The brain dynamics of congenitally blind people seeing faces with sound

Pawel J. Matusz [1,2,#], Lior Reich[6,#], Louise A. Stolz[4,5], Chrysa Retsa[2,4], David A. Tovar[3], Ella Striem-Amit[6,8], Elena Aggius-Vella[7], Amir Amedi [6,7,*], Micah M. Murray [2,4,*]

1. Institute of Health Sciences, School of Health Sciences, HES-SO Valais-Wallis, Sion, Switzerland

2. The Sense Innovation and Research Center, Lausanne and Sion, Switzerland

3. Department of Psychology, Vanderbilt University, Nashville, TN, USA

4. Radiology Department, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

5. Department of Psychiatry at the University of California, San Diego, USA

6. Department of Medical Neurobiology, Faculty of Medicine, Hebrew University Jerusalem, Jerusalem, Israel

7. The Baruch Ivcher Institute for Brain, Cognition & Technology, The Baruch Ivcher School of Psychology, Interdisciplinary Center Herzliya, Reichman University, Herzliya, Israel[#] Equal contributions

8. Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20007, USA.

* Equal contributions and corresponding authors

Correspondence: micah.murray@chuv.ch or amir.amedi@runi.ac.il

Summary

Sensory substitution devices (SSDs) convert images to sounds to equip blind individuals with nominally visual functions, like face or letter sensitivity. Prior studies showed that image-to-sound SSDs engage cortices ordinarily specialised for visual functions. However, the brain dynamics of SSD-supported perception remains unknown. Either visual cortices are the first locus of discrimination of SSD percepts, or their activation is a byproduct of perceptual processes unfurling elsewhere, such as in auditory cortices. Resolving this uncertainty is critical for understanding whether the blind truly "see" via SSDs. Using electrical neuroimaging of EEG data from congenitally blind adults, we show for the first time that inferotemporal visual cortices are the earliest site of face sensitivity when conveyed via SSDs. The blind do indeed "see" with SSDs. By providing the temporal dynamics of SSD perception, our findings provide unique evidence for the theory that cortices are characterised by task-contingent functional organisation.

Keywords : blind, sensory substitution, EEG, electrical neuroimaging, sensory cortices, vision

Highlights

- Sensory substitution devices (SSDs) convert information from one sense to another
- SSDs can provide the blind with alternative access to visual information
- Brain dynamics in visual cortices to SSD-conveyed object images are unknown
- EEG electrical neuroimaging revealed face sensitivity within 500ms post-sound onset
- Localized to fusiform regions, this face sensitivity shows the blind see with sound

eTOC blurb

We show for the first time that inferotemporal visual cortices are the earliest site of face sensitivity when conveyed via SSDs. The blind do indeed "see" with SSDs. Our findings of the temporal dynamics of SSD perception uniquely support the theory that cortices are characterised by task-selective sensory-independent functional organisation.

Introduction

Visual-to-auditory sensory-substitution devices (SSDs) empower the blind to harness auditory information to perform nominally visual functions, like navigation or object recognition. Notably, the SSD-entrained sound stimuli can engage the same brain networks as those activated during visual processing in sighted individuals (Reich et al., 2012; see also van den Hurk et al., 2017; Mahon et al. 2009 for examples of category sensitivity in ventral occipital cortices in the blind). This highlights the brain's natural capacity for plasticity in the service of goal-directed behavior. In the absence of visual sensory inputs, the brain adapts to alternative inputs. A critical question is whether the blind truly "see" with other senses. Brain imaging and mapping methods such as fMRI and EEG are well-poised to address this. While fMRI can delineate *where* SSD-induced vision transpires, it cannot resolve the underlying brain dynamics of where and *when* these processes first occur. Either visual cortices are the initial locus of discrimination of SSD percepts, or their activation is a byproduct of perceptual processes unfurling first elsewhere. Resolving this uncertainty is critical for understanding whether the blind truly "see" via SSDs.

For both blind and sighted participants alike, auditory inputs delivered via auditory-to-visual SSDs enable high-level visual functions. Psychophysical studies have shown that SSDs equip people with skills in recognizing object categories as well as specific object features, e.g., spatial location, size, shape, texture, and orientation (Arno et al., 1999; Bermejo et al., 2015; Levy-Tzedek et al., 2012; Striem-Amit, Cohen, et al., 2012). This evidence in turn motivated the use of neuroimaging to identify the nature of processes underlying the SSD-enabled skills. If the entrained sounds first engage the same visual cortices as those observed in sighted individuals, this would be evidence that visual cortices can be repurposed for seeing in the blind using SSDs. In turn, if SSD-induced perceptual discrimination first occurs outside of visual cortices, this would be more in line with the blind using visual cortices not to see but rather to reinforce functions such as acoustic discrimination or associative learning.

Prior fMRI research has provided evidence for SSD-induced activation of visual cortices (Amedi et al., 2005, 2007; Collignon et al., 2007; Hertz & Amedi, 2014; Striem-Amit, Cohen, et al., 2012). For example, regions of the lateral occipital complex (LOC) that are typically involved in visual shape and object processing are also activated in response to SSD sounds of the same stimuli (Amedi et al., 2007). Regions of the fusiform gyrus have similarly been shown to respond to face stimuli delivered as soundscapes (Arbel et al., 2022). Still other data show that the large-scale functional organization of the visual cortices, as exemplified by dissociation between object and spatial processing, is also observed when the congenitally blind are trained to use SSD sounds (Striem-Amit, Dakwar, et al., 2012). More importantly, this work has shown that such activations are dependent on training; there were no responses in these areas when the participants were unfamiliar with the SSDs (e.g. Amedi et al., 2007; Kim & Zatorre, 2011; Striem-Amit, Cohen, et al., 2012).

This series of works provide the first line of support that SSD-induced activity in these visual areas is linked to perception. Collectively, this evidence has been formalized in the hypothesis that the brain is a *task-contingent machine*, whereby brain areas are characterized by their predominant function rather than the sensory modality inputs they

receive (Reich et al., 2012; also Striem-Amit, 2024). This task-contingent nature of the brain seems so strong, that several hours to weeks of training suffice for brain regions to adapt to a new type of sensory inputs as a source of task-relevant information. To support this task-contingent machine hypothesis more fully, it is necessary to ascertain if visual areas are themselves performing the key perceptual discrimination prior to other brain areas. Presently, this cannot be fully addressed with hemodynamic imaging, but rather requires methods that are time-resolved, such as M/EEG.

The present study applied electrical neuroimaging analyses to high-density EEG to characterize the brain dynamics of face sensitivity conveyed via SSDs in a group of congenitally blind individuals who successfully trained with the vOICe algorithm (Meijer, 1992; see also Methods for results of training). Electrical neuroimaging refers to a pipeline including analyses of response strength and topography as well as time-resolved source reconstructions (e.g. Murray et al., 2008; Michel and Murray 2012; Biasiucci et al., 2019). The vOICe converts pixelated images to sounds by representing vertical position of a pixel by pitch (low to high) and horizontal position of a pixel by time (sweeping left to right). After training and during EEG recordings, participants passively listened to soundscapes of faces, scrambled faces, as well as at letter stimuli (stimulus duration = 1570ms). We analyzed event-related potentials (ERPs) elicited by soundscapes representing line drawings of faces and their scrambled counterparts (**Fig. 1**). Line drawing stimuli are particularly well-suited for conversion by the vOICe algorithm. Because the soundscapes represented a specific, well-studied object category (i.e. faces), we had the advantage of clear predictions regarding the visual regions involved; notably the inferotemporal cortices, including the fusiform face area (Bentin et al., 1996). Specifically, we hypothesized that responses to soundscapes of faces, but not those to scrambled faces, would engage the inferotemporal cortices. Moreover, the timing of these activations, relative to any differential responses in auditory cortices, was crucial for our research question. We would therefore note that ERPs to letters were excluded here as they were acoustically distinct from soundscapes of faces and scrambled faces, confounding the interpretability of any differential responses. For example, the only prior EEG study of SSD training showed that the earliest differential responses to SSD stimuli were first in auditory cortices (Graulty et al., 2018), which runs counter to the task-contingent machine hypothesis. Additionally, they used geometric shapes as stimuli, and their sample consisted of sighted participants. As such, while their data indeed demonstrate brain plasticity following SSD training, the pertinence of their findings to the question of the nature of visual perceptions afforded by SSD in the blind is limited.

Results

Brain dynamics of SSD-enabled face sensitivity: Exemplary waveforms and mass univariate analyses

For demonstrative purposes, and to situate our results alongside publications that use waveform-based single-electrode analyses, we present the ERPs to the faces versus scrambled faces for three canonical electrodes used to report face-related ERP responses (**Fig.2, leftward panels)**. To determine the brain dynamics by which soundscapes of faces and scrambled faces are first discriminated, we conducted a mass univariate analysis across both

the peri-stimulus period (i.e., 100ms pre-stimulus to 900ms post-stimulus onset) and the entire 64-channel scalp electrode montage. This analysis entailed a non-parametric paired randomization test with 1000 shuffles at each sample. The results are displayed in **Figure 2 (rightward upper panel)**. To account for spatio-temporal autocorrelation, we only considered as reliable those effects that were temporally sustained (i.e., ≥10 consecutive data points; Blair and Karniski, 1993) and which encompassed at least 10% of the electrode montage (e.g. Retsa et al., 2024; Maitre et al., 2020; Tivadar et al., 2018). By such criteria, the face vs. scrambled face ERP differences first manifested at 492-620ms post-stimulus onset, with another statistically significant period observed at 632-716ms post-stimulus onset (**Fig.2, rightward upper panel**). Of note, there was no evidence of differential responses to face and scrambled face soundscapes during earlier time periods of the auditory evoked potential (i.e. the P50-N1-P2 complex) that have been shown in other works to be modulated by differences in low-level acoustic parameters and to involve prominent sources along the superior temporal gyrus (e.g. De Santis et al., 2007; Retsa et al., 2018).

Brain dynamics of SSD-enabled face sensitivity: Reference-independent global ERP analyses

Next, we performed multivariate analyses on reference-independent measures of the ERP across the full 64-channel montage; namely the ERP strength, as quantified by Global Field Power, and modulations in the ERP topography as quantified by Global Dissimilarity (Biasiucci et al., 2019; Murray et al., 2008). For both metrics, analyses were conducted across the peri-stimulus period, as detailed above, and included a contiguous temporal criterion of ≥10 samples for an effect to be considered reliable. There was no evidence for statistically reliable modulations in response strength between conditions. By contrast, there was evidence for statistically significant topographic modulations over the 484-756ms period as well as 804-852ms period (**Fig.2, rightward lower panel**). Thus, both univariate and multivariate analyses point to a similar timing for soundscape discrimination. Moreover, the multivariate analyses indicate that the earliest detectable ERP modulation followed from a change in the topography of the electric field at the scalp, which biophysical laws dictate must follow from changes in the underlying configuration of active sources in the brain (Murray et al., 2008). In other words, different brain networks subserved responses to soundscapes of faces versus scrambled faces.

To determine if these brain networks are temporally stable, we performed a hierarchical clustering analysis of the ERP topography, using as input the concatenated group-averaged data in response to soundscapes of faces and scrambled faces. The topographic clustering indicated that 7 different template maps accounted for 97.91% of the global explained variance of the dataset (**Fig.3A**). Until 456ms post-stimulus onset, the same sequence of four topographic patterns were observed in the group-average responses to both soundscapes of faces and scrambled faces. This sequence included topographies consistent with the P50-N1-P2 complex of auditory ERP components. However, over the 460-716ms period, the topographies of ERP responses in those conditions started to differ (**Fig.3B**). Specifically, one topography (**Fig.3B, upper panel, red frame**) appeared to better characterize responses to soundscapes of faces, whereas another topography (**Fig.3B, upper panel, grey frame**) appeared to better characterize ERP responses to soundscapes of scrambled faces. This was

statistically tested by performing a competitive spatial correlation between each template map observed at the group-average level and the single-subject data from both conditions. This procedure yields a measure of how many time samples over the 460-716ms period exhibited a higher spatial correlation with each template map, which was in turn submitted to a Friedman's 2-way non-parametric ANOVA by ranks that yielded a chi-square value that was significant (X^2=9.82; p=0.020). On average, the template map framed in red better characterized responses to soundscapes of faces and another map, framed in dark gray, better characterized responses to soundscapes of scrambled faces (**Fig.3B, lower leftward panel**). This observation was also statistically tested for each individual participant using a Chi-square test comparing the distribution of the number of time samples characterized by each template map for responses to soundscapes of faces (actual distribution) with the distribution observed for responses to soundscapes of scrambled faces (expected distributions). Statistically significant differences were observed for 7 of the 9 participants (all p's ≤0.01; **Fig.3B, lower rightward panel**).

A similar pattern of distinct topographies characterizing responses to soundscapes of faces versus scrambled faces was also observed over the 720-900ms post-sound onset period (**Fig.3C**). As above, this was statistically tested at the group level with Friedman's 2-way non-parametric ANOVA by ranks. This test yielded a chi-square value that was significant (X^2=9.62; p=0.022). On average, one template map better characterized responses to soundscapes of faces and another map (**Fig.3B, upper and lower leftward panel, red and grey frames, respectively**) framed in light gray better characterized responses to soundscapes of scrambled faces. This observation was also statistically tested for each individual participant using a chi-square test comparing the distribution of the number of time samples characterized by each template map for responses to soundscapes of faces (actual distribution) with the distribution observed for responses to soundscapes of scrambled faces (expected distribution). Significant differences were observed for 6 of the 9 participants (all p's ≤0.001; (**Fig.3B, lower panels**).

Brain dynamics of SSD-enabled face sensitivity: Source modelling

Data from the 460-716ms post-sound onset time were then submitted to EEG source modelling (**Fig.4**). The statistical contrast revealed significantly stronger (p<0.05 for >10 contiguous solution points) responses to soundscapes of faces than scrambled faces predominantly within the right inferotemporal cortices as well as within the junction between the right superior temporal cortex and insula (**Fig.4**). Within the inferotemporal cortices, there were 2 clusters: one with a local maximum in the fusiform gyrus (27, -42, -11mm) and the other with a local maximum within the right lateral occipital cortex (42, -72, -4mm). A third cluster was observed within the right insula, extending into the right superior temporal cortex (maximum at 42, -4, -4mm). These results indicate that higher-order visual cortices are involved in the initial discrimination of soundscapes of faces versus scrambled faces.

Discussion

Our study provides the first characterization of the brain dynamics of SSD-mediated face recognition skills in the blind. ERPs to soundscapes of faces differed from those to scrambled faces within the first 460ms post-soundscape onset and followed from differential activation of brain networks that were localized to fusiform and anterior temporal cortices. The spatio-temporal brain dynamics, afforded by electrical neuroimaging analyses of our EEG data, has allowed us to provide the key missing evidence that an image-to-sound SSD indeed enables "seeing with sounds". More generally, our findings uniquely reinforce the extant evidence for the profoundly task-contingent nature of the brain.

Image-to-sound sensory substitution has been shown to enable the recognition of a wide range of object categories from simple shapes through letters to faces and facial emotional expressions and body shape (Arno et al., 1999; Reich et al., 2012; Striem-Amit, Cohen, et al., 2012; Striem-Amit & Amedi, 2014; Arbel et al., 2022). The hemodynamic imaging research, which hitherto dominated this research area, have provided an aggregate purview of the brain networks involved in SSD perceptions. Prior fMRI works have shown that responses in the visual word form area are observed after, but not before, training with the vOICe device (cf. Figure 4 in Striem-Amit, Cohen, et al., 2012). Other studies have documented sensitivity to SSD-conveyed body shapes within extrastriate body areas in the absence of evidence for sensitivity to such in auditory cortices of the superior temporal cortex (Striem-Amit & Amedi, 2014). However, despite subjective reports of the qualia of the substitution-mediated visual skills (Striem-Amit, Cohen, et al., 2012), hemodynamic imaging research cannot provide direct evidence for whether those visual skills activate as first the brain areas traditionally considered as visual, or the latter are activated only as part of a widespread network of brain regions. Recently, dynamic causal modelling has been applied to EEG data from blindfolded sighted participants who learned to use an SSD (Kim et al., 2024). The results support a prominent role therein of visual mental imagery rather than of cross-modal interactions. However, the conducted analyses did not characterize either the temporal dynamics (i.e., only oscillatory responses in a priori defined regions of interest were analyzed) or the capacity to discriminate specific categories of soundscapes (i.e., the data were collapsed across all correctly perceived stimuli). Thus, despite the presence of EEG findings in the area prior to the present study, the precise spatio-temporal brain dynamics of soundscape discrimination remained unknown.

Our study leveraged a combination of fine temporal resolution of EEG with the high spatial resolution of distributed EEG source modelling (collectively referred to as electrical neuroimaging; (Michel & Murray, 2012) to provide key missing evidence regarding the brain dynamics of SSD-conveyed face perception. We capitalized upon the well-understood nature of brain sensitivity to faces. The brain detects faces through a network of regions centered around the fusiform gyrus (Allison et al., 1994; Bentin et al., 1996; Kanwisher et al., 1997) and lateral occipital face areas (Pitcher et al., 2011). Our source localization results situate the earliest discrimination between soundscape faces and soundscape scrambled faces precisely within this network of fusiform and lateral occipital cortices (**Figure 4**).

It is worthwhile to situate the timing of the substitution-mediated face processing observed here alongside the more general findings of auditory object discrimination, as

observed in sighted adults. Prior works have characterized a spatio-temporal hierarchy whereby animacy is discriminated within the initial 70-100ms of brain responses (Murray et al., 2006), followed by discrimination of conspecific vocalizations at ~170–220ms and at ~290–360ms (Charest et al., 2009; De Lucia, Clarke, et al., 2010) in turn followed by discrimination of action-related vs. action-free sounds of objects, at ~300-360ms (De Lucia et al., 2009)(reviewed in (Spierer et al., 2011). Additional work using single-trial EEG classification has shown there to be at least two phases of brain activity involved in perceptual decision-making regarding sound animacy: the first phase, spanning ~110–180ms and is independent of perceptual capabilities, and the second phase, at ~270–350ms, that manifests only during accurate categorizations (De Lucia et al., 2012). The latency of our effects, while slightly later, indicates that also in our study, more complex perceptual decision-making processes had to be engaged, to distinguish faces from scrambled faces. In support with this idea, a study of Philiastides et al. (2006), involving perceptual decision-making with visual stimuli, has shown activations only at ~450ms, when the task conditions were challenging. It is likewise important to consider that the soundscape presentation itself necessitates a temporal unfurling of the stimulus (i.e. the soundscape is rasterized from left to right). As such, the latency of the first effects indicative of sensitivity to soundscapes of faces at 460ms post-sound onset is not unreasonably "late" and is well within the idea that rapidly activated, sensory representations underlie substitution-mediated object recognition skills. Moreover, this latency aligns with the results of an ERP study of the discrimination of soundscapes of geometric forms in the sighted, wherein the recognition effects were observed 420–480 post-soundscape onset for the participants who learnt the vOICe SSD (Graulty et al., 2018).

Our study has several potential limitations. We obtained the data from a small, predominantly female sample, which may potentially curtail the strength of our conclusions. This sample size limitation is offset on the one hand by the large effect sizes that we have observed, and on the other hand by the single-subject analyses that we have performed; significant effects were observed for 7 of our 9 participants (see pie charts in **Figure 3**). The strength of conclusions is further supported by the fact that fMRI data were also obtained from many of these participants (Arbel et al., 2022). More generally, our study highlights the added value of deeper data in situations of special or clinical populations (Ramon & Striem-Amit, 2022). The fact that participants listened passively to the soundscapes might also be considered a potential shortcoming. However, the psychophysical data obtained during a separate session would indicate that participants were able to discriminate soundscapes of different stimulus categories at near-ceiling levels. Thus, the SSD-mediated skills can be regarded as successfully acquired in our study. Additionally, the use of a passive paradigm has removed the risk of contamination of the categorical brain responses of interest by (pre)motor-related activations, which would be especially likely given the relatively "late" latency of the present effects. Furthermore, the category of soundscapes was randomized across trials and multiple exemplars from the same category were presented. These paradigm features make it unlikely that the present effects follow from any potential differences in allocation of attention. Additionally, the timing and sources of our effects are not readily explained in terms of attention modulations, which have been reported to occur

within the initial 400ms post-sound onset (e.g. Hansen & Woldorff, 1991). Another shortcoming of our study was that ERP data were acquired only after, but not before, training with the vOICe SSD. Ideally, the data would have been obtained prior to and during training to show that the discrimination we report was not pre-existing and followed from training. The practical considerations of such a design notwithstanding, recordings prior to and during training might provide insights on how this ability is procured (see Graulty et al., 2018, for an example in sighted participants). However, this point would not detract from the principal observation here, that visual cortices canonically involved in visual face discrimination are the first areas to exhibit discrimination when faces are conveyed as soundscapes even after adequate training. That is, one might even speculate that over-training could shift SSD-induced perceptions into auditory cortices, rather than engaging nominally visual areas. This proposition is in fact evident in the findings of Graulty et al. (2018), where the effects at 150 – 210ms post-soundscape onset correlated with performance accuracy.

More generally, our findings contribute to the accumulating evidence that the brain is fundamentally a task-contingent machine wherein neural architecture is constructed to fulfill a set of functions, rather than being defined by its nominal input sensory modality (Murray et al., 2016; Reich et al., 2012; Striem-Amit, 2024) . Our confirmation of this hypothesis has far-reaching implications. It follows that even brains that have been devoid of canonical sensory experience necessary for the regular functional brain organization to develop can achieve functional organization comparable to that observed in sighted individuals. However, this may be the case only if the missing sensory information becomes substituted with similar/complementary information through another sense. This notion would suggest that many of the findings indicating mass plasticity and cortical repurposing reported in the brains of the congenitally blind (Bedny et al., 2015) do not preclude the recruitment of these same brain regions to perform complex object recognition based on SSD training (see e.g. Striem-Amit 2024 for discussion of these concurrent theories). In conclusion, the current findings build on the extant evidence from the multiple hemodynamic and the scarce EEG studies, while simultaneously extending them critically, to offer new evidence for the spatio-temporal brain dynamics of SSD-elicited object recognition skills and the ability of the blind to see with sound.

Acknowledgments

Portions of this work have been funded by The Swiss National Science Foundation (grant 169206 to MMM), a fellowship from the Merkin Graduate Fellowship to LS, an ERC Consolidator grant "NovelExperiSENSE" to AA, and an Israeli Science Foundation grant to AA. We wish to acknowledge the contributions of Dr. Shlomo Bentin to this work.

Author contributions

PJM: Writing of first draft, data analysis, edits and revisions

LR: Data collection, edits and revisions

LS: Writing of first draft, data curation, data analysis, edits and revision

CR: Data curation, data analysis

DA: Data analysis

ESA: Edits and revisions

EAV: Data collection, edits and revisions

AA: Conceptualization, supervision of data collection, study design, edits and revisions, and funding acquisition.

MMM: Conceptualization, writing of first draft, data curation, data analysis, edits and revisions, funding acquisition.

Declaration of interests

The authors have no conflicts of interest.

Inclusion and diversity statement

As this was a convenience sample of congenitally blind individuals, gender balance and ethics or other types of diversity were challenging to ensure. Nonetheless, we worked to achieve the best balance possible. We worked to ensure that the study questionnaires were prepared in an inclusive way. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Figures and Figure legends

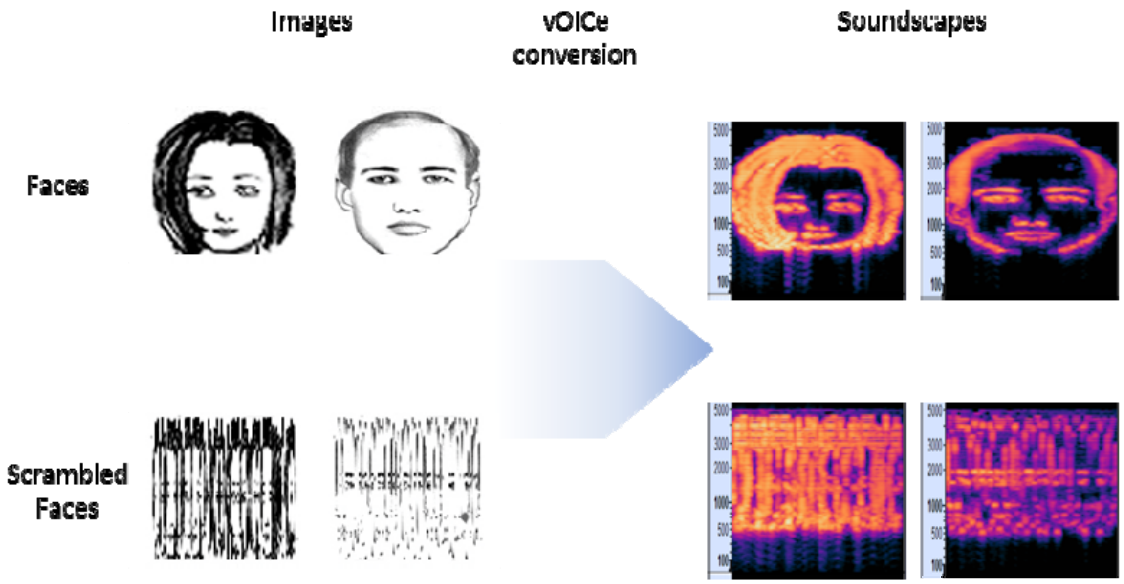

Figure 1. Image stimuli and spectrograms of soundscapes. Images included line drawings of faces and scrambled faces, examples of which are displayed. After conversation via the vOICe algorithm, soundscapes retain image properties in their spectrogram.

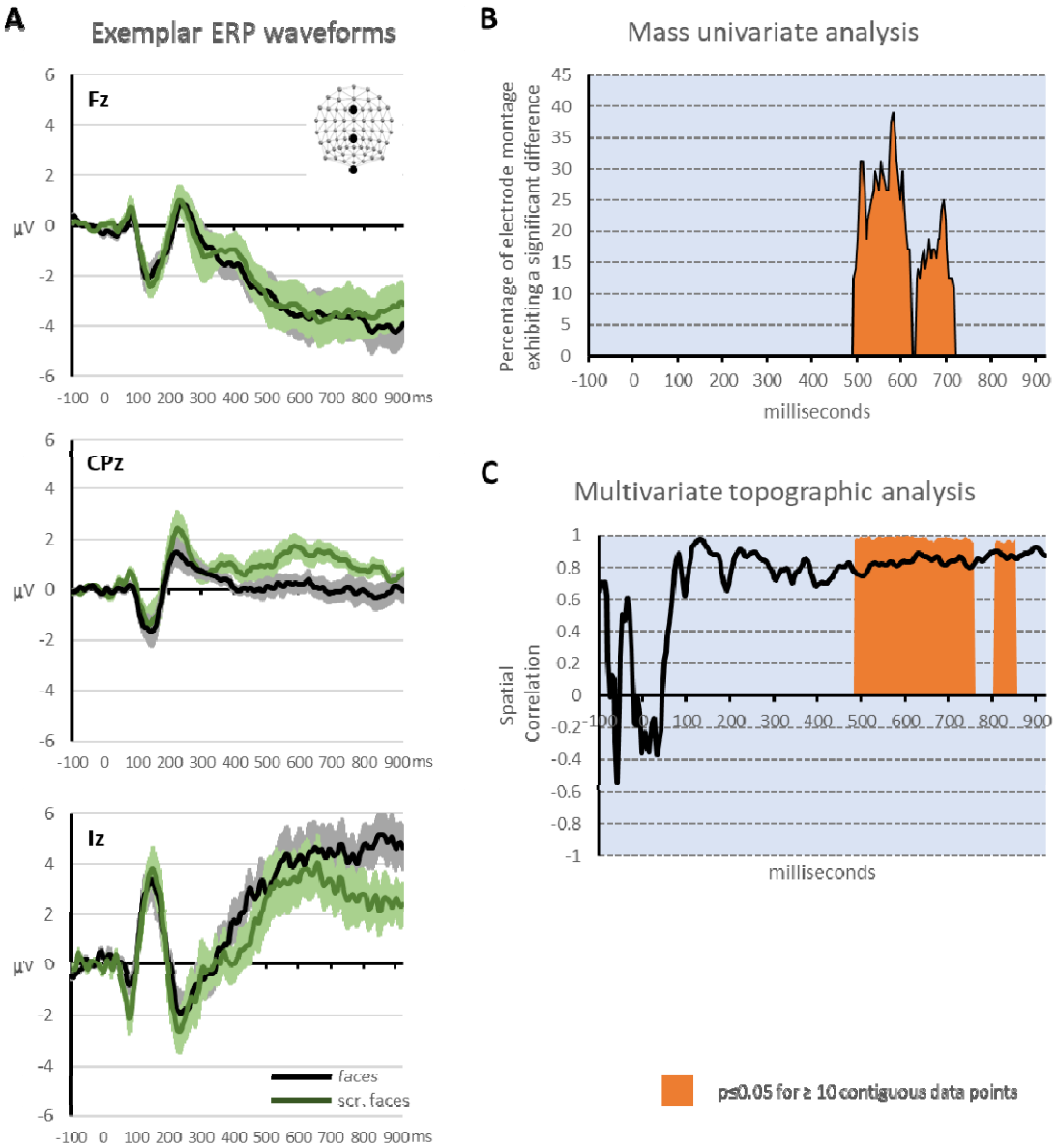

Figure 2. Exemplar voltage ERP waveforms, mass univariate analysis, and multivariate topographic analysis. (A) Mean (±SEM) voltage waveforms at a sample of midline scalp locations (see inset). The black traces are responses to soundscapes of faces, and the green traces are responses to soundscapes of scrambled faces. (B) Results of mass univariate analyses show the percentage of electrodes exhibiting a significant difference across the electrode montage as a function of time. (C) The black trace displays the spatial correlation between brain responses to face and scrambled face soundscapes as a function of time. The orange area displays the significant results of the multivariate topographic analysis (1 minus p-value shown) as a function of time.

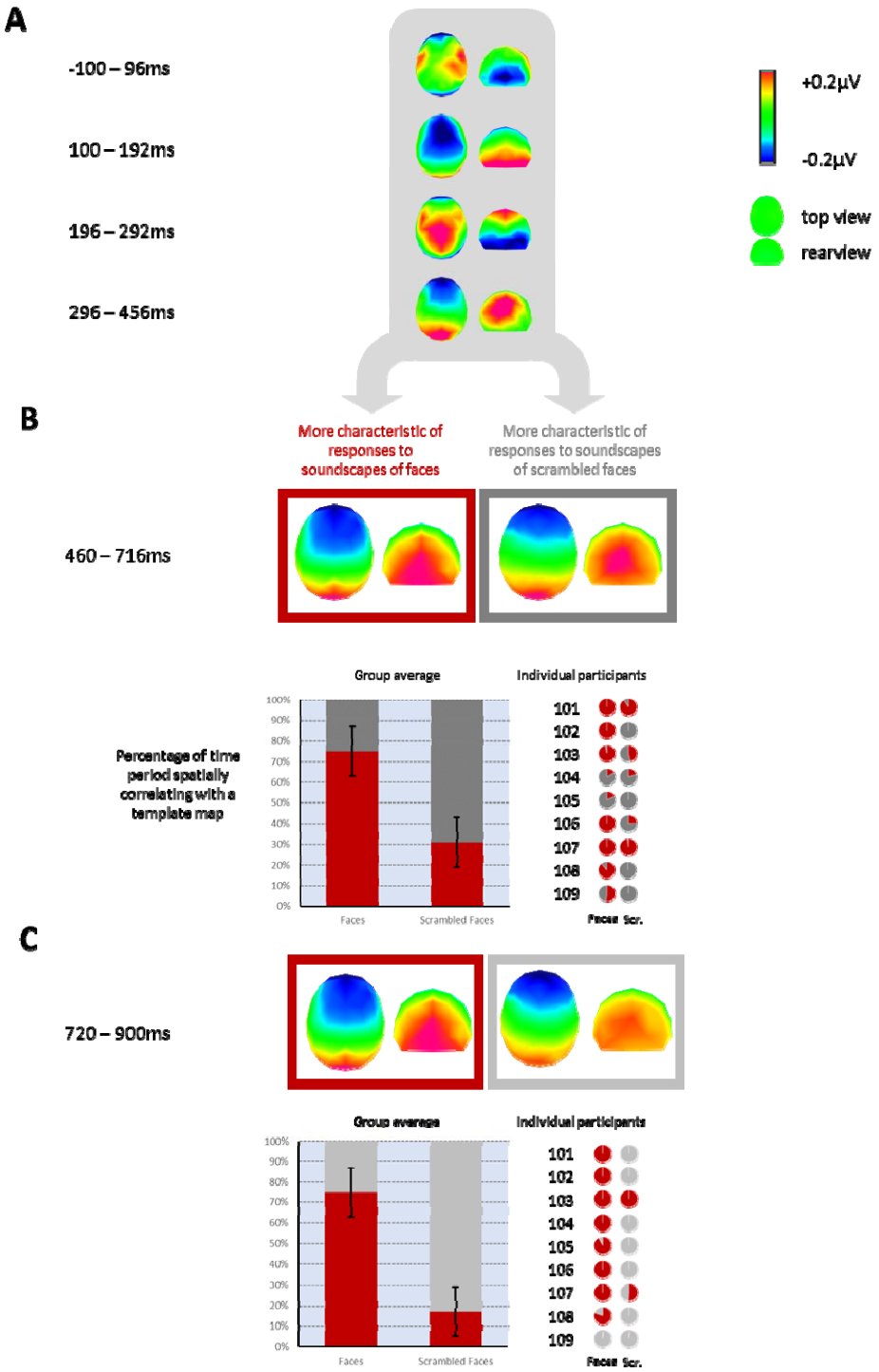

Figure 3. Topographic clustering results. (A) The group-averaged ERPs in response to face and scrambled face soundscapes were submitted to hierarchical clustering. Over several peristimulus onset time windows from -100 to 456ms identical ERP topographies were observed across both conditions. (B) over the 460-716ms post-stimulus time period, distinct maps appeared to characterize responses to face and scrambled face soundscapes (red- and gray- framed topographies, respectively). This observation was confirmed both at the group-average and single-subject levels

(bar and pie charts, respectively). (C) Similar differences were observed over the 720-900ms post-stimulus time period.

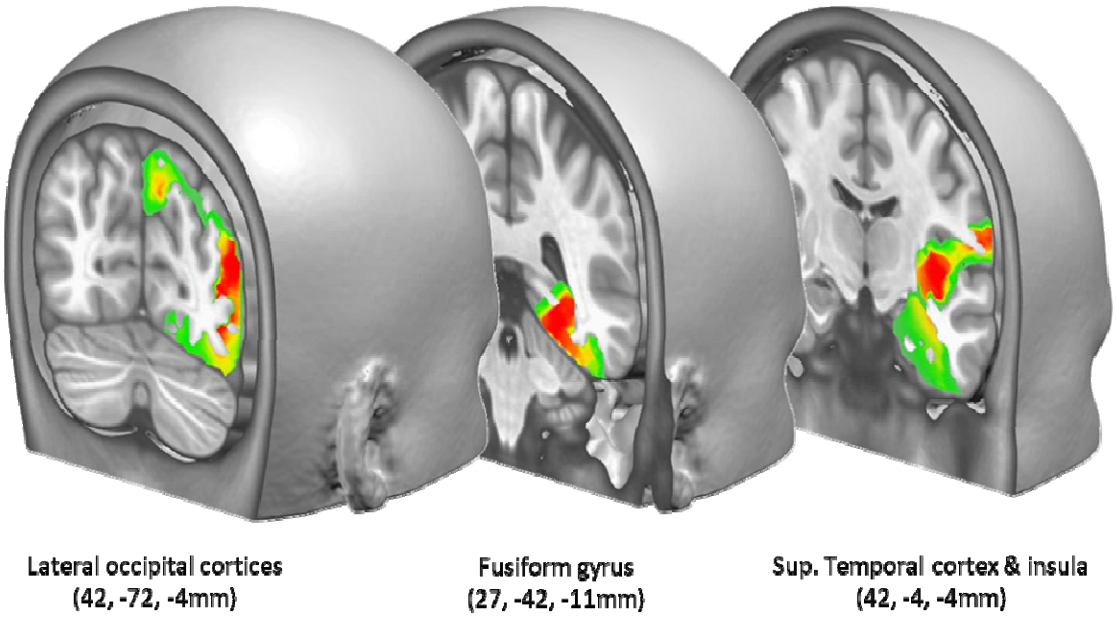

Figure 4. Source modelling analyses. Data from the 460-716ms post-sound onset time were then submitted to distributed linear source modelling. Stronger responses ($p<0.05$ for >10 contiguous solution points) were observed in three clusters; the coordinates of the local maximum are indicated.

Tables

Table 1. Demographic information for each of the nine individuals.

Sex	*Age*	*Condition*	*Light perception*	*Age of blindness onset (years)*
F	30	ROP	None	0
M	37	ROP	None	0
F	23	Retinal detachment	None	0
F	23	Microphthalmia, Retinal detachment	None	1
F	29	Microphthalmia	None	0
F	38	Congenital rubella	None	0
M	54	ROP	None	0
F	35	ROP	None	0
F	30	LCA	Faint	0

ROP = retinopathy of prematurity; LCA = Leber congenital amaurosis

STAR Methods

Participants

Nine congenitally blind individuals (7 women) participated in the study (mean age = 33.2 years, standard error of the mean [SEM] = 3.2yrs.). **Table 1** outlines the demographics of the participants including age, aetiology of blindness, and prior experience with the soundscape apparatus. Participants were recruited for this experiment from the group of participants that underwent long-term training with the vOICe SSD (Striem-Amit et al., 2012). For all participants, consent was obtained prior to testing. The procedure was evaluated and approved by The study was approved by the ethics committee of the Hebrew University of Jerusalem. Data acquisition was completed between January and June 2012.

Procedure & stimuli

Participants were trained on a visual-to-auditory SSD termed the "vOICe (Meijer, 1992). This device converts images into sounds, wherein position along the y-axis is translated into sound pitch (low to high), position along the x-axis is translated into time (left to right equals earlier to later in the sound sweep), and brightness is translated into volume. As the images here were black-and-white, there was no variation in volume. Example images used to generate soundscapes are shown in Figure 1. For more details on the training procedures and protocols, (see Striem-Amit et al., 2012). Before participating in the EEG portion of the study, each individual had completed over 70 hours of training with the vOICe over a period of several months. They then completed the EEG portion of the study with no task required. The EEG portion was comprised of 4 blocks of 60 trials each. In total, each participant was exposed to 240 soundscapes (i.e., 3 categories × 10 exemplars per category × 8 repetitions of each soundscape). Each soundscape had a duration of 1570ms. The inter-stimulus interval was comprised of 2 seconds of silence. During the EEG portion, soundscapes were heard passively. After the EEG, they then completed a 3-AFC discrimination task to ensure that each participant could reliably differentiate soundscapes of faces, scrambled faces, and Hebrew letters. This behavioural task entailed 90 trials (3 categories × 10 exemplars × 3 repetitions of each stimulus). Data from two participants on the behavioural task were corrupted and therefore could not be analysed. The performance of the remaining 7 individuals indicated near-ceiling accuracy (mean±s.e.m. faces = 95.7±1.6%; scrambled faces = 97.1±1.1%; letters = 97.6±1.0%). Reaction times on correct trials differed across conditions (faces = 1079±109ms; scrambled faces = 842±90ms; letters = 958±72ms) and were faster for soundscapes of scrambled faces than faces, though no other contrast was statistically significant.

EEG recording and pre-processing

Continuous EEG was acquired at 256Hz through a 64-channel Biosemi ActiveTwo AD-box (www.biosemi.com), referenced to the common mode sense (CMS, active electrode) and grounded to the driven right leg (DRL; passive electrode), which functions as a feedback loop driving the average potential across the electrode montage to the amplifier zero. Data preprocessing and analyses were performed using the Cartool freeware (version 5.01; Brunet et al., 2011). Prior to epoching, the EEG was filtered (low-pass 40Hz; high-pass 0.1Hz; removed DC; 50Hz notch; using a second-order Butterworth filter with -12dB/octave roll-off

that was computed linearly in both forward and backward directions to eliminate phase shifts). Peri-stimulus epochs, spanning 100ms pre-stimulus to 900ms post-stimulus onset, were averaged from each subject for each condition to compute event-related potentials (ERPs). Epochs were rejected based on automated artefact rejection criterion of ±100µV as well visual inspection for eye blinks and movement or other sources of transient noise. The average (±SD) number of accepted EEG epochs for soundscapes of faces was 78.3 ± 2.06 for face (i.e., 98% of available trials) and for soundscapes of scrambled faces was 79.4 ± 1.33 (i.e., 99.8% of available trials). These values did not significantly differ. Bad channels were identified before averaging and excluded from the artefact rejection. These data at artefact electrodes from each participant were interpolated using 3-D splines prior to group averaging (Perrin et al., 1987). The average number of interpolated channels was 0.25±0.44 (mean±SD). Additionally, we applied a pre-stimulus baseline correction and recalculated the data to the average reference.

ERP analyses design

Differences in the ERP responses to soundscapes of faces and scrambled faces were assessed using a multi-step analysis procedure, referred to as electrical neuroimaging, which involves both local and global measures of the electric field at the scalp. These methods have been described in detail previously (Koenig et al., 2011; Michel & Murray, 2012; Murray et al., 2008; Tzovara et al., 2012)

First, we analysed the ERP voltage waveform data from each scalp electrode as a function of time using a non-parametric paired randomisation test with 1000 shuffles for each datapoint. For this mass univariate analysis a temporal criterion of effects lasting >10 continuous datapoints and a spatial criterion of extending across >10% of the electrode montage were combined to account for autocorrelation along both dimensions (Guthrie & Buchwald, 1991). We would hasten to note, however, that our interpretations here are not based on this reference-dependent univariate analysis, but rather on the multivariate reference-independent analyses detailed below. These analyses are nonetheless included here to assist readers potentially more accustomed to canonical ERP analyses and to facilitate the comparison with multivariate, electrical neuroimaging.

The first reference-independent multivariate metric we analysed was the global field power (GFP), which is the spatial standard deviation across the electrode montage (Lehmann et al., 1987; Lehmann & Skrandies, 1980) A stronger GFP value is indicative of greater and/or more synchronised brain activity, though the root cause (increased neural firing rate, increased numbers of active neurons, etc.) cannot be unequivocally asserted based on this measure alone. However, a modulation of GFP in the absence of reliable evidence for topographic modulations can most parsimoniously be interpreted as a modulation in the strength of responses originating from statistically indistinguishable sources or set of sources. The GFP was statistically analysed using a non-parametric paired randomization, identical to that used for ERP voltage waveforms, above.

The second reference-independent multivariate metric we analysed was the global map dissimilarity (GMD). The GMD is at the core of data-driven topographic clustering (Murray et al., 2008) This clustering was performed on the group-averaged data from both conditions. It is done to identify stable electric field topographies, referred to as “template maps”, and their pattern both in time and across conditions. The clustering is insensitive to pure amplitude modulations across conditions as the data are first normalised by their instantaneous GFP. The optimal number of template maps that explained the whole group-

averaged data set was determined using a meta-criterion (Brechet et al., 2019). The clustering makes no assumption regarding the orthogonality of the derived template maps (De Lucia, Michel, et al., 2010; Pourtois et al., 2008). The pattern of maps identified in the group-average ERPs is used as a hypothesis generation tool that was then tested on single-subject data. was then submitted to a fitting procedure wherein each time point of each individual participant's ERP is labelled according to the template map with which it best correlated spatially (Murray et al., 2008). This yields a measure of relative map presence over a fixed time window (expressed as a percentage) that was then submitted to a Friedman's 2-way non-parametric ANOVA by ranks. This fitting procedure revealed whether responses to one type of soundscape were more often described by one ERP topography versus another, and therefore whether different intracranial generator configurations better accounted for responses to particular soundscapes. That is, biophysical laws indicate that a change in the topography of the electric field at the scalp surface forcibly originate from changes in the configuration of the underlying intracranial sources (reviewed in Michel & Murray, 2012).

Finally, we modelled the underlying intracranial sources of the ERPs in response to the two different types of soundscapes using a distributed linear inverse solution (minimum norm) combined with the LAURA (local autoregressive average) regularisation approach (Grave De Peralta Menendez et al., 2004; see also Michel et al., 2004 for a comparison of inverse solution methods). The solution space was calculated on a realistic head model that included 3000 nodes, selected from a grid equally distributed within the grey matter of the Montreal Neurological Institute's average brain (available from https://github.com/DenisBrunet/Cartool). The head model and lead field matrix were generated with the Spherical Model with Anatomical Constraints (SMAC; Spinelli et al., 2000 as implemented in Cartool version 5.01 (Brunet et al., 2011), using a 6-shell model (CSF, skull compact, skull spongy, skull compact and scalp layers) as well as with an upper skull thickness (5.75mm) and mean conductivity (0.017S/m) values based on the mean age of our sample. As an output, LAURA provides current density measures; their scalar values were evaluated at each solution point. Statistical analysis of source estimations was performed by first averaging the ERPs across time for each participant and condition over the time window exhibiting a stable topography, described above. A non-parametric paired randomisation test was conducted across all solution points (1000 shuffles). The statistical significance criterion at an individual solution point was set at $p<.0.05$. Only clusters with ≥10 contiguous significant solution points were considered reliable in an effort to correct for multiple comparisons and was based on randomization thresholds (see also (De Lucia et al., 2012; Knebel & Murray, 2012; Matusz et al., 2019; Retsa et al., 2018, 2020, 2024) for similar implementations).

Resource availability

Lead contact. Further information and requests for resources should be directed to and will be fulfilled by the lead contacts, Micah Murray (micah.murray@chuv.ch) and Amir Amedi (amir.amedi@runi.ac.il).

Material, data and code availability.

The stimuli, raw data and processed EEG data will be made available upon reasonable requests. The data were analysed using a freely available, GUI-based software for EEG analyses Cartool (https://github.com/DenisBrunet/Cartool).

References

Allison, T., Ginter, H., McCarthy, G., Nobre, A. C., Puce, A., Luby, M., & Spencer, D. D. (1994). Face recognition in human extrastriate cortex. *Journal of Neurophysiology*, *71*(2), 821–825. https://doi.org/10.1152/jn.1994.71.2.821

Amedi, A., Merabet, L. B., Bermpohl, F., & Pascual-Leone, A. (2005). The Occipital Cortex in the Blind: Lessons About Plasticity and Vision. *Current Directions in Psychological Science*, *14*(6), 306–311. https://doi.org/10.1111/j.0963-7214.2005.00387.x

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., Hemond, C., Meijer, P., & Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, *10*(6), 687–689. https://doi.org/10.1038/nn1912

Arbell, R., Heimler, B., & Amedi, A. (2022). Congenitally blind adults can learn to identify face-shapes via auditory sensory substitution and successfully generalize some of the learned features. Scientific Reports, 12(1), 4330. https://doi.org/10.1038/s41598-022-08187-z.

Arno, P., Capelle, C., Wanet-Defalque, M.-C., Catalan-Ahumada, M., & Veraart, C. (1999). Auditory Coding of Visual Patterns for the Blind. *Perception*, *28*(8), 1013–1029. https://doi.org/10.1068/p281013

Bedny, M., Richardson, H., & Saxe, R. (2015). "Visual" Cortex Responds to Spoken Language in Blind Children. *The Journal of Neuroscience*, *35*(33), 11674–11681. https://doi.org/10.1523/JNEUROSCI.0634-15.2015

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551. https://doi.org/10.1162/jocn.1996.8.6.551

Bermejo, F., Di Paolo, E. A., HÃ¼g, M. X., & Arias, C. (2015). Sensorimotor strategies for recognizing geometrical shapes: A comparative study with different sensory substitution devices. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00679

Biasiucci, A., Franceschiello, B., & Murray, M. M. (2019). Electroencephalography. *Current Biology*, *29*(3), R80–R85. https://doi.org/10.1016/j.cub.2018.11.052

Brechet, L., Denis Brunet, Birot, G., Gruetter, R., Michel, C., & Jorge, J. (2019). Capturing the spatiotemporal dynamics of self-generated, task-initiated thoughts with EEG and fMRI. *NeuroImage*, *194*, 82–92. https://doi.org/10.1016/j.neuroimage.2019.03.029

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/813870

Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., Chartrand, J.-P., & Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience*, *10*(1), 127. https://doi.org/10.1186/1471-2202-10-127

Collignon, O., Lassonde, M., Lepore, F., Bastien, D., & Veraart, C. (2007). Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cerebral Cortex*, *17*(2), 457–465. https://doi.org/10.1093/cercor/bhj162

De Lucia, M., Camen, C., Clarke, S., & Murray, M. M. (2009). The role of actions in auditory object discrimination. *NeuroImage*, *48*(2), 475–485. https://doi.org/10.1016/j.neuroimage.2009.06.041

De Lucia, M., Clarke, S., & Murray, M. M. (2010). A Temporal Hierarchy for Conspecific Vocalization Discrimination in Humans. *The Journal of Neuroscience*, *30*(33), 11210–11221. https://doi.org/10.1523/JNEUROSCI.2239-10.2010

De Lucia, M., Michel, C. M., & Murray, M. M. (2010). Comparing ICA-based and Single-Trial Topographic ERP Analyses. *Brain Topography*, *23*(2), 119–127. https://doi.org/10.1007/s10548-010-0145-y

De Lucia, M., Tzovara, A., Bernasconi, F., Spierer, L., & Murray, M. M. (2012). Auditory perceptual decision-making based on semantic categorization of environmental sounds. *NeuroImage*, *60*(3), 1704–1715. https://doi.org/10.1016/j.neuroimage.2012.01.131

Graulty, C., Papaioannou, O., Bauer, P., Pitts, M. A., & Canseco-Gonzalez, E. (2018). Hearing shapes: Event-related potentials reveal the time course of auditory–visual sensory substitution. *Journal of Cognitive Neuroscience*, *30*(4), 498–513. https://doi.org/10.1162/jocn_a_01210

Grave De Peralta Menendez, R., Murray, M. M., Michel, C. M., Martuzzi, R., & Gonzalez Andino, S. L. (2004). Electrical neuroimaging based on biophysical constraints. *NeuroImage*, *21*(2), 527–539. https://doi.org/10.1016/j.neuroimage.2003.09.051

Guthrie, D., & Buchwald, J. S. (1991). Significance Testing of Difference Potentials. *Psychophysiology*, *28*(2), 240–244. https://doi.org/10.1111/j.1469-8986.1991.tb00417.x

Hansen, J.C., & Woldorff, M. (1991) Mechanisms of auditory selective attention as revealed by event-related potentials. *Electroencephalogr Clin Neurophysiol Suppl.* 42, 195-209.

Hertz, U., & Amedi, A. (2014). Flexibility and Stability in Sensory Processing Revealed Using Visual-to-Auditory Sensory Substitution. *Cerebral Cortex (New York, N.Y.: 1991)*, bhu010-. https://doi.org/10.1093/cercor/bhu010

Kanwisher, N., Woods, R. P., Iacoboni, M., & Mazziotta, J. C. (1997). A Locus in Human Extrastriate Cortex for Visual Shape Analysis. *Journal of Cognitive Neuroscience*, *9*(1), 133–142. https://doi.org/10.1162/jocn.1997.9.1.133

Kim, H., Kim, J. S., & Chung, C. K. (2024). Visual Mental Imagery and Neural Dynamics of Sensory Substitution in the Blindfolded Subjects. *NeuroImage*, *295*, 120621. https://doi.org/10.1016/j.neuroimage.2024.120621

Kim, J.-K., & Zatorre, R. J. (2011). Tactile–Auditory Shape Learning Engages the Lateral Occipital Complex. *The Journal of Neuroscience*, *31*(21), 7848–7856. https://doi.org/10.1523/JNEUROSCI.3399-10.2011

Knebel, J. F., & Murray, M. M. (2012). Towards a resolution of conflicting models of illusory contour processing in humans. *NeuroImage*, *59*(3), 2808–2817. https://doi.org/10.1016/j.neuroimage.2011.09.031

Koenig, T., Kottlow, M., Stein, M., & Melie-garc, L. (2011). Ragu: A Free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computational Intelligence and Neuroscience*, *2011*, 1–15. https://doi.org/10.1155/2011/938925

Lehmann, D., Ozaki, H., & Pal, I. (1987). EEG alpha map series: Brain micro-states by space-oriented adaptive segmentation. *Electroencephalography and Clinical Neurophysiology*, *67*, 271–288.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*, *48*, 609–621.

Levy-Tzedek, S., Hanassy, S., Abboud, S., Maidenbaum, S., & Amedi, A. (2012). Fast, accurate reaching movements with a visual-to-auditory sensory substitution device. *Restorative Neurology and Neuroscience*, *30*(4), 313–323. https://doi.org/10.3233/RNN-2012-110219

Matusz*, P. J., Turoman, N., Tivadar, R. I., Retsa, C., & Murray*, M. M. (2019). Brain and Cognitive Mechanisms of Top–Down Attentional Control in a Multisensory World: Benefits of Electrical Neuroimaging. *Journal of Cognitive Neuroscience*, *31*, 431–441. https://doi.org/10.1162/jocn

Meijer, P. B. L. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, *39*(2), 112–121. IEEE Transactions on Biomedical Engineering. https://doi.org/10.1109/10.121642

Michel, C. M., & Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *NeuroImage*, *61*(2), 371–385. https://doi.org/10.1016/j.neuroimage.2011.12.039

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave De Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology*, *115*(10), 2195–2222. https://doi.org/10.1016/j.clinph.2004.06.001

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography*, *20*(4), 249–264. https://doi.org/10.1007/s10548-008-0054-5

Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid Brain Discrimination of Sounds of Objects. *Journal of Neuroscience*, *26*(4), 1293–1302. https://doi.org/10.1523/JNEUROSCI.4511-05.2006

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory Processes: A Balancing Act across the Lifespan. *Trends in Neurosciences*, *39*(8), 567–579. https://doi.org/10.1016/j.tins.2016.05.003

Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology*, *66*(1), 75–81. https://doi.org/10.1016/0013-4694(87)90141-6

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network—ProQuest. *Experimental Brain Research*, 481–493.

Pourtois, G., Delplanque, S., Michel, C., & Vuilleumier, P. (2008). Beyond Conventional Event-related Brain Potential (ERP): Exploring the Time-course of Visual Emotion Processing Using Topographic and Principal Component Analyses. *Brain Topography*, *20*(4), 265–277. https://doi.org/10.1007/s10548-008-0053-6

Ramon, M., & Striem-Amit, E. (2022). The merits and future of data deep-dives for cognitive neuroscience. *Neuropsychologia*, *176*, 108394. https://doi.org/10.1016/j.neuropsychologia.2022.108394

Reich, L., Maidenbaum, S., & Amedi, A. (2012). The brain as a flexible task machine: Implications for visual rehabilitation using noninvasive vs. Invasive approaches. *Current Opinion in Neurology*, *25*(1), 86–95. https://doi.org/10.1097/WCO.0b013e32834ed723

Retsa, C., Matusz, P. J., Schnupp, J. W. H., & Murray, M. M. (2018). What's what in auditory cortices? *NeuroImage*, *176*, 29–40. https://doi.org/10.1016/j.neuroimage.2018.04.028

Retsa, C., Matusz, P. J., Schnupp, J. W. H., & Murray, M. M. (2020). Selective attention to sound features mediates cross-modal activation of visual cortices. *Neuropsychologia*, *144*, 107498. https://doi.org/10.1016/j.neuropsychologia.2020.107498

Retsa, R., Turpin, H., Geiser, E., Ansermet, F., Müller-Nix, C., & Murray, M.M. (2024) Longstanding auditory sensory and semantic differences in preterm born children. *Brain Topography* 37(4), 536-551. https://doi.org/10.1007/s10548-023-01022-2

Spierer, L., Lucia, M. D., Bernasconi, F., Grivel, J., Bourquin, N. M., Clarke, S., & Murray, M. M. (2011). Learning-induced plasticity in human audition: Objects , time , and space. *Hearing Resarch*, *271*, 88–102. https://doi.org/10.1016/j.heares.2010.03.086

Striem-Amit, E. (2024). Can individual differences explain brain plasticity in blindness? *Trends in Cognitive Sciences*, *0*(0). https://doi.org/10.1016/j.tics.2024.09.016

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. https://doi.org/10.1016/j.cub.2014.02.010

Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. *Neuron*, *76*(3), 640–652. https://doi.org/10.1016/j.neuron.2012.08.026

Striem-Amit, E., Dakwar, O., Reich, L., & Amedi, A. (2012). The large-Scale Organization of "Visual" Streams Emerges Without Visual Experience. *Cerebral Cortex*, *22*(7), 1698–1709. https://doi.org/10.1093/cercor/bhr253

Tzovara, A., Murray, M. M., Michel, C. M., & De Lucia, M. (2012). A Tutorial Review of Electrical Neuroimaging From Group-Average to Single-Trial Event-Related Potentials. *Developmental Neuropsychology*, *37*(6), 518–544. https://doi.org/10.1080/87565641.2011.636851