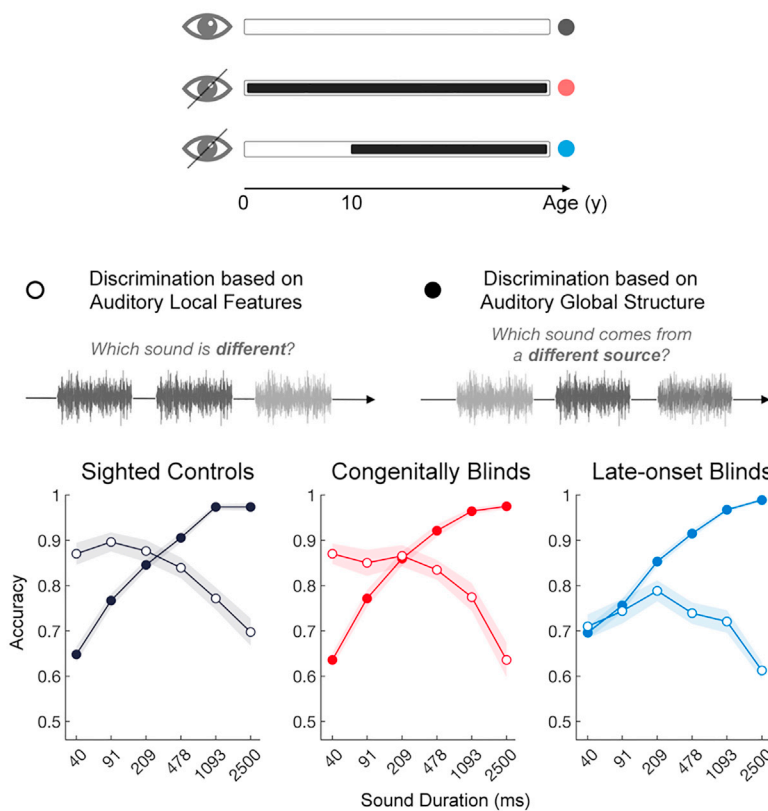**Article**

Interactions between auditory statistics processing and visual experience emerge only in late development

Impact of Visual Experience on Auditory Statistics Processing

Martina Berto,
Emiliano Ricciardi,
Pietro Pietrini,
Davide Bottari

davide.bottari@imtlucca.it
(D.B.)
martina.berto@imtlucca.it
(M.B.)

Highlights

Computational and deprivation models can be combined to assess sensory plasticity

Basic auditory computations develop independently from early visual input

Late-onset sight loss can hamper the efficiency of local auditory processing

Article

Interactions between auditory statistics processing and visual experience emerge only in late development

Martina Berto,[1,*] Emiliano Ricciardi,[1] Pietro Pietrini,[1] and Davide Bottari[1,2,*]

SUMMARY

The auditory system relies on local and global representations to discriminate sounds. This study investigated whether vision influences the development and functioning of these fundamental sound computations. We employed a computational approach to control statistical properties embedded in sounds and tested samples of sighted controls (SC) and congenitally (CB) and late-onset (LB) blind individuals in two experiments. In experiment 1, performance relied on local features analysis; in experiment 2, performance benefited from computing global representations. In both experiments, SC and CB performance remarkably overlapped. Conversely, LB performed systematically worse than the other groups when relying on local features, with no alterations on global representations. Results suggest that auditory computations tested here develop independently from vision. The efficiency of local auditory processing can be hampered in case sight becomes unavailable later in life, supporting the existence of an audiovisual interplay for the processing of auditory details, which emerges only in late development.

INTRODUCTION

The auditory system is specialized in capturing fine-grained details from sound waves (Plomp, 1964). These local temporal features are then integrated over time by mechanisms that retain, summarize, and abstract them into more compact acoustic percepts (McDermott et al., 2013; Yabe et al., 1998). Computational synthesis approaches, derived from information theories (Barlow, 1961), allowed investigation of these processes by implementing mathematical models to describe the set of measurements that the auditory system operates (McDermott and Simoncelli, 2011; McDermott et al., 2013). Such models have revealed that stationary sounds are analyzed by extracting a set of auditory statistics (McDermott and Simoncelli, 2011) whose processing represents the keystone of acoustic features integration. Auditory statistics result from a set of computations strictly dependent on anatomical and functional properties of the auditory pathway (Baumann et al., 2011; Dau et al., 1997; Gygi et al., 2004; Joris et al., 2004; Ruggero, 1992) and can be consequently described only through biologically plausible models (McDermott and Simoncelli, 2011). The auditory system processes these statistics by averaging short-term acoustic events (McDermott and Simoncelli, 2011; McDermott et al., 2013; McWalter and McDermott, 2018) along specific time windows (McWalter and McDermott, 2018) and uses this information to derive compact representations of sound objects. The auditory statistics processing can be broken down into two main modes of representation: (1) local features processing, by which fine-grained temporal details are extracted from a sound and stored; (2) statistical averaging, by which local features are averaged within a time window of integration, resulting in a global representation as local details are no longer retained (McDermott et al., 2013).

Although auditory statistics represent unimodal computations, the auditory system does not develop and function in isolation, and other senses, such as vision, modulate its functional and structural organization. Studies in non-human animal models revealed that, in the early development, the onset of visual input (eyes opening) gates the critical period closure of basic auditory functions. This evidence suggested that visually modulated sensitive periods in the auditory cortex exist (Mowery et al., 2016). In adults, visual events are known to directly modulate auditory neuron responses in both animals (Kayser et al., 2008) and humans (Thorne et al., 2011). At a functional level, visual systems play an important role in auditory features segregation, helping in disambiguating difficult instances (e.g., Golumbic et al., 2013). A common computational

framework for feature extraction might exist between auditory and visual modality (Shamma, 2001). In the same vein, the set of auditory statistics evaluated here were derived by auditory computational models (McDermott and Simoncelli, 2011), which are conceptually very similar to those derived by visual ones (Portilla and Simoncelli, 2000). In the present study, we directly investigated if the auditory statistics development and functioning fall within the exclusive competence of the auditory systems or are instead influenced by vision.

Visual deprivation models have been systematically employed to indirectly uncover the audiovisual (AV) interplay (for review, see Röder et al., 2020). The lack of visual inputs has consistently been associated with altered auditory processing across different functions (for review, see Pavani and Röder, 2012), but whether visual input availability alters specific underlying auditory computations is still unknown. In humans, previous evidence has identified visual inputs availability since birth as a prerequisite for the full development of auditory spatial calibration (Gori et al., 2014; Vercillo et al., 2016). On a different note, a large body of evidence suggests that both congenital and late-onset visual deprivation exert a compensating influence over certain higher-order auditory functions, leading to performance enhancements (see Röder et al., 2020). These compensatory effects have been consistently observed both in the context of spatial processing of auditory stimuli (e.g., Battal et al., 2019) and in tasks requiring spectro-temporal auditory features analyses, such as mnemonic representations of sounds (Röder and Rösler 2003), verbal memory (Amedi et al., 2003), frequency tuning (Huber et al., 2019), speech comprehension (Dietrich et al., 2013; Trouvain, 2007), and auditory temporal resolution (Muchnik et al., 1991).

To address whether the two modes of auditory representation, the local features processing and statistical averaging, are differently influenced by visual input availability, we tested blind populations against sighted individuals in two experiments designed to tap into these two computational modes. By combining a sound synthesis algorithm (McDermott and Simoncelli, 2011; McDermott et al., 2013) with psychophysics, the methodology adopted here provided the opportunity to uncover local and global auditory statistics computations from the discriminative responses elicited by different sound properties. Of importance, comparing samples of individuals who were visually deprived since birth or in late phases of development allowed one to assess whether the selected aspect of sound representation interacts with vision at specific time points along the life span.

We could expect detrimental behavioral effects associated with the lack of vision (e.g., Gori et al., 2014; Vercillo et al., 2016) in one of the samples of blind individuals or both. This outcome would syndicate for a modality interplay between vision and auditory computations in which vision plays a crucial role in supporting specific aspects of basic auditory processing (i.e., acoustic features segregation; Park et al., 2016). Noticeably, some auditory fundamentals, such as periodicity pitch, are innate and not influenced by early experience (Montgomery and Clarkson, 1997). Thus, sensory experience might also not be required for the full development of the functions tested here, but modality interplays could still occur later. However, provided that several auditory processes can benefit from the lack of vision, behavioral compensatory effects in one or both visual deprivation models (congenital and late-onset blindness) could also be expected in our study. Such results would indicate that vision is not necessary for their development or functioning and would provide evidence that a functional adaptation to lack of vision occurs not only for high-order functions but also for selective basic auditory computations. In both scenarios, a difference between the CB and LB groups would characterize the developmental trajectory of AV interplay in the context of local and global auditory statistics processing.

RESULTS

We took advantage of an already corroborated methodological pipeline (McDermott et al., 2013) and produced different synthetic textures built upon original recordings. Thanks to their properties, a specific category of natural sounds, namely, Sound Textures, was used (some examples are the rain, fire, bulldozer, typewriting, waterfall sounds; McDermott and Simoncelli, 2011; Saint-Arnaud and Popat, 2021; Schwarz, 2011). These sounds are rich, ubiquitous, and constant over time (McDermott et al., 2013; McWalter and McDermott, 2018). We selected 54 environmental recordings of Sound Textures (Table S1), among the original set implemented by McDermott et al. (2013). Synthetic stimuli were created using an Auditory Texture Model (McDermott and Simoncelli, 2011), which efficiently simulates computations performed at peripheral stages of the auditory processing. By computing time averages of non-linear functions, it was possible to measure a set of statistics: envelope marginal moments, reflecting

distribution and sparsity of the signal (Lorenzi et al., 1999), envelope cross-correlation between cochlear envelopes, accounting for the presence of broadband events within the signal (McDermott and Simoncelli, 2011), the modulation bands power, providing information about the temporal structure within cochlear channels (Bacon and Wesley Grantham, 1989; Dau et al., 1997; McDermott and Simoncelli, 2011), and their correlations (McDermott and Simoncelli, 2011). Although the model represents a mathematical approximation, cochlear envelope statistics convey most of the perceptually relevant information about the sound (Gygi et al., 2004; McDermott and Simoncelli, 2011; Smith et al., 2002). By using the full set of statistics to generate synthetic sounds, it is possible to obtain compelling exemplars of the same original Sound Texture (McDermott and Simoncelli, 2011). Our experimental stimuli were created accordingly. By imposing the statistics mentioned above on four different white noise samples, we synthesized four different exemplars for each original Sound Texture. Among themselves, synthetic exemplars of the same texture varied only for their local features, whereas their long-term average statistics matched the sound they were derived from. In other terms, this process resulted in four synthetic sounds whose properties were constrained only by the selected average statistics extracted from the original recording (Figure 1A). This provided us with the unique opportunity to test for specific computations within the auditory statistics processing, thanks to the unparalleled level of control exerted over the statistics present in the synthesized sounds (for details about synthesis procedure, see STAR Methods).

We tested sighted and blind participants in two selected experiments that included our synthetic sounds and exploited the two modes of auditory statistics representation.

Blind participants were grouped according to blindness onset, whether from birth or developmentally later (>10 years old; see Table 1). Thus, three groups of participants were recruited: congenitally blind (CB), late-onset blind (LB), and sighted control (SC) individuals. All three groups were matched by sample size, age, and gender (see STAR Methods).

Both experiments consisted of a two-alternative forced-choice (2AFC) oddity; participants had to detect the deviant sound among three acoustic samples, selecting between the first and third intervals. Stimuli were created by cutting each synthetic exemplar into smaller excerpts of different lengths. Each trial included three excerpts of the same length. The duration of the excerpts varied across trials, for a total of six durations (McDermott et al., 2013).

In the first experiment, Exemplar Discrimination, participants were asked to report which sound (the first or the third) was different from the other two. Two excerpts were the exact same sound. By contrast, the odd one was an excerpt extracted from a different synthetic exemplar of the same Sound Texture (Figure 1B). Thus, all of them stemmed from the same sound source (e.g., bulldozer), but one was generated starting from a different white noise and consistently diverged for the local features it encompassed as compared with the other two. Nonetheless, as duration increased, average statistics tended to converge toward the original imposed values (Figure 1D, left panel), making all three sounds perceptually very similar and challenging task performance.

In the second experiment, Texture Discrimination, participants were asked to report which sound came from a different acoustic source. Two excerpts were extracted from two distinct synthetic exemplars of the same Sound Texture (e.g., bulldozer), whereas the deviant was drawn from a synthetic exemplar derived from a different one (e.g., waterfall). Therefore, only the deviant in the triplet comprised both different local features and imposed average statistics (Figure 1C). Since all three sounds represented different excerpts, the local variability between couples was never zero. Still, the average statistics variability between the two sounds originated from the same texture tended to progressively decrease with duration, allowing for the different sound source (the deviant) to emerge perceptually (Figure 1D, right panel).

The employment of these experiments allowed for specific predictions on the outcomes. Consistent with previous findings by McDermott et al. (2013), opposite patterns of results were expected in the two experiments as a function of excerpts duration. For short stimuli, owing to differences in features variability among the three sounds, good performance was expected in Exemplar Discrimination compared with Texture Discrimination. By contrast, for long stimuli, statistical averaging was expected to boost Texture
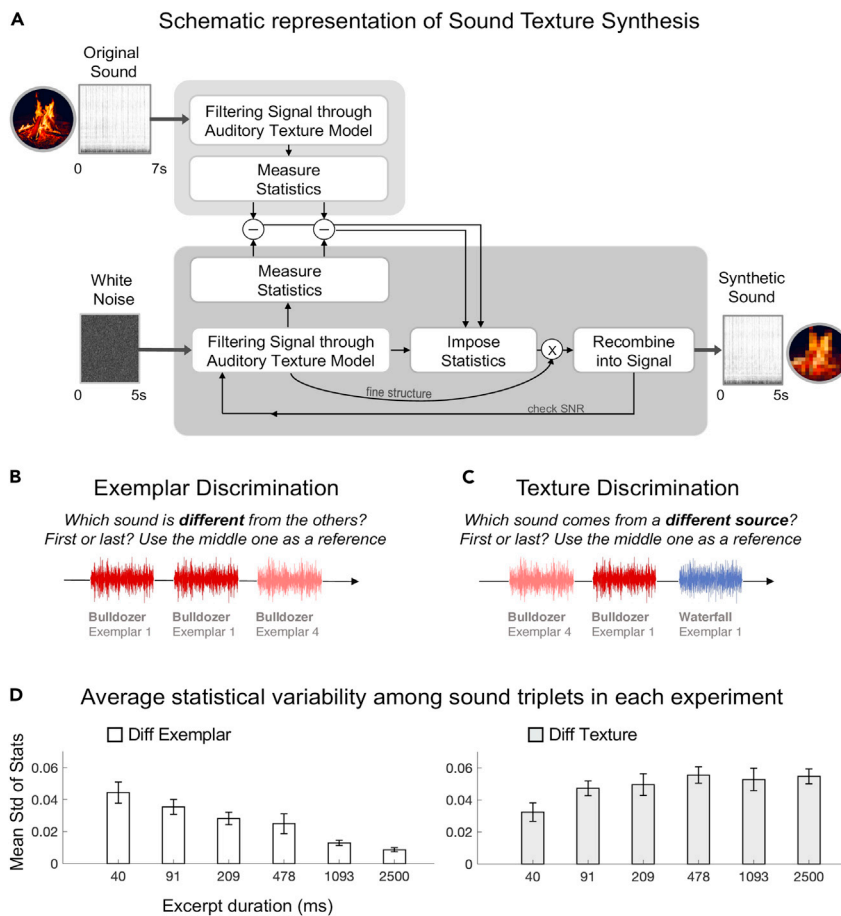
A Schematic representation of Sound Texture Synthesis

B Exemplar Discrimination

*Which sound is **different** from the others?*
First or last? Use the middle one as a reference

Bulldozer Exemplar 1   Bulldozer Exemplar 1   Bulldozer Exemplar 4

C Texture Discrimination

*Which sound comes from a **different source**?*
First or last? Use the middle one as a reference

Bulldozer Exemplar 4   Bulldozer Exemplar 1   Waterfall Exemplar 1

D Average statistical variability among sound triplets in each experiment

Figure 1. Sound texture synthesis, auditory statistics variability, and experimental design

(A) Schematic representation of the Sound Texture Synthesis procedure employed to produce synthetic stimuli. A 7-s Original Recording was passed through the Auditory Texture Model (McDermott and Simoncelli, 2011) to extract average statistics of interest. A white noise sample was then passed again through the model while the Original Recording statistics were imposed on its cochlear envelopes. Modified envelopes were multiplied by their associated fine structure and recombined into the signal. This procedure was iterated until a desirable signal-to-noise ratio is reached. The outcome was a Synthetic Sound Texture Exemplar, a 5-s signal containing statistics that matched the original values and was only constrained by them.

(B) Exemplar Discrimination. Schematic representation of one trial. Participants had to detect the different sound among the three, the other two being identical. In this case, the correct answer is the last sound since it is another exemplar of the "Bulldozer" Sound Texture.

(C) Texture Discrimination. Schematic representation of one trial. Participants had to report which sound was coming from a different source with respect to the other two. In this example, the correct answer is the third sound, it being an excerpt of a different Sound Texture.

(D) Average variability across a set of statistics (envelope mean, skewness, variance, cross-correlation, and modulation power) was computed from couples of excerpts to measure, in both experiments, the objective statistical difference between reference and deviant sounds. Left: average standard deviation (std) across statistics measured in excerpt pairs originated from Different Exemplars of the same Sound Texture (for all textures in column 1, Table S1). When duration increased, average statistics converged to the imposed original values, and variability progressively tended to zero, increasing discrimination difficulty in Exemplar Discrimination. Right: average std across statistics measured in excerpts pairs derived from Different Sound Textures (columns 1 and 2, Table S1). In Texture Discrimination, both scenarios (Different Exemplar and Different Texture) were presented and compared against each other. As duration increased, the difference between the two contexts increased, facilitating the recognition of the deviant one. Error bars represent the standard error of the mean (SEM).

(See also Figure S1 and Table S1).

Table 1. Characteristics of blind participants

1. Late-onset blinds group (N = 18)

	Age	Sex	Hand	Residual	Onset (years)	Etiology	Education	Music Training
LB01	21	M	R	None	18	Congenital glaucoma	Middle school	No
LB02	26	M	R	LP	18	Retinitis pigmentosa	University	No
LB03	44	F	R	LP	20	Retinitis pigmentosa	High school	No
LB04	25	F	R	None	22 LE 14 RE 22	Stevens-Johnson syndrome	University	No
LB05	41	M	R	None	36	Familial exudative vitreoretinopathy	High school	No
LB06	30	F	R	None	20	Eye tumor	High school	No
LB07	38	F	R	None	15	Glaucoma	University	No
LB08	59	M	R	None	43	Optic nerve sheath meningioma	Middle school	No
LB09	55	F	R	LP	30	Retinitis pigmentosa	University	No
LB10	52	F	R	None	27	Retinitis pigmentosa	High school	No
LB11	55	F	R	LP, SP, MP	51	Bietti's crystalline dystrophy	University	No
LB12	46	M	R	LP	18	6mo: Removed congenital cataract; then glaucoma	Middle school	No
LB13	50	F	R	LP	11	Retinitis pigmentosa	High school	No
LB14	55	M	R	LP	38	Retinitis pigmentosa	Middle school	No
LB15	40	F	R	None	20	Glaucoma	High school	No
LB16	32	M	R	LP	10	Optic nerve compression Astrocytoma	University	No
LB17	22	M	R	None	20	Glaucoma	High school	No
LB18	30	M	R	LP in RE	18	Glaucoma (RE) and retinal detachment (LE)	High school	No

2. Congenitally blind group (N = 18)

CB01	32	M	R	LP	0	Congenital glaucoma	University	No
CB02	36	M	R	None	0	Retinitis pigmentosa	High school	No
CB03	44	M	R	None	0	Congenital glaucoma	High school	No
CB04	60	F	R	LP	0	Retinopathy of prematurity	High school	Yes
CB05	45	M	R	LP	0	Congenital glaucoma	University	No
CB06	43	F	R	None	0	Retinopathy of prematurity	High school	No
CB07	28	F	R	LP	0	Microphthalmia	University	No
CB08	29	F	L	None	0	Retinopathy of prematurity	University	No
CB09	27	M	R	LP	0	Retinopathy of prematurity	High school	Yes
CB10	41	M	R	LP	0	Retinitis pigmentosa	University	No
CB11	59	M	R	None	0	Glaucoma	High school	Yes
CB12	37	F	R/L	LP	0	Congenital cataract	High school	No
CB13	20	F	R	None	0	Microphthalmia and aniridia	University	No
CB14	34	F	R	LP	0	Optic nerve hypoplasia	University	No
CB15	32	M	R	LP	0	Retinopathy of prematurity	University	Yes
CB16	30	M	R	LP	0	Leber's congenital amaurosis	University	Yes
CB17	44	F	R	None	0	Virus during pregnancy	High school	No
CB18	28	F	R	None	0	Retinopathy of prematurity	University	No

LP, light perception; SP, silhouette perception; MP, motion perception; M, male; F, female; mo, month old; LE, left eye; RE, right eye.
Music Training: Yes, Professional, studied music for at least 10 years.

Discrimination performance and to hamper Exemplar Discrimination one (McDermott et al., 2013). Any significant deviation from these expected outcomes indicates changes in the processing of local features or statistical averaging.

The attended pattern of results (McDermott et al., 2013) was confirmed in our SC group. The performance was better for short durations (40, 91 ms) in Exemplar Discrimination as compared with Texture Discrimination (all p < 0.002, corrected). Conversely, for long durations (478, 1,093, 2,500 ms), participants' accuracy was higher in Texture Discrimination as compared with Exemplar Discrimination (all p < 0.028, corrected). No difference was observed for trials comprising stimuli that were 209 ms long (p = 0.22, corrected; Figure 2A). Overall, the data replicated previous findings in sighted individuals, despite participants being blindfolded. Thus, these results represent a validated context to assess whether visual experience impacts auditory statistics processing.

The CB group displayed a remarkably similar pattern of results. Better performance was found in Exemplar Discrimination for short durations (40, 91 ms) compared with Texture Discrimination (all p < 0.01, corrected), and, conversely for long durations (478, 1,093, 2,500 ms) in Texture Discrimination compared with Exemplar (all p < 0.004, corrected). As observed in SC, there was no difference between the two experiments for stimuli that were 209 ms long (p = 0.80, corrected; Figure 2B).

By contrast, in the LB group, no significant differences were found between Exemplar Discrimination and Texture Discrimination for all short and intermediate stimuli (40, 91, and 209; all p > 0.05, corrected). For long durations (478, 1,093, 2,500 ms), LB participants performed significantly better in Texture Discrimination as compared with Exemplar Discrimination (all p < 0.001, corrected) (Figure 2C).

Late-onset sight loss hampers the processing of local sound features

For Exemplar Discrimination, when contrasting the performance of the SC group with the one of the CB group, no significant differences could be detected across all durations (all p > 0.29, corrected). Results clearly revealed that the processing of fine-grained temporal details is resilient to the absence of visual input since birth.

Conversely, the LB group performance was impaired as compared with SC group for durations 40, 91, 209, and 478 (all p < 0.04, corrected); at duration 40, 91, and 478, LB's performance was worse also as compared with CB group's one (all p < 0.04, corrected).

Results suggested an altered capability of late blind participants to discriminate sounds when the most efficient strategy is to base performance on local features. These findings reveal that visual input in early phases of development is not a prerequisite to acquiring the ability to discriminate sounds according to their local features. However, this process is encumbered by late-onset blindness (Figure 2D).

Visual deprivation does not impact statistical averaging

By comparing the performances in Texture Discrimination among the three groups, it was possible to address the impact of visual experience on statistical averaging efficiency.

The lack of significant interaction between group and duration (p = 0.08; see STAR Methods) showed that, in Texture discrimination, performance increased with duration in a similar way across all groups (see Figure 2E). This result reveals that visual deprivation does not significantly influence the process by which auditory statistics are computed over time and used to identify different sound sources.

Relative difference between experiments reveals no advantages for LB when local features should support the performance

For each participant, the accuracy scores in Texture Discrimination were subtracted from the ones in Exemplar Discrimination, separately for each duration, to compute the relative mean difference between Experiments (Figure 2F). Results showed that, compared with both SC and CB groups, this difference was significantly smaller for very short durations in LB group (LC versus SC: 40, 91; all p < 0.02, corrected; LC versus CB: 40 p < 0.001, corrected, 91 p = 0.04, uncorrected) and larger for longer ones (LC versus SC: 209, 478, 2,500; all p < 0.03, corrected; LC versus CB: 478 p < 0.03, corrected; see Figure 2F). Conversely,
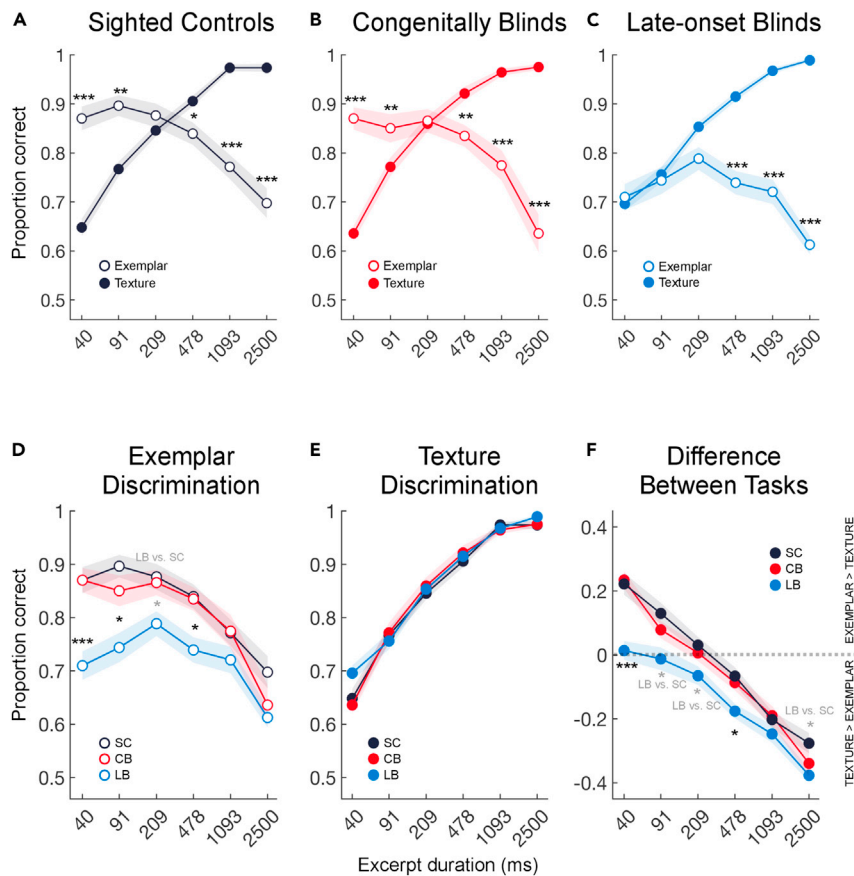
Figure 2. Proportion of correct answers in the two experiments

(A–C) Results for Exemplar Discrimination versus Texture Discrimination in each group. The proportion of correct answers across individuals at the group level is shown as a function of excerpt duration.

(D) Between-group comparisons in Exemplar Discrimination. SC and CB groups' performance did not differ. The LB group showed an impaired performance compared with the other groups: late sight loss has a detrimental impact when performance could benefit from higher local features variability.

(E) Between-group comparisons in Texture Discrimination. No significant difference was observed across groups.

(F) Relative difference between experiments for all three groups. For each group and duration, the averages across participants of the differences between the performance in the two tasks are displayed. Positive values are found when performance in Exemplar is better than in Texture Discrimination, whereas negative values represent better performance in Texture than in Exemplar Discrimination. Unlike the other two groups, LB never showed an advantage in Exemplar Discrimination compared with Texture Discrimination. Shaded regions show interpolated standard error of the mean (SE) at each point. Results were corrected for multiple comparisons using the false discovery rate (FDR). If more than one comparison is significant, stars refer to the lower bound; gray stars indicate that only the labeled comparisons were significant; ***p < .001; **p < .01; *p < .05.

(See also Figure S4).

the relative difference of performance in the two Experiments between CB and SC did not differ for any of the durations (all p > 0.32, corrected; Figure 2F).

These results showed that both CB and SC performed according to predictions showing specific advantages in one experiment or the other according to stimulus duration. On the contrary, the LB group did not display a boost for those conditions in Exemplar Discrimination in which local feature processing could have helped the performance.

Ruling out possible confounds

It could be argued that results in the LB group were idiosyncratic of the participants included in the study. However, the sample size was adequate (N = 18 per group) and relatively large compared with previous investigations

on blind individuals. All groups were matched adequately by several main variables (size, age, and gender); all participants had no documented auditory deficits, cognitive impairments, or neurological disorders (blindness was associated only to peripheral damage; Table 1). Moreover, LB participants' performance was hampered only in one of the two experiments. In contrast, the observed results in Texture Discrimination were indistinguishable from the ones of the remaining groups. If anything, LB group performance was more efficient at 40 ms (mean proportion correct = 0.70; SE = 0.01) as compared with CB (mean proportion correct = 0.64; SE = 0.02; p = 0.01, uncorrected) and SC (mean proportion correct = 0.65; SE = 0.02; p < 0.04, uncorrected). In light of these observations, it was possible to exclude that a general cognitive or acoustic impairment could explain the lack of advantage specific for Exemplar Discrimination. However, given the large age range across participants in all groups (20–62 years), it could have been possible that hampered performance in the LB group could be driven by idiosyncratic age-related hearing impairments in the oldest individuals from that specific sample. That did not seem to be the case, as no significant interaction was found between group and age in our model (p = 0.99). To further rule out this possibility, we also computed correlation between LB participant's age and their performance accuracy for each duration in Exemplar Discrimination (Figure S6). No specific association between age and performance was found in this group (all p > 0.37). In conclusion, the age factor did not seem to explain the deficit observed in this sample.

To rule out other possible confounds, we ran specific tests on our sample data. Providing the temporal presentation of triplets in the employed 2AFC protocol, it could have been possible that a particular disposition bias toward the first or last interval was present and diverged among groups, at least partially accounting for different results. This was not the case, as all groups did not significantly differ in terms of predisposition toward stimulus intervals (Figure S2A).

Another possible confounding effect could have been a between-group difference associated with learning or tiredness within an experimental session. It could have been possible that LB group performance had changed as the experiment proceeded (e.g., progressively diminishing across runs, specifically in the Exemplar Discrimination). However, this was not the case. Comparing accuracy across runs, a similar trend in all groups emerged for each duration (Figure S2B).

DISCUSSION

The present study addressed whether visual experience at different phases of the lifespan exerts a cross-modal influence over specific basic auditory computations. No difference was observed between CB and SC groups. Conversely, we found that LB's performance was selectively impaired compared with both SC and CB, for those conditions in which optimal performance relied on local features. These results provided evidence that visual experience is not a prerequisite for the full emergence of auditory statistics. However, sight loss can negatively impact specific auditory computations only when vision was available in early development and went missing afterward. This outcome is in line with a model in which interactions between auditory local computations and vision emerge only after auditory functions have developed.

Auditory statistics processing develops regardless of visual input availability

Blindness represents a natural state in which it is possible to address cross-modal interdependencies of sensory functions. By comparing CB with SC groups, on the one hand our study investigated whether visual input was necessary for the typical development of the two auditory modes of representation underlying the auditory statistics processing, and on the other hand, whether compensatory mechanisms emerged owing to lack of vision since birth. Therefore, two different outcomes could have been expected.

First, if visual experiences were necessary to properly develop the ability to process local features and/or to compute average statistics, we could have observed an impaired performance of CB participants in one of the two experiments, or both, compared with sighted individuals. For example, evidence exists that CB individuals fail at performing auditory space bisection tasks (Gori et al., 2014), and this observation provides evidence that visual experience is necessary to efficiently encode Euclidian coordinates of sounds (Gori et al., 2014, 2020a, 2020b). Our results provide strong evidence for the relative independence of the development of auditory statistic processing from early visual experience.

On the other hand, previous evidence showed enhanced auditory skills in congenitally or early blind individuals (e.g., Doucet et al., 2005; Muchnik et al., 1991; Röder et al., 1999) and in individuals who experienced a period of visual deprivation since birth (Bottari et al., 2018). Coherently, a congenital absence of vision could have improved

the sensitivity to the auditory information included in sound textures, leading to the emergence of specific compensatory mechanisms. For instance, scattered evidence suggests that blind individuals are able to understand syllables at a greater time speed compared with natural speech (Dietrich et al., 2013; Moos and Trouvain, 2007; Trouvain, 2007), an ability perhaps supported by a higher auditory sampling rate capacity. We could have expected the CB group to be able to retain more local features in time before averaging statistics, showing a better performance for longer durations in Exemplar Discrimination. A compensatory mechanism favoring the statistical averaging mode would have led to better results in Texture Discrimination. Furthermore, a better frequency tuning (Huber et al., 2019) to temporal local features in CB would have enhanced the sensitivity to small acoustic differences, leading to superior-to-sighted performances for short excerpts, both in Exemplar and Texture Discrimination. However, none of these hypotheses was confirmed in our data, as the performance was indistinguishable between CB and SC groups. The absence of compensatory mechanisms, arising owing to the lack of vision since birth, suggests these basic auditory computations cannot be improved by cross-modal adaptations associated to early blindness.

Previous evidence already suggested the existence of auditory functions that develop independently from visual experience since birth (for instance, auditory gap detection threshold; Weaver and Stevens, 2006; pure tone threshold audiometry and acoustic reflex threshold; Starlinger and Niemeyer, 1981). Similarly, it seems that the development of auditory computations tested in this study is not influenced by the availability of other senses.

We can argue that the extraction of the auditory statistics necessary for textures recognition occurs at a very basic level of processing, whereas functions that showed visual input dependencies are mostly the result of higher-order cortical operations. Evidence suggests that multisensory functions performed at subcortical levels rely less on experience compared with cortical multisensory functions (e.g., Putzar et al., 2010). As the auditory model used in our study to extract and impose statistics represented peripheral computations, from the cochlea up to the midbrain (McDermott and Simoncelli, 2011), it is possible that in early phases of development, the functional emergence of the set of auditory statistics included in our stimuli is exclusive competence of the auditory system and does not rely on other modalities, such as vision. Further studies including statistics resulting from computations beyond primary cortex (Chi et al., 2005; Norman-Haignere and McDermott, 2018) and involving other types of naturalistic stimuli will be able to shed light on potential visual influences over the development of higher-order auditory statistics.

Functional interplay between selective auditory computations and vision

By comparing the performance of CB, SC, and LB groups, we can affirm that, although early vision is not necessary for the typical development of the auditory statistics processing, the lack of vision at later developmental phases can still influence the processing of local features. No compensatory effects were found in the LB group for any of the tested computational modes, whereas a detrimental effect was present for the acoustic local features analysis.

Two main explanations for these results can be forwarded. The first one is specific for the altered mode (local features processing) and its dependencies with vision changing throughout developmental phases; the second one considers adaptations at a higher level of processing, which would prompt, in case vision is lost at a late developmental phase, a strategic employment of one mode (statistical averaging) at the expense of the other (local processing).

Consistent evidence shows that, even considering basic auditory functions, human performance keeps improving gradually for over nearly a decade. Frequency discrimination, such as perceiving differences between two tones presented sequentially, does not mature until roughly 10 years of age for low frequencies (Jensen and Neff, 1993; Maxon and Hochberg, 1982; Moore et al., 2011). Similarly, the thresholds for detecting amplitude modulations become adult-like only after 10 years of age (Banai et al., 2011). The use of local features comprised a number of operations, including the measurement of amplitude modulations over time (Dau et al., 1997). Thus, the first explanation of the present results is that these functions could be designed to develop independently from visual input and only after their functional development is completed, interactions across senses are allowed. Supporting this framework, previous studies have revealed that certain aspects of multisensory integration can occur only by the age of 8–10 years (Gori et al., 2008). In the same vein, basic AV multisensory facilitations have been found immature until the age of 9 years, whereas similar patterns have been observed between adolescents and adults (Brandwein

et al., 2011). Similarly, the development of multisensory speech perception has been found to progress until late childhood (Ross et al., 2011). Auditory statistics processing represents a unimodal function, strictly dependent on biological and anatomical properties of the auditory system, and vision is not required to perform any of the experimental tasks; in fact, the performance of our SC sample replicated the one in McDermott et al. (2013), in the total absence of retinal input (participants were blindfolded). However, late-onset sight loss hampers participants' ability to rely on acoustic local details, suggesting that the availability of visual input in early developmental phases (as it occurred in the LB group) may trigger the typical development of AV facilitations in auditory features segregation. As a result, the sudden and permanent loss of visual input could, in turn, alter auditory processing that had typically developed.

The absence of associations between the accuracy in both tasks and duration or onset of blindness (Figure S5) suggested that the performance (1) was not influenced by the number of years people were visually deprived (in our LB sample, duration of blindness spanned from 2 years up to 28 years); (2) was similar for people who became blind during childhood (10 years of age) as well as during adulthood (up to 51 years of age, in our LB sample). However, we had the chance to additionally test two early blind individuals whose blindness onset occurred during their third year of life (Figure S3). Remarkably, their results systematically overlapped with the CB and SC groups, providing further support to the evidence that the dampening of local features processing manifests only if sight loss occurs relatively late in the development.

Shamma (2001) suggested that there is a correspondence between certain auditory and visual percepts and their underpinning computational principles, that is, essential auditory percepts such as timbre, pitch, and location can be derived through computational approaches typically employed to model visual processing (see also McDermott and Simoncelli., 2011). As an example, the computations required for extracting the profile of sound spectrum seem to correspond to the type of neural computations necessary for the extraction of the form of an image. The common principle could rely on lateral inhibition for the enhancement of peaks or edges (Lyon and Shamma, 1996; Shamma, 1985). The present results might suggest that, if a unified plan for basic auditory and visual computations exists, it might develop, for certain functions, independently within each sensory modality and only later in the development cross-modal interactions may emerge.

Alternatively, the second possible explanation accounts for an adjustment of the reliance on local and global auditory processing induced by late-onset sight loss. The deficit observed in the LB group does not exclude that limiting local features accessibility can provide ecological advantages. Identifying a Sound Texture equals to recognizing the statistics included in its sound waveform (McDermott and Simoncelli, 2011). Thus, it is possible that relying on statical averaging only when information is consistent, at the expense of temporal details, prompts sound-object recognition in an everyday environment. Local and global processes can be favored depending on the task demands and thus represent flexible modes that can be interchanged strategically. The results in our LB group could suggest a form of adaptive perceptual learning (Watanabe and Sasaki, 2015) aimed at facing a remarkable loss in the overall available sensory input (de Villers-Sidani and Merzenich, 2011). Perceptual learning in the adult brain does not emerge merely from bottom-up mechanisms but is instead driven by feedback regulations (Gilbert et al., 2009; Schäfer et al., 2007) and by the relevance of the stimulus feature for ecological purposes (Polley et al., 2006). Studies on rodents showed that the selective task (e.g., discrimination based on frequency versus intensity) can drive plastic adaptations, disregarding bottom-up properties of the presented sounds (Polley et al., 2006). Thus, it is possible that late-onset sight loss is responsible for relying more on one auditory mode (statistical averaging) as compared with the other (local processing), in order to obtain ecological advantages and boost stimulus recognition. However, this strategical functional adaptation may reveal detrimental when it comes to discriminate sounds all containing the same set of statistics for which local details would be more informative.

Further investigations are required to validate one explanation or the other, understanding how visual input availability interacts with basic auditory computations and their developmental trajectories and from which level of processing this effect originates (low or high level).

Conclusions

Overall, this evidence has several important implications. First, it discloses how basic auditory computations can develop independently from early visual input. Second, it shows a selective detrimental effect induced by a late-onset sight loss over selective, non-spatial aspects of the auditory processing. Third, it

has the potential to expand our approaches in several fields, including auditory, visual, multisensory, and brain disorders research. Moreover, it provides novel ground for applied science such as sensory substitution device and auditory rehabilitation strategies for people who lost vision compared with those born blind. Finally, we proved that combining computational methods with human models of sensory deprivation (Bottari and Berto, 2021) provides the context to assess the degree of plasticity of specific computations performed by the sensory systems. Although the present study focused on the auditory domain, similar approaches could be employed for other modalities. Providing the resemblance of the Auditory Texture Model (McDermott and Simoncelli, 2011) with a previously validated Visual Texture Model (Portilla and Simoncelli, 2000), and the presence of textures in other modalities, such as touch (Picard et al., 2003; Weber et al., 2013), it can be possible to address with similar approaches the development and sensory interplays across computations performed by other senses.

Limitations of the study

The main limitation of this study is the lack of a precise temporal characterization of adaptations observed in late blind individuals. Blind participants were divided according to blindness onset being at birth or after 10 years of age. Because of this broad distinction, it is not possible to precisely track the developmental trajectory of AV interplays. A further sample of blind individuals covering early-onset blindness (<10 years of age), and possibly developmental studies, may help answering this unsolved question. Finally, we did not combine any neuroimage technique (i.e., electroencephalography, magnetoencephalography, or functional MRI) with our psychophysics protocol, so the neural correlates of the tested auditory computations and of the adaptation observed in LB are still unknown. Moreover, the lack of behavioral effects in CB individuals does not imply that no alteration at the brain level exists, even though it does not affect performance. Further studies are required to explore these issues.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECTS DETAILS
  - Participants
- METHODS DETAILS
  - Synthetic texture synthesis
  - Stimuli
  - Experimental design
  - Exemplar discrimination
  - Texture discrimination
  - Average statistical variability
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Difference between exemplar and texture
  - Testing for disposition bias
  - Testing for learning effect
  - Correlations

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS

REFERENCES

Amedi, A., Raz, N., Pianka, P., Malach, R., and Zohary, E. (2003). Early 'visual'cortex activation correlates with superior verbal memory performance in the blind. Nat. Neurosci. *6*, 758–766.

Bacon, S.P., and Grantham, D.W. (1989). Modulation masking: effects of modulation frequency, depth, and phase. J. Acoust. Soc. Am. *85*, 2575–2580.

Banai, K., Sabin, A.T., and Wright, B.A. (2011). Separable developmental trajectories for the abilities to detect auditory amplitude and frequency modulation. Hearing Res. *280*, 219–227.

Barlow, H.B. (1961). Possible principles underlying the transformation of sensory messages. Sensory Commun. *1*, 217–234.

Battal, C., Rezk, M., Mattioni, S., Vadlamudi, J., and Collignon, O. (2019). Representation of auditory motion directions and sound source locations in the human planum temporale. J. Neurosci. *39*, 2208–2220.

Baumann, S., Griffiths, T.D., Sun, L., Petkov, C.I., Thiele, A., and Rees, A. (2011). Orthogonal representation of sound dimensions in the primate midbrain. Nat. Neurosci. *14*, 423–425.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B *57*, 289–300.

Bottari, D., and Berto, M. (2021). Three factors to characterize plastic potential transitions in the visual system. Neurosci. Biobehav. Rev. *126*, 444–446.

Bottari, D., Kekunnaya, R., Hense, M., Troje, N.F., Sourav, S., and Röder, B. (2018). Motion processing after sight restoration: no competition between visual recovery and auditory compensation. Neuroimage *167*, 284–296.

Brandwein, A.B., Foxe, J.J., Russo, N.N., Altschuler, T.S., Gomes, H., and Molholm, S. (2011). The development of audiovisualaudiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. Cereb. Cortex. *21*, 1042–1055.

Chi, T., Ru, P., and Shamma, S.A. (2005). Multiresolution spectrotemporal analysis of complex sounds. J. Acoust. Soc. Am. *118*, 887–906.

Cramer, A.O., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R.P., Waldorp, L.J., and Wagenmakers, E.J. (2016). Hidden multiplicity in exploratory multiway ANOVA: prevalence and remedies. Psychon. Bull. Rev. *23*, 640–647.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. *102*, 2892–2905.

Dietrich, S., Hertrich, I., and Ackermann, H. (2013). Ultra-fast speech comprehension in blind subjects engages primary visual cortex, fusiform gyrus, and pulvinar–a functional magnetic resonance imaging (fMRI) study. BMC Neurosci. *14*, 74.

Doucet, M.E., Guillemot, J.P., Lassonde, M., Gagné, J.P., Leclerc, C., and Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. Exp. Brain Res. *160*, 194–202.

Gilbert, C.D., Li, W., and Piech, V. (2009). Perceptual learning and adult cortical plasticity. J. Physiol. *587*, 2743–2751.

Golumbic, E.Z., Cogan, G.B., Schroeder, C.E., and Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". J. Neurosci. *33*, 1417–1426.

Gori, M., Del Viva, M., Sandini, G., and Burr, D.C. (2008). Young children do not integrate visual and haptic form information. Curr. Biol. *18*, 694–698.

Gori, M., Sandini, G., Martinoli, C., and Burr, D.C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. Brain *137*, 288–293.

Gori, M., Amadeo, M.B., and Campus, C. (2020a). Spatial metric in blindness: behavioural and cortical processing. Neurosci. Biobehav. Rev. *109*, 54–62.

Gori, M., Amadeo, M.B., and Campus, C. (2020b). Temporal cues trick the visual and auditory

cortices mimicking spatial cues in blind individuals. Hum. Brain Mapp. *41*, 2077–2091.

Gygi, B., Kidd, G.R., and Watson, C.S. (2004). Spectral-temporal factors in the identification of environmental sounds. J. Acoust. Soc. Am. *115*, 1252–1265.

Huber, E., Chang, K., Alvarez, I., Hundle, A., Bridge, H., and Fine, I. (2019). Early blindness shapes cortical representations of auditory frequency within auditory cortex. J. Neurosci. *39*, 5143–5152.

Jensen, J.K., and Neff, D.L. (1993). Development of basic auditory discrimination in preschool children. Psychol. Sci. *4*, 104–107.

Joris, P.X., Schreiner, C.E., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. Physiol. Rev. *84*, 541–577.

Kayser, C., Petkov, C.I., and Logothetis, N.K. (2008). Visual modulation of neurons in auditory cortex. Cereb. Cortex. *18*, 1560–1574.

Lorenzi, C., Berthommier, F., Apoux, F., and Bacri, N. (1999). Effects of envelope expansion on speech recognition. Hear. Res. *136*, 131–138.

Lyon, R., and Shamma, S. (1996). Auditory representations of timbre and pitch. In Auditory Computation, H.L. Hawkins, T.A. McMullen, A.N. Popper, and R.R. Fay, eds. (Springer), pp. 221–270.

Maxon, A.B., and Hochberg, I. (1982). Development of psychoacoustic behavior: sensitivity and discrimination. Ear Hear. *3*, 301–308.

McDermott, J.H., and Simoncelli, E.P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron *71*, 926–940.

McDermott, J.H., Schemitsch, M., and Simoncelli, E.P. (2013). Summary statistics in auditory perception. Nat. Neurosci. *16*, 493–498.

McWalter, R., and McDermott, J.H. (2018). Adaptive and selective time averaging of auditory scenes. Curr. Biol. *28*, 1405–1418.

Moore, D.R., Cowan, J.A., Riley, A., Edmondson-Jones, A.M., and Ferguson, M.A. (2011).

Development of auditory processing in 6-to 11-yr-old children. Ear Hear. *32*, 269–285.

Moos, A., and Trouvain, J. (2007). Comprehension of ultra-fast speech–blind vs.'Normally Hearing'Persons. In Proceedings of the 16th International Congress of Phonetic Sciences, *1*Proceedings of the 16th International Congress of Phonetic Sciences (Saarland University Saarbrücken), pp. 677–680.

Mowery, T.M., Kotak, V.C., and Sanes, D.H. (2016). The onset of visual experience gates auditory cortex critical periods. Nat. Commun. *7*, 1–11.

Muchnik, C., Efrati, M., Nemeth, E., Malin, M., and Hildesheimer, M. (1991). Central auditory skills in blind and sighted subjects. Scand. Audiol. *20*, 19–23.

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

Park, H., Kayser, C., Thut, G., and Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. Elife *5*, e14521.

Pavani, F., and Röder, B. (2012). Crossmodal plasticity as a consequence of sensory loss: insights from blindness and deafness. In The New Handbook of Multisensory Processes, B. Stein, ed. (MIT Press), pp. 737–759.

Picard, D., Dacremont, C., Valentin, D., and Giboreau, A. (2003). Perceptual dimensions of tactile textures. Acta Psychol. *114*, 165–184.

Plomp, R. (1964). Rate of decay of auditory sensation. J. Acoust. Soc. Am. *36*, 277–282.

Polley, D.B., Steinberg, E.E., and Merzenich, M.M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. J. Neurosci. *26*, 4970–4982.

Portilla, J., and Simoncelli, E.P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. Int. J. Comput. Vis. *40*, 49–70.

Putzar, L., Hötting, K., and Röder, B. (2010). Early visual deprivation affects the development of face recognition and of audio-visualaudio-visual

speech perception. Restor. Neurol. Neurosci. *28*, 251–257.

Röder, B., and Rösler, F. (2003). Memory for environmental sounds in sighted, congenitally blind and late blind adults: evidence for cross-modal compensation. Int. J. Psychophysiol. *50*, 27–39.

Röder, B., Teder-SaÈlejaÈrvi, W., Sterr, A., Rösler, F., Hillyard, S.A., and Neville, H.J. (1999). Improved auditory spatial tuning in blind humans. Nature *400*, 162–166.

Röder, B., Kekunnaya, R., and Guerreiro, M.J. (2020). Neural mechanisms of visual sensitive periods in humans. Neurosci. Biobehav. Rev. *120*, 86–99.

Rogers Montgomery, C., and Clarkson, M.G. (1997). Infants' pitch perception: masking by low- and high-frequency noises. J. Acoust. Soc. Am. *102*, 3665–3672.

Ross, L.A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., and Foxe, J.J. (2011). The development of multisensory speech perception continues into the late childhood years. Eur. J. Neurosci. *33*, 2329–2337.

Ruggero, M.A. (1992). Responses to sound of the basilar membrane of the mammalian cochlea. Curr. Opin. Neurobiol. *2*, 449–456.

Saint-Arnaud, N., and Popat, K. (2021). Analysis and synthesis of sound textures. In Computational Auditory Scene Analysis, D.F. Rosenthal and H.G. Okuno, eds. (CRC Press), pp. 293–308.

Schäfer, R., Vasilaki, E., and Senn, W. (2007). Perceptual learning via modification of cortical top-down signals. PLoS Comput. Biol. *3*, e165.

Schwarz, D. (2011). State of the art in sound texture synthesis. Digital Audio Effects (DAFx), 221–232.

Shamma, S.A. (1985). Speech processing in the auditory system II: lateral inhibition and the central processing of speech evoked activity in the auditory nerve. J. Acoust. Soc. Am. *78*, 1622–1632.

Shamma, S. (2001). On the role of space and time in auditory processing. Trends Cogn. Sci. *5*, 340–348.

Smith, Z.M., Delgutte, B., and Oxenham, A.J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. Nature *416*, 87–90.

Starlinger, I., and Niemeyer, W. (1981). Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness I. Peripheral functions. Audiology *20*, 503–509.

Thorne, J.D., De Vos, M., Viola, F.C., and Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. J. Neurosci. *31*, 3853–3861.

Trouvain, J. (2007). Comprehension of synthetic speech. Saarland Working Papers in Linguistics (SWPL) *1*, 5–13.

Vercillo, T., Burr, D., and Gori, M. (2016). Early visual deprivation severely compromises the auditory sense of space in congenitally blind children. Dev. Psychol. *52*, 847.

de Villers-Sidani, E., and Merzenich, M.M. (2011). Lifelong plasticity in the rat auditory cortex: basic mechanisms and role of sensory experience. In Enhancing Performance for Action and Perception Multisensory Integration, Neuroplasticity and Neuroprosthetics, Part I, *191*, A.M. Green, E. Chapman, J.F. Kalaska, and F. Lepore, eds (Elsevier), pp. 119–131.

Watanabe, T., and Sasaki, Y. (2015). Perceptual learning: toward a comprehensive theory. Annu. Rev. Psychol. *66*, 197–221.

Weaver, K.E., and Stevens, A.A. (2006). Auditory gap detection in the early blind. Hear. Res. *211*, 1–6.

Weber, A.I., Saal, H.P., Lieber, J.D., Cheng, J.W., Manfredi, L.R., Dammann, J.F., and Bensmaia, S.J. (2013). Spatial and temporal codes mediate the tactile perception of natural textures. Proc. Natl. Acad. Sci. U S A *110*, 17107–17112.

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA *310*, 2191–2194.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R.J., and Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. Psychophysiology *35*, 615–619.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MATLAB 2018b	Mathworks	http://www.mathworks.com
IBM SPSS statistics, version 26.0	IBM Corp., Armonk, N.Y., USA	https://www.ibm.com/
Sound texture synthesis toolbox	McDermott and Simoncelli (2011)	http://mcdermottlab.mit.edu/
Deposited data		
Raw behavioral data	This paper	https://doi.org/10.17632/56kj5ngn78.1

RESOURCE AVAILABILITY

Lead contact

Further information should be directed and will be fulfilled by the lead contact, Davide Bottari (davide.bottari@imtlucca.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw behavioral data have been deposited at Mendeley Data and are publicly available as of the date of publication. The DOI is listed in the key resources table.

- This paper does not report original code. Synthesis algorithm is publicly available at the link listed in the key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECTS DETAILS

Participants

Data from three groups of participants, matched for size, gender, and age, were analyzed in this study. We recruited a group of congenitally blind individuals (CB; N = 18; F = 9; mean age = 37.06 years; std = 10.75). Data of the 18 CB individuals were used as reference for the recruitment of a group of late blind individuals and a group of sighted controls. Each CB individual in our sample was matched with a sighted individual and an LB individual of same gender and similar age (mean age of groups was within 2 std of difference), for a total of 18 participants for each group. Late onset blind individuals (LB; N = 18; F = 9; mean age = 40.11; std = 12.68); range of blindness onset = 10-51 years; range of blindness duration = 2-28 years), and sighted controls (SC; N = 18; F = 9; mean age = 38.06 years; std = 12.93).

Before experimental session began, all blind participants underwent a short interview to gather several information, especially about onset, cause, and duration of blindness, together with other anamnestic information.

All participants in the final sample were healthy and fully understood the task requests. During recruitment, exclusion criteria comprised documented hearing impairment (i.e., acoustic implants, tinnitus) and neurological disorders. The data of 3 CB individuals, 1 early blind participant (blindness onset: 7 months) and of 2 SC were not included in the final sample and thus were not analyzed due to their inability to perform/terminate one or both sessions or being easily distracted during experimental sessions (i.e., participant often asked questions and talked during the task). For all the blind participants in the sample, blindness was total and caused only by peripherical pathologies (Table 1).

Two early blind (EB) participants were also tested (EB1, gender = M; age = 26, blindness onset = 3years; EB2, gender = M; age = 24, blindness onset = 3years). Results from these participants were excluded from the analyses, as blindness onset was borderline between blind groups. Their data are plotted in supplemental information (Figure S3).

Beforehand, all participants were informed about the procedures and purpose of the study and signed a written informed consent prior to testing. The study was approved by the regional Ethical Committee (CEAVNO protocol n 24,579). The study protocol adhered to the guidelines of the Declaration of Helsinki (World Medical Association, 2013).

METHODS DETAILS
Experimental procedures and methods are inspired by the work of McDermott et al. (2013).

Synthetic texture synthesis
Synthetic textures were synthesized from the model described in a previously published paper (McDermott and Simoncelli, 2011). For synthesis, we used the MATLAB-based Sound Texture Synthesis Toolbox, available here: http://mcdermottlab.mit.edu.

Auditory statistics were computed from different signals of 7-s length, each one being an Original Recording of a natural Sound Texture, with a sampling rate of 20kHz. These sounds are also available on the website previously cited. The Original Recordings we used for synthesis and experiment were a subset of the ones used by McDermott et al. (2013).

For each original recording, cochlear envelopes and modulation bands were extracted from the signal, by filtering it through the Auditory Texture Model (McDermott and Simoncelli, 2011). Precisely, extracted statistics included marginal moments (mean, variance, and skew) of each cochlear envelope, envelope cross-correlation, the normalize power in each modulation band, and two correlations between modulation bands (C1 and C2). Kurtosis was omitted from the computation (see toolbox-authors' suggestion). Statistics were measured with a temporal weighting function that faded to zero over the first and last second of the original recordings to avoid boundary artifacts (McDermott and Simoncelli, 2011).

After setting the parameters, the process was initialized from a 5-s white noise sample. The noise signal was again filtered through the model and the original sound statistics were imposed on its cochlear envelopes, with circular boundary conditions. The resulting envelopes were multiplied by their associated fine structure and recombined into the signal. The procedure was repeated in an iterative way, until the imposed statistics measured from the synthetic signal reached a desirable signal-to-noise-ratio (SNR) of 30-dB, which represents the ratio of the squared error of a statistics class, summed across all statistics in the class, to the sum of squared statistics values of that class. The average of each statistics class was at least 20-dB (McDermott and Simoncelli, 2011). The resulting Synthetic Sound is a 5-s random signal constraint only by the over-imposed statistics of interest (Figure 1A). The procedure was repeated four times on four different white noise samples to obtain four different synthetic exemplars for each Sound Texture (adapted from McDermott et al., 2013).

Stimuli
Each synthetic exemplar was cut into excerpts of different durations, equally spaced on a logarithmic scale (40, 91, 209, 478, 1093, and 2500 ms). A 10-ms Hanning window was applied to the beginning and the ending segments of each excerpt, to smooth signal onset and offset. All excerpts were equalized to the same root-mean-squared level (rms = 0.1).

Experimental design
Participants sat in front of a computer and performed the task using the mouse. Experiments were implemented in MATLAB. All subjects were blindfolded by a mask that could filter almost 100% of the light and kept their eyes closed. Stimuli were played on a Macbook Pro 2017, with a built-in sound card with a frequency sampling rate of 44.1 Hz, through a headphone set Audio Technica Pro ATH-M50X, at a volume of c.a. 75 dB SPL, which was kept constant for each stimulus.

The task was very similar to the one in the original paper by McDermott et al. (2013) but modified to be suitable for visually deprived individuals. In fact, no visual interaction with the screen was required and audio instructions were provided through the headphones in Italian or English, according to participant's preferred language.

Participants performed two sessions, each comprising either version of the two experiments (Exemplar Discrimination or Texture Discrimination), presented in a counterbalanced order across subjects. Both sessions were performed in the same day with a half-an-hour break in between; each session lasted approximately 35 - 40 min 54 Sound Textures were employed in Texture Discrimination and 36 in Exemplar (see Table S1); each experiment comprised 216 trials. Every 54 trials, corresponding to about 6-7 min of stimulation, participants were allowed to take a few-minutes break, for a total of four runs per each session, separated by four small breaks.

In both sessions, each trial consisted of a triplet of sounds of the same duration. Stimuli duration could vary across trials, and six durations were employed (either 40, 91, 209, 478, 1093 or 2500 ms). The number of trials per duration was equal across all the six possible durations (36 trials for each one) and stimuli were presented in a randomized order. To control for a stimulus expectancy effect, inter-stimulus interval (ISI) could vary between 400, 500, 600, 700, and 800-ms. Participants were asked to be as accurate as possible in their choice and there was no time-limit to answer. Once the participant had responded, the presentation of the next triplet occurred after a short pause lasting between 2 and 2.5-s, with steps of 100 ms.

Test sessions were performed before both of the actual experimental sessions, to make sure participants understood the tasks and to have them familiarize with the type of stimuli. Test stimuli were drawn randomly across the 36 texture pairs (Table S1); excerpts used in the trial session were then excluded by the actual experiment. Three trials for each duration were presented in a random order during test sessions (for a total of 18 trials per duration). Feedback was provided only during the trial session and consisted in an audio message stating if their response was correct or incorrect. Feedback was not provided during experimental sessions, following the protocol by McDermott et al. (2013).

Exemplar discrimination

Two excerpts coming respectively from two exemplars of the same Sound Texture were selected. Both excerpts were extracted at the same time point along the 5-s segments. Excerpts were then presented as triplet of sounds in a trial: consequently, one of the two excerpts was repeated twice, while the third sound was the other. The different sound could be located as the first one or the last one in the triplet, and this varied randomly across trials.

Participants were informed that two stimuli will be identical and were asked to indicate which stimulus was different from the other two. If they thought it was the first one, they would click the left button of the mouse, otherwise the right one. The middle sound was used as the reference (Figure 1C).

Texture discrimination

Three excerpts were drawn among three different exemplars coming from pairs of Original Sound Textures: two of the excerpts came from two different exemplars of the same Sound Texture while the third one from an exemplar of another Sound Texture. Sound Textures were paired according to similarity, as done by McDermott et al. (2013); Table S1) and the different exemplar was a synthetic sound derived from the same white noise of one of the other two exemplar, so their original associated fine structure was the same, while only the imposed statistics were different. All three excerpts were extracted at the same time point along the 5-s segments. Again, the three excerpts were presented in a trial so that all the sounds were different, but two of them would originate from the same Sound Texture, while the other one had a different sound source. Participants were made aware of the fact that all sounds could be different and were asked to report which one came from the diverging acoustic source. Some examples were provided in order to facilitate task comprehension (i.e., "Two sounds can be the sound of a fireplace, the other one is the sound of the rain"). The correct answer could be either the first or last sound in the triplet, while middle one had to be used has a reference. Middle excerpt and deviant excerpts were derived from the same white noise sample. If participants thought the deviant was the first sound, they would click the left button of the mouse, otherwise the right one (Figure 1B).

Average statistical variability

Standard deviation of a set of employed envelope statistics (envelope mean, variance, skew, cross-correlation, and modulation power) was measured between couples of excerpts presented in both experiments. The pair could be made of excerpts coming from different exemplar of the same Sound Texture (column 1, Table S1) or excerpts coming from different Sound Textures (column 1 and 2, Table S1; Figure 1D). The average of all statistics variability among couples of sounds was computed to make predictions about performance. When both excerpts originated from different exemplars of the same Sound Textures, their variability tended toward zero as duration increased, whereas opposite trend was observed when stimuli came from different Sound Textures. Moreover, differences in variability between the two conditions (Different Exemplar vs. Different Texture) increased with duration (Figure 1D). Separately for each of the aforementioned statistics, we performed pairwise comparisons (t-tests, FDR corrected) between the two conditions (Different Exemplar vs. Different Texture) and observed that variability between couples of excerpts coming from different Sound Textures, as compared to ones originating from different exemplar of the same Texture in envelope mean was larger for all durations (all $p > 0.05$, corrected); in envelope skewness in was larger at duration 91, 478, 1093, 2500 ms (all $p > 0.05$, corrected); variability significantly diverged only at long durations in envelope variance, envelope cross-correlation (478ms, 1093, and 2500ms; all $p < 0.001$, corrected), and modulation power (1093, 2500 ms; all $p < 0.01$, corrected; Figure S1). Analyses were performed using MATLAB.

QUANTIFICATION AND STATISTICAL ANALYSIS

Proportion of correct responses for each individual was used as dependent measure for statistical analyses. Main statistical analyses were performed using IBM SPSS Statistics for Macintosh, Version 26.0; Post-hoc analyses and False Discovery Rate (Benjamini and Hochberg, 1995) correction were performed using MATLAB. To assess that sample data were normally distributed, we performed both Kolmogorov-Smirnov and Shapiro-Wilk tests with the average of proportion of correct answers in each experiment (separately) as dependent variables. In each test, dependent variable was split according to group. Both tests indicated that data were normally distributed in both experiments and for all the groups, giving significance values higher than 0.05 (see Table S2).

We performed an ANCOVA. The model included the between-participants factor Group (SC, CB, LB) and two within-participant factors: Experiment (Exemplar Discrimination vs. Texture Discrimination) and Duration (40, 91, 209, 478, 1093, 2500); given the relatively large age-range across the entire sample (20–62 years old), age was included as a covariate. Since groups were matched, age unlikely accounted for between-groups effects, but it could still have had an impact at the individual level. Mauchly's test indicated that the assumptions of sphericity had been violated for the interaction Experiment * Duration ($c2(2) = 25.62$, $p < 0.05$). Thus, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.81$).

There was a significant main effects of Duration, $F_{(5, 250)} = 9.45$, $p < 0.001$, $\eta^2 = 0.16$, and its interaction with the factor Experiment, $F_{(4.07, 203.35)} = 15.91$, $p < 0.001$, $\eta^2 = 0.24$. Also, there was a significant main effect of the between-participants factor Group, $F_{(2, 50)} = 4.99$, $p = 0.01$, $\eta^2 = 0.17$, and a significant interaction between factors Group and Experiment, $F_{(2,50)} = 8.35$, $p = 0.001$, with a large effect size of $\eta^2 = 0.25$. Moreover, there was a significant three-way interaction effect across the factors Experiment, Duration, and Group, $F_{(10, 250)} = 3.29$, $p = 0.001$, with a medium-to-large eta-squared of $\eta^2 = 0.12$. Participant's age and its interactions with all other independent variables in the model were non-significant (all $p > 0.14$).

To break down the three-way interaction, we performed two-ways ANOVAs.

To test how the performance changed between the two experiments within each sample (either SC, CB, or LB), we performed a two-ways ANOVA with factors Experiment and Duration (Figures 2A–2C).

For all groups, a significant interaction between Experiment and Duration emerged (SC: $F_{5,85} = 119.81$, $p < 0.001$, $\eta^2 = 0.88$; CB: $F_{2.97, 50.37} = 86.09$, $p < 0.001$, $\eta^2 = 0.88$, degrees of freedom were corrected for Greenhouse-Geisser estimate, $\varepsilon = 0.59$; LB: $F_{5,85} = 46.46$, $p < 0.001$, $\eta^2 = 0.73$). Then, for each duration, the performance in Exemplar and Texture Discrimination was compared. To this end, post-hoc pairwise comparisons were performed using two-sided t-tests (Cramer et al., 2016). All p values were corrected using the

False Discovery Rate (Benjamini and Hochberg, 1995), with a q-value of 0.05. In both SC and CB, data replicated the ones from McDermott et al. (2013). In SC, the performance diverged between experiments at duration 40 (mean proportion correct: Exemplar = 0.87; SE = 0.02; Texture = 0.65; SE = 0.02; $p < 0.001$, corrected), 91 (Exemplar = 0.89; SE = 0.02; Texture = 0.77; SE = 0.02; $p = 0.001$, corrected), 478 (Exemplar = 0.84; SE = 0.02; Texture = 0.91; SE = 0.01; $p = 0.03$, corrected), 1093 (Exemplar = 0.77; SE = 0.03; Texture = 0.97; SE = 0.01; $p < 0.001$), and 2500 (Exemplar = 0.70; SE = 0.03; Texture = 0.97; SE = 0.01; $p < 0.001$, corrected), while no difference was observed for the intermediate duration 209 (Exemplar = 0.88; SE = 0.02; Texture = 0.85; SE = 0.01; $p = 0.22$, corrected; see Figure 2A). The same pattern of results were observed in CB group, with performance being better in Exemplar Discrimination for duration 40 (mean proportion correct = 0.87; SE = 0.02), and 91 (mean proportion correct = 0.85; SE = 0.03) as compared to Texture Discrimination (mean proportion correct: 40 = 0.64; SE = 0.02; $p < 0.001$, corrected; 91 = 0.77; SE = 0.02; $p < 0.01$, corrected), and in Texture Discrimination for duration 478 (mean proportion correct = 0.92; SE = 0.01), 1093 (mean proportion correct = 0.96; SE = 0.01), and 2500 (mean proportion correct = 0.98; SE = 0.01) as compared to Exemplar Discrimination (mean proportion correct: 478 = 0.83; SE = 0.02; $p = 0.004$, corrected; 1093 = 0.77; SE = 0.03; $p < 0.001$, corrected; 2500 = 0.64; SE = 0.04; $p < 0.001$, corrected). No difference was found at duration 209 (Exemplar = 0.87; SE = 0.02; Texture = 0.86; SE = 0.02; $p = 0.80$, corrected; see Figure 2B).

Conversely, in LB, the performance between Exemplar and Texture did not significantly differ for the short durations 40 (Exemplar: 40 = 0.71; SE = 0.03; Texture = 0.70; SE = 0.01; $p = 0.72$, corrected), 91 (Exemplar = 0.74; SE = 0.03; Texture = 0.76; SE = 0.02; $p = 0.72$, corrected), and 209 (Exemplar = = 0.79, SE = 0.02; Texture = 0.85, SE = 0.01; $p = 0.05$, corrected). Significant different performances were observed only for long durations: 478 (Exemplar: 0.74; SE = 0.02; Texture = 0.92; SE = 0.01; $p < 0.001$, corrected), 1093 (Exemplar = 0.72; SE = 0.03; Texture = 0.97; SE = 0.01; $p < 0.001$, corrected), and 2500 (Exemplar = 0.61, SE = 0.02; Texture = 0.99; SE = 0.00; $p < 0.001$, corrected) with participants performing better in Texture as compared to Exemplar Discrimination (see Figure 2C).

To investigate between-groups differences within each experiment, we conducted two separate two-ways ANOVAs (one for Texture and one for Exemplar Discrimination) with two factors, Duration and Group.

Within Exemplar Discrimination, we observed a significant effect of Group ($F_{2, 51} = 7.42$; $p = 0.001$; $\eta^2 = 0.23$), Duration ($F_{3.92, 199.78} = 55.18$; $p < 0.001$; $\eta^2 = 0.52$), and their interaction ($F_{10, 255} = 2.77$; $p = 0.003$; $\eta^2 = 0.10$); degree of freedom for factor Duration were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.78$). Post-hoc pairwise comparisons were run using two-sided t-tests for independent samples comparing the performance between groups in Exemplar Discrimination, at each duration. All p values were corrected using the same FDR procedure employed for within-group comparisons. Results revealed that, for any duration, performance did not significantly differ between SC and CB (all $p > 0.29$, corrected). For short durations, LB group preformed significantly worse than SC (40, $p < 0.001$, corrected; 91, $p < 0.001$, corrected; 209 $p = 0.03$, corrected; 478, $p = 0.02$, corrected), and CB group (40, $p < 0.001$, corrected; 91, $p = 0.03$, corrected; 478, $p = 0.02$, corrected). No difference was observed between LB and both SC and CB at duration 1093 and 2500 (all $p > 0.05$, corrected), nor between LB and CB at duration 209 ($p = 0.05$, corrected; see Figure 2D).

Within Texture Discrimination, only a significant effect of Duration was found $F_{3.95, 201,443} = 362.296$, $p > 0.01$, $\eta^2 = 0.88$ (degree of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = 0.79$) while the effect Group and its interaction with Duration were not significant (all $p > 0.10$). As expected, in this experiment, performance progressively increased with duration, in a manner that did not significantly differ across groups (Figure 2E).

Difference between exemplar and texture

For every duration, each participant's accuracy-score in Texture discrimination was subtracted from the scores in Exemplar Discrimination (Figure 2F). To test whether there was a significant difference among groups, we ran an ANOVA for repeated measures with the within-subjects factor Duration and the between-subjects factor Group. The dependent variable was the relative difference between the accuracy scores in Exemplar Discrimination and Texture Discrimination. Main effects of Group, $F_{(2, 51)} = 8.72$, $p < 0.001$, $\eta^2 = 0.26$, and Duration, $F_{(5, 255)} = 230.42$, $p < 0.001$, $\eta^2 = 0.82$, were significant, together with their interaction, $F_{(10, 255)} = 3.22$, $p < 0.001$, $\eta^2 = 0.11$. We ran FDR corrected pairwise comparisons (two-tailed

t-tests; q-value = 0.05) on pre-selected contrasts of interest highlighting the differences between groups within each duration. All the comparisons between CB and SC were not significant (all $p > 0.32$, corrected). Comparisons between LB and CB were significant at duration 40 ($p < 0.001$, corrected), 478 ($p < 0.03$ corrected) and not significant for other durations (all $p > 0.09$, corrected). Finally, comparisons between SC and LB were significant for most of the durations (40, 91, 209, 478, 2500; all $p < 0.03$, corrected) but 1093 ($p = 0.32$, corrected) (Figure 2D). Absolute values of the average across participants of the differences at the Group level are displayed separately for each group in supplemental information, Figures S4A–S4C. For both SC (Figure S4A) and CB (Figure S4B) we observed a U-shaped trend, consisting of higher values at short and long durations and almost zero at intermediate one. On the other hand, in LB groups (Figure S4C) a different trend was observed, with values being almost zero at short durations (40, 91) and progressively increasing at longer ones.

Testing for disposition bias

As participants were asked to choose between first and third intervals, the temporal connotation of this 2AFC protocol could have led to a response disposition bias (e.g., listeners could have shown a tendency toward reporting one of the two intervals, for example the last one). In order to rule out that systematic trends for bias could differ across groups and, to some extent, could account for the impaired performance in LB, for each participant we calculated how many times they overestimated stimuli in one position, as compared to the other. Separately for each experiment and for each duration, the total number of times participants pressed the mouse's left button, stating that deviant sound was the first interval, was subtracted from the total number of times that correct answer was actually the left one. Positive numbers would indicate an overestimation of stimuli in the last interval, whereas negative values would refer to an overestimation of first intervals. In order to check for significant differences, we ran a Repeated-Measure ANOVA with total number of overestimated stimuli in one position as dependent variable, Group as between-subjects factor, and Experiment and Duration as within-subjects factors. As Mauchly's test indicated that the assumptions of sphericity had been violated for the effect Duration ($c2(2) = 76.28$, $p < 0.001$), the interactions Experiment*Duration ($c2(2) = 78.76$, $p < 0.001$), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (Duration, $\varepsilon = 0.58$; Experiment*Duration, $\varepsilon = 0.56$). No significant effects were found for any within-subjects factors, nor for the between-subject factor Group and their interactions (all $p > 0.07$). Statistics were carried out using IBM SPSS Statistics for Macintosh, Version 26.0. Data are displayed in Figure S2A.

Testing for learning effect

Sessions were divided into four runs of 54 trials each. In order to check for the occurrence of divergent learning and/or tiredness effects between groups, within each experiment and duration, total number of correct answers in run 1 was used as a baseline and was subtracted from the number of correct answers in the next runs (run 2, 3, and 4). Positive values meant that performance increased compared to the first run, showing a learning effect, whereas negative values were associated with a decrease in the performance, possibly a tiredness effect. We ran a Repeated-Measure ANOVA using IBM SPSS Statistics for Macintosh, Version 26.0, with baseline-subtracted correct answers as the dependent variable, Group as between-subjects factor, and Experiment, Duration, and Run within-subjects factors. Mauchly's test indicated that the assumptions of sphericity had been violated for the main effect of Duration ($c2(2) = 67.35$, $p < 0.001$), the interactions Experiment*Duration ($c2(2) = 38.80$, $p < 0.001$), Duration*Run ($c2(2) = 150.31$, $p < 0.001$), and Experiment*Duration*Run ($c2(2) = 212.37$, $p < 0.001$). Thus, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (Duration, $\varepsilon = 0.68$; Experiment*Duration, $\varepsilon = 0.77$; Duration*Run, $\varepsilon = 0.62$; Experiment*Duration*Run, $\varepsilon = 0.49$). We observed significant effects of Duration, $F_{(3.4, 173.47)} = 20.11$, $p < 0.001$, $\eta^2 = 0.28$, Run, $F_{(2, 102)} = 12.48$, $p < 0.001$, $\eta^2 = 0.20$, and the interactions Duration*Run, $F_{(10, 315.08)} = 13.81$, $p < 0.001$, $\eta^2 = 0.21$, and Duration*Run*Group, $F_{(12.36, 315.08)} = 20.11$, $p < 0.05$, $\eta^2 = 0.08$. Pairwise comparisons were carried out, to test whether between groups differences (SC vs. CB, CB vs. LB, SC vs. LB) existed for each run and duration. No significant difference was observed between the 54 contrasts of interest (all $p > 0.05$, corrected). Data are plotted in Figure S2B, showing similar trends across groups for all durations.

Correlations

Pearson's correlation coefficient (RHO) was computed, together with p values between LB participants' onset of blindness and duration of blindness with their overall performance in each experiment (Figure S5A)

and their performance for each duration in each experiment (Figures S5B and S5C). We observed no significant correlation for most of the conditions and the variable tested (all p > 0.05). To rule out a possible age-related confounding effect, linear correlation was computed also between performance in Exemplar Discrimination with LB's participants age (Figure S6). No correlation was found significant for any of the tested durations (all p > 0.37; see Figure S6). Correlations were performed and plotted with MATLAB.