stream, besides the sound editing procedures associated to the video editing ones described above, we included a voice-over audio track. We adapted the commercial Italian audio-description of the movie (transmitted on TV by the Radiotelevisione Italiana) to our custom-made shortened version of the film. Therefore, this process meant re-writing several parts of the original script not only to better bridge the tale gaps we introduced through scene cutting but also to ensure a satisfactory verbal depiction of those aspects of the visual scenery that are caught by neither characters' dialogues nor music valence but still, are essential for understanding. This work was done in collaboration with a professional Italian actor[1] with extensive experience in the field of dubbing, whose competence was fundamental to make the description effective and appropriate for the samples under investigation. We asked him to make the voice-over of the story to guarantee uniform pitch and cadence, avoid any dialectal inflections and be sure to provide the proper expressivity and intonation in conveying the emotional content. The voice-over was recorded in a studio[2] insulated from environmental noise and provided with professional hardware (Neumann U87 ai microphone, Universal Audio LA 610 mk2 preamplifier, Apogee Rosetta converter, Apple MacOS) and software (Logic Pro) equipment comprising a

---

[1] Alessandro Bertolucci

[2] DAVstudio di Lazzarini Riccardo www.davstudio.it

set of microphones and filters to manipulate sounds rendering. The voice track was then adequately combined with the movie original soundtracks and dialogues. We introduced fade-in and fade-out effects to smooth the auditory content at the beginning and end of each run to better manage the transitions among the subsequent segments of the film. Music and voice tracks were mixed and adjusted to tailor the speaker's voice according to the movie soundscape and to optimize the overall volume as a function of the MRI scanner noise.

Finally, we focused our effort in synchronizing the written text to the auditory track. Since line segmentation does not interfere with either reading and story comprehension or image processing (Perego et al., 2010), we modified the subtitle pattern in subsequent visual displays upon necessity, appearing in both two-lines and one-line format.

Thus, we ended up with a paradigm consisting of 7 runs: 6 for the movie plus a run used as control condition. All the runs were administered to five groups of participants employing only the sound stream (i.e., congenitally blind individuals and their controls), only the video stream (i.e., congenitally deaf individuals and their controls) and the complete audiovisual experience in a fifth independent sample.

All the work on the movie, allowed us to craft a multidimensional stimulus set that comprises two single-modality versions (visual-only and auditory-only) as well as a complementary multimodal one, represented by the audio-visual live action movie.

Figure 1. Experimental Conditions.
Depictions of the three experimental settings are reported in (A), (B) and (C). (A) The audio-visual condition comprises both visual (movie landscape plus subtitles) and acoustic (original soundtrack and voice-over descriptions) stimulation, while in the two unimodal conditions (i.e., auditory-only and visual-only) represented in (B) and (C), information is conveyed by a single sensory channel (i.e., audition and vision respectively).

Experimental procedure

All participants were asked to rate their general knowledge of 101 Dalmatians plot and, more specifically, their acquaintance with the *action movie* per se. Stimulus familiarity was examined by means of a simple questionnaire in which participants were asked to rate their knowledge of the movie from 1 (not at all / never seen it) to 5 (very well) (see Appendix, Supplemental Behavioral assessment). Since the movie and the cartoon plots are similar and they are both adaptations of the 1956 children's novel *The Hundred and One Dalmatians* by Dodie Smith it could happen that subjects had never seen the movie before but still, know the story. For this reason, we asked people not just to report their familiarity with the movie itself but also, more broadly, with the narration, that they could have acquired in a different way. The participants that affirmed to have seen the movie at least once, were requested to verbally summarize what they remembered about it, while their voices were digitally recorded. In the case of deaf individuals, the interpreter just wrote down what they told us by sign language (Appendix, Supplemental Behavioral assessment).

Functional and structural data acquisition were performed in a single session per participant. The experiment consisted in the presentation of 1h long movie split in seven runs. In the break between the end of a run and the starting of the following, the experimenter communicated with the subject through an interphone (for the auditory and the audio-visual conditions) or by written questions displayed into the goggle screens (for the deaf individuals) to check people compliance and ascertain any problems have occurred while scanning. After the movie was finished, participants were asked to remain focused and keep being attentive throughout the presentation of the control stimulus. Subjects were requested to remain as still as possible for the entire duration of the acquisition, during the structural scans and while watching the movie/listening to the story. The very same recommendation about movement inhibition was repeated during each break to limit unwanted body/head movement. While much attention was put in preventing excessive and undesirable movements, volunteers did not receive any specific indication concerning eye-movements: although a red fixation cross was drawn in the middle of the screen, they were free to gaze at different parts of the visual display following the major actions at play. Since the experiment does not have a structured task, no specific instructions were provided to the participants that were just told to "follow the plot and enjoy the movie". Since for this kind of experimental paradigm there is no straightforward way to assess task performance or participant engagement, we prepared a post-scan true/false questionnaire which accuracy score reflects subjects' tale comprehension and compliance during the experiment (see Appendix, Supplemental Behavioral assessment). Questions were made to address unique aspects of the movie as much as possible trying to include those peculiar aspects which answer was not arguable from previous knowledge of the story plot. Despite our effort, a bit of overlap with the book tale and the cartoon was still recognizable, since the movie derived from the novel itself. Although simple and focused on key, relevant parts of the plot, each question contained tricky elements that require good understanding of facts, actions unfolding and interaction among characters. For this reason, we believed that performance in the test represents an index of subjects' engagement in the task meaning we may speculate about the goodness and reliability of the respective fMRI data basing on the accuracy in this final debriefing. Aside, we administered other surveys: one concerning manual dexterity, assessed through the Edinburgh Handedness Inventory (Oldfield R., 1971) and a set of questionnaires related to sensory imagery: the shortened version of the Bett's Questionnaire upon Mental Imagery (Sheehan, 1967) the Visual Vividness Imagery

Questionnaire (VVIQ) (Marks, 1973) and the Plymouth Sensory Imagery Questionnaire (Andrade et al., 2014).

Stimulation setup

Participants were presented with different experimental conditions basing on the experimental group they belong to. Therefore, the stimulation setup we used varied accordingly to the sample under investigation and the characteristics of the associated experimental paradigm. Audio and visual stimulation were delivered through MR-compatible LCD goggles and headphones (VisualStim Resonance Technology, video resolution 800x600 at 60 Hz, visual field 30° × 22°, 5″, audio 30 dB noise-attenuation, 40 Hz to 40 kHz frequency response). Both goggles and headphones were prescribed irrespectively of the experimental condition and group membership, meaning that each subject worn always both devices. They were turn off when not useful or switched on depending on the stimulation we wanted to deliver. Therefore, in the audio-only condition goggles were shut down, in the video-only scenario headphones were silenced while in the complete audio-video condition both were kept running. In all cases, we asked subject to put on earplugs (even when sound was played) to ensure ear protection against MRI noise. Instructions were provided verbally whenever it was possible and communication with the subject was carried through a two-way communication system implemented in VisualStim apparatus technology. For what concerns deaf subjects, communication was possible thanks to the support of an interpreter of the Italian Sign Language (affiliated with the Associazione Interpreti di Lingua dei Segni Italiana (Anios) and working for the Ente Nazionale Sordi (ENS)) that carefully introduced the participants to the MR-environment and instructed them about the experiment before starting the scanning session.

During the acquisition the interaction was maintained projecting directly in the goggles written instructions and waiting for the participant reply through sign language gestures that we carefully monitored from the MRI control room. The video and audio clips were administered through PsychoPy "Builder" Graphical User Interface (GUI) and then refined compiling the relative python script.

Functional MRI data acquisition

Brain activity was recorded with Philips 3T Ingenia scanner. Functional images were acquired using gradient recall echo planar imaging (GRE-EPI; TR = 2000 ms; TE = 30 ms; FA = 75°; FOV = 240 mm; acquisition matrix (in plane resolution) = 80 × 80; acquisition slice thickness = 3 mm; acquisition voxel size =3x3x3 mm; reconstruction voxel size =3x3x3 mm; 38 sequential axial ascending slicestotal volumes 1614 for the six runs of the movie, plus 256 for the control run). In the same session, three-dimensional high-resolution anatomical image of the brain was also acquired using a magnetization-prepared rapid gradient echo (MPRAGE) sequence (TR =7 ms; TE = 3.2 ms; FA = 9°; FOV= 224, acquisition matrix = 224 x 224; slice thickness = 1mm; voxel size = 1x1x1 mm; 156 sagittal slices).

Functional MRI data preprocessing

fMRI data preprocessing was performed following the standard steps with AFNI_17.1.12 software package (Cox, 1996). First, we removed scanner-related noise correcting the data by spike removal (*3dDespik*e). Then, all volumes comprising a run were temporally aligned (*3dTshift*) and successively corrected for head motion using as base the first run (*3dvolreg*). We apply a spatial smoothing with a Gaussian kernel (*3dBlurToFWHM*, 6mm, Full Width at Half Maximum) and then data of each run underwent percentage normalization. Aside, we performed detrending applying Savitzky-Golay filtering in MATLAB (function *sgolayfilt*, polynomial order: 3, frame length: 200 timepoints) onto the normalized runs to smooth the corresponding time series and cleaning them from unwanted trends and outliers (Cukor et al., 2013). Runs were then concatenated, and multiple regression analysis was performed (*3dDeconvolve*) to remove BOLD signal related to head motion parameters and movement spike regressors (frame wise displacement above 0.3). Afterwards, single subject fMRI volumes were nonlinearly (*3dQWarp*) registered to the MNI-192 standard space (Fonov et al., 2009).We perform a quality check of the data through visual inspection and automated checks. Two subjects from the blind group were discarded for excessive motion.

Stimulus modeling and feature spaces

To understand how deprived sensory cortices compute complex, continuous information, we used a series of computational models to describe our stimulus and to measure their associations to BOLD brain activations. Consistently with the theoretical framework of hierarchical sensory processing (Heeger et al., 1996; Di Carlo et al., 2012; de Heer et al., 2017), we made advantage of the richness of naturalistic paradigms to investigate stimulus-driven brain responses to low-level, high-level and categorical movie features. Therefore, we employed models of early visual and auditory computations to characterize low-level perceptual features of the movie (i.e. image gist and motion energy; sound power spectrum and auditory envelope) along with high-level descriptors such as a word embedding space for the analysis of the movie script (word2vec algorithm, Mikolov et al., 2013), a validated neural net algorithm for image segmentation and category discrimination (Zheng et al., 2015), as well as manual tagging of the most commonly

used categorial features (e.g., faces, body-parts, animals). We aimed to get a complete set of data-driven stimulus descriptors that grasp the complexity of the stimulus naturalistic stimulation. Indeed, one limit of naturalistic paradigms was that they did not allow a full control over the cognitive states elicited by the stimulation as well as a clear control of the subjects' vigilance and attentional states. Thus, using our models we were able to provide a finer characterization of the elicited brain activity.
The subsequent paragraphs introduced the computational modeling along with the semantic analysis of the movie script, the description of the features related to the movie editing and finally, the categorial tagging.

Low-level computational models

In the following sections the data-driven models we adopt to extract the movie low-level features will be reviewed. We modeled the frequential signal properties and the envelope domain of both the visual and auditory movie stimuli.

Visual computational models

In the last years a great amount of work in the field of computer vision has shown that it is possible to infer what a person is looking at (the so called 'brain reading') by appropriately modeling the brain activity pattern elicited by the presentation of both simplified, artificial geometric visual arrays and complex natural scenes. Aside from image classification (Haxby et al.,2001; Kamitani and Tong, 2005), decoding techniques can be used also for stimulus reconstruction (Kriegeskorte & Douglas, 2019) and this is particularly relevant when studying vision in natural contexts. Indeed, being able to trace back the information encoded into voxels activity to the actually viewed natural pictures could provide a deeper and more valuable understanding of the kind of computations behind everyday life scenes perception (imagery and dreaming). Modeling natural images features has been an open field of research itself given the difficulties related to providing in-depth descriptions both in terms of the embedded rich statistical structure and the presence of a meaningful, semantic content. However, the complexity of natural vision cannot be reduced to the computations of static geometric patterns or life-like scene pictures taken in isolation since it implies the processing of a continuous stream of information, that unfold over time. To this regard, a movie is a good resemblance of reality: although being essentially a collection of static frames, it presents a pictorial structure which changes as a function of time. Therefore, a good description of our visual stimulus should comprise at least the modeling of basic image statistics and at the same time keeping track of its temporal structure dynamics. In the following sections we will describe the models we used to describe the visual-movie features.

GIST feature space

We used a scene GIST model (Oliva and Torralba, 2006) to quantify the spatial properties of the movie frames convolving a set of Gabor filters with a specific frequency and orientation to the image. We segmented each movie frame into a 4x4 grid and sampled the responses to Gabor filters having four different sizes and four orientations, resulting in a model comprising 256 features for each frame (Lettieri et al., 2019). Each feature represented the total energy at a particular orientation and spatial frequency, averaged over a position of the visual field. Subsequently, GIST descriptors across 50 frames within 2 second were averaged to match the temporal resolution of fMRI.

Motion energy feature space

We computed the total motion energy for each movie second through a set of 4,715 motion energy descriptors consisting of a quadrature-pair of space-time Gabor filters (e.g., Gabor wavelets with three different temporal frequencies at 0, 2, and 4 Hz as in Nishimoto et al., 2011). Such model described each movie frame by a set of preferred spatial frequencies, orientations and temporal frequencies that grasp fast-changing visual information. For further details please refer to the publicly available code (https://github.com/gallantlab/motion_energy_matlab).

Auditory computational models

The movie audio description gives us the opportunity to study a complex and highly ecological stimulus that includes meaningful natural narrative speech and environmental sounds simultaneously. Indeed, akin to the studies in the visual field, auditory processes have been commonly addressed through simplified stimuli such as isolated sounds or speech fragments that, although providing a fundamental understanding of auditory perception in the brain, shed a little light upon real life sounds perception. Indeed, natural environments usually consist of complex auditory scenes in which sounds with heterogeneous acoustic characteristics originate from multiple sources and overlap in time. The ability to identify individual sources segregating the corresponding sound from the broad auditory stream depends heavily on spectral information and temporal changes in the auditory waveform amplitude. Therefore, the movie soundtrack was modeled using two low-level feature spaces in order to take account for the frequency and intensity modulations of the auditory signal over time. Here, with the term soundtrack we denote the entire soundscape of the movie thus referring to the composite audio signal in which environmental sounds, music, dialogues and narrator monologues are mixed together. Speech words meaning and anecdotic referential sounds (i.e. those produced by nature, animals or human activities. Jean-Louis Di Santo, 2012) will be used for the construction of the semantic features space discussed below.

Power spectrum feature space

Spectral features extraction was carried out following the method described by de Heer and colleagues (2017). We estimated the signal power spectrum for each run through the Welch's power spectral density estimate (Welch, 1967) with a Gaussian window (SD of 5 ms, length 30 ms, 1 ms spacing between window) over portions of the signal of 2 seconds length (in order to match the fMRI TR). The output is a 449-dimensional vector that summarizes the signal power spectrum (expressed in dB units) in the range of 0 Hz to ~15000 Hz computed in bands of 33.5 Hz. For further details about the parameters used please refer to de Heer et al., 2017 and Lettieri et al., 2019.

Signal envelope feature space

To model sound amplitude changes over time, we extracted the soundtrack envelope power spectral density. First we first evaluated the the upper and lower root-mean-square envelopes of the raw sound signal averaged over the two channels, through the MATLAB function *envelope* (option '*rms*') with a sliding window of 10 ms length. We estimated the signal power spectrum over 2 seconds signal bins with the Welch's power spectral density estimate (Gaussian window, SD 800 ms, length 1 s, 0.5 s spacing between windows). The output is a 49-dimensional vector that summarizes the envelope power spectrum (expressed in dB units) in the range of 1 Hz to 99 Hz computed in bands of 2 Hz.

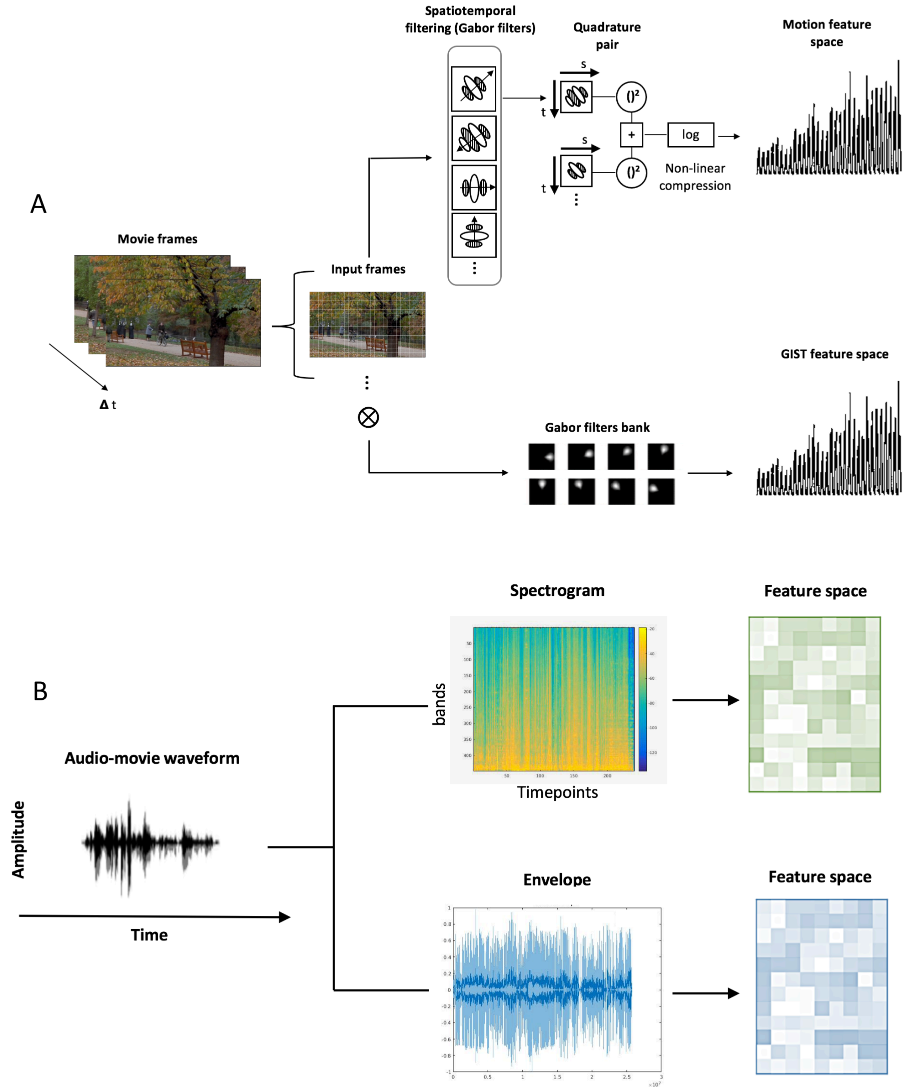


Figure 2. Low-level features modeling.
Low-level visual features extraction for visual (A) and acoustic (B) features. A. top row shows the computations for the extraction of the GIST features while the bottom row illustrates the motion energy

calculations. B. Power spectrum and envelope extraction depiction from the movie auditory track.

High-level models

Semantic feature space

Mental representations of meanings can be expressed through language. To investigate how the brain encodes and manages world conceptual knowledge, we built a proxy of the semantics of the stimulus relying on the annotations of the movie natural language. We manually transcribed the stimulus storyline recording all the spoken parts (narrator descriptions, characters' monologues and dialogues) along with the environmental sounds onomatopoeia. It's important to mention that the presence of the narrator descriptions of scene appearance, actions characters' feelings and emotions is an added value when dealing with the representation of the movie semantic content. The written text pertaining to each run was broke up into single words that were in turn paired with their correspondent timing (seconds resolution). Note that only nouns, verbs, adverbs and onomatopoeic words were spared and then used to derive the semantic representation of the story. We decided to include also onomatopoeic lexes because they directly mimic specific non-speech and non-musical sounds (produced by nature, animals or human activities) whose source is easily recognizable. Indeed, such vivid expressions are also classified as anecdotic referential words (Jean-Louis Di Santo, 2012) since they inherently *signify* what they refer to. The semantic representation of each term in the transcript was derived through word embeddings, a technique developed in the field of Natural Language Processing (Mikolov et al., 2013), in order to delineate the movie "meaning" feature space based on the context and the distributional properties of words in large corpora of text. In the semantic space, each word is vectorized and vector cosine similarity is taken as a measure of conceptual proximity. Indeed, in natural language, words with comparable meaning are often interchangeable and tend to appear in analogous linguistic contexts. Thus, to identify the movie vocabulary semantic features, we used the itWaC corpus (Baroni et al., 2009; Dell'Orletta et al., 2018). The itWaC corpus consisted of 2 billion Italian words extracted from the Web (http://wacky.sslmit.unibo.it/doku.php?id=corpora), and it was the largest Italian corpora currently available. Therefore, we used the word2vec algorithm to calculate the co-occurrence between each word in the story and a set of 128-sized word embedding space (window 5, cbow architecture). As results, we obtained for each 2 seconds interval (fMRI temporal resolution) a 128-sized vector obtained by averaging all the word vectors included in the time frame.

Categorial features spaces

Category selective regions in the brain are known to be tuned for processing specific classes of stimuli (Kanwisher & Yovel, 2006; Epstein 2008; McCandliss, Cohen, & Dehaene, 2003; Martin, 2007; Peelen and Downing, 2017; van den Hurk et al., 2017). A rich naturalistic paradigm allows us to investigate brain category selectivity across different sensory conditions (audio-video; audio-only; video-only) and sample specificities. Through a data-driven approach we aimed to estimate the contribution of the stimulus visual and auditory categorial information in explaining shared brain responses. In order to do that, from each run of the two unisensory conditions, we manually extracted the timing, at second resolution, of all the events belonging to a set of a priori defined categories. This allowed us to understand the specific contribution of non-linguistic, high-level semantic information in modulating the observed pattern of brain activity within and across groups. We relied on previous literature on the topic (Grill-Spector and Weiner, 2014) for the definition of the visual categories that, after being validated through comparison with the image segmentation output of an automatic algorithm (see Visual categorial model validation), where then applied to the auditory context too. Thus, for the visual condition we classified the continuous stream of information in seven categories: *Animals, Body-parts, Faces, Houses, Objects, Person, and Vehicles*. For every category, we annotated manually the timing of appearance of visual items on the screen writing down the start and end time (second resolution) of each event along with a brief description of the specific item presented. Notably, only the elements in foreground or those whose visual features are salient (color, size, motion) were labeled. The tagged items were used to build the correspondent categorial timecourse vector filled with binary values: 1 for marking the presence of a class entry and 0 for the remaining timepoints. This procedure allowed us to obtain a complete and detailed modeling of the movie categorial content in the form of regressors timeseries to be associated with the fMRI BOLD signal. We applied the very same tagging procedure to the auditory stimulus but instead of using all the seven visual categories, we used the notation based on five major classes derived from the validation procedure. Indeed, to allow comparability across the manual and the automatic image segmentation models, three of the original classes were collapsed together in superordinate categories: *Body-parts, Faces,* and *Person* were grouped in a unique *Whole-person* descriptor (see below Visual categorial model validation and the Appendix Supplemental Experimental procedure paragraph for further details). Note that, the two stimulation conditions intrinsically vary in the degree of detail conveyed by the diverse sensory modalities: although the presence of the narrator is meant to describe visual information through speech, the acoustic content still communicates information at a broader scale than the visual. Think about a dialogue between two characters: in the visual condition is possible to appreciate either the *Face* appearance, the *Person* silhouette, posture, clothing or a specific *Body-part* while in the auditory setting we can just say, globally speaking, that a *Person* is present (a general idea/mental representation of a man or a woman which appearance features are up to the listener). These modality-dependent aspects and the need of consistency in the methods across stimulus conditions motivated the choice of the following five auditory categories: *Animals, Houses, Objects, Person, and Vehicles*. Everyday life hearing depends on the selection of informative sounds among less-relevant background noise. Therefore, we focused our classification mainly on non-stationary foreground sounds, namely those sounds whose signal statistics change over time and result to be more informative of the world around us. However, previous work showed that the presence background environmental "noise" differentially affects primary and non-primary auditory areas responses to concurrent foreground sounds (Kell et al., 2019). Additionally, we can assume that stationary natural sounds represent a reliable source of information for blind individuals. For these reasons, we extended the classification also to the, statistically speaking, noise-like, background audio signals (nature, animals, man-made objects, human activities). For further details about the manual tagging procedure, categories description and models generation refer to the Appendix chapter (Supplemental Experimental procedure). We strongly believed that accurate annotations would help in the interpretations of observed brain patterns especially with naturalistic paradigms where stimuli cannot be selected beforehand and many features co-occur and overlap over time.

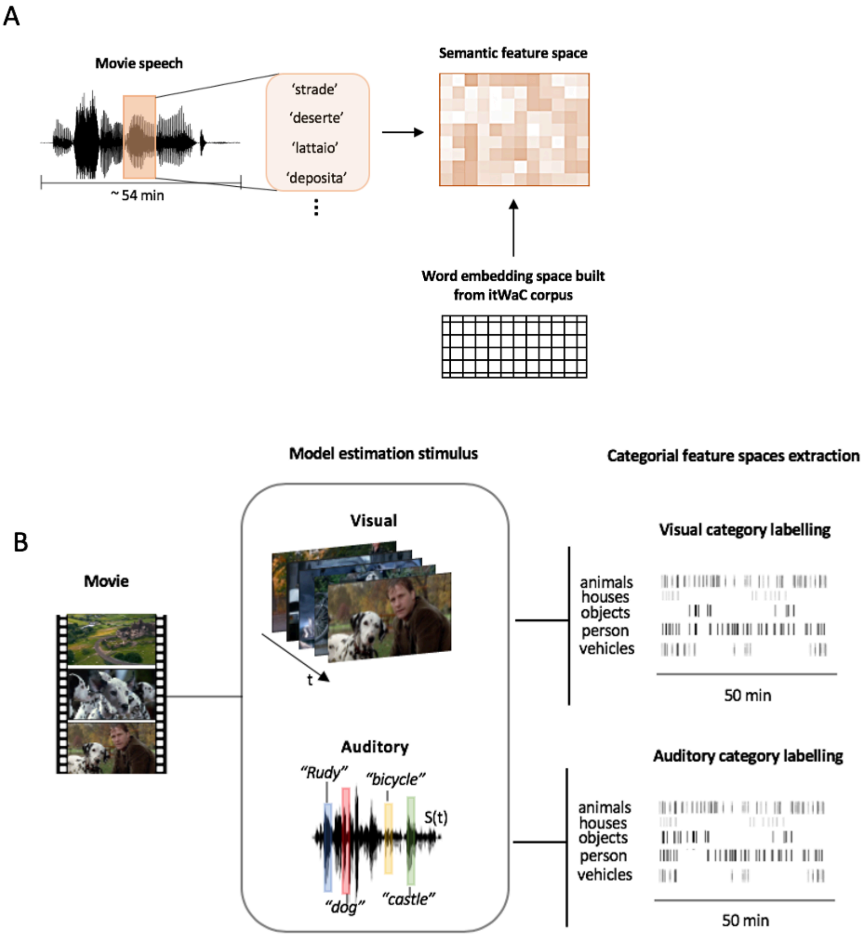


Figure 3. High-level features modeling.
A. Semantic features modelling via word embedding space generated through the word2vec algorithm. **B.** Categorical visual and auditory spaces were built from manual tagging of visual and auditory categories.

Visual categorial model validation

Although done with great accuracy, manual tagging is a time-consuming procedure that can be error prone. Hence, to ensure accuracy we checked and validated the original tagging at different consecutive moments in time with the aim to spot errors and integrate missing information. This procedure was carried out by the same person to preserve consistency in the labelling notation and avoiding the difficulties related to inter-rater labelling discrepancies. Moreover, to check the quality of the categorial tagging, we run an automatic labeling of the content of the visual scenery and test the degree of classification similarity between the two methods. We relied on a specific kind of a pre-trained convolutional neural network (CRF-RNN) (Zheng et al., 2015) to solve image segmentation and classify the elements of the visual display into five categories (*Animals, Houses, Objects, Person, Vehicles)*. In this approach, pixel-level labels are predicted combining the strenghts of Convolutional Neural Networks (CNNs) technique with Conditional Random Fields (CRFs)-based probabilistic graphical modeling. The model comprises two stages: an intial full convolutional deep network followed by a CRF-RNN step, that can be effectively used to accomplish categorial image segmentation tasks. Thus, we grouped the original classes in shared superordinate categories to allow models comparability and, at the same time, improve automatic segmentation performance reducing the percentage of errors due to misclassification (e.g. dog vs cow, where both belong to the ensemble of "animals"). For additional details about the similarity in classification performance of the two models refer to the Appendix section (Supplemental Experimental procedure).

Movie editing feature space

Movies are complex stimuli not only for the multifaceted information they convey but also because of their formal architecture, as it results from the work of the film editor. Indeed, the stylistic choices (e.g. camera's cuts selection, scenes arrangement and duration) build up the peculiar features of the movie framework that, possibly, influence brain activity. Moreover, we contributed to the process of editing as well, shortening the original duration and modifying both the auditory and the visual streams. To investigate whether these formal aspects influence movie perception, rather than focusing on the content of the stimulus, we explored the structure of the film in order to model what we called the *movie editing* features. With this term, we thus refer not only to the editor choices already present in the original version but also to the major modifications we made, namely the inclusion of the audio descriptions and subtitles. We proceeded in the very same way as for the category labeling in extracting the visual and auditory building blocks of the movie outline. Briefly, the *movie editing* features comprised the temporal definition of cuts, scenes, subtitles, text embedded in the frames, audio descriptions, music and dialogues. Detailed description of *movie-editing* features are provided in the Appendix (Additional Experimental procedure).

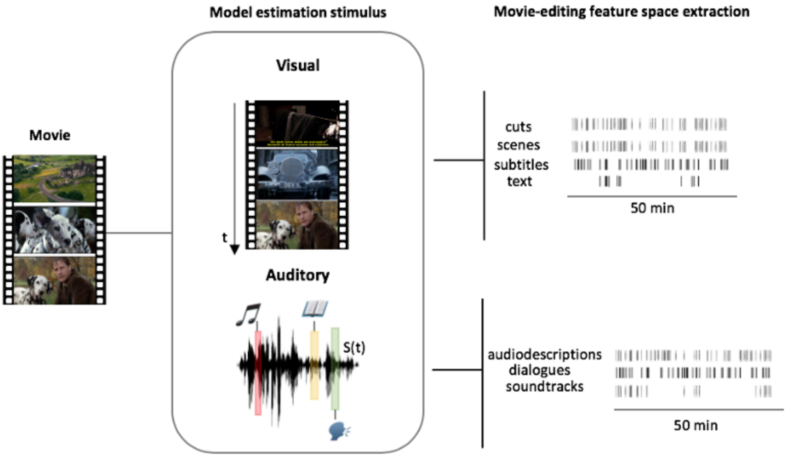


Figure 4. Movie-editing features modeling
The movie-editing features space has been created from visual and auditory properties of the movie structure (i.e, cuts, scenes, text, dialogues and soundtracks, subtitles and audio descriptions).

Feature spaces post-processing steps

We came up with eight concurrent descriptors of the movie stimulus: GIST and motion energy to model visual low-level information; power spectrum and envelope as auditory low-level descriptors, the word2vec semantic and the categorial (both visual and auditory) as high-level models, and finally the movie-editing features. Our set of low-level and high-level feature spaces provided a fine-grained characterization of stimulus representation in each experimental condition. Thus, every stimulus descriptor timecourse (features x seconds) was downsampled when necessary to match the fMRI temporal resolution. Given the different dimensionality among features, ranging from few columns for the categorial models to thousands in the motion energy one, we ran principal components (PC) analysis across each model vector dimensions in order to obtain a lower dimensional space that retains only the principal components that explains up to 90% of the total variance. This allowed us to both reduce model dimensionality and to conveniently identify a set of orthogonal components for each model.

One of the main issues when dealing with naturalistic stimulation is the collinearity across models. Indeed, this kind of paradigms does not allow to control for stimulus features independence, meaning that a certain amount of information is redundant and shared across auditory, visual streams as well as the high-level semantic representations. This collinearity issue undermines the interpretation of the role of each model as an effective predictor of brain activity, since brain activity could simply depend on portions of shared variance across different models. To take this problem into account, we decided to clean out from each model at least the portion of common variance related to the movie-editing feature space. Indeed, we reasoned that the movie-editing feature space encoded stimulus characteristics that were cross-sectional, since it described properties of the movie itself. For this reason, we orthogonalized through a multiple regression approach all the models for the movie-editing feature space. This procedure allowed us to compute the portion of “unique” variance explained by each model discarding a large percentage of common variance (Figure S1). After this procedure, we measured the residual collinearities across models by using a multiple regression analysis. Specifically we compared each pairing of models where each model acted as a predictor of the other and viceversa. To do this, we combined the multiple regression with a bootstrapping procedure (10000 iterations) to randomly sample columns from the predictor and the predicted models. Ultimately this procedure generated a predicted model which was

compared to the original one by means of R^2 to obtain a final unbiased estimation of collinearity between models of different dimensionality. Note, that after having cleaned the movie-editing features from all the other models, as expected, they still retained some degree of collinearity one with the others (Figure S2).
Finally, all the models were convoluted with a canonical gamma haemodynamic response function to account for the physiological characteristics of the BOLD signal.

Whole-brain Inter-Subject Correlation Analysis (ISC)

We computed the voxel-wise synchronization of brain activity during stimulus presentation for all the subjects belonging to the same experimental sample. Synchronicity was assessed calculating the average Pearson correlation coefficient (r) between fMRI timeseries pertaining to each pairing of subjects (Hasson et al., 2004). This procedure was repeated for every possible pairing and the resulting ISC values were averaged within each sample to obtain a group level measure of synchronicity. To test the statistical significance of the ISC values obtained, we run a non-parametric permutation test by generating surrogate voxel timeseries splicing the original data in twelve chunks which were randomly rearranged (1000 permutations). This procedure allowed to generate a null distribution which shared the same distributional parameters (e.g., mean, standard deviation) of the original data, as well as similar (but non identical) temporal dynamics (e.g., power spectrum characteristics in mid e high frequency ranges). Results were corrected for multiple comparisons using False Discovery Rate Method (FDR, $q < 0.01$).

ROI selection

Since our aim was to test the effect of each model in modulating ISC, for computational reasons, we limited these further analyses in a predefined set of regions of interest (ROIs). Thus, we selected visual, auditory and multimodal areas by means of a previously published and freely available, population-based brain atlas (Glasser et al., 2016). Taking advantage of the vast pool of multimodal magnetic resonance images from the human connectome project (HCP), the authors defined 180 brain parcels per hemisphere by the combination of different brain “areal features” such as cortical architecture, function, connectivity and topography. We defined bilateral visual, auditory and multimodal ROIs from the atlas itself, following the description of the 22 major spatially contiguous regions they provided in the paper along with their constituent parcels. This allowed us to create comprehensive ROIs including nearby pieces of cortex that share common functional properties. Hence, for the purpose of the present work, we outlined bilaterally seven visual and three auditory and one multimodal ROIs in each hemisphere separately. For the visual ROIs we selected: V1 (primary visual), early visual cortex, the dorsal and ventral stream, MT+ Complex, lateral occipital complex (LOC) plus neighboring areas and parahippocampal gyrus/sulcus (PHG). Auditory ROIs included: A1, early auditory cortex, and auditory association areas. We added also a multimodal patch of cortex mainly overlapping with the temporo-parietal-occipital junction. For a complete and detailed description of the atlas parcels names and indices included in each ROI, please refer to Table S3-S5 (Supplemental Data) and to the original publication (Glasser et al., 2016).

Model-based Inter-subject correlation (ISC) analysis

To assess the impact of each model, we developed a new algorithm which was conceptually based on mediation analysis (MacKinnon et al., 2002; Chén et al., 2017; Hayes 2017). The idea behind mediation analysis relies on the fact that a mediating factor intervenes in the relationship between the independent and the dependent variables. Here, we used each model as a mediating factor during ISC. Specifically, before computing the ISC as described below, we first removed through a multiple regression in each subject separately the model contribution in the prediction of the BOLD signal. This procedure generated an ISC value for each model which represented the residual synchronization among subjects independent from our stimulus descriptors. As example, in a voxel which showed high ISC, a model that was able to predict all its neural activity would generate a model-based ISC of zero. Thus the synchronization across subjects would critically depend on the features represented in that model. Conversely, in a voxel which showed high ISC and which retained the same ISC after the mediation analysis would be interpretable as a voxel with an elevated synchronization across subjects driven by unspecified neural activity.

This methodological approach had the main advantage to be able to statistically measure the relative contribution of models with different dimensionalities without necessary relying on a machine learning procedure which required a larger amount of available data (Huth et al, 2016) and which critically depend on cross-validation strategies (Varoquaux, 2018; Poldrack et al., 2020).For the reasons discussed above, the model-based ISC analysis was restricted to a set of predefined ROIs and to the voxels that retained a significant ISC.

To obtain a statistical measure on the mediation effects of our computational models, we performed a permutation test through the generation of 1000 null descriptors for each model. Surrogate data were created using the IAAFT procedure (Iterative Amplitude Adjusted Fourier Transformed (Schreiber et al., 1996; Lettieri et al., 2019); Chaotic System Toolbox), in order to preserve the temporal autocorrelation and spectral density features of the original models. This allowed us to have a null distribution of model-based ISC values for each voxel and model. The ISC and model-based ISC values and their null distributions were averaged across voxels in each ROI and results were then corrected for multiple comparisons using the False Discovery Rate (FDR) approach ($q < 0.05$).

Results

Whole-brain Inter-Subject Correlation Analysis (ISC)

Overall description

Whole-brain voxel-wise inter-subject correlation (ISC) analysis was computed to evaluate within-group synchronicity of brain activity over time and allow between-group comparisons. Note that the inter-subject correlation measure used here specifically means that a given region is consistently recruited across individuals under certain experimental conditions: it follows that the information about the stimulus encoded in that area is reliable for the observed group of subjects. Therefore, an ISC value greater than zero is not synonym of "increased activation" but refers to collective tuning of cortical brain fluctuations (either in the form of increased or decreased activity) to be considered in "absolute terms" without any implications about the directionality (positive or negative) of the underlying brain activity.

Results in the blind, in the deaf and in the three control samples highlighted a well-known network of regions involved in processing auditory, visual and multimodal information (voxel-wise ISC, $q<0.01$, Figure 1). As expected, exposure to the two unimodal conditions (i.e., auditory- and visual- movies) synchronized brain activity across subjects in areas traditionally defined as predominantly tuned for sounds and images processing (i.e., temporal and occipital regions respectively). Indeed, listening to the auditory movie elicits significant ISC values in the superior temporal gyrus/sulcus and angular gyrus (STG, STS, AG) both in blind and sighted controls. On the other hand, film watching (i.e., the visual condition) elicited a much more spread recruitment of the cortical mantle comprising not only the whole occipital cortex but also temporal regions (STS, STG). Moreover, both auditory and visual conditions drive synchronized activity in 'hub' areas of the Default Mode Network (DMN) such as in the posterior cingulate cortex (PCC) and the precuneus while significant ISC values in middle prefrontal cortex (mPFC) and ventrolateral prefrontal cortex (vmPFC) were more pronounced for the visual setting than the auditory one. As already reliably shown from previous literature (Hasson et al., 2008), the complete audiovisual setting, generated a dramatic, distributed recruitment of almost the entire cortical mantle, with spreading of synchronized activity across occipito-temporal and fronto-parietal areas. Direct inspection of the Figure 1 reveals that, overall, control individuals exhibited a broader recruitment of the cortical mantle than deprived individuals in both auditory and visual conditions. Indeed, in the former setting, synchronous brain activity in sighted subjects was found also in inferior frontal gyrus (IFG) and covered temporal regions more broadly than in the blind group, spreading in a rostro-caudal direction from the anterior temporal pole (ATP) to the temporo-parieto-occipital junction (TPOJ). The same phenomenon is visible in hearing subjects as well, that synchronously engaged temporal (ATP, VTC), prefrontal and frontal cortices to a larger extent than deaf people. Moreover, visual inspection of ISC maps reveals that, during naturalistic listening, blind, but not paired sighted controls, exhibited significant ISC values bilaterally in the mid-calcarine sulcus, lateral occipital and inferior occipito-temporal cortices (voxel-wise ISC, $q<0.01$, Figure 1).

Overall, deaf and hearing individuals showed strikingly similar and almost overlapping ISC maps throughout the entire brain encompassing frontal and occipital cortices and covering temporal areas as well. Indeed, while watching the mute movie, both groups significantly synchronized Heschl's gyrus, STG and STS regions.

Conjunction analysis depicts the regions reliably synchronized across different conditions: namely STS/STG and precuneus (Figure 5).

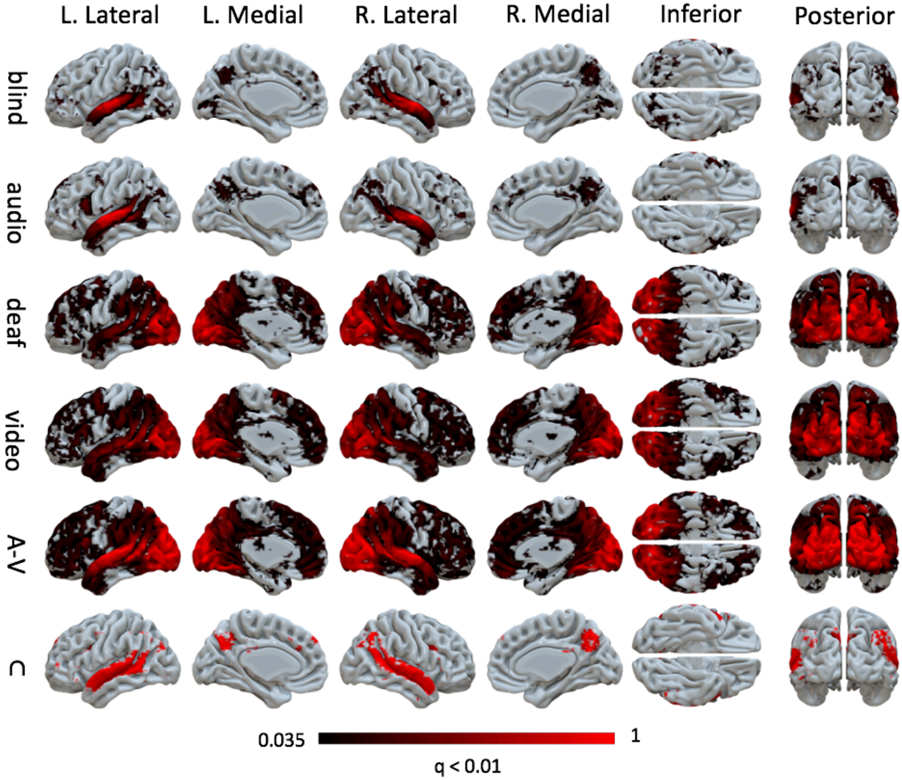


Figure 5. Whole-brain ISC (Inter-Subject Correlation)
Voxels of synchronized activity are displayed for the different groups and conditions along with the conjunction analysis. a) Progressing from the top- to the bottom- row, we can observe an increasingly greater recruitment of the cortical mantle in relation to the administered experimental condition and experimental group (i.e. unimodal auditory, unimodal visual, multimodal audiovisual). Results are corrected for multiple comparisons using False Discovery Rate method ($q < 0.01$; voxelwise permutation test; n=1614 timepoints) and mapped on the

cortical surface. Blind subjects, but not sighted controls, showed significant ISC bilaterally in the mid-calcarine sulcus, lateral and inferior occipito-temporal cortex during the auditory presentation of the movie. Deaf individuals instead, presented significant ISC values in Heschl's gyrus, STS and STG. b) The conjunction analysis highlights regions reliably recruited across the three different conditions (i.e. STS, STG and precuneus).

ROIs-based ISC and mediation analysis

ROIs selection

Since we are interested in understanding how sensory deprived cortices (i.e. visual and auditory) deal with complex naturalistic information, we focused the analysis on patches of cortex spanning the occipital and temporal areas. Indeed, this approach allowed us to test in unbiased functionally defined brain structures both the group differences on ISC and, more importantly, the effects of computational models on neural synchronization. To this aim, we relied on the HCP Atlas (Glasser et al., 2016) and we selected a set of Regions of Interest (ROIs) to cover the cortical mantle from primary visual cortex (V1) to primary auditory cortex (A1). We delineated a cortical path moving from V1 to high order visual regions along the cortical hierarchy, reaching multimodal (visual and auditory) areas and then descending to the auditory domain in the reversed order from high level regions to Heschl's gyrus (A1). Hence, we picked 11 ROIS in each hemisphere (7 visual, 1 multimodal and 3 auditory) according to the clusters description provided by Glasser et al. (2016). The visual ROIs comprise primary visual (V1) and extrastriate (V2, V3, V4) cortex responsible for low-level stimulus features processing along with the higher order visual areas listed below. Hence, among the latter, we chose the ventral stream (V8, VVC, VMV1, VMV2, VMV3, PIT Complex, FFC) recruited for stimulus classification, the phg (PHA1, PHA3, PHA2) responsible for scene representation and spatial navigation, the dorsal stream (V6, V6A, V7, IPS1, V3A, V3B) employed for objects localization in space and visually guided actions planning, the LOC (LO1, LO2, LO3, V3CD, V4t, FST, PH) involved in object recognition and the MT+ Complex (MT, MST) mainly engaged by motion perception.

For the auditory cortex, we selected three ROIs: the primary auditory (A1) and the early auditory (Mbelt, Lbelt, Pbelt, RI) areas devoted to sounds frequency and amplitude processing plus the auditory association cortex (A4, A5, STSdp, STSda, STSvp, STSva, STGa, TA2) recruited for speech comprehension and language processing. The multimodal ROI encloses TPOJ1, TPOJ2, TPOJ3, STV and PSL areas responsible for integrating multimodal inputs and engaged in several cognitive tasks involving attention, memory and emotional understanding. For further details about the ROIs included in the analysis and discussed in the following sections please refer to the Appendix (Supplemental Data, ROIs selection).

ROI- and model- based Inter Subject Correlation

Figure notation

The following part will deepen the ISC analysis at the level of single ROIs and take advantage of stimulus features modeling to evaluate the impact of visual and auditory movie properties in driving the observed group synchronizations. Since we wanted to estimate which characteristics of the stimulus modulate ISCs within and between groups, we restricted the model fitting analysis only to those voxels (for each ROI) that show ISC values significantly greater than zero at voxel level, as depicted in Figure 1 ($ISC > 0$, $q<0.01$). This choice allowed us to look just at “meaningful” voxels (those that are synchronously engaged across subjects) with the aim to assess which specific content of information, from those we account for through stimulus modeling, is responsible of tuning subject’s brain activity over time. Additionally, this kind of analysis performs group comparisons and gets interesting insights about regional stimulus processing differences across samples.

The following figures (Figures 5, 6, 7, 8) will show the results of ISC analysis at ROIs level, per hemisphere. Then, each region will be discussed according to the dominant stimulus features driving the observed group ISCs. We chose to use the polar plot representation as a way to convey both ROIs-based and model-mediated ISC for group pairings (i.e. blind and sighted; deaf and hearing) in order to provide, with a single figure, a complete overview of the main results.

The central, bigger polar plot represents the ROIs-based ISC: each colored bar refers to a specific ROI (brain surface mapping and label list in Figure 5a, 6a, 7a, 8a) arranged hierarchically and in a clockwise direction from V1 (at the top) to A1 (at the bottom) (Figure 5A, 6A, 7A, 8A). Selected ROIs are mapped on the cortical surface with a color scale going from red (V1) to blue (A1). Corresponding ROIs labels are reported in the table for greater clarity (Figure 5a, 6a, 7a, 8a). The color-coded arrow indicates how to navigate the central polar plot moving from primary visual to primary auditory cortex, region by region. Matching in color coding guides in image reading and aids the understanding.

ISC values are reported for group pairings in order to allow easy comparisons among conditions. Imagine to divide the central circular plot in two symmetric halves tracing the vertical diameter of the circle: the bars in solid colors painted on the right side refer to the deprived group (blind in Figure 5A, 6A; deaf in Figure 7A, 8A), while those drawn to the left with dashed colors relate to the non-deprived control samples (sighted in Figure 5A, 6A; hearing in Figure 7A, 8A respectively). Statistically significant differences in ROIs ISC values between groups (Deprived – Controls > $\text{Deprived}_{\text{null}}$ - $\text{Controls}_{\text{null}}$) are reported through an asterisk, whose location in the polar plot indicates the directionality of the comparison (the group showing higher ISC values). Thus, an asterisk placed in the right hemi-plot means that the deprived group exhibited greater ISC coefficients (for that specific ROI) than the matched control group, while the alternative scenario (asterisk on the left) indicates that controls synchronized significantly more that brain area during the presentation of the movie than the corresponding group of deprived individuals.

Model-based ISC is reported, for coherence, through the usage of polar plots with some minor, but crucial, differences relative to what has been explained so far concerning ROI-based ISC. All the ROIs represented in the central larger circle (Figure 5A, 6A, 7A, 8A) are now "magnified" and depicted through dedicated smaller polar plots that show the feature models driving the observed ISC for that area (Figures 5b, 6b, 7b, 8b). Thin grey lines connect the specific ROI (center of the figure) with the corresponding model-based plots on the outer part of the images (Figures 5B, 6B, 7B, 8B). Hence, we got 11 smaller polar plots that are arranged circularly around the central one and show in details the information content coded in each selected brain area (Figure 5b/B; 6b/B; 7b/B; 8b/B). Inside each graph, only significant models are reported in bold, plain colors while non-significant models are depicted in transparency. Note that the colors here, code for the specific feature model while in the central plot we used them to indicate a specific ROI. We introduced a matching in the two color legends to simplify figures reading and aid results understanding. Indeed, a given model color representation is paired as much as possible to the one used for the brain area known to process the modeled feature itself (e.g. red: V1-gist; dark blue A1-power spectrum). Experimental groups are represented using the same notation explained before: deprived groups cover the right side of the plots, non-deprived lays on the left. Asterisks report significant group differences and are placed on the same side of the group presenting greater values ($\text{Deprived}_{\text{modelmediated}} - \text{Controls}_{\text{modelmediated}} > \text{Deprived}_{\text{modelmediatednull}} - \text{Controls}_{\text{modelmediatednull}}$).

Keeping these considerations in mind, the following paragraphs will focus on the description of the results for each hemisphere and experimental conditions/groups through polar plots reading. First, the ROI-based ISC results will be explained and then we will focus in the characterization of the stimulus features driving the observed ISC in each ROI. The presented results (Pearson's r ± SE) are corrected for multiple comparisons using False Discovery Rate method (FDR; $q < 0.05$).

ROIs- and model- based ISC: blind vs sighted controls, left hemisphere.

Results revealed that blind subjects, but not sighted controls, showed significant ISC values in V1 during naturalistic listening (Figure 5A, Table 9). Direct comparison of the two groups indicates that only blind subjects synchronously recruited the left V1 while sighted individuals did not reliably engage the left primary visual cortices while exposed to the auditory movie. This result was further corroborated when considering the percentages of ISC overlap in V1 for the two groups (blind: 28% vs sighted 0%; $q < 0.01$) reported in Table S8. Additionally, ISC values in other "visual" areas, namely the extrastriate regions, ventral stream, LOC and MT/MST were significantly higher in the blind relative to the controls (Figure 5A, Table 9). Neither blind nor sighted synchronized their brain activity in the parahippocampal (phg) region. Conversely, auditory areas were synchronously recruited in both groups, provided the acoustic nature of the stimulation. However, non-deprived subjects engaged those regions significantly more than the deprived individuals during the audio-movie listening (Figure 5A; Table 9).

Model-based ISC revealed that blind synchronous activity in V1 was driven by categorial visual, movie-editing, envelope and power spectrum models, with movie-editing end envelope retaining the highest effect size. Low-level auditory models (power spectrum and envelope) along with the movie-editing model were also significantly different between groups with blind showing higher values than controls (Figure 5B; Table 17, power spectrum: $r_{(diff)}=0.0015$ $q<0.05$; envelope: $r_{(diff)}=0.0023$ $q<0.05$; movie-editing: $r_{(diff)}=0.0024$ $q<0.05$; blind>controls). ISC in extrastriate areas was guided by envelope, movie-editing and visual categorial models in blind with the movie-editing model higher in blind than controls (Figure 5A; Table 17, $r_{(diff)}$=0.0028 blind>controls, $q<0.05$). Envelope and movie-editing models modulated subjects' synchronization also in ventral stream in the blind group, and were significantly higher in deprived than non-deprived individuals (Table 17). Of note, among the other features, ISC in LOC for blind subjects was driven by the semantic (word2vec) model (Figure 5B; Table 13).

Overall, the behavior of auditory areas was fairly comparable across groups with the not surprisingly dominance of low-level acoustic features (envelope and power spectrum) and aspecific movie-editing properties in explaining the observed ISC (Figure 5B).

Main results for the left hemisphere in blind and sighted participants

Blind, but not sighted individuals, showed significant ISC in V1 (Figure 5A). This synchronization was driven mainly by sound envelope, power spectrum and movie-editing (Figure 5B). While listening to the audio-movie blind did also significantly synchronize other higher order "visual" areas such as extrastriate, ventral stream, LOC, MT/MST and the values of such tunings were significantly higher than those founded in sighted controls (Figure 5A). ISC values were significant for both groups in auditory areas but higher in controls than visually deprived individuals (Figure 5A).

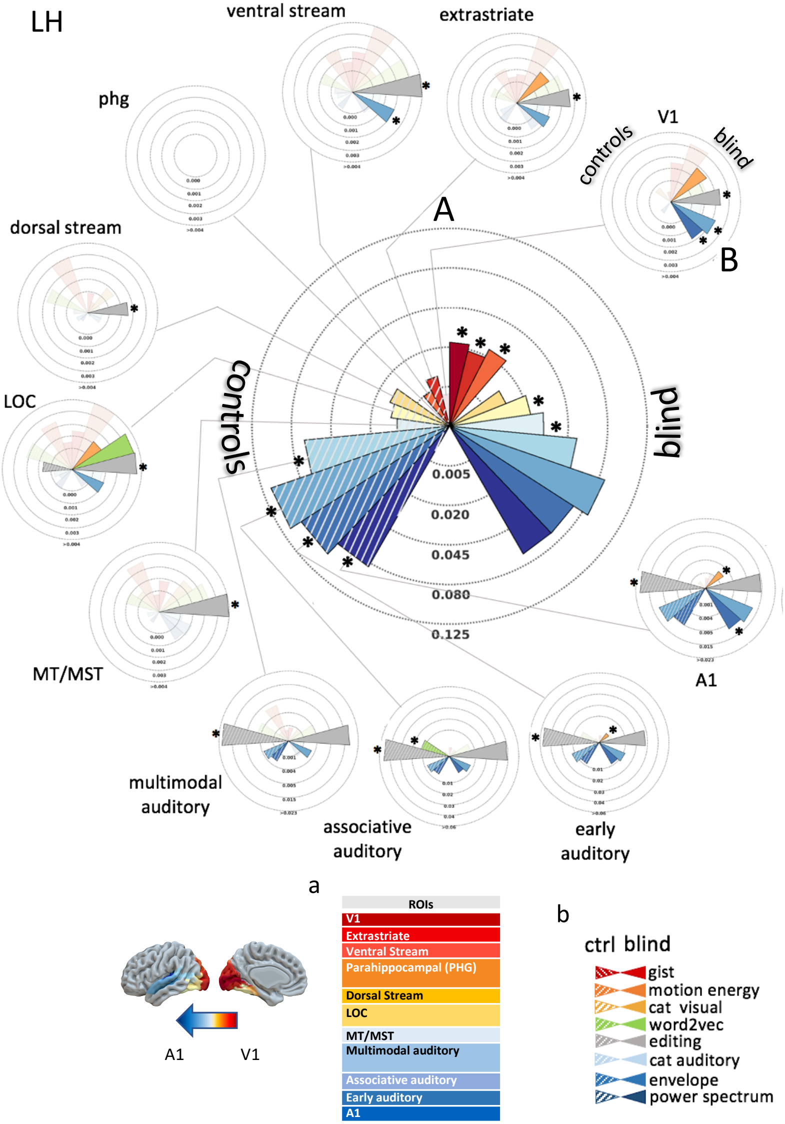


Figure 6. ROIs- and model- based ISC: blind vs sighted controls, left hemisphere.

(A) Central polar plot shows the ROI-based ISC analysis for blind (right side, plain color) and sighted controls (left side, dashed colors). Image reading proceeds from the top to the bottom in a clockwise manner going to V1 (first dark red bar) to A1 (last dark blue bar). Asterisks denote statistical differences in the magnitude of ROI-based ISCs across groups. Blind, but not sighted controls, presented significant ISC in V1 while listening to the auditory movie (ISC>0; q<0.05). Note that the corresponding bar in V1 for the sighted is missing, meaning that control subjects did not synchronize their left primary visual cortex at all while listening to the auditory stimulation. All the other ROIs, but phg, are significantly synchronized in both groups (ISC>0; q<0.05). Group comparisons reveals that blind individuals synchronously recruited visual areas (V1, extrastriate and ventral stream) significantly more than paired controls (blind>controls, q<0.05). On the other hand, sighted individuals synchronously engaged auditory and multimodal areas (A1, early auditory, associative, auditory, multimodal auditory, MT/MST) significantly more than blind subjects when processing the acoustic movie (sighted>blind, q<0.05). (**B)** Representation of model-based ISC at ROIs level: each ROI used in the ROI-based ISC analysis (large central polar plot) was further investigated to assess which specific movie properties (among those we modeled) was processed, resulting in a smaller polar plot for each ROI. Color transparency was modulated to mean statistical significance (solid color for significant models, transparent for non-significant models). Asterisks mean statistical difference among groups and their positions convey the directionality of group comparisons: they are overlaid next to the group with significantly higher values. ISC in V1 in blind subjects was mainly driven by acoustic features of the movie such as power spectrum and envelope but also by information concerning visual categories and movie-editing properties. This finding was specific for the visually deprived subjects since sighted individuals did not synchronize V1 to acoustic stimulation in the first place. The importance of sound envelope features in driving synchronous responses in the "visual" areas of the blind brain was appreciated also in extrastriate and ventral stream ROI (blind>sighted, q<0.05). The behavior of non-deprived auditory areas is fairly similar across groups. **(a**) Color coding (red to blue) of the selected ROIs: labels are listed in the table and mapped on the cortical surface (lateral and medial view). **(b)** Color coding for the features models used in the smaller polar plots mirrors the ROIs functional specificity (e.g. red is used for V1 and gist model).

ROIs- and model- based ISC: blind and sighted controls, right hemisphere

Results showed that, similarly to what observed for the left hemisphere, blind individuals synchronously engaged "visual" areas during the presentation of the naturalistic audio-movie (Figure 6A). Indeed, correlations of brain activities across subjects was significant in V1 for blind but not sighted controls (Table 9). The ISCs percentages in V1 reported in Table S8 clearly confirmed this observation (blind: 24% vs sighted: 0%). In the blind, statistically significant ISC values extend also to other "visual" ROIs comprising the nearby extrastriate region, the ventral and dorsal stream, LOC and MT/MST (Table 9). Brain activity in the parahippocampal ROI (phg) was not synchronously tuned either by blind or the sighted subjects. Auditory cortices were fully recruited by the acoustic nature of the stimulus in both deprived and non-deprived subjects (Figure 6A; Table 9). However, congenitally visual deprived people exceeded sighted controls for what concerns the magnitude of the computed ISC in "auditory" (A1, early auditory, associative auditory, multimodal auditory) and "visual" ROIs (V1, extrastriate, ventral stream, LOC, MT/MST) (Figure 6A; Table 9, blind > controls, $q < 0.05$).

Model-mediation analysis acknowledged about the content of information guiding ISC in each ROI. Synchronization in V1 in blind was driven by the low-level acoustic features of the stimulus, namely the power spectrum, the envelope, by visual categorial and movie-editing models (Figure 6B, Table 13). Specifically, the fitting in V1 of the sound envelope was significantly different across groups (Table 17, $r_{(diff)}=0.0035$, blind > controls, $q < 0.05$). Semantic model (word2vec) was significant for blind in extrastriate areas and for sighted in dorsal stream regions (Figure 6B, Table).

Similar to the left hemisphere, movie-editing model modulated ISC in each ROI. This broad effect did not refer to a specific feature of the stimulus but rather to an ensemble of different movie properties capturing slow-varying stimulus characteristics across multiple models.

Main results for the right hemisphere in blind and sighted participants

Blind individuals, but not sighted subjects, showed significant ISC in V1 while listening to the audio-movie (Figure 6A; ISC > 0, $q < 0.05$). The main stimulus feature driving this synchronization of brain timeseries was the sound envelope model whose modulation was observed also in extrastriate and ventral stream ROI (Figure 6B, Table 13, $q < 0.05$). The movie-editing model modulated ISC patterns in almost every brain area we explored (Figure 6B, Table 12-13).

Auditory regions did process, as expected, the acoustic information in both groups but blind people exhibited significantly greater ISC values than controls (Figure 6A; Table 9, 17; blind > controls, $q < 0.05$).

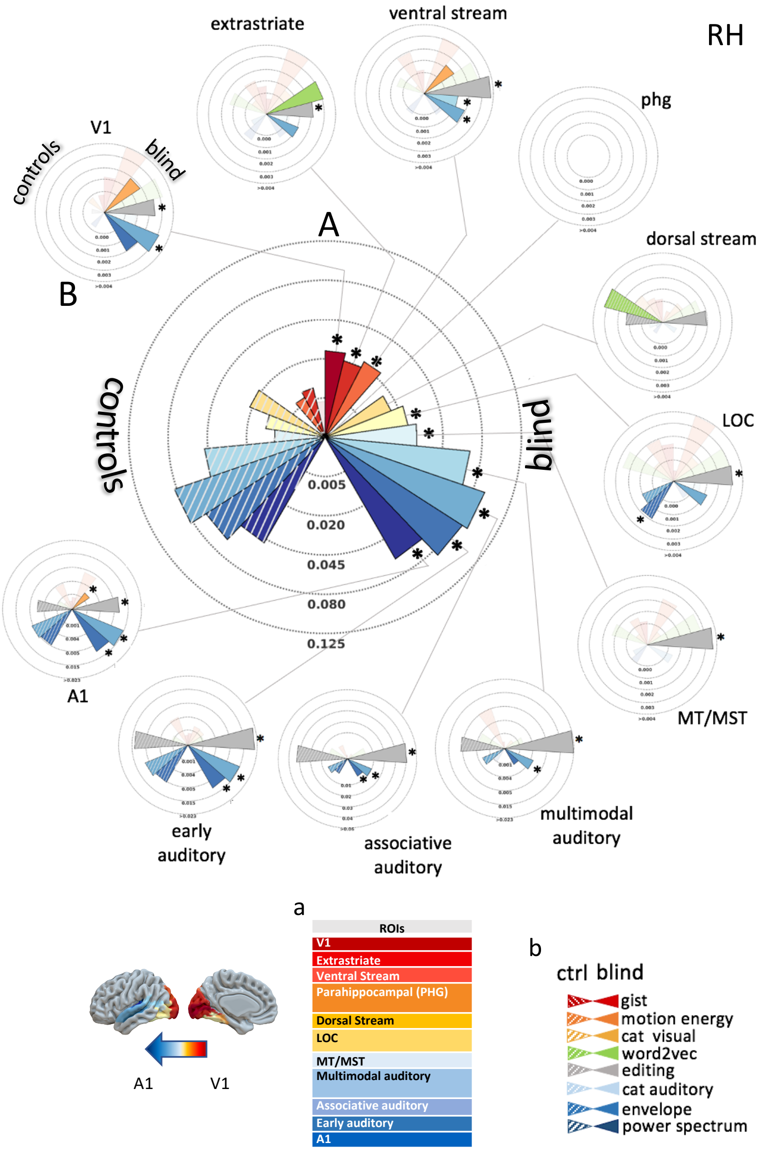


Figure 7 ROIs- and model- based ISC: blind vs sighted controls, right hemisphere.

(A) Central polar plot shows the ROI-based ISC analysis for blind (right side, plain color) and sighted controls (left side, dashed colors). Image reading proceeds from the top to the bottom in a clockwise manner going to V1 (first dark red bar) to A1 (last dark blue bar). Asterisks denote statistical differences in the magnitude of ROIs-based ISC across groups. Bind, but not sighted controls, exhibited significant ISC in V1 during movie-listening (ISC > 0, q < 0.05). Both blind and sighted subjects showed significant ISC in the other visual regions (extrastriate, ventral stream, dorsal-stream, LOC, MT/MST) and auditory areas (multimodal auditory, associative auditory, early auditory, A1) but the former showed significantly greater ISC values in all of these ROIs than the controls (blind >controls, q < 0.05). Neither blind nor sighted did synchronize the parahippocampal area (phg) while listening to the movie track.
(B) Representation of model-based ISC at ROIs level: each ROI used in the ROI-based ISC analysis (large central polar plot) was further investigated to assess which specific movie properties (among those we modeled) were responsible for the subjects' synchronization. Color transparency is modulated to mean statistical significance (solid for significant models, transparent for non-significant models). Asterisks signify statistical difference among groups and their position convey the directionality of group comparisons: they are overlaid next to the group with significantly higher values. ISC in V1 in blind was driven by power spectrum, envelope, movie-editing and visual categorial models (q < 0.05). The sound envelope features were significantly encoded also in extrastriate and ventral stream areas in the blind brain (q < 0.05). Movie-editing model was ubiquitous in all the selected ROIs (q < 0.05). (**a)** Color coding (red to blue) of the selected ROIs: labels are listed in the table and mapped on the cortical surface (lateral and medial view). **(b)** Color coding for the features models used in the smaller polar plots mirrors the ROIs functional specificity (e.g. red is used for V1 and gist model).

Overall findings for both hemispheres in blind and sighted controls

Blind subjects showed significant ISC (ISC > 0; q< 0.05) bilaterally in V1 while presented with the naturalistic sounds of the audio-movie (Figures 5A, 6A, Table 9). Synchronization of primary "visual" cortex to a non-visual stimulus (i.e., auditory movie listening) was specific for the visually deprived subjects: sighted controls did not synchronously engage V1 cortices in processing natural-like soundscapes (Figures 5A, 6A, Table 9). We observed a different hemispheric dominance in the two groups with non-deprived individuals being synchronized more strongly on the left hemisphere, while congenitally blind subjects exhibited greater brain ISC on the right side of the brain (Figures 5A and 6A, Table 17). The information encoded bilaterally in V1 and driving the synchronization across blind subjects was mainly related to low-level properties of the stimulus, namely the sound signal envelope and the power spectrum (Figures 5B, 6B, Table 12-13).

The movie-editing features drove the observed ISCs in both hemispheres for almost all the explored ROIs (Figures 5B, 6B, Table 12-13).

ROIs- and model- based ISC: deaf and hearing controls, left hemisphere

During movie watching both deaf and hearing subjects showed significant ISC values in visual, "auditory" and multimodal ROIs (Figure 7A; ISC>0; q<0.05). Noteworthy, synchronization of primary auditory cortex (A1) was significant and strikingly similar between non-deprived and deprived individuals and the two groups were not statistically different (Figure 7A, Table 10). Indeed, the percentages of ISC overlap in A1 are fairly similar across the two groups (deaf: 66% vs hearing 71%) (Table S8, ISCs>0; q<0.01). Hearing and deaf participants did synchronize multimodal auditory and associative auditory ROIs (Table 10). On the other hand, even if the two groups still retained a significant ISC, deaf exhibited greater ISC values in visual phg and MT/MST ROIs (Figure 7A, Table 10).

Moving from V1 to A1 (Figure 7a), we can appreciate the contribution of stimulus features in modulating temporal synchronization in brain activity across subjects (Figure 7B). The GIST features were well represented in V1 and extrastriate areas (Table 14-15) and became less and less processed by higher order visual regions along the cortical hierarchy. Indeed, as we departed from early visual regions the complexity of the encoded stimulus features increased: motion energy and categorial visual properties were encoded up to the multimodal auditory ROI. Note that motion energy model reached the maximum effect in modulating ISC in MT/MST in both groups (Figure 7B; Table 14-15).

Deaf individuals exhibited an important modulation in high-order and early auditory areas from visual stimulus features. Indeed, in multimodal auditory ROI they showed significantly greater values than controls for motion energy (Table 18; $r_{(diff)}$=0.0053 deaf>hearing, q<0.05) and visual categorial models (Table 18, $r_{(diff)}$=0.0018, deaf>hearing, q<0.05). The same holds for associative auditory and early auditory areas where the fitting of visual categorial model is statistically greater in deaf than in controls (Table 18, $r_{(diff)}=0.0032$, $r_{(diff)}=0.0010$ respectively; deaf>hearing, $q<0.05$). Associative auditory region did process semantic information in both groups (Figure 7B, Table 14-15). The synchronization observed in A1 for deaf and control subjects was not ascribed to any of our models, except from the movie-editing one (Figure 7B; Table 14-15) that, still, did not reflect an univocal visual or acoustic stimulus feature being made by an ensemble of different visual or acoustic properties of the movie (see Material and Methods).

Main results for the left hemisphere in deaf and hearing participants

Deaf and hearing subjects did show significant ISC in all the selected ROIs (Figure 7A). We did not observe any significant difference in ISC magnitude in (deprived) primary "auditory" cortex (A1) across groups: deprived and non-deprived individuals appeared surprisingly similar in the way they process naturalistic (spared) visual input (Figure 7A, Table 10). Synchronizations in A1 were significantly modulated by none of the individually modeled low-level and high-level acoustic and visual features but the movie-editing descriptor (Figure 7B, Table 14-15). Moreover, deaf individuals as compared to controls exhibited a higher modulation of the visual categorial model in multimodal, associative and early auditory areas.

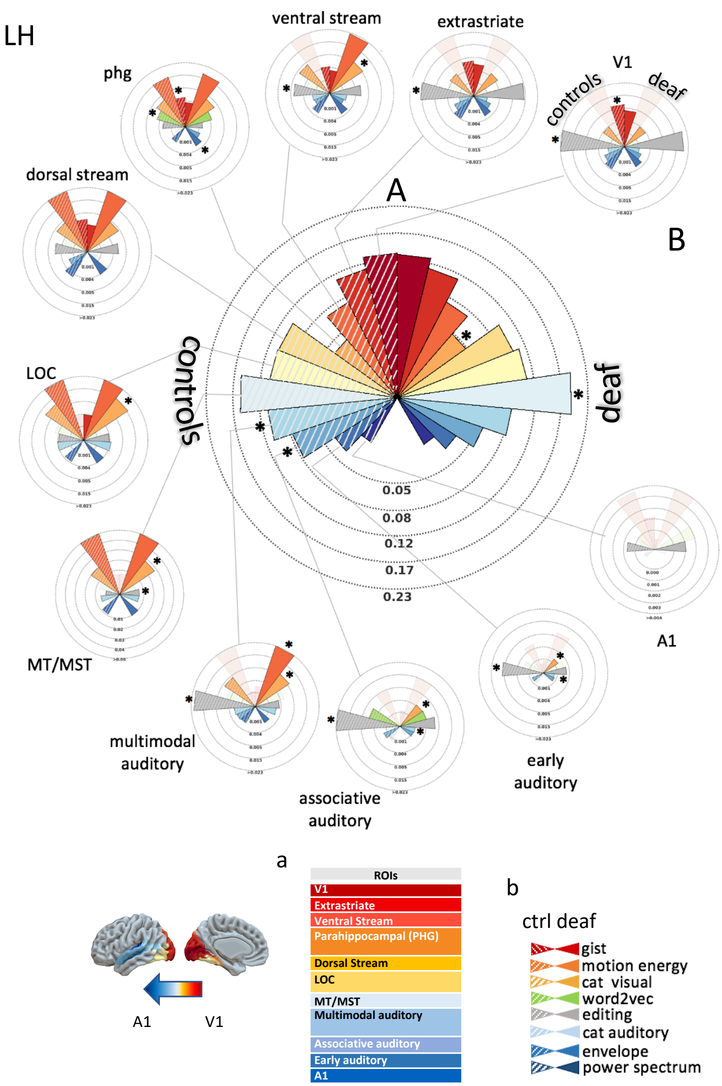


Figure 8. ROIs- and model- based ISC: deaf vs hearing controls, left hemisphere.

(A) Central polar plot shows the ROI-based ISC analysis for deaf (right side, plain color) and hearing controls (left side, dashed colors). Image reading proceeds from the top to the bottom in a clockwise manner going to V1 (first dark red bar) to A1 (last dark blue bar). Asterisks denote statistical differences in the magnitude of ROIs-based ISC across groups. Deaf and hearing controls did exhibit significant ISC values (ISC > 0, q < 0.05) for all the ROIs included in the analysis. When looking at (deprived) primary auditory cortex (A1) we observed strikingly similar synchronizations across groups (ISC > 0, q < 0.05). Hearing individuals synchronized strongly than deaf subjects in multimodal and associative auditory ROIs (controls > deaf, q < 0.05), while we observed a higher synchronization in hearing controls for dorsal stream and MT/MST regions (deaf > controls, q < 0.05). **(B)** Gist and motion energy models modulated V1, extrastriate and MT/MST areas respectively. Significant ISC in A1 were not accounted for any of our set of models except the ubiquitous and aspecific movie-editing. (a) Color coding (red to blue) of the selected ROIs: labels are listed in the table and mapped on the cortical surface (lateral and medial view). (b) Color coding for the features models used in the smaller polar plots mirrors the ROIs functional specificity (e.g. red is used for V1 and gist model).

ROIs- and model- based ISC: deaf vs hearing controls, right hemisphere

The two groups significantly synchronized all the selected ROIs while watching the movie with higher ISC values, not surprisingly, in visual areas (Figure 8A; ISC > 0, q < 0.05). Deaf subjects expressed significantly greater ISC values than controls in V1, extrastriate, dorsal stream, LOC and MT/MST regions (Figure 8A, Table 10).
Model fitting shows that, similarly to the left hemisphere, GIST features were encoded mainly in V1 and extrastriate (Table 14-15) regions while motion energy and visual categorial models were represented by higher-order areas along the visual hierarchy: namely the ventral and dorsal stream, phg, LOC, MT/MST areas (Figure 8B, Table 14-15). Movie-editing properties were able to modulate ISC in almost every ROIs in both groups (Figure 8B, Table 14-15). Noteworthy, none of the models we computed in the present study was sufficient to account for the observed ISC in A1, aside from the movie-editing, that was significant and statistically greater than in control subjects (Figure 8B).

Main results for the right hemisphere in deaf and hearing controls

Watching the visual movie significantly synchronized brain activity for both groups in all the ROIs, included the (deprived) primary auditory cortex (Figure 8A; ISC > 0; q<0.05). Deaf subjects expressed higher synchronization coefficients than controls in several ROIs: V1, extrastriate, dorsal stream, LOC, MT/MST (Table 18, deaf>controls, q<0.05) (Figure 8A). Interestingly, ISC values in A1 were comparable across deprived and non-deprived subjects (Figure 8A, Table S8). However, our stimulus modelling did not spot a particular low- or high- level movie property related to the assessed groups ISCs. In fact, only the aspecific movie-editing descriptor was significant in the deaf sample in A1 (Figure 8B).

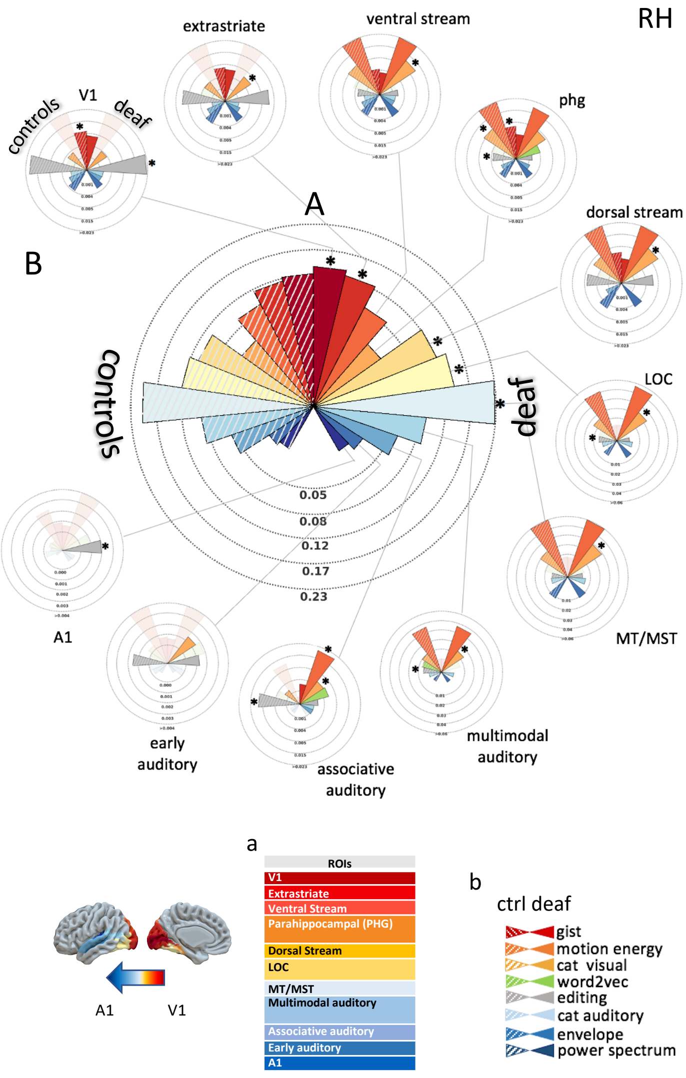


Figure 9. ROIs- and model- based ISC: deaf vs hearing controls, right hemisphere.

(A) Central polar plot shows the ROI-based ISC analysis for deaf (right side, plain color) and hearing controls (left side, dashed colors). Image reading proceeds from the top to the bottom in a clockwise manner going to V1 (first dark red bar) to A1 (last dark blue bar). Asterisks denote statistical differences in the magnitude of ROIs-based ISC across groups. Both groups showed statistically significant ISC values in visual, multimodal and auditory ROIs (ISC > 0, q<0.05). Synchronization of brain BOLD timeseries across subjects in V1, extrastriate, dorsal stream, LOC and MT/MST were higher for deaf than controls (deaf > controls, q < 0.05). ISC coefficients in V1 were not statistically different across the two groups. **(B)** Results of model modulation of ISC showed, as expected, a significant contribution of visual models in visual regions but did not provide insights on the observed A1 synchronization in both groups about the movie properties guiding such an effect, except from those conveyed through movie-editing. (**a)** Color coding (red to blue) of the selected ROIs: labels are listed in the table and mapped on the cortical surface (lateral and medial view). **(b)** Color coding for the features models used in the smaller polar plots mirrors the ROIs functional specificity (e.g. red is used for V1 and gist model).

Overall findings for both hemispheres in deaf and hearing controls

The visual movie elicited broad cortical ISC encompassing occipital and frontotemporal cortices. Indeed, deaf and control subjects significantly synchronized bilaterally visual, auditory and multimodal regions (Figures 7A and 8A; Table 10) when presented with complex naturalistic visual information. Both deprived and non-deprived individuals engaged, bilaterally, the (deafferented) primary auditory cortices while watching the movie (Table 10, ISC>0, q<0.05). Note that the magnitude of ISC values in this area was, surprisingly, fairly similar across the two groups for both hemispheres, as the percentages of ISC overlap in A1 region suggested ($\text{deaf}_{\text{left}}$: 66% vs $\text{hearing}_{\text{left}}$: 71%; $\text{deaf}_{\text{right}}$: 61% vs $\text{hearing}_{\text{right}}$: 75%; FDR, q<0.01) (Table S8).

Similarly to the evidence found in blind individuals, the group of acoustically deprived participants presented stronger synchronizations values in the right hemisphere than the controls, while the latter showed greater ISC coefficients than the deaf in the left hemisphere. Model-based mediation analysis revealed a significant modulation of our computational models bilaterally for the selected regions. When looking at the information content driving significant correlations across different subjects in A1 we found that neither sensory (visual and acoustic) nor sematic/categorial models were sufficient to explain the measured synchronization brain activity in both groups. The only reliable and robust model in primary auditory area turned out to be the movie-editing descriptor, which although significant, did not allow us to attribute this finding to a unique and unequivocal stimulus feature.

ROIs- and model- based ISC: audiovisual condition

The acquisition of the audiovisual movie in an independent sample of subjects, served as further test to measure the reliability of our ISC approach. The results of the audiovisual setting are depicted in Figure 6. Note that each polar plot here, differently from all the figures discussed before, refers just to one sample (i.e., control subjects) with the two symmetrical halves representing the two hemispheres (with the left side corresponding to the left hemisphere and viceversa).

During the presentation of the audiovisual movie subjects synchronized vast portions of the cortical mantle encompassing the (visual and auditory) sensory cortices, and expanding to fronto-temporal and parietal areas responsible for high-order computations. The visual input tuned all the visual regions along the cortical hierarchy from V1 to MT/MST area (Figure 9A). V1 and extrastriate regions encoded mainly low-level GIST features whose modulation decreased in higher-level ROIs, while motion energy was preferentially encoded in bilateral MT/MST area (Table 16). Similarly, the acoustic stimulation provided by the movie soundscape engaged the auditory cortices from A1 to the associative auditory cortex (Figure 9A). Auditory areas retained exclusively low-level acoustic features, namely the envelope and the power spectrum (Table 16).

On the other hand multimodal auditory areas modulated ISC mainly with categorial information in both visual and auditory modalities (Table 16). Movie editing properties were processed by almost every ROI examined in the study as a further demonstration of the aspecific nature of the stimulus features captured in it.

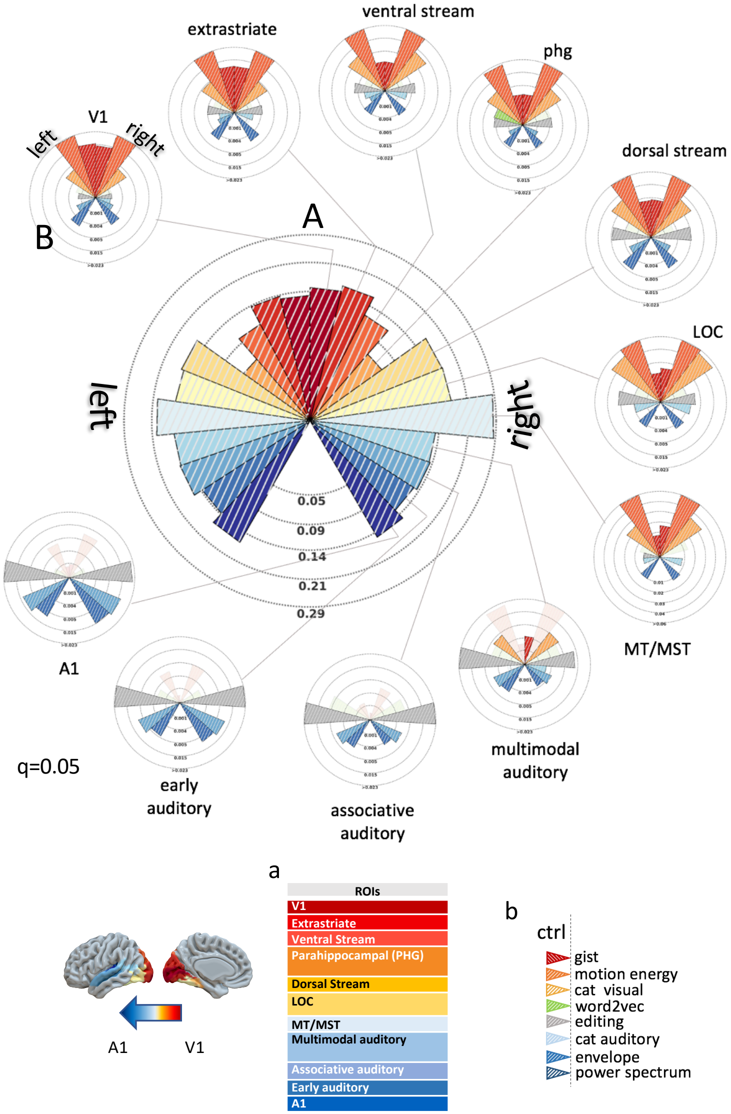


Figure 10 ROIs- and model- based ISC: audiovisual condition in left and right brain hemispheres for control subjects.

Left and right side of each polar plot depicted in (A) and (B) represent left and right brain hemisphere respectively. **(A)** Central polar plot shows the ROI-based ISC analysis for control subjects presented with the multimodal, audiovisual version of the movie. Image reading proceeds from the top to the bottom in a clockwise manner going to V1 (first dark red bar) to A1 (last dark blue bar). Complex naturalistic stimulation synchronized subjects' brain activity over visual and acoustic sensory cortices in both left and right hemispheres. **(B)** Model-based modulation of ISC showed, as expected, significant effects of visual models in visual regions and acoustic models in auditory areas according to the well-known hierarchy in information processing (i.e., from primary/early to high-order regions with the increase in stimulus features complexity). (**a)** Color coding (red to blue) of the selected ROIs: labels are listed in the table and mapped on the cortical surface (lateral and medial view). **(b)** Color coding for the features models used in the smaller polar plots mirrors the ROIs functional specificity (e.g. red is used for V1 and gist model).

Highlights of major findings

- Listening to the audio narrative evoked significant ISC in temporal areas (STS, STG, AG, precuneus) in both blind and sighted individuals
- Blind but not sighted subjects exhibited significant ISC across the occipital cortices while listening to the audio-movie (bilateral medial, lateral and inferior portions of the occipital cortex)
- Deaf and hearing subjects showed strikingly similar ISC patterns bilaterally in occipital, temporal and frontal areas. Watching the movie elicited synchronized activity in Heschl's gyrus and neighbouring cortex in both groups
- Across groups and experimental conditions, synchronization commonly occurred in STG/STS and precuneus (language network)
- Stimulus feature modelling at ROIs level allowed a finer description of the observed ISC, revealing the specific "content" of information driving synchronization
- Significant ISC in calcarine cortex (V1) for blind subjects was mainly driven by acoustic features related to the sound envelope
- None of our stimulus models (except the non-specific movie-editing) was sufficient to account for observed synchronization of primary auditory cortex (A1) in the deaf and control subjects
- The movie-editing descriptor was ubiquitous in explaining ROIs-level synchronized activity. This observation poses questions about the role of bottom-up and top-down mechanisms in modulating every-day life perception and cognition

Discussion

Here we aimed at measuring brain synchronization in two sensory deprived groups (i.e., congenitally deaf and blind individuals) and their controls, during the same naturalistic stimulation. Results demonstrated that a large extent of cortex comprising high-order, modality-independent areas mainly related to semantic processing and internal state/external environment monitoring, was commonly synchronized across sensory deprived and control individuals. This preservation of a global, large-scale brain functional organization did not include deafferented primary cortices, which indeed showed a distinctive behaviour in the systematic encoding of different stimulus features across blind and deaf individuals.

The present study aimed to investigate how congenital sensory deprivation affects the brain cortical function in real-world perception, cognition and behavior. Sensory loss from birth represents an optimal scenario to study how innate, biologically predetermined factors from one hand and experience-dependent plasticity from the other, shape brain functioning (for recent reviews on the topic, refer to the recent Special Issue 'Rethinking the sensory-deprived brain: hints from the Blind Brain Workshop 2018' on Neuroscience & Biobehavioral Reviews https://www.sciencedirect.com/journal/neuroscience-and-biobehavioral-reviews/special-issue/103T8D281NN). Indeed, congenital blindness and deafness offer the possibility to get important insights about the key principles guiding brain functional architecture through weighting the relative contribution of (lack of) experience (i.e., visual or auditory) and intrinsic, genetic heritage in carving the measurable brain cortical phenotype.

We used a naturalistic stimulation, in the form of a long movie, to convey complex, multimodal information

(Maguire et al., 2012; Lahnakoski et al., 2012, Hasson & Honey 2012) and elicit brain responses at multiple granularities in stimulus representation: from a fine-grained low-level feature processing through their integration up to the creation of high-level coherent and meaningful percepts of the surrounding environment. Experimental conditions of movie presentation via different sensory modalities (i.e. visual-movie and auditory-movie) were used to assess within-group brain activity synchrony in congenitally blind and congenitally deaf individuals, and to compare the extent and magnitude of the observed correlations in BOLD timeseries to those of sighted and hearing control subjects. Note that synchronization of brain activity in a given cortical area means nothing but the content of information is similarly encoded across the studied subjects. Hence, it represents an informative tool to highlight shared mechanisms of input processing across different individuals. However, ISC alone does not permit to identify the specific computations performed at local level. Thus, we performed a mediation analysis in order to understand the specific perceptual and cognitive processes carried out in synchronized cortices at regional (ROI) level in both deprived individuals and matched controls. This procedure "weights" the contribution of different perceptual and cognitive mechanisms and turns out to be particularly fruitful in understanding how sensory deprivation challenges encoding of natural information. The analysis proceeded from general to specific: we worked 'peeling the onion layer by layer' from whole-brain voxel-wise ISC evaluation through specific ROI-level ISC assessment, and finally digging up the content of information (i.e. stimulus features) encoded in each selected brain region and driving synchronization. This step-by-step process allowed us not only to look at shared, systematic activations across subjects and differences among deprived and non-deprived people but also to go further and identify which stimulus properties (i.e., information content) drove the observed commonalities or idiosyncrasies in natural perception and cognition.

Naturalistic stimulation synchronizes subjects' brain activity across different experimental conditions

Naturalistic stimulation drove synchronous fluctuations in subjects' brain activity in all the experimental conditions (i.e. visual, auditory and audiovisual) and groups. This finding is consistent with prior evidence revealing stimulus-locked tuning of individuals' brain responses over time for natural vision (Hasson et al., 2004, 2010; Thomas et al., 2018), spoken stories listening (Lerner et al., 2011) and narrative speech comprehension (Wilson et al., 2008). ISCs spatial cortical extent was differentially affected by the experimental condition (i.e. the version of the movie administered) and topographically specific, at least for what concerns low-level sensory areas, to the presented stimulus modality with (audio)visual-movie systematically entraining greater portions of the cortical mantle than the auditory-movie (Hasson et al., 2008). Hence, although matched as much as possible in their content, the video and audio stimuli do not present the same richness in the degree of detail provided to the listener/viewer. As a matter of fact, the amount of information conveyed through the auditory track results to be necessarily impoverished when compared with the higher perceptual load of crowded, visual frames. A lot of visually presented items, in fact, are lost in the auditory domain either because they do not produce any sounds or because they are not sufficiently relevant to be verbally reported by the narrator voice. On the other hand, the visual, silent movie loses, to some extent, the emotional coloring, emphasized in the auditory track by paralinguistic features such as the narrator prosody and the characters' nuances of voice (including pitch, loudness and timbre) or carried by environmental sound and background music (Weninger et al., 2013). Although most of the natural sounds were included into the displayed subtitles through the usage of onomatopoeias, their affective load is somehow lost in the translation from the acoustic realm into words (Bruti and Zanotti, 2017). Henceforth, the visual and auditory conditions overlap each other only superficially, in relation to the global story they convey, while differ on specific attributes of the movie itself that result to be magnified/weakened relative to the complete, multimodal audiovisual version.

Aside from modality-specific cortical recruitment, ISC maps revealed commonly synchronized, spatially distributed response patterns across all the conditions/groups bilaterally in language related areas and default mode hubs (STS, STG, AG and precuneus respectively) that, indeed, survived the conjunction analysis.

Our findings are compatible with previous evidence showing a shared, synchronous engagement of high-order regions belonging to language circuitry and DMN in complex naturalistic information encoding and retrieval (Chen et al., 2017) and spoken stories content understanding (Honey et al., 2012). Hence, conveying naturalistic content through qualitatively different modalities allowed us to isolate brain areas responsible for computing high-level, modality-invariant properties of the stimulus (Regev et al., 2013) related to attentional mechanisms, language processing and memory, ultimately leading to narrative understanding (Ferstl et al., 2008).

Recent evidence showed synchronous recruitment of DMN hubs and language network regions such as STG/STS, precuneus and PCC across people engaged in intelligible verbal communication (Stephens et al., 2010). Critically, coupling in speaker-listener brain activity patterns relates both to narrative (Dikker et al., 2014; Silbert et al., 2014) and technical content comprehension (Nguyen et al., 2020).

Studies investigating natural language representation and processing in the brain have usually conceived speech as a unitary entity since acoustic and visual signals (i.e., sounds, lip-reading and gestures) in everyday communication, generally converge and complement each-other in order to aid intelligibility (Giordano et al., 2017). However, from a behavioural point of view, the ability to understand auditory speech does not require the additional visual cues provided by lip-reading or bodily gestures interpretation (Summerfield, 1992). A phone call is just a common example of how auditory speech understanding might be independent from visual perception. On the other hand, we could provide comparable evidence for language comprehension in the deaf where the discourse is carried out proficiently without any acoustic information. To this regard, a recent study by Keitel and colleagues (2019) dissociated the acoustic and visual components of naturalistic language processing showing that while the angular gyrus (AG) and the inferior frontal gyrus (IFG) support both auditory and visual comprehension, other regions appear to be modality-specific (i.e., specifically encoding acoustic or visual information). Interestingly, acoustic speech understanding relied on brain activity in parietal and "visual" middle occipital cortex, perhaps due to the influence of top-down, feedback projections aiding the contextualization of incoming sensory information (Muckli et al., 2013; Vetter et al., 2014; Petro et al., 2017; Van Ackeren et al., 2018). However, the capability of primary and early sensory cortices to process cross-modal, content-specific information, has been proven even when controlling for the influence of top-down mechanisms (i.e., attention, imagery, spontaneous recall) suggesting that these areas, traditionally thought as unimodal, might be instead inherently multimodal (Gu et al., 2019). Yet, incoming sensory input would be encoded locally by distinct and modality-dependent activity patterns whose information content is not sufficient to allow cross-classification across senses (Gu et al., 2019). On the other hand, temporal and superior frontal regions were found to be specific for speech comprehension via lip-reading. Thus, everyday communication via speech comprehension appears to be a rather complex activity, involving the concerted recruitment of partially distinct neural networks dedicated to the processing of either acoustic and/or visual stimulus features.

Furthermore, spoken and written information provided by the movie auditory track and the subtitles respectively propagates across Deafault Mode Networks (DMN) hubs such as TPJ, AG, retrosplenial cortex, PCC and precuneus in our results. This evidence is in line with previous findings showing that default mode regions participate in spoken and written information processing via attentive mechanisms that allow the spreading of neural responses from early sensory regions to higher-order (extra)linguistic areas (Regev et al., 2018). Hence, attention gates and directs the flow of information to secondary-order areas prioritizing further processing of attended sensory input while the encoding of unattended stimuli remains confined to earlier stages of the cortical hierarchy (Regev et al., 2018). Therefore, attention modulates the activity patterns in the brain networks involved in multimodal, real-life language understanding (comprising both speech and written text comprehension) driving the routing of incoming sensory inputs across different brain areas. Furthermore, DMN nodes has been proven to support the creation of (the same) event representation in other people' brains through imagery (Zadbood et al., 2017; Chen et al., 2017). Hence, verbal communication is fundamental in bridging individual experiences and allows the creation of shared, collective representation. Therefore, real-world natural perception and imagination of complex and dynamic events are sustained, to a certain extent, by the recruitment of the same neural substrates (Chichy et al., 2012; Vetter et al., 2014) that ultimately led humans to make sense of the surrounding world. And in our protocol, these same neural substrates to represent surrounding world result to be 'immune' from sensory experience.

Although traditionally thought to support internal state/external environment monitoring rather than being engaged by specific tasks (thus the name "task-negative" network), these findings revealed that DMN, along with other high-order cortices, is also crucial in integrating information over time, constantly updating accumulated stimulus knowledge with fresh perceptual input and previous experience (enabling imagery, predictions, theory of mind - ToM). Importantly, the results of the present study fit this theoretical framework and go beyond it showing DMN and high-level semantic areas recruitment for complex naturalistic information processing in different communicative settings both containing and lacking spoken language (i.e., silent visual movie). This observation generalizes the previous evidence of speaker-listener neural coupling during orally-mediated interactions to other forms of communicative means, namely written language and paraverbal behaviors characterizing the visual and audiovisual movie conditions. Hence, we found modality-independent shared engagement of DMN and language-related (STS/STG, AG) regions across (sensory deprived and control) individuals and stimulus settings, likely subtending story encoding and understanding. These findings are in line with previous studies showing an engagement of the DMN during naturalistic stimulation (Van Praag et al., 2017) and television programs viewing (Raichle et al., 2000), and converge with the idea of its involvement in temporal integration of meaningful events in audiovisual narrative comprehension (Ames et al., 2015). Moreover, engagement of DMN regions has been associated with coherent filmic editing but not with the random presentation of (the same) shots (Anderson et al., 2006).

Not surprisingly, modality-specific responses were driven by the stimulus intrinsic low-level features as revealed by the measured brain activity synchronizations in the two unimodal conditions (i.e., visual and auditory) and in the complete audiovisual scenario. Indeed, as expected, tuning of neuro-vascular fluctuations in control subjects reflected the nature of the stimulus with temporal and occipital cortices responding to complex auditory and visual information respectively. Local, sensory-specific synchronizations were boosted during the exposure to the audiovisual stimulation that, akin to real-life contexts, requires the integration of inputs coming from multiple sources. While synchronization patterns in control individuals confirmed a well-known circuitry of areas involved in (multi)sensory processing, blindness and deafness presented specific and distinctive mechanisms of naturalistic information processing.

Naturalistic information processing in sensory deprivation: evidence from blindness and deafness

Congenitally deprived subjects did show recruitment of the deafferented cortices while processing naturalistic information conveyed through the spared modalities. However, the deprived sensory cortices (i.e., "visual" and "auditory" in blindness and deafness respectively) exhibited different behaviors when compared to the typical functioning of non-deprived, controls individuals.

While congenitally blind idiosyncratically recruited "visual" regions during natural listening of complex sounds, deaf individuals relied on deafferented "auditory" regions similarly to how hearing controls did when processing rich streams of visual information. Nevertheless, in both cases the deprived sensory cortices underwent some kind of changes (e.g., re-routing, potentiation, weakening of cortico-cortical connections) that allow to exploit otherwise "unused" neural resources for processing spared input signals. Taken together, these evidences suggest that congenital sensory deprivation drives robust and reliable plastic reorganization mechanisms in the deprived sensory cortex (Bavalier et al., 2002).

The present work, along with previous evidence (Amedi et al., 2017; Rezk et al., 2020; Ricciardi et al., 2014), challenges the conceptual framework reignign in the last decades of the last century about the organization and functioning of sensory areas conceptualized as devoted to the processing of a preferential sensory input (Felleman and van Essen, 1991) in favor of a more "systemic" and integrated view of brain functioning in natural contexts based on task-relevance rather than on information modality (e.g., "visual" areas may process non-visual information provided it is relevant for the task at hand). Anyhow, the strikingly similarity in synchronization maps between deaf and hearing subjects poses questions about the nature of such, at least global, correspondence in brain responses. Indeed, the observed ISC patterns in the two groups could possibly emerge as a fortuitous epiphenomenon somehow masking different underlying mechanisms in stimulus features encoding. We addressed this hypothesis by looking at the contribution of specific stimulus characteristics, modeled in terms of low- (i.e., visual and acoustic) and high-level (i.e., semantic and categorial) features, in driving shared responses in selected ROI covering sensory and multimodal cortices (from V1 to A1, see Supplemental Material Table S3-S5).

Stimulus features encoding at regional level in typical perception

Synchronizations of brain activity elicited in control subjects by the two unimodal conditions (i.e., visual and auditory) allowed to evaluate stimulus features modeling goodness-of-fit in each ROI. Hence, the low-level properties of the visual display computed here, namely the GIST (Oliva and Torralba, 2001) and motion energy (Nishimoto et al., 2011), drove tuning of subjects' brain oscillations predominantly in V1/extrastriate cortex and MT/MST ROIs respectively. This finding is in line with the state of the art on scene low-level statistics processing (i.e., spatial frequency and orientation) and motion perception. Indeed, previous works showed that activity in V1, V2, V3 areas (primary and extrastriate) preferentially encodes information about image spatial frequency and orientation (i.e., the so called *spatial envelope*) and that activity in those areas could be in turn used to predict and to identify responses to natural images in the ventral visual cortex (Kay et al., 2008; Naselaris et al., 2009; Andrews et al., 2015). To this regard, we found that while GIST information reduced along the visual hierarchy from V1 up to LOC region, there was a parallel increase of categorial and semantic information encoding at later stages of the visual cortical hierarchy (ventral and dorsal stream, LOC). What we have found here converges with recent evidence showing that as long as the information related to category increases going from V1 to VTC, the capacity to map stimulus shape features (ultimately emerging by GIST features) decreases (Zeman et al., 2020).

Along the visual hierarchy, we found that, as expected, the MT/MST region was specifically tuned for motion processing as already well documented in the literature on vision in both animals (Snowden et al., 1991; Zeki, 1974) and humans (Bartels et al., 2008; Durant et al., 2011; Nishimoto et al., 2011; Furlan & Smith, 2016 and Pitzalis et al., 2020).

For what concern the acoustic movie, similarly to what happens in the visual condition, encoding of sound low-level properties in control individuals (i.e., power spectrum and envelope) occurred in primary and early auditory cortices and diminished as long as we walked the cortical pathway towards multimodal and visual areas. Previous researches have already shown such gradient in acoustic information processing with primary and early auditory regions mapping low-level sound features (i.e., frequency and amplitude) and high-order regions (STS/STG) performing more abstract semantic and categorial computations (Leaver & Rauschecker, 2010; Norman-Haignere & McDermott, 2018). To this regard, evidence reliably reported that early areas of the auditory pathways are spatially organized in cochleotopic or tonotopic maps according to the spectral tuning profiles encoded locally by different populations of neurons (Formisano et al., 2003; Humphries et al., 2010; Moerel et al., 2012, 2013). Moreover, differences in spectral processing sensitivity among the two hemispheres have been systematically reported with left auditory cortex computing higher temporal rates than the homologous regions on the right side (Zatorre and Belin, 2001; Boemio et al., 2005). This preferential tuning for sounds spectral features has been argued to be a crucial factor in determining the (left-) lateralization of linguistic functions in the brain (Zatorre et al., 2002; Hickok and Poeppel, 2007). Relatively to sound amplitude modulation, previous finding demonstrated coding of speech-related envelope features in core auditory cortex (Nourski et al., 2009) and belt areas (Kubanek et al., 2013). Taken together, findings in typical sensory development, speak in favor of a hierarchically structured continuum of information processing, that supports the formation of coherent and meaningful representations of the surrounding environment. In light of current literature on the topic, with the current work, we looked at the behavior of sensory deprived cortices in processing (spared) complex visual and auditory information.

Stimulus features encoding at regional level in congenital blindness

Listening to the auditory movie generated reliable, synchronous brain activity fluctuations in primary (V1) visual cortex of congenitally blind, but not of sighted individuals. Evidence of V1 activation by acoustic information have been already reported in blindfolded subjects listening complex soundscapes (Muckly et al., 2015). Crucially, as demonstrated during typical vision (Kay et al., 2008), the authors proved that the measured activity in early visual areas was informative enough to successfully predict auditory-scene category. Importantly, stimulus-locked "resonance" in blind subjects' neuro-vascular response, occurred also in higher-order regions located on the lateral, medial and ventral aspects of the "visual" cortex (ventral and dorsal pathways, phg, LOC and MT/MST areas). The results of the present study converge with the findings of a recent publication by Loiotile et al. (2019) proving robust and systematic repurposing of visually deprived areas in congenitally blind individuals for processing naturalistic auditory information presented in the form of audio-movie clips.

Note that although we used a naturalistic auditory movie as well, both the stimulus and the theoretical approach adopted here are substantially different from previous studies. Indeed, we presented subjects with a rich, long and engaging audio-movie (about 50 min) containing realistic scenes, events and characters; thus resulting familiar to the listeners. The length and the richness of the stimulus used here, combined with a thorough stimulus modeling, allowed us to go beyond the mere observation of common tuning of brain activities over time and to question the behavior of sensory (deprived) cortices in depth. This was a crucial step since it permitted not only to appreciate the reliability and the extent of cortical repurposing in blindness (and deafness) but also, and more interestingly, to evaluate the kind of computations performed by deafferented areas on the incoming, non-typical inputs. Moreover, the meaningful, continuous flow of information provided by the movie allowed us to look at different granularities of stimulus processing levels over space (i.e., in the selected ROIs). Finally, since computational models were tuned to specific stimulus features which had characteristic time scales, from instantaneous processing of spatial and temporal information (i.e., GIST and acoustic power spectrum) to long temporal contexts (e.g., motion, acoustic envelope, semantic models), we indirectly characterized also the temporal dynamics of stimulus processing (Baldassano et al., 2017). On this regard, congenital blind retained an encoding of low-level features in early sensory areas and a mapping of acquired (modality-independent) knowledge in high-order regions.

Noteworthy, as we showed here, the observed intersubject tuning (i.e., ISC) of "visual" cortex to meaningful audio-narratives is a specific "fingerprint" of congenitally blind subjects since it was not found in sighted controls.
Crucially, we demonstrated that the shared engagement of V1 for processing structured acoustic information was driven by the envelope features of the experienced soundscape. This observation converges with the findings of a recent study by Martinelli et al. (2020) reporting the capability of the striate cortex to map, in absence of visual input, the envelope features of speech-related and environmental sounds. Moreover, these resuls are in line with previous evidence in blind individuals, showing that acoustic envelope fluctuations in human speech were processed in early visual areas (Van Ackeren et al., 2018).
In addition, correlation values while showing similarities in information processing across different subjects (i.e., engagement of V1 for processing complex, dynamic acoustic information through sound envelope features encoding) gave us a measure of reliability and robustness of the observed results. In fact, although our sample was composed exclusively by congenitally blind individuals, the socio-demographic characteristics, especially related to the etiology of sensory loss, were quite heterogeneous across individuals (for further details refer to Supplemental Material, Table S2). To this regard, the findings of the present study showed that visual areas deafferented of their typical input from birth, are systematically coopted to process spared auditory information (Figure S4), no matter the cause of deprivation. Variability in demographic and clinical variables have been representing a major concern in the field and often lead to inconsistent findings especially for what concerns the study of brain structural changes following sensory deprivation. However, in the theoretical approach

of the study presented here, individual idiosyncrasies somehow boost the robustness of the findings themselves, indicating that despite individual differences the brain systematically and reliably compensates sensory loss through cross-modal reorganization and preservation of local functional specificity.

Moreover, we observed that story listening entailed, in both visually deprived and non-deprived subjects, distributed patterns of synchronized brain activity spatially spanning areas related to auditory processing per se, as well as regions entrained by the concurrent deployment of many cognitive operations at once. Indeed, we observed the recruitment of areas devoted to language processing and comprehension (STS, STG), sustained and selective attention, working memory (for constant update of acquired knowledge with new incoming information and future events prediction), emotional processing, ToM, and "default mode" hubs (AG, and TPJ, PCC, precuneus). Taken together, these results are in line with previous evidence in the field reporting, in blind people, the recruitment of deprived visual areas by a wide gamut of task-based experimental designs using auditory and tactile stimuli (among others Wanet-Defalque et al, 1988; Buchel et al., 1998; de Volder et al., 2001; Bevalier & Neville, 2002; Gougoux et al., 2005; Collignon et al., 2011). Indeed, engagement of visually deafferented cortex has been proven for a series of perceptual tasks and mental processes crossing the borders between different cognitive domains and likely deployed in a concerted manner during daily-life interactions. To this regard, blind subjects rely on "visual" cortices while performing motion detection, sound processing (Vetter et al., 2020), memory retrieval (Amedi et al., 2003, 2004) as well as language-related operations such as Braille-reading (Sadato et al., 1996; Cohen et al., 1997) and speech comprehension (Röder et al., 2002; Amedi et al., 2003; Raz et al., 2005; Bedny et al., 2011; Lane et al., 2015). A recent publication by Vetter et al. (2020) proved that calcarine cortex in congenitally blind individuals processes natural sounds information and that brain activity recorded in this region can be successfully used for accurate sound decoding. Interestingly, the authors found an eccentricity gradient in the blind V1 similarly to what has been described in sighted individuals. This finding suggests that the typical organization of early visual cortex is preserved despite the absence of visual input from birth. However, the multimodal audio-visual recruitment of V1 in the typical brain is not new in the literature and the debate about the unisensory nature of primary visual cortex is still nowadays a matter of intense discussion in the field. Following the first observations of multimodal processing in the striate cortex of animals (Wang et al., 2008; Ibrahim et al., 2016) comparative studies in humans started to flourish. In the last decade, a lot of work showed multisensory "contamination" of primary visual areas by auditory information in typical development (Rohe and Noppeney, 2016; Mercier et al., 2013; Vetter et al., 2014). Therefore, the state of the art on the topic, along with the findings of the present study, highlighted the role of both experience and innate, genetically predetermined factors in shaping brain functional architecture. Indeed, while experience-depended plasticity allows cross-modal processes to take place (i.e., sensory deprivation prompts recruitment of deafferented cortices to process spared information), the brain functional organization is, to large extent, preserved (i.e., modality independent) and constrained by pre-existing, biologically-rooted aspects.

Stimulus features encoding at regional level in congenital deafness

During silent move watching, deaf and control individuals synchronized the cortical surface to a large extent spanning occipital, temporal and frontal regions devoted to sensory specific, perceptual and semantic processing as well as to modality-invariant cognitive computations (Hasson et al., 2008). Both hearing and deaf subjects presented systematic tuning of brain activity in bilateral A1 but, strikingly, the

calculated ISC values were not statistically different across groups. This means that the two samples comparably relied on primary auditory cortices to compute information relative to complex and dynamic natural visual scenes. Although prior evidence had shown cross-modal activation of Heschl's gyrus in deaf individuals (Scott et al., 2014), the extent by which the recruitment of temporal areas by top-down sign-language processing bewilders the effects of auditory deprivation per se, still needs to be fully understood. To this regard, a recent study by Cardin et al. (2016) showed that cross-modal plasticity in Heschl's gyrus is guided by general visuo-spatial computations rather than the specific processing of sign language gestures. However, the capability of primary auditory cortex to accommodate non-typical information through cross-modal adaptation mechanisms is still object of intense debate (Karns et al., 2012; Scott et al., 2014). On this subject, the stimulus used in the present study conjointly includes linguistic and non-linguistic content making impossible to directly evaluate, from ISCs maps alone, their differential impact on primary auditory areas. Moreover, although we aimed to unravel, through stimulus feature modelling, the types of computations carried by "host (auditory) cortices" as a result of sensory deprivation, we were not able to provide a fine characterization of the stimulus properties encoded in A1.
Indeed, we found that the recruitment of primary auditory cortex was driven only by the movie-editing descriptor, that, inherently, does not refer to a single, unequivocal stimulus feature (see Material and Methods). Hence, in the present study it includes both visual features deriving from the editor's choices and representing the so called *editing* (i.e., cuts and scenes) and the linguistic content we added at a later time (i.e., subtitles for the visual condition). Noteworthy, movie-editing properties ubiquitously influenced synchronous brain response patterns across deprived (deaf and blind) and non-deprived participants for almost all the studied ROIs. This observation suggests that

this descriptor captures some slow-changing stimulus characteristics, possibly mediated by bottom-up feed-forward mechanisms and aspecific top-down modulations (e.g., attentional and memory-related). Such hypothesis follows from the idea that movie understanding requires concurrent deployment of several perceptual and cognitive processes necessary for the creation of a meaningful and coherent representation (i.e., the story) of the action sequences unfolding over time and space.

In line with this interpretation we observed that both deprived and non-deprived viewers strongly correlated bilaterally the visual cortices and language related areas (STS, STG, AG) perhaps involved in more abstract operations, such as narrative understanding, (subtitles) reading (Hirshorn et al., 2014), visuo-spatial linguistic (e.g., lip reading) and non-linguistic processing (e.g., gestures interpretation). Previous works supports the results presented here for both deaf and hearing individuals. Indeed, evidence from sensory deprivation studies (Nishimura et al., 1999; MacSweeney et al., 2002, 2006; Cardin et al., 2013, 2016; Vachon et al., 2013; Bola et al. 2017), converge with our observations reporting a cross-modal reorganization of high-order auditory areas in processing non-acoustic information.

For what concern typical sensory processing, confirmatory findings come from studies in the field of media psychology showing that movie editing, in terms of the characteristics of video editing, affects the reliability of brain responses across subjects (Anderson et al., 2009). Indeed, the authors showed that edited silent videos, made by coherent shots, but not those built from random/scrambled visual sequences, elicited synchronous brain activity among viewers in a predominantly right-sided network of brain areas encompassing extrastriate cortices, temporal and prefrontal regions (ITG, STG, IFG) and bilateral areas in posterior cingulate cortex (PCC). Importantly, this finding has been recently demonstrated also for auditory movie, with the observation that disruption of audiodescriptions meaning is detrimental for subjects' brain activity synchronization (Loiotile et al., 2019). Taken together, along with the results of the present work, the movie editing seems to orchestrate spatially distributed synchronizations of different brain areas recruited in a hierarchical fashion according to their response latency profiles of information processing (Hasson et al., 2008; Overath et al., 2008; Baldassano et al., 2017; Santoro et al., 2014; Berezutskaya et al., 2020). An alternative possibility to explain the behavior of primary auditory areas we found here, is simply that we failed to model some stimulus features that could potentially be relevant in driving brain fluctuations across deprived and non-deprived subjects. However, basing on the existing literature on visual (Kamitani and Tong 2005; Oliva and Torralba, 2006; Miyawaki et al., 2008; Nishimoto et al., 2011), acoustic (Chi et al., 2005; Giraud et al., 2000; Martinelli et al., 2020) and semantic (Kanwisher et al., 1997; Haxby et al., 2001; Tsao et al., 2006; Peelen et al., 2009; Bartels and Zeki 2004; Hasson et al., 2004, 2008) stimulus modeling, we believe that our panel of computational and manual models is sufficiently inclusive to grasp the major stimulus properties encoded in the brain patterns of activity.

Limitations

One of the main limits of the present study is the small sample size of blind and deaf groups (n=9). However many studies in the literature on sensory deprivation had small sample sizes (Lewis et al., 2010; Watkins et al., 2013; Bedny et al., 2011; Collignon et al., 2011; Kim et al., 2014) due to the difficulties related to participant recruitment. Moreover, we acknowledge that deaf participants were not matched for age with the other experimental groups, with deaf individuals being on average, younger than the others (deaf ≅ 24±4 years; blind ≅ 44±14 years; controls ≅ 37±14 years). Another possible weakness of the research resides in the fact that, due to the fact that all the groups wore the fMRI-compatible goggles, we were not able to track whether participant eyes were open or closed and, more importantly, their changes in vigilance or alertness. However, the results of the survey performed at the end of the fMRI sessions reveled on average an high accuracy in responding to movie-related questions (mean score=15,4±1,8; lowest score: 10/18, highest score: 18/18). Finally, we have to mention that subtitles and lip-reading did not match since they rely on two different languages (i.e., Italian for subtiles and English for acting). Nevertheless, selective attention span hardly allows to focus on both features at once: it's extremely difficult to read the text and look at mouth movements at the same time.

Conclusions

Overall, the findings presented here, showed a preserved large-scale organization of sensory deprived cortices both in blind and deaf individuals exposed to spared naturalistic information. However, deafferented primary "visual" (V1) and auditory areas (A1) presented different behavior in processing non-typical inputs. Indeed, while the reliable, cross-modal recruitment of V1 for processing the envelope features of natural sounds is a specific characteristic of the (congenital) "blind brain", we did not find an analogous signature in the deaf primary auditory cortex for processing continuous flows of visual scenes. Indeed, we were not able to spell out the details about the content of information encoded in deaf and hearing primary auditory cortex. Indeed, primary auditory cortices appeared more susceptible than visual cortices to top-down and bottom-up processes concurring to shape perception and cognition in natural settings. Future investigations should address this point with different naturalistic stimuli and/or trying to devise a more complete modelling of stimulus properties.

Appendix

List of Supplementary Figures

List of Supplementary Tables

Supplemental Experimental procedures

Methods S1. Manual annotation of sounds, visual scenery and speech in the movie 101 Dalmatians

Introduction

Naturalistic stimulation has been gaining a lot of attention in these years in neuroscientific research given that it permits to reproduce common, real-life scenarios through the faithful depiction of characters, object, environments and actions people experience in daily living. This comes with two major benefits: it allows to study a broad range of cognitive functions within the very same experiment in a more ecological way and it induces reproducible and robust synchronous brain activity across individuals (Hasson, 2004). Indeed, naturalistic protocols make use of temporally continuous stimuli such as movies or audio descriptions to convey complex, structured information which processing spreads from local circuitries in charge of low-level stimulus analysis to distributed patterns of brain activity recruited for high level computations. Therefore, conversely to conventional paradigms with simplistic artificial stimuli, movies are a powerful tool to observe brain activity into a rich multisensory spatiotemporal context. From the other hand, increased stimulus complexity comes with unavoidable costs related to the difficulty to control for regressors of interest collinearity, thus posing questions about the interpretations of the observed brain activity. One way to deal with this problem, is to complement fMRI data with time-locked annotations of stimulus features as a valid support in results reading (Häusler & Hanke, 2019).

Materials and methods

Stimulus

The stimulus used for the study is a shortened version (approximately 55 minutes) of the live-action movie 101 Dalmatians (Stephen Herek, Walt Disney Pictures & Great Oaks Entertainment, 1996) paired with the Italian audio description voiced by a native speaker actor with extensive experience in the field of dubbing. The movie occurs in three different versions: the *audio-visual* implying a multimodal scenario and two single modality conditions (*auditory-only and visual-only*). In the latter case, the video is enriched by subtitles that match the auditory track (comprising voice-over, dialogues and environmental sounds).

Annotations content

The annotations we present here are meant to complete the 101-dataset with additional information about the stimulus that can be helpful for other research availing the very same dataset. Given the richness of a naturalistic stimulus, it has been possible to benefit its complexity to investigate many different aspects of multi-sensory processing. For this reason, tagging has been carried out for a broad spectrum of movie characteristics ranging from the annotation of the visual and auditory features to the analysis of the semantic content of speech. We manually labeled several visual categories largely exploited in the field of vision research comprising animals, body parts (hands or feet in isolation), faces, houses, objects, person (human bodies/silhouettes), subtitles, text, vehicles (See Visual categories). Moreover, we traced the temporal sequence of scenes and camera's cuts that mark the rhythm of the story unfolding. We took note of the exact timing, second by second (given the low temporal resolution of the BOLD signal), in which an element of the visual field pertaining to the above-mentioned classes shows up on the screen. In this way, each category annotation consists of a timeserie defined by a sequence of starting and ending points that mark the visual appearance and the duration of a specific event of interest. We applied the very same procedure to the auditory soundscape of the movie, writing down the timing (onset, offset) of audio descriptions, dialogues and soundtracks (See Auditory Categories).

Annotation procedure

One of the authors (FS) of this study carried out the tagging procedure.

The stimulus was explored run by run, and the relevant features were manually annotated along with their timecode at the time resolution of one second. To increase accuracy and preserve consistency in the annotation, the tagging procedure was repeated a second time, later, by the same person with the aim to spot flaws and integrate possible gaps. This multi-step process worked also as validation procedure of the annotation itself. Other than this, to test the goodness of the visual manual tagging, we performed also an automatic classification based on a convolutional neural network algorithm (see Visual categorial model validation).

Visual category annotations

For simplicity and considering the sluggishness of the fMRI signal in time, instead of working at the resolution of frames, we treated the movie as a sequence of still images of one second. For each image the observer wrote down the elements located in foreground and when present, also additional items that for their appearance (color, size, change in position) or story relevance (key information, main characters) capture the viewer attention. Each entry is then properly classified according to the specific category it belongs to (see Visual categories).

Category legend

We made a separate annotation table for every run the movie. Each text file contains the timestamps of event occurrence per category and a brief description of the specific item appearing on the visual screen. Thus, the first two columns refer to the *Start* time and *End* time of the event of interest respectively, while the third column contains a label for a detailed *description* of the current item (Table 1). This allows not only to track the category at hand but also to provide an accurate and exhaustive report of which specific element of that class appears on the screen along with its time of occurrence. In this way, every row contains the full characterization of the stimulus by its visual and auditory content (see Visual categories and Auditory categories).

Visual categories

We classified depicted items according to seven categories that capture the main features of the movie visual scenery event space.

Animals: all the different kinds and species of animals of the story were classified here.
Body-parts: it applied only to a given part of the body appearing in *isolation and foreground*, except from faces. A leg, toe or hand were example of body-parts.
Faces: all the faces represented *in foreground* regardless viewpoint and lighting. Faces visible from distance and in presence of the trunk or the entire body were not considered falling in this category (see person).
***Houses*:** it referred both to houses façade in isolation and groups of edifices (cityscape). It comprised also other kinds of structures that are not "houses" in a narrow sense but still, pertained to the more general concept of "buildings" (e.g. a farm, a castle).
Objects: it represented man-made objects and tools.
Person: it included the images in which the complete silhouette of the body was visible, or the head and the upper body. Faces in isolation were not falling to this category.

Vehicles: this category was meant to include all "means of transportation" represented (for example a car, a bicycle, a truck) or parts of them sufficiently big and detailed to be recognized as pertaining to a specific kind of vehicle (i.e. a bicycle handlebar or tire; a car hood).

RUN1: Faces		
Start	**End**	**Description**
00:01:01	00:01:10	Rudy's face
00:02:00	00:02:01	Anita's face
00:02:03	00:02:12	TV journalist's face
00:02:17	00:02:20	Anita's face
00:03:06	00:03:09	Rudy's face
00:03:11	00:03:13	Rudy's face
00:03:17	00:03:21	Rudy's face
00:03:58	00:03:59	Butler face
00:04:03	00:04:04	Butler face
00:04:19	00:04:23	Alonso's face

Table S1. Manual tagging
Example of annotations for the first run relative to the visual category of Faces. Each row represents an event: it contains information about the timecode along with the description of the specific item appearing on the screen.

Visual categorial model validation

To check the quality of the categorial tagging, we ran an automatic labeling of the content of the visual scenery and test the degree of classification similarity between the two methods. We relied on a specific kind of convolutional neural network (CRF-RNN) (Zheng et al., 2015) to solve image segmentation and classify the elements of the visual display into categories. Since the neural net labels slightly differred from our manual classes, we grouped some of them to allow within category classification comparability across approaches. For what concerns the manual labels we aggregated person, faces and body-parts in a single descriptor called *whole-person* to adjust for the absence of such finer distinctions in the neural network classes (where *person* referred all the previous). On the contrary, regarding the net labels we put together those referring to the same superordinate category: airplane, bicycle, boat, bus, car, motorbike, train into *vehicles*; bird, cat, cow, dog, horse, sheep into *animals*; bottle, chair, dining table, potted plant, sofa, tv monitor into *objects* (Table 2). We then used the classification output to generate category regressors in the very same way we did for the manual ones (see Regressors generation).

Table S2. Manual labels and neural net label correspondence

Manual Labels	animals	objects	vehicles	whole person
Neural Net Labels	bird, cat, cow, dog, horse, sheep	bottle, chair, dining table, potted plant, sofa, tv monitor	airplane, bicycle, boat, bus, car, motorbike, train	person

To look at the coherence between the results of the two classification methods, we computed the rank correlation between regressors generated by the automatic image segmentation algorithm with those obtained by the manual labelling. Here, the Spearman's ρ coefficients for the four categories with the 95% confidence interval (CI): *animals* (+0.51878, CI: +0.47665 +0.55913), *objects* (+0.15358, CI: +0.09820 +0.21174), *vehicles* (+0.38455, CI: +0.32185 +0.44218); *whole person* (+0.61654, CI: +0.58084 +0.65056) (Figure S1).

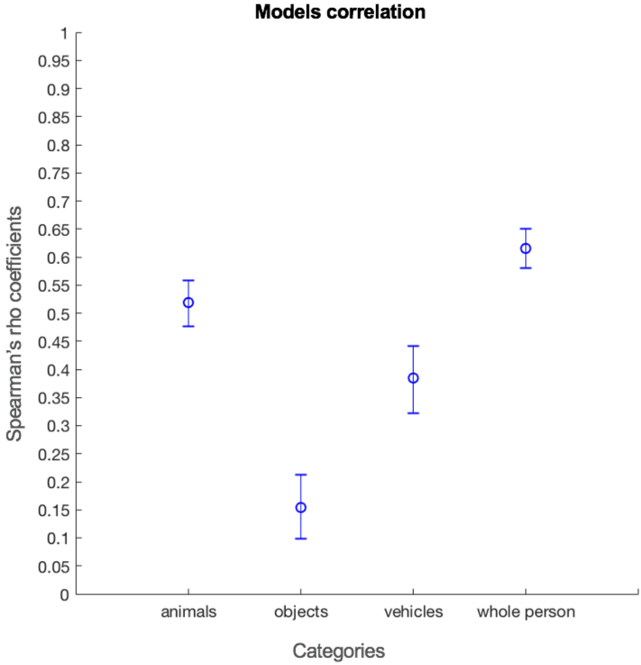


Figure S1. Automatic and manual tagging coherence
Correlation values (Spearman's ρ) between automatic classification and manual tagging. Correlation coefficients are represented with 95% CI.

Auditory category annotations

We described the categorial content of the auditory movie following the same method we applied for the visual stimulus. Analogously to the visual manual tagging, we sampled the auditory track of each run second by second and we manually labeled the foreground and salient sounds according to the category they pertained. Taking in mind the intrinsic differences among the two kinds of sensory stimulations but aiming to be consistent across models, we used here the same categories we used for the labeling of the visual movie. Therefore, we classified each sound as belonging to one of the following clusters: *Animals, Houses, Person, Objects* or *Vehicles*. The material was organized as explained in the Category legend paragraph.

Auditory categories

Below, a description of the inclusion criteria for each auditory class.

Animals: all the different kind of animal sounds which were audible in foreground or still clearly recognizable from the background audio track.
Houses: it referred to the descriptions of houses, buildings or cityscape appearance. Usually made by the narrator voice, such portrayals made explicit reference to the presence of a building (not necessarily "houses" strictly speaking). Examples are: "in front of Anita's house", "outside, from the castle gate".
Objects: it collected overt reference to objects and sounds that were easily acknowledged and attributed to man-made objects or tools. Examples are the bell ringing, the sound of the shower, the teacups chinking.
Person: the presence of a person was mainly denoted by characters speech or dialogues. Moreover, this category applied to any moment in time the narrator speaks of a person and to those sounds that could be unmistakably

attributed to a human being (e.g. sound of footsteps, cough, background chattering, screaming).

Vehicles: this category was meant to include all the auditory features related to vehicles: sounds, descriptions of vehicles by the narrator voice or characters speech. All the onomatopoeic sounds recalling a "means of transportation" were included in this category (e.g. "wrooooom", "beep", "slam", "screech").

Movie editing features

Other than encoding the visual and auditory categories we wanted to take into account the technical aspects of the movie itself and try to understand their contribution in shaping the observed brain activity. For this purpose, rather than focusing on the content of the stimulus, we explored the structure of the film in order to model what we called the *movie editing* features. With this term we referred not only to the editor choices already present in the original version but also to the modifications we made, namely the inclusion of the audio descriptions and subtitles. We proceeded in the very same way as for the category labeling (see Annotation procedure) in extracting the visual and auditory building blocks of the movie outline.

Visual movie editing features

Indeed, movie editing did not reduce to simple cutting and pasting of footage, since the editor manipulated shots comprising visual, auditory and sound track information into coherent sequences with a temporal and spatial continuity. This procedure generated the pacing of narrative unfolding in the movie. At the same time, film editing generated collinearities among the visual and auditory streams, as well as with the high-level semantic information of the movie script.

Hence, we modeled film editing features to represent movie structure. On our side, we did an editing procedure too, that in the case of the visual stimulus consisted in the inclusion of subtitles. Thus, we also marked the timecode of subtitles appearance on the screen since they are a salient feature in the visual display. Aside from that, we wanted also to account for the presence of written text *in* the movie itself, by creating a specific class called "text" (see below Visual movie editing classes).

Visual movie editing classes

Cuts: this term is used to define sudden changes in camera angle, location, placement from one shot to the following. These events occurred frequently during the narration and can be easily spotted at a glance.

Scenes: this label referred to the major changes in the story setting (location, characters, actions and time). Thus, we considered a scene as a story unit that took place in a specific location and in a defined period of time. Compared with cuts these events happen on a slower timescale.

Subtitles: it reflected the story script including all the spoken parts: voice-over, dialogues and environmental sound (mostly animal sounds).

Text: all the written text readable on the screen *and* belonging to the original movie (it did not include the subtitles that we added afterwards to the only-video version).

Auditory movie-editing features

We manually scored three auditory classes that reflected the editing work done for the movie soundscape: audio description, dialogues and soundtracks. In the case of dialogues, we just took note of the starting and ending points of speeches among characters while for soundtracks, we marked all the parts accompanied by background music. We applied the same procedure we used for the categorial features: thus, for each run separately, we took note of the timecode of each event of interest and we matched it with the relative content description. Regarding audio description and dialogues the word "description" indicated the transcript of the narrator script and characters conversations respectively. For what concerns soundtracks, we tried to associate each piece of music with a meaningful title based on the story context in which it was played. For this purpose, we took the soundtracks list of the original movie and we adapted it to our shortened version. All three auditory labels applied to events that, for obvious reasons, were never punctual as it could happen for visual features (one second of permanence on the screen was sufficient for a visual item to be tagged) but rather continuous and prolonged in time.

Auditory movie-editing classes

Audio description: all the parts of scripts spoken by the narrator voice-over. These descriptions were meant to convey the salient aspects of the movie that cannot be inferred solely by listening to the original audio track (but that were generally communicated with the aid of the video). Therefore, they contained mostly scene descriptions and characters actions/emotional state depiction. Note that this category did not include dialogues, environmental sounds and music.

Dialogues: it refers only to the speech related part of the movie script. This includeed both conversations, and "monologues" (i.e. the journalist reporting news to the spectators, the priest speaking to the audience in a church, a character "talking" to an animal).

Soundtracks: background music tracks.

Methods S2. Stimulus modelling

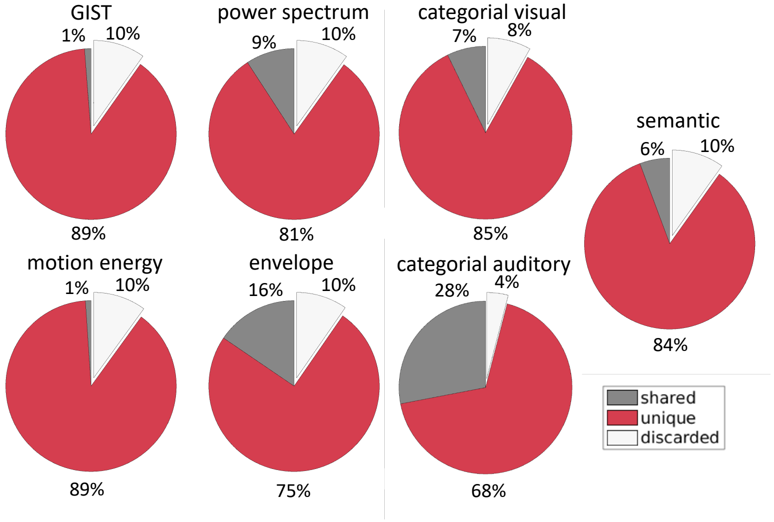
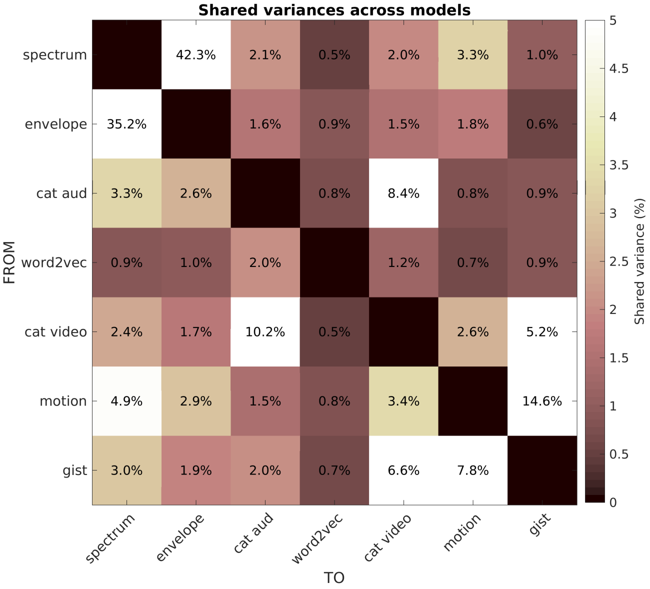


Figure S2. Pie charts representing models explained variance.
Each pie-chart depicts the percentage of variance accounted for by each model. Color-coded areas represent the amount of discarded (white) (about 10%), shared (grey) and unique (red) variance. The shared variance is the percentage of common variance between each model and the movie-editing model while unique variance is what remains after models have been cleaned from movie-editing. This unique portion is the one used to measure model-based ISC. Models were low-level visual (GIST, motion energy), auditory (power spectrum, envelope), high-level visual and auditory models (categorial visual, categorial auditory and semantic).



FROM \ TO	spectrum	envelope	cat aud	word2vec	cat video	motion	gist
spectrum		42.3%	2.1%	0.5%	2.0%	3.3%	1.0%
envelope	35.2%		1.6%	0.9%	1.5%	1.8%	0.6%
cat aud	3.3%	2.6%		0.8%	8.4%	0.8%	0.9%
word2vec	0.9%	1.0%	2.0%		1.2%	0.7%	0.9%
cat video	2.4%	1.7%	10.2%	0.5%		2.6%	5.2%
motion	4.9%	2.9%	1.5%	0.8%	3.4%		14.6%
gist	3.0%	1.9%	2.0%	0.7%	6.6%	7.8%	

Figure S3. Models collinearities.
Matrix showing the percentage of shared variance for pairings of models.. Reading from rows to columns (FROM-TO), percentages reflect the degree of collinearity across models after each one has been cleaned from the movie-editing features.

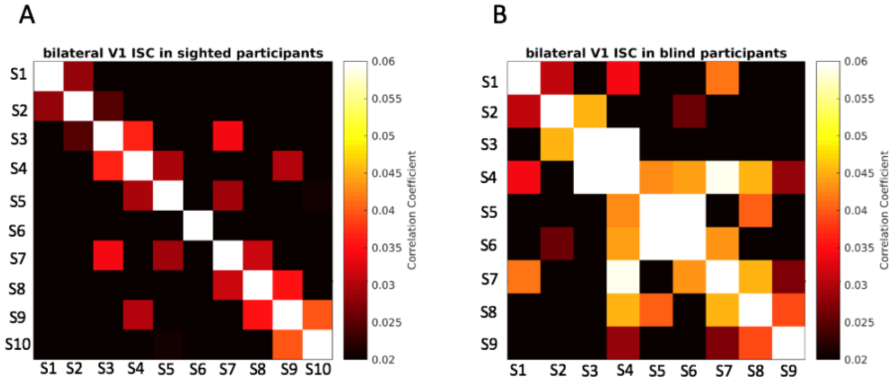


Figure S4. Single subject correlations in bilateral V1.
Correlations of brain activity in V1 for control subjects in panel A and for blind individuals in panel B. Subjects are listed in rows and columns. Blind, but not sighted, showed higher synchronization values while listening to the movie.

Supplemental Sample Demographics

Table S3. Characteristics of deaf participants

Subject	**Age**	**Gender**	**Hearing loss age**	**Cause of deafness**	**First language Sign/Oral**	**Hearing aid use**	**Residual hearing ability**
ID1 MS	24	M	congenital	Hereditary deafness	Sign	Used in the past (Childhood)	Hears only vibrations
ID2 SB	21	F	congenital	Hereditary deafness	Sign +oral	Used in the past (Childhood)	Hears only very loud sounds (through vibrations)
ID3 LB	24	M	congenital	Hereditary deafness	Sign	Used in the past (Childhood)	Hears only vibrations
ID4 SB	26	M	congenital	Hereditary deafness	Sign	Used in the past (Childhood)	Hears only vibrations
ID5 SD	22	M	congenital	Hereditary deafness	Sign	Used in the past (Childhood)	Hears only vibrations

ID9 GC	**ID8 LC**	**ID7 JV**	**ID6 CF**
22	32	28	18
F	F	F	F
congenital	congenital	1 year of age	congenital
Hereditary deafness	Hereditary deafness	Sensorineural hearing loss caused by high fever	Hereditary deafness
Sign	Sign	Sign	Sign
Used in the past	Used in the past (until 17 years old)	Uses currently	Used in the past
Hears only very loud sounds (through vibrations)	Hears only very loud sounds (through vibrations)	Hears only very loud sounds (through vibrations)	Hears only vibrations

Table S4. Characteristics of blind participants

Subject	**Age**	**Gender**	**Vision loss age**	**Cause of blindness**	**Residual light perception**	**Age of Braille reading learning**	**Handedness**
ID1 NP	69	F	congenital	Optic nerve atrophy	NLP	6	Right-handed
ID2 FC	37	F	congenital	Leber Congenital amaurosis (LCA)	NLP	6	Right-handed
ID3 AP	50	M	congenital	Optic nerve atrophy	NLP	6	Right-handed
ID4 AR	32	M	congenital	Retinal detachment	NLP	6	Right-handed
ID5 GA	43	M	congenital	Retinitis pigmentosa	NLP	6	Right-handed
ID6 AG	50	M	congenital	Retinopathy of prematurity (hyperbaric oxygen too high)	NLP	6	Right-handed

ID7 AP	19	M	congenital	Retinal detachment	NLP	6	Right-handed
ID8 EB	41	F	congenital	Bilateral retinoblastoma	NLP	6	Right-handed
ID9 BB	57	M	congenital	Retinopathy of prematurity (hyperbaric oxygen too high)	NLP	6	Right-handed

*NLP: No Light Perception

Supplemental Data

Regions of interest (ROIs) definition

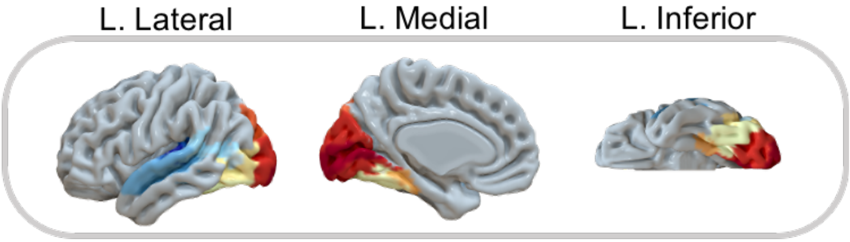


Figure S5. ROIs definition.
ROIs were extracted following the HCP Atlas parcellization (Glasser et al., 2016). Parcels (as defined by the authors) of each ROIs are listed below (Table S3, S4, S5).

Table S5. Visual ROIs.
Visual ROIs are listed along with the specific parcels (name and index) as defined in Glasser et al., 2016.

ROIs	**Area names**	**Parcels index (Glasser et al., 2016)**
Primary visual	V1	1
Extrastriate	V2, V3, V4	4, 5, 6
Dorsal Stream	V6, V6A, V7, IPS1, V3A, V3B	3, 152, 16, 17, 13, 19
Ventral Stream	V8, VVC, VMV1, VMV2, VMV3, PIT Complex, FFC	7, 163, 153, 160, 154, 22, 18
MT + Complex	MT, MST	23, 2
LOC and neighboring areas	LO1, LO2, LO3, V3CD, V4t, FST, PH	20, 21, 159, 158, 156, 157, 138
Parahippocampal	PHA1, PHA3, PHA2	126, 127, 155

Table S6. Auditory ROIs
Auditory ROIs are listed along with the specific parcels (name and index) as defined in Glasser et al., 2016.

ROIs	Area names	Parcels index (Glasser et al., 2016)
Primary auditory	A1	24
Early auditory	MBelt, LBelt, PBelt, RI	173, 174, 124, 104
Auditory association cortex	A4, A5, STSdp, STSda, STSvp, STSva, STGa, TA2	175, 125, 129, 128, 130, 176, 123, 107

Table S7. Multimodal ROIs
Multimodal ROIs are listed along with the specific parcels (name and index) as defined in Glasser et al., 2016.

ROIs	Area names	Parcels index (Glasser et al., 2016)
Multimodal	TPOJ1, TPOJ2, TPOJ3, STV, PSL	139, 140, 141, 28, 25

Table S8. ISC and ROIs overlap.
The table reports the percentage of significant voxels (q<0.01) considering the whole volume of each ROI in four groups (blind, deaf and their controls).

hemisphere	HCP_all_rois_bil.nii	ISC blind>0, q<0.01	ISC audio>0, q<0.01	ISC deaf>0, q<0.01	ISC video>0, q<0.01
		%	%	%	%
L	V1	28	0	93	97
L	extrastriate	15	2	95	97
L	dorsal stream	3	4	98	100
L	PHA	0	0	71	66
L	loc complex	28	11	99	97
L	ventral stream	28	0	95	96
L	MST MT	35	0	99	95
L	multimodal auditory	67	72	99	97
L	associative auditory	82	90	89	91
L	early auditory	81	79	73	52
L	A1	100	100	66	71
R	V1	24	0	99	100
R	extrastriate	13	3	100	99
R	dorsal stream	3	7	99	100
R	PHA	0	0	68	76
R	loc complex	8	1	100	100
R	ventral stream	24	0	98	99
R	MST MT	9	0	100	100
R	multimodal auditory	70	57	98	98
R	associative auditory	85	89	91	89
R	early auditory	83	75	67	68
R	A1	96	89	61	75

Table S9. Correlation values for the discussed ROIs in blind and sighted controls.
q values are reported to highlight the absence of significant ISC in sighted V1 (second row).

ROI	hemisphere	group	r	SE	q	difference
V1	L	blind	0.0226	0.0039	=0.00000	blind>ctrl
V1	L	ctrl	0.0033	0.0035	=0.11648	
extrastriate	L	blind	0.0192	0.0028	=0.00000	blind>ctrl
extrastriate	L	ctrl	0.0085	0.0034	=0.00254	
loc_complex	L	blind	0.0215	0.0025	=0.00000	blind>ctrl
loc_complex	L	ctrl	0.0112	0.0025	=0.00009	
MST_MT	L	blind	0.0287	0.0043	=0.00000	blind>ctrl
MST_MT	L	ctrl	0.0089	0.0033	=0.00895	
ventral_stream	L	blind	0.0239	0.0027	=0.00000	blind>ctrl
ventral_stream	L	ctrl	0.0049	0.0025	=0.02059	
dorsal_stream	L	blind	0.0118	0.0025	=0.00000	
dorsal_stream	L	ctrl	0.0132	0.0035	=0.00000	
A1	L	blind	0.0695	0.0036	=0.00000	blind<ctrl
A1	L	ctrl	0.0842	0.0039	=0.00000	
early_auditory	L	blind	0.0697	0.0044	=0.00017	blind<ctrl
early_auditory	L	ctrl	0.0997	0.0042	=0.00000	
associative_auditory	L	blind	0.0864	0.0066	=0.00000	
associative_auditory	L	ctrl	0.1143	0.0053	=0.00000	
multimodal_auditory	L	blind	0.0524	0.0048	=0.00000	blind<ctrl
multimodal_auditory	L	ctrl	0.0682	0.0054	=0.00000	
V1	R	blind	0.0243	0.0044	=0.00001	blind>ctrl
V1	R	ctrl	0.0042	0.0039	=0.08463	
extrastriate	R	blind	0.0201	0.0033	=0.00001	blind>ctrl
extrastriate	R	ctrl	0.0083	0.0035	=0.00435	
loc_complex	R	blind	0.0227	0.0037	=0.00000	blind>ctrl

loc_complex	R	ctrl	0.0117	0.0033	=0.00079	
MST_MT	R	blind	0.0276	0.0062	=0.00000	blind>ctrl
MST_MT	R	ctrl	0.0083	0.0039	=0.02383	
ventral_stream	R	blind	0.0235	0.0036	=0.00000	blind>ctrl
ventral_stream	R	ctrl	0.0065	0.0027	=0.01021	
dorsal_stream	R	blind	0.0164	0.0037	=0.00000	
dorsal_stream	R	ctrl	0.0214	0.0034	=0.00000	
A1	R	blind	0.0642	0.0043	=0.00000	blind>ctrl
A1	R	ctrl	0.0483	0.0058	=0.00000	
early_auditory	R	blind	0.0864	0.0041	=0.00000	blind>ctrl
early_auditory	R	ctrl	0.0651	0.0046	=0.00000	
associative_auditory	R	blind	0.0926	0.0039	=0.00000	blind>ctrl
associative_auditory	R	ctrl	0.0825	0.0062	=0.00000	
multimodal_auditory	R	blind	0.0688	0.0067	=0.00000	blind>ctrl
multimodal_auditory	R	ctrl	0.0476	0.0043	=0.00000	

Table S10. Correlation values for the discussed ROIs in deaf and hearing controls.
Only significant (q<0.05) differences among groups are reported in the last column.

ROI	hemisphere	group	r	SE	difference
V1	L	deaf	0.1291	0.0070	
V1	L	ctrl	0.1311	0.0074	
extrastriate	L	deaf	0.1115	0.0063	
extrastriate	L	ctrl	0.1102	0.0060	
loc_complex	L	deaf	0.1084	0.0057	
loc_complex	L	ctrl	0.1042	0.0054	
MST_MT	L	deaf	0.1933	0.0091	deaf>ctrl
MST_MT	L	ctrl	0.1560	0.0071	
ventral_stream	L	deaf	0.0729	0.0034	

ventral_stream	L	ctrl	0.0723	0.0041	
dorsal_stream	L	deaf	0.0985	0.0062	
dorsal_stream	L	ctrl	0.1031	0.0054	
PHA	L	deaf	0.0448	0.0039	deaf>ctrl
PHA	L	ctrl	0.0371	0.0028	
A1	L	deaf	0.0183	0.0031	
A1	L	ctrl	0.0168	0.0027	
early_auditory	L	deaf	0.0312	0.0026	
early_auditory	L	ctrl	0.0342	0.0024	
associative_auditory	L	deaf	0.0522	0.0034	deaf<ctrl
associative_auditory	L	ctrl	0.0774	0.0043	
multimodal_auditory	L	deaf	0.0836	0.0054	deaf<ctrl
multimodal_auditory	L	ctrl	0.1064	0.0052	
V1	R	deaf	0.1358	0.0054	deaf>ctrl
V1	R	ctrl	0.1224	0.0070	
extrastriate	R	deaf	0.1259	0.0055	deaf>ctrl
extrastriate	R	ctrl	0.1157	0.0053	
loc_complex	R	deaf	0.1399	0.0083	deaf>ctrl
loc_complex	R	ctrl	0.1219	0.0049	
MST_MT	R	deaf	0.2295	0.0139	deaf>ctrl
MST_MT	R	ctrl	0.2030	0.0098	
ventral_stream	R	deaf	0.0862	0.0040	
ventral_stream	R	ctrl	0.0919	0.0042	
dorsal_stream	R	deaf	0.1214	0.0080	deaf>ctrl
dorsal_stream	R	ctrl	0.1058	0.0056	
PHA	R	deaf	0.0472	0.0033	
PHA	R	ctrl	0.0532	0.0031	
A1	R	deaf	0.0180	0.0027	
A1	R	ctrl	0.0163	0.0030	
early_auditory	R	deaf	0.0239	0.0023	

early_auditory	R	ctrl	0.0190	0.0024	
associative_auditory	R	deaf	0.0549	0.0036	
associative_auditory	R	ctrl	0.0518	0.0037	
multimodal_auditory	R	deaf	0.0890	0.0071	
multimodal_auditory	R	ctrl	0.0874	0.0045	

Table S11. Correlation values for the discussed ROIs in controls (audiovisual condition).

ROI	hemisphere	r	SE
V1	L	0.1317	0.0084
extrastriate	L	0.1378	0.0072
loc_complex	L	0.1550	0.0074
MST_MT	L	0.1985	0.0125
ventral_stream	L	0.0988	0.0050
dorsal_stream	L	0.1608	0.0100
PHA	L	0.0557	0.0034
A1	L	0.1648	0.0089
early_auditory	L	0.1376	0.0046
associative_auditory	L	0.1584	0.0044
multimodal_auditory	L	0.1564	0.0077
V1	R	0.1487	0.0076
extrastriate	R	0.1612	0.0072
loc_complex	R	0.1703	0.0065
MST_MT	R	0.2845	0.0095
ventral_stream	R	0.1189	0.0049
dorsal_stream	R	0.1691	0.0086
PHA	R	0.0650	0.0031
A1	R	0.1512	0.0070
early_auditory	R	0.1278	0.0050

associative_auditory	R	0.1354	0.0034
multimodal_auditory	R	0.1331	0.0064

Table S12. Model-based ISC correlation values for sighted controls. Only significant r values are reported along with the SE.

Model	ROI	hemisphere	r	SE
editing	loc_complex	L	0.0008	0.0002
editing	A1	L	0.0203	0.0011
envelope	A1	L	0.0119	0.0009
spectrum	A1	L	0.0085	0.0008
editing	early_auditory	L	0.0367	0.0016
envelope	early_auditory	L	0.0100	0.0007
spectrum	early_auditory	L	0.0069	0.0005
word2vec	associative_auditory	L	0.0097	0.0007
editing	associative_auditory	L	0.0492	0.0027
envelope	associative_auditory	L	0.0064	0.0004
spectrum	associative_auditory	L	0.0048	0.0003
editing	multimodal_auditory	L	0.0211	0.0020
envelope	multimodal_auditory	L	0.0034	0.0004
spectrum	multimodal_auditory	L	0.0025	0.0003
envelope	loc_complex	R	0.0010	0.0003
spectrum	loc_complex	R	0.0017	0.0003
word2vec	dorsal_stream	R	0.0033	0.0008
editing	dorsal_stream	R	0.0012	0.0004
editing	A1	R	0.0058	0.0007
envelope	A1	R	0.0090	0.0014
spectrum	A1	R	0.0082	0.0014
editing	early_auditory	R	0.0138	0.0009

envelope	early_auditory	R	0.0103	0.0010
spectrum	early_auditory	R	0.0088	0.0009
editing	associative_auditory	R	0.0305	0.0028
envelope	associative_auditory	R	0.0052	0.0005
spectrum	associative_auditory	R	0.0033	0.0003
editing	multimodal_auditory	R	0.0093	0.0014
envelope	multimodal_auditory	R	0.0028	0.0005

Table S13 Model-based ISC correlation values for blind subjects. Only significant models r values are reported along with the SE.

Model	ROI	hemisphere	r	SE
video_cat	V1	L	0.0016	0.0005
editing	V1	L	0.0022	0.0008
envelope	V1	L	0.0021	0.0005
spectrum	V1	L	0.0016	0.0005
video_cat	extrastriate	L	0.0014	0.0004
editing	extrastriate	L	0.0026	0.0005
envelope	extrastriate	L	0.0011	0.0003
video_cat	loc_complex	L	0.0011	0.0003
word2vec	loc_complex	L	0.0036	0.0007
editing	loc_complex	L	0.0038	0.0006
envelope	loc_complex	L	0.0011	0.0003
editing	MST_MT	L	0.0080	0.0015
editing	ventral_stream	L	0.0054	0.0007
envelope	ventral_stream	L	0.0018	0.0003
editing	dorsal_stream	L	0.0014	0.0005
video_cat	A1	L	0.0022	0.0003
editing	A1	L	0.0147	0.0015

envelope	A1	L	0.0124	0.0008
spectrum	A1	L	0.0104	0.0008
video_cat	early_auditory	L	0.0016	0.0003
editing	early_auditory	L	0.0249	0.0015
envelope	early_auditory	L	0.0091	0.0008
spectrum	early_auditory	L	0.0070	0.0006
editing	associative_auditory	L	0.0394	0.0032
envelope	associative_auditory	L	0.0061	0.0006
spectrum	associative_auditory	L	0.0041	0.0002
editing	multimodal_auditory	L	0.0178	0.0018
envelope	multimodal_auditory	L	0.0029	0.0004
video_cat	V1	R	0.0017	0.0005
editing	V1	R	0.0023	0.0010
envelope	V1	R	0.0032	0.0004
spectrum	V1	R	0.0018	0.0005
word2vec	extrastriate	R	0.0031	0.0009
editing	extrastriate	R	0.0020	0.0007
envelope	extrastriate	R	0.0011	0.0002
editing	loc_complex	R	0.0032	0.0008
envelope	loc_complex	R	0.0012	0.0004
editing	MST_MT	R	0.0039	0.0017
video_cat	ventral_stream	R	0.0012	0.0005
editing	ventral_stream	R	0.0040	0.0008
audio_cat	ventral_stream	R	0.0010	0.0004
envelope	ventral_stream	R	0.0017	0.0003
editing	dorsal_stream	R	0.0017	0.0005
video_cat	A1	R	0.0021	0.0005
editing	A1	R	0.0107	0.0010
envelope	A1	R	0.0147	0.0009
spectrum	A1	R	0.0115	0.0007

editing	early_auditory	R	0.0215	0.0015
envelope	early_auditory	R	0.0151	0.0007
spectrum	early_auditory	R	0.0118	0.0005
editing	associative_auditory	R	0.0413	0.0015
envelope	associative_auditory	R	0.0080	0.0005
spectrum	associative_auditory	R	0.0047	0.0003
editing	multimodal_auditory	R	0.0250	0.0035
envelope	multimodal_auditory	R	0.0046	0.0005
spectrum	multimodal_auditory	R	0.0016	0.0003

Table S14. Model-based ISC correlation values for hearing subjects. Only significant models r values are reported along with the SE

Model	ROI	hemisphere	r	SE
gist	V1	L	0.0101	0.0007
video_cat	V1	L	0.0036	0.0003
editing	V1	L	0.0251	0.0014
audio_cat	V1	L	0.0015	0.0002
envelope	V1	L	0.0026	0.0002
spectrum	V1	L	0.0035	0.0004
gist	extrastriate	L	0.0071	0.0004
video_cat	extrastriate	L	0.0045	0.0003
editing	extrastriate	L	0.0159	0.0009
audio_cat	extrastriate	L	0.0012	0.0001
envelope	extrastriate	L	0.0024	0.0002
spectrum	extrastriate	L	0.0037	0.0004
motion	loc_complex	L	0.0352	0.0017
video_cat	loc_complex	L	0.0118	0.0006
editing	loc_complex	L	0.0035	0.0006
audio_cat	loc_complex	L	0.0039	0.0003

spectrum	loc_complex	L	0.0034	0.0003
motion	MST_MT	L	0.0662	0.0035
video_cat	MST_MT	L	0.0161	0.0007
editing	MST_MT	L	0.0026	0.0006
audio_cat	MST_MT	L	0.0048	0.0004
spectrum	MST_MT	L	0.0071	0.0005
gist	ventral_stream	L	0.0041	0.0003
video_cat	ventral_stream	L	0.0074	0.0004
editing	ventral_stream	L	0.0072	0.0007
audio_cat	ventral_stream	L	0.0017	0.0002
envelope	ventral_stream	L	0.0015	0.0002
spectrum	ventral_stream	L	0.0028	0.0003
gist	dorsal_stream	L	0.0060	0.0003
motion	dorsal_stream	L	0.0399	0.0019
video_cat	dorsal_stream	L	0.0075	0.0004
editing	dorsal_stream	L	0.0058	0.0008
envelope	dorsal_stream	L	0.0022	0.0002
spectrum	dorsal_stream	L	0.0047	0.0003
gist	PHA	L	0.0049	0.0003
motion	PHA	L	0.0155	0.0016
video_cat	PHA	L	0.0066	0.0005
word2vec	PHA	L	0.0047	0.0005
editing	PHA	L	0.0028	0.0002
spectrum	PHA	L	0.0017	0.0002
editing	A1	L	0.0008	0.0005
editing	early_auditory	L	0.0095	0.0007
envelope	early_auditory	L	0.0008	0.0001
word2vec	associative_auditory	L	0.0058	0.0006
editing	associative_auditory	L	0.0262	0.0017
envelope	associative_auditory	L	0.0017	0.0002

video_cat	multimodal_auditory	L	0.0077	0.0004
editing	multimodal_auditory	L	0.0213	0.0015
audio_cat	multimodal_auditory	L	0.0025	0.0003
envelope	multimodal_auditory	L	0.0022	0.0003
spectrum	multimodal_auditory	L	0.0020	0.0002
gist	V1	R	0.0095	0.0006
video_cat	V1	R	0.0031	0.0003
editing	V1	R	0.0203	0.0012
audio_cat	V1	R	0.0016	0.0002
envelope	V1	R	0.0023	0.0001
spectrum	V1	R	0.0034	0.0004
gist	extrastriate	R	0.0072	0.0004
video_cat	extrastriate	R	0.0038	0.0003
editing	extrastriate	R	0.0114	0.0008
audio_cat	extrastriate	R	0.0010	0.0001
envelope	extrastriate	R	0.0023	0.0002
spectrum	extrastriate	R	0.0040	0.0004
motion	loc_complex	R	0.0416	0.0021
video_cat	loc_complex	R	0.0118	0.0007
editing	loc_complex	R	0.0048	0.0004
audio_cat	loc_complex	R	0.0032	0.0002
spectrum	loc_complex	R	0.0043	0.0003
motion	MST_MT	R	0.0893	0.0048
video_cat	MST_MT	R	0.0166	0.0014
editing	MST_MT	R	0.0037	0.0006
audio_cat	MST_MT	R	0.0041	0.0005
spectrum	MST_MT	R	0.0090	0.0003
gist	ventral_stream	R	0.0041	0.0003
motion	ventral_stream	R	0.0337	0.0016
video_cat	ventral_stream	R	0.0088	0.0004

editing	ventral_stream	R	0.0030	0.0004
audio_cat	ventral_stream	R	0.0015	0.0002
envelope	ventral_stream	R	0.0024	0.0002
spectrum	ventral_stream	R	0.0036	0.0004
gist	dorsal_stream	R	0.0060	0.0003
motion	dorsal_stream	R	0.0380	0.0018
video_cat	dorsal_stream	R	0.0070	0.0005
editing	dorsal_stream	R	0.0075	0.0009
envelope	dorsal_stream	R	0.0017	0.0002
spectrum	dorsal_stream	R	0.0052	0.0004
gist	PHA	R	0.0068	0.0004
motion	PHA	R	0.0241	0.0015
video_cat	PHA	R	0.0092	0.0005
editing	PHA	R	0.0033	0.0002
envelope	PHA	R	0.0016	0.0002
spectrum	PHA	R	0.0029	0.0003
editing	early_auditory	R	0.0014	0.0004
video_cat	associative_auditory	R	0.0020	0.0002
editing	associative_auditory	R	0.0107	0.0009
motion	multimodal_auditory	R	0.0343	0.0020
video_cat	multimodal_auditory	R	0.0080	0.0005
word2vec	multimodal_auditory	R	0.0060	0.0005
editing	multimodal_auditory	R	0.0049	0.0004
audio_cat	multimodal_auditory	R	0.0023	0.0002
spectrum	multimodal_auditory	R	0.0017	0.0002

Table S15. Model-based ISC correlation values for deaf subjects. Only significant models r values are reported along with the SE

Model	ROI	hemisphere	r	SE
gist	V1	L	0.0075	0.0007
video_cat	V1	L	0.0036	0.0003
editing	V1	L	0.0199	0.0016
audio_cat	V1	L	0.0013	0.0002
envelope	V1	L	0.0019	0.0003
spectrum	V1	L	0.0029	0.0003
gist	extrastriate	L	0.0059	0.0005
video_cat	extrastriate	L	0.0047	0.0002
editing	extrastriate	L	0.0145	0.0017
audio_cat	extrastriate	L	0.0010	0.0001
envelope	extrastriate	L	0.0019	0.0003
spectrum	extrastriate	L	0.0029	0.0003
gist	loc_complex	L	0.0038	0.0004
motion	loc_complex	L	0.0351	0.0028
video_cat	loc_complex	L	0.0158	0.0007
editing	loc_complex	L	0.0037	0.0009
audio_cat	loc_complex	L	0.0042	0.0004
spectrum	loc_complex	L	0.0039	0.0003
motion	MST_MT	L	0.0686	0.0034
video_cat	MST_MT	L	0.0242	0.0009
editing	MST_MT	L	0.0054	0.0008
audio_cat	MST_MT	L	0.0052	0.0003
spectrum	MST_MT	L	0.0082	0.0004
gist	ventral_stream	L	0.0030	0.0004
motion	ventral_stream	L	0.0225	0.0020
video_cat	ventral_stream	L	0.0089	0.0004

editing	ventral_stream	L	0.0056	0.0005
audio_cat	ventral_stream	L	0.0017	0.0003
envelope	ventral_stream	L	0.0013	0.0002
spectrum	ventral_stream	L	0.0031	0.0003
gist	dorsal_stream	L	0.0041	0.0005
motion	dorsal_stream	L	0.0323	0.0026
video_cat	dorsal_stream	L	0.0079	0.0004
editing	dorsal_stream	L	0.0055	0.0008
spectrum	dorsal_stream	L	0.0038	0.0004
gist	PHA	L	0.0034	0.0005
motion	PHA	L	0.0177	0.0018
video_cat	PHA	L	0.0069	0.0007
word2vec	PHA	L	0.0043	0.0006
editing	PHA	L	0.0017	0.0003
envelope	PHA	L	0.0015	0.0003
spectrum	PHA	L	0.0027	0.0003
editing	A1	L	0.0010	0.0004
video_cat	early_auditory	L	0.0018	0.0002
editing	early_auditory	L	0.0030	0.0005
audio_cat	early_auditory	L	0.0007	0.0002
envelope	early_auditory	L	0.0008	0.0002
video_cat	associative_auditory	L	0.0041	0.0003
word2vec	associative_auditory	L	0.0042	0.0006
editing	associative_auditory	L	0.0070	0.0010
audio_cat	associative_auditory	L	0.0013	0.0003
envelope	associative_auditory	L	0.0012	0.0002
motion	multimodal_auditory	L	0.0330	0.0024
video_cat	multimodal_auditory	L	0.0095	0.0007
editing	multimodal_auditory	L	0.0033	0.0008
audio_cat	multimodal_auditory	L	0.0025	0.0004

spectrum	multimodal_auditory	L	0.0020	0.0003
gist	V1	R	0.0075	0.0005
video_cat	V1	R	0.0036	0.0003
editing	V1	R	0.0224	0.0017
audio_cat	V1	R	0.0018	0.0002
envelope	V1	R	0.0017	0.0003
spectrum	V1	R	0.0027	0.0003
gist	extrastriate	R	0.0065	0.0005
video_cat	extrastriate	R	0.0060	0.0004
editing	extrastriate	R	0.0122	0.0007
audio_cat	extrastriate	R	0.0012	0.0002
envelope	extrastriate	R	0.0017	0.0002
spectrum	extrastriate	R	0.0033	0.0003
motion	loc_complex	R	0.0508	0.0035
video_cat	loc_complex	R	0.0197	0.0010
editing	loc_complex	R	0.0025	0.0007
audio_cat	loc_complex	R	0.0038	0.0004
spectrum	loc_complex	R	0.0057	0.0003
motion	MST_MT	R	0.0881	0.0048
video_cat	MST_MT	R	0.0258	0.0015
editing	MST_MT	R	0.0034	0.0009
audio_cat	MST_MT	R	0.0050	0.0004
spectrum	MST_MT	R	0.0086	0.0003
gist	ventral_stream	R	0.0031	0.0004
motion	ventral_stream	R	0.0310	0.0021
video_cat	ventral_stream	R	0.0116	0.0004
editing	ventral_stream	R	0.0021	0.0004
audio_cat	ventral_stream	R	0.0016	0.0003
envelope	ventral_stream	R	0.0016	0.0003
spectrum	ventral_stream	R	0.0040	0.0003

gist	dorsal_stream	R	0.0038	0.0005
motion	dorsal_stream	R	0.0442	0.0033
video_cat	dorsal_stream	R	0.0122	0.0007
editing	dorsal_stream	R	0.0064	0.0007
spectrum	dorsal_stream	R	0.0046	0.0005
gist	PHA	R	0.0039	0.0003
motion	PHA	R	0.0186	0.0013
video_cat	PHA	R	0.0079	0.0006
word2vec	PHA	R	0.0040	0.0004
editing	PHA	R	0.0017	0.0003
envelope	PHA	R	0.0014	0.0003
spectrum	PHA	R	0.0030	0.0003
editing	A1	R	0.0018	0.0003
video_cat	early_auditory	R	0.0013	0.0002
editing	early_auditory	R	0.0012	0.0002
gist	associative_auditory	R	0.0025	0.0003
motion	associative_auditory	R	0.0202	0.0019
video_cat	associative_auditory	R	0.0052	0.0005
word2vec	associative_auditory	R	0.0056	0.0004
editing	associative_auditory	R	0.0021	0.0005
audio_cat	associative_auditory	R	0.0013	0.0004
envelope	associative_auditory	R	0.0013	0.0003
motion	multimodal_auditory	R	0.0354	0.0033
video_cat	multimodal_auditory	R	0.0108	0.0012
audio_cat	multimodal_auditory	R	0.0023	0.0004
spectrum	multimodal_auditory	R	0.0021	0.0003

Table S16. Model-based ISC correlation values for control subjects (audiovisual condition).
Only significant models r values are reported along with the SE

Model	ROI	hemisphere	r	SE
gist	V1	L	0.0156	0.0010
motion	V1	L	0.0493	0.0029
video_cat	V1	L	0.0067	0.0004
editing	V1	L	0.0016	0.0002
envelope	V1	L	0.0022	0.0002
spectrum	V1	L	0.0057	0.0002
gist	extrastriate	L	0.0116	0.0008
motion	extrastriate	L	0.0454	0.0027
video_cat	extrastriate	L	0.0082	0.0004
editing	extrastriate	L	0.0038	0.0003
audio_cat	extrastriate	L	0.0012	0.0001
envelope	extrastriate	L	0.0017	0.0001
spectrum	extrastriate	L	0.0052	0.0002
gist	loc_complex	L	0.0046	0.0003
motion	loc_complex	L	0.0460	0.0020
video_cat	loc_complex	L	0.0184	0.0007
editing	loc_complex	L	0.0096	0.0006
audio_cat	loc_complex	L	0.0037	0.0003
spectrum	loc_complex	L	0.0043	0.0003
gist	MST_MT	L	0.0062	0.0005
motion	MST_MT	L	0.0666	0.0033
video_cat	MST_MT	L	0.0203	0.0012
editing	MST_MT	L	0.0035	0.0006
audio_cat	MST_MT	L	0.0036	0.0003
spectrum	MST_MT	L	0.0081	0.0008
gist	ventral_stream	L	0.0046	0.0003

motion	ventral_stream	L	0.0314	0.0018
video_cat	ventral_stream	L	0.0104	0.0004
editing	ventral_stream	L	0.0050	0.0003
audio_cat	ventral_stream	L	0.0023	0.0002
envelope	ventral_stream	L	0.0015	0.0001
spectrum	ventral_stream	L	0.0040	0.0002
gist	dorsal_stream	L	0.0074	0.0005
motion	dorsal_stream	L	0.0606	0.0030
video_cat	dorsal_stream	L	0.0120	0.0006
editing	dorsal_stream	L	0.0082	0.0005
envelope	dorsal_stream	L	0.0030	0.0002
spectrum	dorsal_stream	L	0.0074	0.0003
gist	PHA	L	0.0051	0.0005
motion	PHA	L	0.0219	0.0015
video_cat	PHA	L	0.0097	0.0006
word2vec	PHA	L	0.0047	0.0004
editing	PHA	L	0.0044	0.0002
envelope	PHA	L	0.0012	0.0001
spectrum	PHA	L	0.0033	0.0003
editing	A1	L	0.0384	0.0032
envelope	A1	L	0.0135	0.0011
spectrum	A1	L	0.0106	0.0011
editing	early_auditory	L	0.0439	0.0018
envelope	early_auditory	L	0.0108	0.0006
spectrum	early_auditory	L	0.0078	0.0006
editing	associative_auditory	L	0.0628	0.0016
envelope	associative_auditory	L	0.0070	0.0003
spectrum	associative_auditory	L	0.0054	0.0002
video_cat	multimodal_auditory	L	0.0074	0.0006
editing	multimodal_auditory	L	0.0365	0.0020

envelope	multimodal_auditory	L	0.0060	0.0003
spectrum	multimodal_auditory	L	0.0053	0.0003
gist	V1	R	0.0144	0.0008
motion	V1	R	0.0537	0.0028
video_cat	V1	R	0.0071	0.0004
editing	V1	R	0.0014	0.0002
audio_cat	V1	R	0.0012	0.0002
envelope	V1	R	0.0019	0.0002
spectrum	V1	R	0.0051	0.0002
gist	extrastriate	R	0.0115	0.0007
motion	extrastriate	R	0.0513	0.0027
video_cat	extrastriate	R	0.0083	0.0004
editing	extrastriate	R	0.0042	0.0002
audio_cat	extrastriate	R	0.0020	0.0002
spectrum	extrastriate	R	0.0058	0.0002
gist	loc_complex	R	0.0062	0.0004
motion	loc_complex	R	0.0606	0.0021
video_cat	loc_complex	R	0.0207	0.0006
editing	loc_complex	R	0.0065	0.0005
audio_cat	loc_complex	R	0.0050	0.0002
spectrum	loc_complex	R	0.0049	0.0002
gist	MST_MT	R	0.0129	0.0005
motion	MST_MT	R	0.1165	0.0026
video_cat	MST_MT	R	0.0323	0.0013
audio_cat	MST_MT	R	0.0066	0.0005
spectrum	MST_MT	R	0.0093	0.0002
gist	ventral_stream	R	0.0046	0.0003
motion	ventral_stream	R	0.0397	0.0019
video_cat	ventral_stream	R	0.0131	0.0006
editing	ventral_stream	R	0.0049	0.0003

audio_cat	ventral_stream	R	0.0027	0.0003
spectrum	ventral_stream	R	0.0042	0.0002
gist	dorsal_stream	R	0.0077	0.0005
motion	dorsal_stream	R	0.0642	0.0027
video_cat	dorsal_stream	R	0.0124	0.0006
editing	dorsal_stream	R	0.0088	0.0006
envelope	dorsal_stream	R	0.0024	0.0002
spectrum	dorsal_stream	R	0.0066	0.0004
gist	PHA	R	0.0052	0.0004
motion	PHA	R	0.0251	0.0015
video_cat	PHA	R	0.0112	0.0006
editing	PHA	R	0.0046	0.0003
envelope	PHA	R	0.0014	0.0002
spectrum	PHA	R	0.0036	0.0003
editing	A1	R	0.0215	0.0020
envelope	A1	R	0.0167	0.0008
spectrum	A1	R	0.0150	0.0008
editing	early_auditory	R	0.0260	0.0019
envelope	early_auditory	R	0.0126	0.0005
spectrum	early_auditory	R	0.0109	0.0005
editing	associative_auditory	R	0.0468	0.0012
envelope	associative_auditory	R	0.0063	0.0003
spectrum	associative_auditory	R	0.0044	0.0002
gist	multimodal_auditory	R	0.0043	0.0003
video_cat	multimodal_auditory	R	0.0088	0.0007
editing	multimodal_auditory	R	0.0136	0.0012
audio_cat	multimodal_auditory	R	0.0026	0.0004
envelope	multimodal_auditory	R	0.0045	0.0004
spectrum	multimodal_auditory	R	0.0033	0.0003

Table S 17. Group ISC differencences for blind and sighted subjects, Only significant group differences are reported for each model and ROI.

Model	ROI	hemisphere	difference between groups
editing	V1	L	BLIND>CTRL r=0.0024
envelope	V1	L	BLIND>CTRL r=0.0023
spectrum	V1	L	BLIND>CTRL r=0.0015
editing	extrastriate	L	BLIND>CTRL r=0.0028
editing	loc_complex	L	BLIND>CTRL r=0.0030
editing	MST_MT	L	BLIND>CTRL r=0.0081
editing	ventral_stream	L	BLIND>CTRL r=0.0051
envelope	ventral_stream	L	BLIND>CTRL r=0.0019
editing	dorsal_stream	L	BLIND>CTRL r=0.0013
video_cat	A1	L	BLIND>CTRL r=0.0019
editing	A1	L	BLIND<CTRL r=0.0056
spectrum	A1	L	BLIND>CTRL r=0.0020
video_cat	early_auditory	L	BLIND>CTRL r=0.0013
editing	early_auditory	L	BLIND<CTRL r=0.0118
word2vec	associative_auditory	L	BLIND<CTRL r=0.0040
editing	associative_auditory	L	BLIND<CTRL r=0.0098
editing	multimodal_auditory	L	BLIND<CTRL r=0.0033
editing	V1	R	BLIND>CTRL r=0.0022
envelope	V1	R	BLIND>CTRL r=0.0035
editing	extrastriate	R	BLIND>CTRL r=0.0021
editing	loc_complex	R	BLIND>CTRL r=0.0030
spectrum	loc_complex	R	BLIND<CTRL r=0.0013
editing	MST_MT	R	BLIND>CTRL r=0.0041
editing	ventral_stream	R	BLIND>CTRL r=0.0041
audio_cat	ventral_stream	R	BLIND>CTRL r=0.0012
envelope	ventral_stream	R	BLIND>CTRL r=0.0017
video_cat	A1	R	BLIND>CTRL r=0.0017

editing	A1	R	BLIND>CTRL r=0.0049
envelope	A1	R	BLIND>CTRL r=0.0057
spectrum	A1	R	BLIND>CTRL r=0.0033
editing	early_auditory	R	BLIND>CTRL r=0.0076
envelope	early_auditory	R	BLIND>CTRL r=0.0048
spectrum	early_auditory	R	BLIND>CTRL r=0.0030
editing	associative_auditory	R	BLIND>CTRL r=0.0108
envelope	associative_auditory	R	BLIND>CTRL r=0.0028
spectrum	associative_auditory	R	BLIND>CTRL r=0.0014
editing	multimodal_auditory	R	BLIND>CTRL r=0.0157
envelope	multimodal_auditory	R	BLIND>CTRL r=0.0019

Table S18. Group ISC differencences for deaf and hearing subjects, Only significant group differences are reported for each model and ROI.

Model	ROI	hemisphere	difference between groups
gist	V1	L	DEAF<CTRL r=0.0026
editing	V1	L	DEAF<CTRL r=0.0051
editing	extrastriate	L	DEAF<CTRL r=0.0014
video_cat	loc_complex	L	DEAF>CTRL r=0.0041
video_cat	MST_MT	L	DEAF>CTRL r=0.0081
editing	MST_MT	L	DEAF>CTRL r=0.0028
video_cat	ventral_stream	L	DEAF>CTRL r=0.0015
editing	ventral_stream	L	DEAF<CTRL r=0.0017
gist	PHA	L	DEAF<CTRL r=0.0015
editing	PHA	L	DEAF<CTRL r=0.0011
spectrum	PHA	L	DEAF>CTRL r=0.0010
video_cat	early_auditory	L	DEAF>CTRL r=0.0010

editing	early_auditory	L	DEAF<CTRL r=0.0064
audio_cat	early_auditory	L	DEAF>CTRL r=0.0007
video_cat	associative_auditory	L	DEAF>CTRL r=0.0032
editing	associative_auditory	L	DEAF<CTRL r=0.0191
audio_cat	associative_auditory	L	DEAF>CTRL r=0.0016
motion	multimodal_auditory	L	DEAF>CTRL r=0.0053
video_cat	multimodal_auditory	L	DEAF>CTRL r=0.0018
editing	multimodal_auditory	L	DEAF<CTRL r=0.0180
gist	V1	R	DEAF<CTRL r=0.0020
editing	V1	R	DEAF>CTRL r=0.0021
video_cat	extrastriate	R	DEAF>CTRL r=0.0022
video_cat	loc_complex	R	DEAF>CTRL r=0.0079
editing	loc_complex	R	DEAF<CTRL r=0.0023
video_cat	MST_MT	R	DEAF>CTRL r=0.0092
video_cat	ventral_stream	R	DEAF>CTRL r=0.0028
video_cat	dorsal_stream	R	DEAF>CTRL r=0.0052
gist	PHA	R	DEAF<CTRL r=0.0029
video_cat	PHA	R	DEAF<CTRL r=0.0013
editing	PHA	R	DEAF<CTRL r=0.0017
editing	A1	R	DEAF>CTRL r=0.0016
gist	associative_auditory	R	DEAF>CTRL r=0.0022
motion	associative_auditory	R	DEAF>CTRL r=0.0088
video_cat	associative_auditory	R	DEAF>CTRL r=0.0032
editing	associative_auditory	R	DEAF<CTRL r=0.0086
audio_cat	associative_auditory	R	DEAF>CTRL r=0.0011
video_cat	multimodal_auditory	R	DEAF>CTRL r=0.0028
editing	multimodal_auditory	R	DEAF<CTRL r=0.0037

Supplemental Behavioural assessment

The questionnaires are in Italian, according to the sample mother tongue.

Behavioral S1: Movie Familiarity Checklist

Subject ID:

- Quanto ritieni di conoscere *la storia* de "La carica dei 101?" *

  1 per niente
  2 poco
  3 abbastanza
  4 molto bene
  5 quasi a memoria

- Hai mai visto il film "La carica dei 101" prodotto dalla Walt Disney nel 1996 con protagonista Glenn Close?
- Se si, quante volte (all'incirca)?
- In base alla tua esperienza quanto ritieni di conoscere il film (scala da 1 a 5)? *

1 per niente

2 poco

3 abbastanza

4 molto bene

5 quasi a memoria

*Se il soggetto risponde di conoscere la storia:

REGISTRAZIONE: IL SOGGETTO HA 30" DI TEMPO PER DESCRIVERE LA TRAMA DELLA STORIA.

Behavioural S2. True/false final assessment

These questions were meant to be asked right after the end of the experiment, as soon as the MRI experiment ended.

Run 1:

- Crudelia interagisce in modo affabile e gentile con i suoi collaboratori.
- Crudelia è proprietaria di una casa di moda in cui si producono abiti di seta e lino
- Crudelia è interessata alla pelliccia maculata ispirata dal dalmata di Anita

Run 2:

- Anita possiede un cane bulldog di nome Peggy.
- Anita aggredisce Rudy quando lui, per sbaglio, afferra Peggy al posto di Pongo
- Entrambi i protagonisti vengono trascinati dai loro cani nel laghetto del parco.

Run 3:

- Per riscaldarsi, Rudy e Anita bevono una tazza di thè
- Nilla è la vicina di casa di Anita
- A seguito di una discussione accanita per la vendita dei cani, Crudelia licenzia Anita.

Run 4:

- Rudy e Anita al rientro della passeggiata serale, colgono i due scagnozzi di Crudelia intenti a rapire i cuccioli.
- I cuccioli vengono portati in un vecchio castello dismesso dai due scagnozzi di Crudelia
- Attraverso una sorta di telegrafo canino, Pongo riesce a chiedere aiuto agli altri cani della città per ritrovare i suoi cucciol

Run 5:

- Anita capisce che i cani sono stati rapiti a causa del suo figurino di pelliccia maculato
- Pongo riesce a far fuggire tutti i cagnolini rubati dal castello

- Crudelia, non fidandosi dei suoi collaboratori, è felice di poter pensare all’inseguimento dei cuccioli da sola.

Run 6

- Nella casa di Crudelia viene rinvenuta la pelliccia della tigre siberiana ma la polizia non la ritiene colpevole.
- Crudelia, è vittima della trappola di due procioni e cade in un tino pieno di melassa.
- 101 cuccioli riescono a fuggire dalle grinfie di Crudelia e raggiungere le auto di Rudy e Anita che li stanno cercando alla periferia della città.

Behavioural S3: Movie knowledge evaluation and examples of story narrations.

Familiarity with the movie per se (not only the story in general) was assessed through verbal/written description of the movie tale content. All the participants that affirmed to have seen the movie at least once, were requested to verbally summarize what they remembered about it, while their voices were digitally recorded. In the case of deaf individuals, the interpreter just wrote down what they told us by sign language. Quite surprisingly, none of the blind reported to know the movie, thus we did not have any story description for the people belonging to that group. However, most of them claimed being not totally new to the content of the tale, as they confirmed with high scores in rating the general familiarity with the plot. For what concerns the other groups only 4 subjects reported no experience with the movie: 3 among the controls and 1 for the deaf group. Examples of written descriptions of what subjects recalled about the story were reported below (in Italian) for the group of deaf subjects.

Subject 44
"Donna cattiva cerca di vendere i cani. Erano 100. Uno resta fuori. Ci voleva fare una pelliccia"

Subject 45
"C'è una donna che cerca di rapire dei cani, due cani hanno avuto un cucciolo che è stato rapito. La signora ne ha rapiti cento, ora sono 101"

Subject 46
"Qualcuno ruba i 101 cani e i genitori cercano i cuccioli e poi li trovano. Nella fattoria ci sono gli animali e c'è anche un tasso (procione?)"

Subject 48
“I padroni dei due cani si innamorano e si sposano, i cani hanno 101 cuccioli, poi c’è una donna cattiva che ama le pellicce e vuole rapire i cuccioli.”

Subject 49
“Ci sono due cattivi che vogliono rubare i cuccioli oppure una donna, non mi ricordo. E poi c’è una famiglia che ha 101 cuccioli”

Subject 50
“Una coppia che prima litiga e poi si innamora, una donna cattiva che vuole rubare i cani, poi una casa piccola che non va bene, ne cercano una più grande”

Subject 51
“Due proprietari di cani dalmata, un maschio e una femmina, si innamorano. La capa di lei al lavoro, Crudelia Demon, è ossessionata dalle pellicce e vuole fare una pelliccia di cani dalmata. Quindi fa rapire i cuccioli della coppia dai suoi scagnozzi e altri cuccioli fino ad arrivare a 101. Ma poi li salvano!”

Subject 52
“Ci sono un uomo e una donna che hanno due cani dalmata , Pongo e Peggy. Si incontrano e si innamorano. I cani hanno dei cuccioli e c’è una donna cattiva, Crudelia, che vuole rubarli per farsi una pelliccia e quindi manda i suoi scagnozzi a rubare i cani”

Bibliography

Amedi, A., Floel, A., Knecht, S., Zohary, E., & Cohen, L. G. (2004). Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. *Nature neuroscience*, *7*(11), 1266-1270.

Amedi, A., Hofstetter, S., Maidenbaum, S., & Heimler, B. (2017). Task selectivity as a comprehensive principle for brain organization. Trends in cognitive sciences, 21(5), 307-310.

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. Nature neuroscience, 4(3), 324-330.

Amedi, A., Raz, N., Pianka, P., Malach, R., & Zohary, E. (2003). Early 'visual'cortex activation correlates with superior verbal memory performance in the blind. *Nature neuroscience*, *6*(7), 758-766.

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., ... & Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature neuroscience*, *10*(6), 687-689.

Ames, D. L., Honey, C. J., Chow, M. A., Todorov, A., & Hasson, U. (2015). Contextual alignment of cognitive and neural dynamics. *Journal of cognitive neuroscience*, *27*(4), 655-664.

Anderson, D. R., Fite, K. V., Petrovich, N., & Hirsch, J. (2006). Cortical activation while watching video editing: An fMRI study. *Media Psychology*, *8*(1), 7-24.

Andrade, J., May, J., Deeprose, C., Baugh, S. J., & Ganis, G. (2014). Assessing vividness of mental imagery: the plymouth sensory imagery questionnaire. *British Journal of Psychology*, *105*(4), 547-563.

Angelucci, A., Clascá, F., & Sur, M. (1998). Brainstem inputs to the ferret medial geniculate nucleus and the effect of early deafferentation on novel retinal projections to the auditory thalamus. *Journal of Comparative Neurology*, *400*(3), 417-439.

Araneda, R., Renier, L. A., Rombaux, P., Cuevas, I., & De Volder, A. G. (2016). Cortical plasticity and olfactory function in early blindness. *Frontiers in systems neuroscience*, *10*, 75.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709-721.

Baldwin, M. K., & Krubitzer, L. (2018). Architectonic characteristics of the visual thalamus and superior colliculus in titi monkeys. *Journal of Comparative Neurology*, *526*(11), 1760-1776.

Barone, P., Lacassagne, L., & Kral, A. (2013). Reorganization of the connectivity of cortical field DZ in congenitally deaf cat. *PloS one*, *8*(4), e60093.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. Language resources and evaluation, 43(3), 209-226.

Bartels, A., & Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. *Human brain mapping*, *21*(2), 75-85.

Bartels, A., Zeki, S., & Logothetis, N. K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cerebral cortex*, *18*(3), 705-717.

Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: where and how?. *Nature Reviews Neuroscience*, *3*(6), 443-452.

Bavelier, D., Brozinsky, C., Tomann, A., Mitchell, T., Neville, H., & Liu, G. (2001). Impact of early deafness and early exposure to sign language on the cerebral organization for motion processing. *Journal of Neuroscience*, *21*(22), 8931-8942.

Beaulieu, C., & Cynader, M. (1990). Effect of the richness of the environment on neurons in cat visual cortex. I. Receptive field properties. *Developmental Brain Research*, *53*(1), 71-81.

Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., & Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Sciences*, *108*(11), 4429-4434.

Bedny, M., Richardson, H., & Saxe, R. (2015). "Visual" cortex responds to spoken language in blind children. *Journal of Neuroscience*, *35*(33), 11674-11681.

Berardi, N., Pizzorusso, T., Ratto, G. M., & Maffei, L. (2003). Molecular basis of plasticity in the visual cortex. Trends in neurosciences, 26(7), 369-378.

Berezutskaya, J., Freudenburg, Z. V., Güçlü, U., van Gerven, M. A., & Ramsey, N. F. (2020). Brain-optimized extraction of complex sound features that drive continuous auditory perception. *PLOS Computational Biology*, *16*(7), e1007992.

Betti, V., Della Penna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G. L., & Corbetta, M. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. Neuron, 79(4), 782-797.

Bishop, C. W., & Miller, L. M. (2009). A multisensory cortical network for understanding speech in noise. *Journal of cognitive neuroscience*, *21*(9), 1790-1804.

Blakemore, C., & Cooper, G. F. (1970). Development of the brain depends on the visual environment. *Nature*, *228*(5270), 477-478.

Blakemore, C., & Cooper, G. F. (1971). Modification of the visual cortex by experience. *Brain research*, *31*(2), 366.

Blakemore, C., & Van Sluyters, R. C. (1974). Reversal of the physiological effects of monocular deprivation in kittens: further evidence for a sensitive period. *The Journal of physiology*, *237*(1), 195-216.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature neuroscience*, *8*(3), 389-395.

Bola, Ł., Zimmermann, M., Mostowski, P., Jednoróg, K., Marchewka, A., Rutkowski, P., & Szwed, M. (2017). Task-specific reorganization of the auditory cortex in deaf humans. *Proceedings of the National Academy of Sciences*, *114*(4), E600-E609.

Bottari, D., Heimler, B., Caclin, A., Dalmolin, A., Giard, M. H., & Pavani, F. (2014). Visual change detection recruits auditory cortices in early deafness. *Neuroimage*, *94*, 172-184.

Bottenhorn, K. L., Flannery, J. S., Boeving, E. R., Riedel, M. C., Eickhoff, S. B., Sutherland, M. T., & Laird, A. R. (2018). Cooperating yet distinct brain networks engaged during naturalistic paradigms: A meta-analysis of functional MRI results. *Network Neuroscience*, *3*(1), 27-48.

Brang, D., Towle, V. L., Suzuki, S., Hillyard, S. A., Di Tusa, S., Dai, Z., ... & Grabowecky, M. (2015). Peripheral sounds rapidly activate visual cortex: evidence from electrocorticography. *Journal of neurophysiology*, *114*(5), 3023-3028.

Bruti, S., & Zanotti, S. (2017). Exploring the sensory dimension of translated films: An analysis of The King's Speech. *Palimpsestes. Revue de traduction*, (30), 154-177.

Büchel, C., Price, C., Frackowiak, R. S., & Friston, K. (1998). Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain: a journal of neurology*, *121*(3), 409-419.

Bueti, D., & Macaluso, E. (2010). Auditory temporal expectations modulate activity in visual cortex. *Neuroimage*, *51*(3), 1168-1183.

Burton, H., Sinclair, R. J., & Dixit, S. (2010). Working memory for vibrotactile frequencies: comparison of cortical activity in blind and sighted individuals. *Human brain mapping*, *31*(11), 1686-1701.

Butler, B. E., & Lomber, S. G. (2013). Functional and structural changes throughout the auditory system following congenital and early-onset deafness: implications for hearing restoration. *Frontiers in systems neuroscience*, *7*, 92.

Butler, B. E., Chabot, N., & Lomber, S. G. (2016). Quantifying and comparing the pattern of thalamic and cortical projections to the posterior auditory field in hearing and deaf cats. *Journal of Comparative Neurology*, *524*(15), 3042-3063.

Butler, B. E., Chabot, N., Kral, A., & Lomber, S. G. (2017). Origins of thalamic and cortical projections to the posterior auditory field in congenitally deaf cats. *Hearing research*, *343*, 118-127.

Butler, B. E., de la Rua, A., Ward-Able, T., & Lomber, S. G. (2018). Cortical and thalamic connectivity to the second auditory cortex of the cat is resilient to the onset of deafness. *Brain Structure and Function*, *223*(2), 819-835.

Campus, C., Sandini, G., Amadeo, M. B., & Gori, M. (2019). Stronger responses in the visual cortex of sighted compared to blind individuals during auditory space representation. *Scientific reports*, *9*(1), 1-12.

Capek, C. M., MacSweeney, M., Woll, B., Waters, D., McGuire, P. K., David, A. S., ... & Campbell, R. (2008). Cortical circuits for silent speechreading in deaf and hearing people. *Neuropsychologia*, *46*(5), 1233-1241.

Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. Trends in cognitive sciences, 7(8), 354-361.

Cardin, V., Orfanidou, E., Rönnberg, J., Capek, C. M., Rudner, M., & Woll, B. (2013). Dissociating cognitive and sensory neural plasticity in human superior temporal cortex. *Nature communications*, *4*(1), 1-5.

Cardin, V., Smittenaar, R. C., Orfanidou, E., Rönnberg, J., Capek, C. M., Rudner, M., & Woll, B. (2016). Differential activity in Heschl's gyrus between deaf and hearing individuals is due to auditory deprivation rather than language modality. *Neuroimage*, *124*, 96-106.

Cecchetti, L., Kupers, R., Ptito, M., Pietrini, P., & Ricciardi, E. (2016). Are supramodality and cross-modal plasticity the yin and yang of brain development? From blindness to rehabilitation. *Frontiers in Systems Neuroscience*, *10*, 89.

Chang, W. T., Jääskeläinen, I. P., Belliveau, J. W., Huang, S., Hung, A. Y., Rossi, S., & Ahveninen, J. (2015). Combined MEG and EEG show reliable patterns of electromagnetic brain activity during natural viewing. *NeuroImage*, *114*, 49-56.

Chebat, D. R., Schneider, F. C., & Ptito, M. (2020). Neural networks mediating perceptual learning in congenital blindness. *Scientific Reports*, *10*(1), 1-10.

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature neuroscience*, *20*(1), 115-125.

Chén, O. Y., Crainiceanu, C., Ogburn, E. L., Caffo, B. S., Wager, T. D., & Lindquist, M. A. (2018). High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*, *19*(2), 121-136.

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, *118*(2), 887-906.

Cichy, R. M., Heinzle, J., & Haynes, J. D. (2012). Imagery and perception share cortical representations of content and location. *Cerebral cortex*, *22*(2), 372-380.

Cimino, A., De Mattei, L., & Dell'Orletta, F. (2018). Multi-task learning in deep neural networks at evalita 2018. *Proceedings of the 6th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA'18)*, 86-95.

Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Faiz, L., Dambrosia, J., ... & Hallett, M. (1997). Functional relevance of cross-modal plasticity in blind humans. *Nature*, *389*(6647), 180-183.

Collignon, O., Champoux, F., Voss, P., & Lepore, F. (2011). Sensory rehabilitation in the plastic brain. In *Progress in brain research* (Vol. 191, pp. 211-231). Elsevier.

Collignon, O., Davare, M., Olivier, E., & De Volder, A. G. (2009). Reorganisation of the right occipito-parietal stream for auditory spatial processing in early blind humans. A transcranial magnetic stimulation study. *Brain topography*, *21*(3-4), 232-240.

Collignon, O., Dormal, G., Albouy, G., Vandewalle, G., Voss, P., Phillips, C., & Lepore, F. (2013). Impact of blindness onset on the functional organization and the connectivity of the occipital cortex. *Brain*, *136*(9), 2769-2783.

Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., & Lepore, F. (2011). Functional specialization for auditory–spatial processing in the occipital cortex of congenitally blind humans. *Proceedings of the National Academy of Sciences*, *108*(11), 4435-4440.

Coq, J. O., & Xerri, C. (1998). Environmental enrichment alters organizational features of the forepaw representation in the primary somatosensory cortex of adult rats. *Experimental brain research*, *121*(2), 191-204.

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. Nature neuroscience, 16(6), 763-770.

De Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *Journal of Neuroscience*, *37*(27), 6539-6557.

De Heer, Wendy A., et al. "The hierarchical cortical organization of human speech processing." Journal of Neuroscience 37.27 (2017): 6539-6557.

De Volder, A. G., Toyama, H., Kimura, Y., Kiyosawa, M., Nakano, H., Vanlierde, A., ... & Senda, M. (2001). Auditory triggered mental imagery of shape involves visual association areas in early blind humans. *Neuroimage*, *14*(1), 129-139.

Dempsey-Jones, H., Themistocleous, A. C., Carone, D., Ng, T. W., Harrar, V., & Makin, T. R. (2019). Blocking tactile input to one finger using anaesthetic enhances touch perception and learning in other fingers. *Journal of Experimental Psychology: General*, *148*(4), 713.

Di Santo, J. L. (2012). Referential Sounds, Symbolism and Semiotics. In *Ponencia presentada en el simposio Electroacoustic Music Studies Network. Electroacoustic Music Studies 2012 Proceedings. Disponible en: http://www. ems-network. org/IMG/pdf_EMS12_disanto. pdf (consultado el 24/11/16)*.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition?. *Neuron*, *73*(3), 415-434.

Dormal, G., Rezk, M., Yakobov, E., Lepore, F., & Collignon, O. (2016). Auditory motion in the sighted and blind: Early visual deprivation triggers a large-scale imbalance between auditory and "visual" brain regions. *Neuroimage*, *134*, 630-644.

Durant, S., Wall, M. B., & Zanker, J. M. (2011). Manipulating the content of dynamic natural scenes to characterize response in human MT/MST. *Journal of Vision*, *11*(10), 5-5.

Elton, A., & Gao, W. (2015). Task-positive functional connectivity of the default mode network transcends task domain. *Journal of Cognitive Neuroscience*, *27*(12), 2369-2381.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in cognitive sciences*, *12*(10), 388-396.

Espinosa, J. Sebastian, and Michael P. Stryker. "Development and plasticity of the primary visual cortex." *Neuron* 75, no. 2 (2012): 230-249.

Fairhall, S. L., & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. Journal of Neuroscience, 33(25), 10552-10558.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. In *Cereb cortex*.

Feng, W., Störmer, V. S., Martinez, A., McDonald, J. J., & Hillyard, S. A. (2014). Sounds activate visual cortex and improve visual discrimination. *Journal of Neuroscience*, *34*(29), 9817-9824.

Ferstl, E. C., Neumann, J., Bogler, C., & Von Cramon, D. Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Human brain mapping*, *29*(5), 581-593.

Finney, E. M., Fine, I., & Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nature neuroscience*, *4*(12), 1171-1173.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, *10*(4), 507-521.

Formisano, E., Kim, D. S., Di Salle, F., Van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*(4), 859-869.

Frasnelli, Johannes, Olivier Collignon, Patrice Voss, and Franco Lepore. "Crossmodal plasticity in sensory loss." In *Progress in brain research*, vol. 191, pp. 233-249. Elsevier, 2011.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological reviews*, *91*(4), 1357-1392.

Frost, D. O., Boire, D., Gingras, G., & Ptito, M. (2000). Surgically created neural pathways mediate visual pattern discrimination. *Proceedings of the National Academy of Sciences*, *97*(20), 11068-11073.

Fullerton, B. C., & Pandya, D. N. (2007). Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. *Journal of Comparative Neurology*, *504*(5), 470-498.

Furlan, M., & Smith, A. T. (2016). Global motion processing in human visual cortical areas V2 and V3. *Journal of Neuroscience*, *36*(27), 7314-7324.

Gagnon, Lea, Maurice Ptito, and Ron Kupers. "Superior orthonasal but not retronasal olfactory skills in congenital blindness." *PloS one* 10, no. 3 (2015b): e0122567.

Gagnon, Léa, Ron Kupers, and Maurice Ptito. "Neural correlates of taste perception in congenital blindness." *Neuropsychologia* 70 (2015a): 227-234.

Giordano, B. L., Ince, R. A., Gross, J., Schyns, P. G., Panzeri, S., & Kayser, C. (2017). Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *Elife*, *6*, e24763.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of neurophysiology*, *84*(3), 1588-1598.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... & Smith, S. M. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171-178.

Gödecke, I., & Bonhoeffer, T. (1996). Development of identical orientation maps for two eyes without common visual experience. *Nature*, *379*(6562), 251-254.

Gordon, J. A., & Stryker, M. P. (1996). Experience-dependent plasticity of binocular responses in the primary visual cortex of the mouse. *Journal of Neuroscience*, *16*(10), 3274-3286.

Gougoux, F., Zatorre, R. J., Lassonde, M., Voss, P., & Lepore, F. (2005). A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals. *PLoS biology*, *3*(2).

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536-548.

Gu, J., Liu, B., Li, X., Wang, P., & Wang, B. (2019). Cross-modal representations in early visual and auditory cortices revealed by multi-voxel pattern analysis. *Brain Imaging and Behavior*, 1-13.

Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., ... & Pietrini, P. (2016). How concepts are encoded in the human brain: a modality independent, category-based cortical organization of semantic knowledge. Neuroimage, 135, 232-242.

Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F. R., ... & Stadler, J. (2016). A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Scientific data*, *3*(1), 1-15.

Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., ... & Stadler, J. (2014). A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Scientific data*, *1*, 140003.

Hasson, U., & Honey, C. J. (2012). Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, *62*(2), 1272-1278

Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., & Behrmann, M. (2009). Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Research*, *2*(4), 220-231.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends in cognitive sciences*, *19*(6), 304-313.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The neuroscience of film. *Projections*, *2*(1), 1-26.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, *14*(1), 40-48.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *science*, *303*(5664), 1634-1640.

Haufe, S., DeGuzman, P., Henin, S., Arcaro, M., Honey, C. J., Hasson, U., & Parra, L. C. (2018). Elucidating relations between fMRI, ECoG, and EEG through a common natural stimulus. *NeuroImage*, *179*, 79-91.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425-2430.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (1996). Computational models of cortical visual processing. *Proceedings of the National Academy of Sciences*, *93*(2), 623-627.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience*, *6*(11), 877-888.

Henschke, J. U., Oelschlegel, A. M., Angenstein, F., Ohl, F. W., Goldschmidt, J., Kanold, P. O., & Budinger, E. (2018). Early sensory experience influences the development of multisensory thalamocortical and intracortical connections of primary sensory cortices. *Brain Structure and Function*, *223*(3), 1165-1190.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, *8*(5), 393-402.

Hirshorn, E. A., Dye, M. W., Hauser, P. C., Supalla, T. R., & Bavelier, D. (2014). Neural networks mediating sentence reading in the deaf. *Frontiers in human neuroscience*, *8*, 394.

Hubel, D. H., & Wiesel, T. N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of neurophysiology*, *26*(6), 994-1002.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology*, *28*(2), 229-289.

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage*, *50*(3), 1202-1211.

Hunt, D. L., Yamoah, E. N., & Krubitzer, L. (2006). Multisensory plasticity in congenitally deaf mice: how are cortical areas functionally specified?. *Neuroscience*, *139*(4), 1507-1524.

Huth, A. G., Lee, T., Nishimoto, S., Bilenko, N. Y., Vu, A. T., & Gallant, J. L. (2016). Decoding the semantic content of natural movies from human brain activity. Frontiers in systems neuroscience, 10, 81.

Hyvärinen, J., Carlson, S., & Hyvärinen, L. (1981). Early visual deprivation alters modality of neuronal responses in area 19 of monkey cortex. *Neuroscience letters*, *26*(3), 239-243.

Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., ... & Medini, P. (2012). Sound-driven synaptic inhibition in primary visual cortex. Neuron, 73(4), 814-828.

Kaas, J. (2000). The Reorganization of Sensory and Motor Maps after Injury. *The new cognitive neurosciences*, 223.

Kahn, D. M., & Krubitzer, L. (2002). Massive cross-modal cortical plasticity and the emergence of a new cortical area in developmentally blind mammals. *Proceedings of the National Academy of Sciences*, *99*(17), 11429-11434.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature neuroscience*, *8*(5), 679-685.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109-2128.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience*, *17*(11), 4302-4311.

Karns, C. M., Dow, M. W., & Neville, H. J. (2012). Altered cross-modal processing in the primary auditory cortex of congenitally deaf adults: a visual-somatosensory fMRI study with a double-flash illusion. *Journal of Neuroscience*, *32*(28), 9626-9638.

Kauttonen, J., Hlushchuk, Y., & Tikka, P. (2015). Optimizing methods for linking cinematic features to fMRI data. *Neuroimage*, *110*, 136-148.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352-355.

Keitel, A., Gross, J., & Kayser, C. (2019). Largely distinct networks mediate perceptually-relevant auditory and visual speech representations. *bioRxiv*, 661405.

Kell, A. J., & McDermott, J. H. (2019). Invariance to background noise as a signature of non-primary auditory cortex. *Nature communications*, *10*(1), 1-11.

Kim, E., Kang, H., Lee, H., Lee, H. J., Suh, M. W., Song, J. J., ... & Lee, D. S. (2014). Morphological brain network assessed using graph theory and network filtration in deaf adults. *Hearing research*, *315*, 88-98.

Kim, J. K., & Zatorre, R. J. (2011). Tactile–auditory shape learning engages the lateral occipital complex. *Journal of Neuroscience*, *31*(21), 7848-7856.

Klinge, C., Eippert, F., Röder, B., & Büchel, C. (2010). Corticocortical connections mediate primary visual cortex responses to auditory stimulation in the blind. *Journal of Neuroscience*, *30*(38), 12798-12805.

Klinke, R., Kral, A., Heid, S., Tillein, J., & Hartmann, R. (1999). Recruitment of the auditory cortex in congenitally deaf cats by long-term cochlear electrostimulation. *science*, *285*(5434), 1729-1733.

Kok, M. A., Chabot, N., & Lomber, S. G. (2014). Cross-modal reorganization of cortical afferents to dorsal auditory cortex following early-and late-onset deafness. *Journal of Comparative Neurology*, *522*(3), 654-675.

Kral, A., Hubka, P., Heid, S., & Tillein, J. (2013). Single-sided deafness leads to unilateral aural preference within an early sensitive period. *Brain*, *136*(1), 180-193.

Kral, A., Schröder, J. H., Klinke, R., & Engel, A. K. (2003). Absence of cross-modal reorganization in the primary auditory cortex of congenitally deaf cats. *Experimental Brain Research*, *153*(4), 605-613.

Kral, A., Yusuf, P. A., & Land, R. (2017). Higher-order auditory areas in congenital deafness: Top-down interactions and corticocortical decoupling. *Hearing Research*, *343*, 50-63.

Kriegeskorte, N., & Douglas, P. K. (2019). Interpreting encoding and decoding models. *Current opinion in neurobiology*, *55*, 167-179.

Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., & Schalk, G. (2013). The tracking of speech envelope in the human cortex. *PloS one*, *8*(1), e53398.

Kujala, T., Alho, K., Paavilainen, P., Summala, H., & Näätänen, R. (1992). Neural plasticity in processing of sound location by the early blind: an event-related potential study. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *84*(5), 469-472.

Kupers, R., Beaulieu-Lefebvre, M., Schneider, F. C., Kassuba, T., Paulson, O. B., Siebner, H. R., & Ptito, M. (2011). Neural correlates of olfactory processing in congenital blindness. *Neuropsychologia*, *49*(7), 2037-2044.

Kupers, R., Chebat, D. R., Madsen, K. H., Paulson, O. B., & Ptito, M. (2010). Neural correlates of virtual route recognition in congenital blindness. *Proceedings of the National Academy of Sciences*, *107*(28), 12716-12721.

Kupers, R., Fumal, A., De Noordhout, A. M., Gjedde, A., Schoenen, J., & Ptito, M. (2006). Transcranial magnetic stimulation of the visual cortex induces somatotopically organized qualia in blind subjects. *Proceedings of the National Academy of Sciences*, *103*(35), 13256-13260.

Lahnakoski, J. M., Glerean, E., Salmi, J., Jääskeläinen, I. P., Sams, M., Hari, R., & Nummenmaa, L. (2012). Naturalistic FMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Frontiers in human neuroscience*, *6*, 233.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of cognitive neuroscience*, *14*(3), 420-429.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *Journal of Neuroscience*, *30*(22), 7604-7612.

Leonard, M. K., Ramirez, N. F., Torres, C., Travis, K. E., Hatrak, M., Mayberry, R. I., & Halgren, E. (2012). Signed words in the congenitally deaf evoke typical late lexicosemantic responses with no early visual responses in left superior temporal cortex. *Journal of Neuroscience*, *32*(28), 9700-9705.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, *31*(8), 2906-2915.

Lettieri, G., Handjaras, G., Ricciardi, E., Leo, A., Papale, P., Betta, M., ... & Cecchetti, L. (2019). Emotionotopy in the human right temporo-parietal cortex. *Nature communications*, *10*(1), 1-13.

Lewis, L. B., Saenz, M., & Fine, I. (2010). Mechanisms of cross-modal plasticity in early-blind subjects. *Journal of neurophysiology*, *104*(6), 2995-3008.

Loiotile, R. E., Cusack, R., & Bedny, M. (2019). Naturalistic Audio-Movies and Narrative Synchronize "Visual" Cortices across Congenitally Blind But Not Sighted Individuals. *Journal of Neuroscience*, *39*(45), 8940-8948.

Lomber, S. G., Meredith, M. A., & Kral, A. (2010). Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. *Nature neuroscience*, *13*(11), 1421-1427.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. Annu. Rev. Psychol., 58, 593-614.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological methods*, *7*(1), 83.

MacSweeney, M., Capek, C. M., Campbell, R., & Woll, B. (2008). The signing brain: the neurobiology of sign language. *Trends in cognitive sciences*, *12*(11), 432-440.

MacSweeney, M., Woll, B., Campbell, R., McGuire, P. K., David, A. S., Williams, S. C., ... & Brammer, M. J. (2002). Neural systems underlying British Sign Language and audio-visual English processing in native users. *Brain*, *125*(7), 1583-1593.

Maguire, E. A. (2012). Studying the freely-behaving brain with fMRI. *Neuroimage*, *62*(2), 1170-1176.

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. Neuron, 63(3), 397-405.

Mao, Y. T., Hua, T. M., & Pallas, S. L. (2011). Competition and convergence between auditory and cross-modal visual inputs to primary auditory cortical areas. *Journal of Neurophysiology*, *105*(4), 1558-1573.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British journal of Psychology*, *64*(1), 17-24.

Marshall, J., Atkinson, J., Smulovitch, E., Thacker, A., & Woll, B. (2004). Aphasia in a user of British Sign Language: Dissociation between sign and gesture. *Cognitive neuropsychology*, *21*(5), 537-554.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25-45.

Martinelli, A., Handjaras, G., Betta, M., Leo, A., Cecchetti, L., Pietrini, P., ... & Bottari, D. (2020). Auditory features modelling demonstrates sound envelope representation in striate cortex. bioRxiv.

Matteau, I., Kupers, R., Ricciardi, E., Pietrini, P., & Ptito, M. (2010). Beyond visual, aural and haptic movement perception: hMT+ is activated by electrotactile motion stimulation of the tongue in sighted and in congenitally blind individuals. *Brain research bulletin*, *82*(5-6), 264-270.

Mattioni, S., Rezk, M., Battal, C., Bottini, R., Mendoza, K. E. C., Oosterhof, N. N., & Collignon, O. (2020). Categorical representation from sound and sight in the ventral occipito-temporal cortex of sighted and blind. Elife, 9, e50732.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in cognitive sciences*, *7*(7), 293-299.

McDonald, J. J., Störmer, V. S., Martinez, A., Feng, W., & Hillyard, S. A. (2013). Salient sounds activate human visual cortex automatically. *Journal of Neuroscience*, *33*(21), 9194-9201.

Merabet, L. B., & Pascual-Leone, A. (2010). Neural reorganization following sensory loss: the opportunity of change. *Nature Reviews Neuroscience*, *11*(1), 44-52.

Merabet, L. B., Swisher, J. D., McMains, S. A., Halko, M. A., Amedi, A., Pascual-Leone, A., & Somers, D. C. (2007). Combined activation and deactivation of visual cortex during tactile sensory processing. *Journal of neurophysiology*, *97*(2), 1633-1641.

Merabet, L., & Pascual-Leone, A. (2008). Studies of crossmodal functions with TMS. *Oxford handbook of transcranial Stimulation*, 447-462.

Merabet, L., Thut, G., Murray, B., Andrews, J., Hsiao, S., & Pascual-Leone, A. (2004). Feeling by sight or seeing by touch?. *Neuron*, *42*(1), 173-179.

Meredith, M. A., & Allman, B. L. (2009). Subthreshold multisensory processing in cat auditory cortex. *Neuroreport*, *20*(2), 126.

Meredith, M. A., & Allman, B. L. (2012). Early hearing-impairment results in crossmodal reorganization of ferret core auditory cortex. *Neural plasticity*, *2012*.

Meredith, M. A., & Lomber, S. G. (2011). Somatosensory and visual crossmodal plasticity in the anterior auditory field of early-deaf cats. *Hearing research*, *280*(1-2), 38-47.

Meredith, M. A., Clemo, H. R., Corley, S. B., Chabot, N., & Lomber, S. G. (2016). Cortical and thalamic connectivity of the auditory anterior ectosylvian cortex of early-deaf cats: Implications for neural mechanisms of crossmodal plasticity. *Hearing research*, *333*, 25-36.

Meredith, M. A., Keniston, L. R., Dehner, L. R., & Clemo, H. R. (2006). Crossmodal projections from somatosensory area SIV to the auditory field of the anterior ectosylvian sulcus (FAES) in Cat: further evidence for subthreshold forms of multisensory processing. *Experimental brain research*, *172*(4), 472-484.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006-1010.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M. A., Morito, Y., Tanabe, H. C., ... & Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, *60*(5), 915-929.

Moerel, M., De Martino, F., & Formisano, E. (2012). Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *Journal of Neuroscience*, *32*(41), 14205-14216.

Moerel, M., De Martino, F., Santoro, R., Ugurbil, K., Goebel, R., Yacoub, E., & Formisano, E. (2013). Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. *Journal of Neuroscience*, *33*(29), 11888-11898.

Moreno, A., Limousin, F., Dehaene, S., & Pallier, C. (2018). Brain correlates of constituent structure in sign language comprehension. *NeuroImage*, *167*, 151-161.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage*, *13*(4), 684-701.

Mozolic, J. L., Joyner, D., Hugenschmidt, C. E., Peiffer, A. M., Kraft, R. A., Maldjian, J. A., & Laurienti, P. J. (2008). Cross-modal deactivations during modality-specific selective attention. *BMC neurology*, *8*(1), 35.

Muckli, L., & Petro, L. S. (2013). Network interactions: Non-geniculate input to V1. *Current opinion in neurobiology*, *23*(2), 195-201.

Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science*, *309*(5736), 951-954.

Müller, F., Niso, G., Samiee, S., Ptito, M., Baillet, S., & Kupers, R. (2019). A thalamocortical pathway for fast rerouting of tactile information to occipital cortex in congenital blindness. *Nature communications*, *10*(1), 1-9.

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. *Neuropsychologia*, *83*, 161–169.

music and speech. *Trends in cognitive sciences*, *6*(1), 37-46.

Nakai, T., Koide-Majima, N., & Nishimoto, S. (2018, October). Encoding and decoding of music-genre representations in the human brain. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 584-589). IEEE.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63*(6), 902-915.

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation.

Neville, H., & Bavelier, D. (2002). Human brain plasticity: evidence from sensory deprivation and altered language experience. In *Progress in brain research* (Vol. 138, pp. 177-188). Elsevier.

Nguyen, M., Chang, A., Micciche, E., Meshulam, M., Nastase, S. A., & Hasson, U. (2020). Teacher-student neural coupling during teaching and learning. *bioRxiv*.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. PLoS computational biology, 10(4).

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. Current Biology, 21(19), 1641-1646.

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

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., ... & Brugge, J. F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience*, *29*(49), 15564-15574.

Nummenmaa, L., Lahnakoski, J. M., & Glerean, E. (2018). Sharing the social world via intersubject neural synchronisation. *Current opinion in psychology*, *24*, 7-14.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, *42*(3), 145-175.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. Progress in brain research, 155, 23-36.

Overath, T., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2008). Encoding of spectral correlation over time in auditory cortex. *Journal of Neuroscience*, *28*(49), 13268-13273.

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in brain research*, *134*, 427-446.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annu. Rev. Neurosci.*, *28*, 377-401.

Pattamadilok, C., Planton, S., & Bonnard, M. (2019). Spoken language coding neurons in the Visual Word Form Area: Evidence from a TMS adaptation paradigm. *NeuroImage*, *186*, 278–285.

Pavani, F., & Röder, B. (2012). Crossmodal plasticity as a consequence of sensory loss: insights from blindness and deafness. *The new handbook of multisensory processes*, 737-759.

Peelen, M. V., & Downing, P. E. (2017). Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia*, *105*, 177-183.

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94-97.

Penhune, V. B., Cismaru, R., Dorsaint-Pierre, R., Petitto, L. A., & Zatorre, R. J. (2003). The morphometry of auditory cortex in the congenitally deaf measured using MRI. *Neuroimage*, *20*(2), 1215-1225.

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media psychology*, *13*(3), 243-272.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., ... & Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. Nature communications, 9(1), 1-13.

Petro, L. S., Paton, A. T., & Muckli, L. (2017). Contextual modulation of primary visual cortex by auditory signals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160104.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. H. C., Cohen, L., ... & Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences*, *101*(15), 5658-5663.

Pitzalis, S., Serra, C., Sulpizio, V., Committeri, G., de Pasquale, F., Fattori, P., ... & Galati, G. (2020). Neural bases of self-and object-motion in a naturalistic vision. *Human Brain Mapping*, *41*(4), 1084-1111.

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of best practices for evidence for prediction: a review. JAMA psychiatry, 77(5), 534-540.

Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *Journal of Neuroscience*, *32*(18), 6263-6274.

Ptito, M., & Kupers, R. (2005). Cross-modal plasticity in early blindness. *Journal of integrative neuroscience*, *4*(04), 479-488.

Ptito, M., Fumal, A., De Noordhout, A. M., Schoenen, J., Gjedde, A., & Kupers, R. (2008). TMS of the occipital cortex induces tactile sensations in the fingers of blind Braille readers. *Experimental Brain Research*, *184*(2), 193-200.

Ptito, M., Giguère, J. F., Boire, D., Frost, D. O., & Casanova, C. (2001). When the auditory cortex turns visual. In *Progress in brain research* (Vol. 134, pp. 447-458). Elsevier.

Ptito, M., Kupers, R., Lomber, S., & Pietrini, P. (2012). Sensory deprivation and brain plasticity.

Ptito, M., Moesgaard, S. M., Gjedde, A., & Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain*, *128*(3), 606-614.

Ptito, M., Moesgaard, S. M., Gjedde, A., & Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain*, *128*(3), 606-614.

Raichle, M. E. (2000). The neural correlates of consciousness: An analysis of cognitive skill. In M. Gazzaniga (Ed.). The new cognitive neurosciences (pp. 1305–1318). Cambridge, MA: MIT Press.

Rampon, C., Tang, Y. P., Goodhouse, J., Shimizu, E., Kyin, M., & Tsien, J. Z. (2000). Enrichment induces structural changes and recovery from nonspatial memory deficits in CA1 NMDAR1-knockout mice. *Nature neuroscience*, *3*(3), 238-244.

Raz, N., Amedi, A., & Zohary, E. (2005). V1 activation in congenitally blind humans is associated with episodic retrieval. *Cerebral Cortex*, *15*(9), 1459-1468.

Regev, M., Simony, E., Lee, K., Tan, K. M., Chen, J., & Hasson, U. (2019). Propagation of information along the cortical hierarchy as a function of attention while reading and listening to stories. *Cerebral Cortex*, *29*(10), 4017-4034.

Renier, L. A., Anurova, I., De Volder, A. G., Carlson, S., VanMeter, J., & Rauschecker, J. P. (2010). Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. *Neuron*, *68*(1), 138-148.

Renier, L., Cuevas, I., Grandin, C. B., Dricot, L., Plaza, P., Lerens, E., ... & De Volder, A. G. (2013). Right occipital cortex activation correlates with superior odor processing performance in the early blind. *PLoS One*, *8*(8), e71907.

Rezk, M., Cattoir, S., Battal, C., Occelli, V., Mattioni, S., & Collignon, O. (2020). Shared Representation of Visual and Auditory Motion Directions in the Human Middle-Temporal Cortex. Current Biology.

Ricciardi, E., & Pietrini, P. (2011). New light from the dark: what blindness can teach us about brain function. *Current opinion in neurology*, *24*(4), 357-363.

Ricciardi, E., Basso, D., Sani, L., Bonino, D., Vecchi, T., Pietrini, P., & Miniussi, C. (2011). Functional inhibition of the human middle temporal cortex affects non-visual motion perception: a repetitive transcranial magnetic stimulation study during tactile speed discrimination. *Experimental Biology and Medicine*, *236*(2), 138-144.

Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014). Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture?. *Neuroscience & Biobehavioral Reviews*, *41*, 64-77.

Ricciardi, E., Bottari, D., Ptito, M., Röder, B., & Pietrini, P. (2020). The sensory-deprived brain as a unique tool to understand brain development and function.

Ricciardi, E., Handjaras, G., & Pietrini, P. (2014). The blind brain: How (lack of) vision shapes the morphological and functional architecture of the human brain. *Experimental Biology and Medicine*, *239*(11), 1414-1420.

Ricciardi, E., Vanello, N., Sani, L., Gentili, C., Scilingo, E. P., Landini, L., ... & Pietrini, P. (2007). The effect of visual experience on the development of functional architecture in hMT+. *Cerebral Cortex*, *17*(12), 2933-2939.

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature communications*, *9*(1), 1-12.

Röder, B., Rösler, F., & Neville, H. J. (2000). Event-related potentials during auditory language processing in congenitally blind and sighted people. *Neuropsychologia*, *38*(11), 1482-1502.

Röder, B., Stock, O., Bien, S., Neville, H., & Rösler, F. (2002). Speech processing activates visual cortex in congenitally blind humans. *European Journal of Neuroscience*, *16*(5), 930-936.

Roe, A. W., Pallas, S. L., Hahm, J. O., & Sur, M. (1990). A map of visual space induced in primary auditory cortex. *Science*, *250*(4982), 818-820.

Roe, A. W., Pallas, S. L., Kwon, Y. H., & Sur, M. (1992). Visual projections routed to the auditory pathway in ferrets: receptive fields of visual neurons in primary auditory cortex. *Journal of Neuroscience*, *12*(9), 3651-3664.

Romei, V., Murray, M. M., Cappe, C., & Thut, G. (2009). Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Current biology*, *19*(21), 1799-1805.

Saari, P., Burunat, I., Brattico, E., & Toiviainen, P. (2018). Decoding musical training from dynamic processing of musical features in the brain. *Scientific reports*, *8*(1), 1-12.

Sadato, N., Okada, T., Honda, M., Matsuki, K. I., Yoshida, M., Kashikura, K. I., ... & Yonekura, Y. (2005). Cross-modal integration and plastic changes revealed by lip movement, random-dot motion and sign languages in the hearing and deaf. *Cerebral Cortex*, *15*(8), 1113-1122.

Sadato, N., Pascual-Leone, A., Grafman, J., Deiber, M. P., Ibanez, V., & Hallett, M. (1998). Neural networks for Braille reading by the blind. *Brain: a journal of neurology*, *121*(7), 1213-1229.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, *380*(6574), 526-528.

Saito, D. N., Okada, T., Honda, M., Yonekura, Y., & Sadato, N. (2006). Practice makes perfect: the neural substrates of tactile discrimination by Mah-Jong experts include the primary visual cortex. *BMC neuroscience*, *7*(1), 79.

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., & Formisano, E. (2014). Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput Biol*, *10*(1), e1003412.

Schreiber, T., & Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Physical review letters*, *77*(4), 635.

Scott, G. D., Karns, C. M., Dow, M. W., Stevens, C., & Neville, H. J. (2014). Enhanced peripheral visual processing in congenitally deaf humans is supported by multiple brain regions, including primary auditory cortex. *Frontiers in human neuroscience*, *8*, 177.

Sengpiel, F., & Kind, P. C. (2002). The role of activity in development of the visual system. *Current Biology*, *12*(23), R818-R826.

Sengpiel, F., Stawinski, P., & Bonhoeffer, T. (1999). Influence of experience on orientation maps in cat visual cortex. *Nature neuroscience*, *2*(8), 727-732.

Sharma, J., Angelucci, A., & Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature*, *404*(6780), 841-847.

Sheehan, P. W. (1967). A shortened form of Betts' questionnaire upon mental imagery. *Journal of clinical psychology*.

Singh, A. K., Phillips, F., Merabet, L. B., & Sinha, P. (2018). Why does the cortex reorganize after sensory loss?. *Trends in cognitive sciences*, *22*(7), 569-582.

Siuda-Krzywicka, K., Bola, Ł., Paplińska, M., Sumera, E., Jednorog, K., Marchewka, A., ... & Szwed, M. (2016). Massive cortical reorganization in sighted Braille readers. *Elife*, *5*, e10762.

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience*, *11*(9), 2768-2785.

Sorokowska, A., Sorokowski, P., Karwowski, M., Larsson, M., & Hummel, T. (2019). Olfactory perception and blindness: a systematic review and meta-analysis. *Psychological research*, *83*(8), 1595-1611.

Stanton, S. G., & Harrison, R. V. (2000). Projections from the medial geniculate body to primary auditory cortex in neonatally deafened cats. *Journal of Comparative Neurology*, *426*(1), 117-129.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425-14430.

Striem-Amit, E., & Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people “seeing” by using sounds. *Current Biology*, *24*(6), 687-692.

Striem-Amit, E., Almeida, J., Belledonne, M., Chen, Q., Fang, Y., Han, Z., ... & Bi, Y. (2016). Topographical functional connectivity patterns exist in the congenitally, prelingually deaf. *Scientific reports*, *6*, 29375.

Striem-Amit, E., Guendelman, M., & Amedi, A. (2012). ‘Visual’acuity of the congenitally blind using visual-to-auditory sensory substitution. *PloS one*, *7*(3).

Striem-Amit, E., Hertz, U., & Amedi, A. (2011). Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding FMRI. *PLoS One*, *6*(3).

Striem-Amit, E., Ovadia-Caro, S., Caramazza, A., Margulies, D. S., Villringer, A., & Amedi, A. (2015). Functional connectivity of visual cortex in the blind follows retinotopic organization principles. Brain, 138(6), 1679-1695.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 71-78.

Sur, M. (2004). Rewiring cortex: Cross-modal plasticity and its implications for cortical development and function. *Handbook of multisensory processing*, 661-694.

Sur, M., & Rubenstein, J. L. (2005). Patterning and plasticity of the cerebral cortex. *Science*, *310*(5749), 805-810.

Sur, M., Garraghty, P. E., & Roe, A. W. (1988). Experimentally induced visual projections into auditory thalamus and cortex. *Science*, *242*(4884), 1437-1441.

Thomas, R. M., De Sanctis, T., Gazzola, V., & Keysers, C. (2018). Where and how our brain represents the temporal structure of observed action. *NeuroImage*, *183*, 677-697.

Toldi, J., Feher, O., & Wolff, J. R. (1996). Neuronal plasticity induced by neonatal monocular (and binocular) enucleation. *Progress in neurobiology*, *48*(3), 191-218.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(5761), 670-674.

Vachon, P., Voss, P., Lassonde, M., Leroux, J. M., Mensour, B., Beaudoin, G., ... & Lepore, F. (2013). Reorganization of the auditory, visual and multimodal areas in early deaf individuals. *Neuroscience*, *245*, 50-60.

Van Ackeren, M. J., Barbero, F. M., Mattioni, S., Bottini, R., & Collignon, O. (2018). Neuronal populations in the occipital cortex of the blind synchronize to the temporal dynamics of speech. ELife, 7, e31640.

Van den Hurk, J., Van Baelen, M., & de Beeck, H. P. O. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proceedings of the National Academy of Sciences*, *114*(22), E4501-E4510.

Van Praag, C. D. G., Garfinkel, S. N., Sparasci, O., Mees, A., Philippides, A. O., Ware, M., ... & Critchley, H. D. (2017). Mind-wandering and alterations to default mode network connectivity when listening to naturalistic versus artificial sounds. *Scientific Reports*, *7*, 45273.

Vanderwal, T., Eilbott, J., & Castellanos, F. X. (2019). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental cognitive neuroscience*, *36*, 100600.

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., & Castellanos, F. X. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *Neuroimage*, *122*, 222-232.

Varoquaux, G. (2018). Cross-validation failure: small sample sizes lead to large error bars. Neuroimage, 180, 68-77.

Vetter, P., Bola, Ł., Reich, L., Bennett, M., Muckli, L., & Amedi, A. (2020). Decoding Natural Sounds in Early "Visual" Cortex of Congenitally Blind Individuals. *Current Biology*.

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Current Biology*, *24*(11), 1256-1262.

Vodrahalli, K., Chen, P. H., Liang, Y., Baldassano, C., Chen, J., Yong, E., ... & Arora, S. (2018). Mapping between fMRI responses to movies and their natural language annotations. *Neuroimage*, *180*, 223-231.

Von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, *404*(6780), 871-876.

Voss, P., & Zatorre, R. J. (2012). Organization and reorganization of sensory-deprived cortex. *Current Biology*, *22*(5), R168-R173.

Voss, P., Alary, F., Lazzouni, L., Chapman, C. E., Goldstein, R., Bourgoin, P., & Lepore, F. (2016). Crossmodal processing of haptic inputs in sighted and blind individuals. *Frontiers in systems neuroscience*, *10*, 62.

Wanet-Defalque, M. C., Veraart, C., De Volder, A., Metz, R., Michel, C., Dooms, G., & Goffinet, A. (1988). High metabolic activity in the visual cortex of early blind human subjects. *Brain research*, *446*(2), 369-373.

Wang, J., Ren, Y., Hu, X., Nguyen, V. T., Guo, L., Han, J., & Guo, C. C. (2017). Test–retest reliability of functional connectivity networks during naturalistic fMRI paradigms. *Human brain mapping*, *38*(4), 2226-2241.

Watkins, K. E., Shakespeare, T. J., O'Donoghue, M. C., Alexander, I., Ragge, N., Cowey, A., & Bridge, H. (2013). Early auditory processing in area V5/MT+ of the congenitally blind brain. *Journal of Neuroscience*, *33*(46), 18242-18246.

Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L. G., ... & Rauschecker, J. P. (2000). A positron emission tomographic study of auditory localization in the congenitally blind. *Journal of Neuroscience*, *20*(7), 2664-2672.

Weninger, F., Eyben, F., Schuller, B. W., Mortillaro, M., & Scherer, K. R. (2013). On the acoustics of emotion in audio: what speech, music, and sound have in common. *Frontiers in psychology*, *4*, 292.

Wiesel, T. N., & Hubel, D. H. (1965). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *Journal of neurophysiology*, *28*(6), 1029-1040.

Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cerebral cortex*, *18*(1), 230-242.

Wirthlin, M., Chang, E. F., Knörnschild, M., Krubitzer, L. A., Mello, C. V., Miller, C. T., ... & Yartsev, M. M. (2019). A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait. *Neuron*, *104*(1), 87-99.

Wong, C., Chabot, N., Kok, M. A., & Lomber, S. G. (2015). Amplified somatosensory and visual cortical projections to a core auditory area, the anterior auditory field, following early-and late-onset deafness. *Journal of Comparative Neurology*, *523*(13), 1925-1947.

Yaka, R., Yinon, U., & Wollberg, Z. (1999). Auditory activation of cortical visual areas in cats after early visual deprivation. *European Journal of Neuroscience*, *11*(4), 1301-1312.

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How we transmit memories to other brains: constructing shared neural representations via communication. *Cerebral cortex*, *27*(10), 4988-5000.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral cortex*, *11*(10), 946-953.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in cognitive sciences*, *6*(1), 37-46.

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *The Journal of physiology*, *236*(3), 549-573.

Zeman, A. A., Ritchie, J. B., Bracci, S., & de Beeck, H. O. (2020). orthogonal Representations of object Shape and category in Deep convolutional neural networks and Human Visual cortex. *Scientific Reports*, *10*(1), 1-12.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., ... & Torr, P. H. (2015). Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 1529-1537).