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

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

[1]Molecular Mind Lab, IMT School for Advanced Studies, Lucca, Italy

*Correspondence: davide.bottari@imtlucca.it, (D.B.)

Summary

The auditory system relies on both detailed and summarized representations to recognize different sounds. As the amount of local features can exceed storage capacity, average statistics are computed over time to generate more compact representations at the expense of temporal details availability. It is unknown whether local features processing and statistical averaging are restricted to the auditory system or interact with other modalities, such as vision. To answer this question, we tested sighted (SC), congenitally blind (CB), and late-onset blind (LB) individuals in two auditory experiments: one relied on the availability of local features at specific time points; the other benefitted from computing average statistics over longer durations. SC performed as expected: when sound duration increased, detailed representation gave way to summary statistics. CB individuals showed similar patterns of behavior, revealing that auditory processing of local features, as well as statistical averaging, are not altered by absence of vision since birth. Surprisingly, interactions between auditory computations and sight emerge later in development: LB individuals performed poorly compared to both SC and, crucially, CB individuals. Only the performance relying on local features analysis was degraded, with no impact on statistical averaging. These results suggest that early development is characterized by the segregation between basic auditory computations and sight, while at later developmental phases a functional interplay between the access to local auditory details and availability of visual input emerges.

INTRODUCTION

The auditory system is characterized by the capture of fine-grained details of 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 (Yabe *et al.*, 1998; McDermott, Schemitsch,and Simoncelli, 2013). Computational synthesis approaches, derived by information theories (Barlow, 1961), allowed to investigate these processes by implementing mathematical models to describe the set of measurements that the auditory system operates (McDermott and Simoncelli, 2011; McDermott, Schemitsch and Simoncelli, 2013). Such models have revealed that sounds are analyzed through the extraction of a set of auditory statistics (McDermott and Simoncelli, 2011) whose processing represents the keystone of auditory features integration. Auditory statistics result from a set of computations strictly dependent on anatomical and functional properties of the auditory pathway (Ruggero, 1992; Dau, Kollmeier and Kohlrausch, 1997; Gygi, Kidd and Watson, 2004; Joris, Schreiner and Rees, 2004; Baumann *et al.*, 2011), 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, Schemitsch, and Simoncelli, 2013; McWalter and McDermott, 2018), along specific time-windows (McWalter and McDermott, 2018) and uses this information to derive compact representations of sound objects.

While 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. For instance, studies in non-human animal models revealed that, in the early development, the onset of visual input gates the critical period closure of basic auditory functions (Mowery, Kotak and Sanes, 2016). Equally, studies in humans have also suggested that visual inputs availability since birth might be a prerequisite for the full development of specific auditory functions (e.g., auditory spatial calibration is impaired in congenitally blinds; Gori *et al.*, 2014). On the other hand, the lack of vision is known to induce compensatory effects in hearing. Consistent evidence exists that congenitally blind individuals equate, if not outperform, their sighted counterpart in several auditory tasks (for review, see Pavani & Röder, 2012). Compensatory mechanisms have also been described for those individuals who have lost sight later in development (e.g., Fieger et al., 2006; Föcker et al., 2015; Röder & Rösler, 2003). However, none of these studies in vision-deprived individuals adopted computational approaches allowing to unveil subtending auditory computations.

Whether auditory statistics develop and function exclusively within the influence of the auditory systems, or whether they are influenced by other sensory modalities is as yet unknown. This study implemented a rigorous approach whose main advantage is to uncover auditory

statistics computations from behavioral performances. 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*, the process by which local features are averaged within an integration time-window when the amount of information exceeds memory span and local details can no longer be retained (McDermott, Schemitsch and Simoncelli, 2013).

To address whether these main modes of representation are influenced by visual input availability, we tested blind populations against sighted individuals in two experiments designed to tap into these computational modes. Importantly, comparing populations that were visually deprived since birth or in late phases of the development allowed us to assess whether the selected auditory computations interact with vision at specific time points along the life span.

RESULTS

To understand whether and how visual experience affects the auditory statistics processing, we combined the study of human models of permanent visual deprivation with a computational approach. We took advantage of an already corroborated methodological pipeline (McDermott, Schemitsch and Simoncelli, 2013) and we 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 & Simoncelli, 2011; Saint-Arnaud & Popat, 2006; Schwarz, 2011). These sounds are rich, ubiquitous, and constant over time (McDermott, Schemitsch and Simoncelli, 2013; McWalter and McDermott, 2018). We selected 54 environmental recordings of Sound Textures (**Table S1**), among the original set implemented by McDermott, Schemitsch, and Simoncelli, 2013. Synthetic stimuli were created by 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, Kollmeier and Kohlrausch, 1997; McDermott and Simoncelli, 2011), and their correlations (McDermott and Simoncelli, 2011). Although the model represents a mathematical approximation, cochlear envelope statistics are able to convey

most of the perceptually relevant information about the sound (Smith, Delgutte and Oxenham, 2002; Gygi, Kidd and Watson, 2004; McDermott and Simoncelli, 2011). 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 all the aforementioned statistics 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, while 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 aspects of 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).

First, to address which specific aspects within the auditory statistic processing undergo cross-modal visual influence, we compared sighted and blind participants' performances in two selected experiments designed to exploit the two modes of auditory statistics representation.

Second, to assess whether the interplay between visual input availability and the two computational modes changes along the life cycle, we selected blind participants according to blindness onset, being it from birth or developmentally later (>10 years old; see **Table S2).** 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.

Both experiments consisted of a two-alternative forced choice oddity (2AFC); participants had to detect the deviant sound among three acoustic samples, selecting between 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. Duration of the excerpts varied across trials, for a total of six durations (McDermott, Schemitsch and Simoncelli, 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 to the other two.

Nonetheless, as duration increased, average statistics tended to converge towards 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), while 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, but 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 perceptually emerge (**Figure 1D, Right panel**).

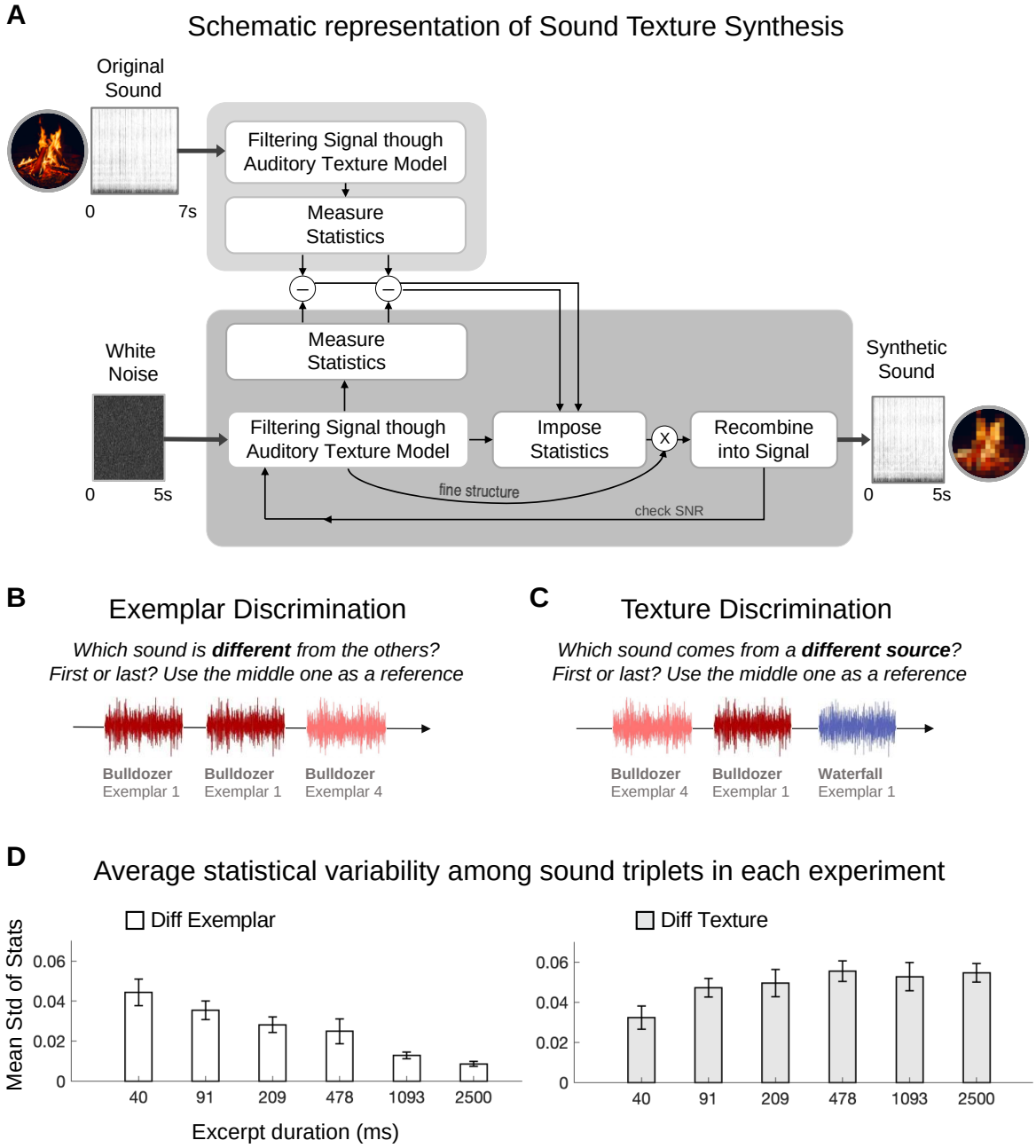

Figure 1. Sound Texture Synthesis, Auditory Statistics Variability, and Experimental Design
(A) Schematic representation of the Sound Texture Synthesis procedure employed to produce synthetic stimuli. A 7-s Original Recording was passed through the Auditory Texture Model [5] to extract average statistics 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 match the original values and that was only constrained by them. Adapted from McDermott and Simoncelli, 2011.
(B) Exemplar Discrimination. Schematic representation of one trial. Participants had to detect the different sound among three, being the other two 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, being it 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.
(See also Figure S3 and Table S1)

The employment of these experiments allowed to make specific predictions on the outcomes. Consistently with previous findings by McDermott, Schemitsch, and Simoncelli, 2013, opposite patterns of results were expected in the two experiments, as a function of excerpts duration. In summary, for short stimuli, due to differences in features variability among the three sounds, good performance was expected in Exemplar Discrimination as compared to Texture Discrimination, while for long stimuli statistical averaging was expected to boost performance in Texture Discrimination and to hamper it in Exemplar Discrimination (McDermott, Schemitsch and Simoncelli, 2013). Any significant deviation from these expected outcomes would indicate changes in the processing of local features and/or statistical averaging.

The attended pattern of results (McDermott, Schemitsch and Simoncelli, 2013) was confirmed in our SC group. The performance was better for short durations (40, 91ms) in Exemplar Discrimination as compared to Texture Discrimination (all $p < 0.003$, corrected). Conversely, for long durations (478, 1093, 2500ms) participants' accuracy was higher in Texture Discrimination as compared to Exemplar Discrimination (all $p < 0.03$, corrected). No difference was observed for trials comprising stimuli that were 209ms long ($p = 0.31$, corrected; see **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, 91ms) compared to Texture Discrimination (all $p < 0.02$, corrected), and, conversely for long durations (478, 1093, 2500ms) in Texture Discrimination compared to Exemplar (all $p < 0.006$, corrected). As observed in SC, there was no difference between the two experiments for stimuli that were 209ms long ($p = 0.86$, corrected; see **Figure 2B).**

By contrast, in the LB group, no significant differences were found between the performance in Exemplar Discrimination and in Texture Discrimination for all short stimuli (40, 91, 209ms; all $p > 0.05$, corrected)**.** However, for long durations (478, 1093, 2500ms), LB participants performed better in Texture Discrimination as compared to Exemplar Discrimination (all adjusted $p < 0.001$), in line with what was observed in the other groups (**Figure 2C**).

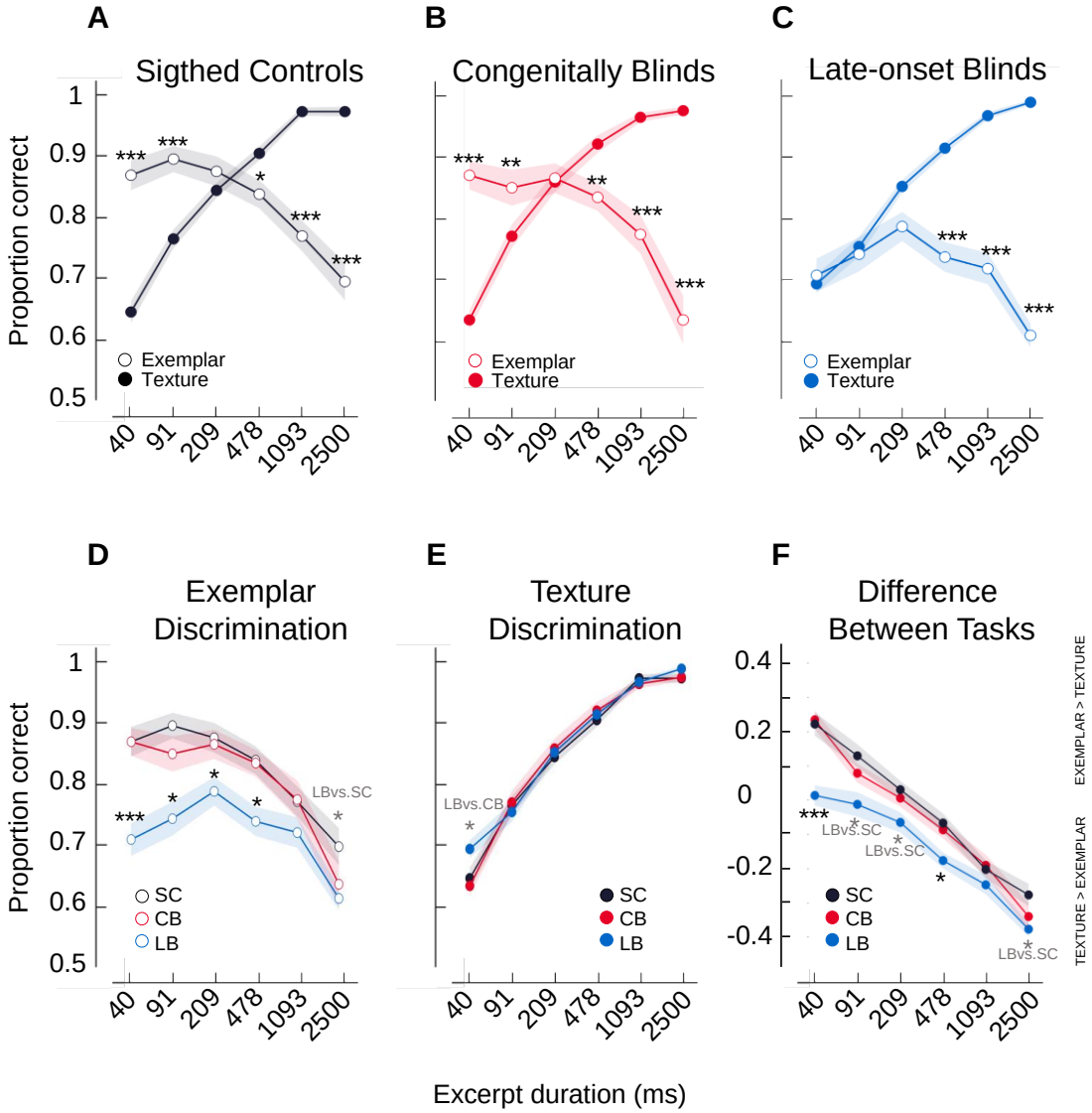

Figure 2. Proportion of correct answers in the two Experiments
(A, B, C) Results for Exemplar Discrimination vs. Texture Discrimination in each group. Proportion of correct answers across individuals at the group level are shown as a function of excerpt duration. (D) Between Group comparisons in Exemplar Discrimination. SC and CB groups' performance did not differ. LB group showed an impaired performance as compared to 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 differences were observed between SC and CB groups. LB performed better for 40ms stimuli compared to CB.
(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, while negative values represent better performance in Texture compared to

Exemplar Discrimination. Differently from the other two groups, LB never showed an advantage in Exemplar Discrimination compared to 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 was 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 S2 and Figure S4).

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

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 temporal fine-grained details is resilient to the absence of visual input since birth.

Conversely, the LB group performance was impaired for almost all of the tested durations as compared to both SC group (40, 91, 209, 478, 2500ms; all $p < 0.03$, corrected) and CB group (40, 91, 209, 478ms; all $p < 0.001$, 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, although this process is encumbered by late-onset blindness (**Figure 2D).**

Visual deprivation does not impact on statistical averaging

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

The performance of the SC group for any of the selected duration (40, 91, 209, 478, 1093, 2500ms) did not differ from both the CB group (all $p > 0.47$, corrected) and the LB group (all $p > 0.06$, corrected). When comparing the accuracy of the CB group with the one of the LB group, LB performed better for trials where stimuli were 40ms long (adjusted $p = 0.01$). No other difference was observed for all the remaining durations (91, 209, 478, 1093, 2500ms; all $p > 0.09$, corrected; **Figure 2E).** Altogether, these results reveal that visual deprivation does not exert a significant influence on 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 to both SC and CB groups, this difference was significantly smaller for very short durations in LB group (LC vs. SC: 40, 91; all $p < 0.02$, corrected; LC vs. CB: 40 $p < 0.001$, corrected, 91 $p = 0.04$, uncorrected) and larger for longer ones (LC vs. SC: 209, 478, 2500; all $p < 0.03$, corrected; LC vs. CB: 478 $p < 0.03$, corrected; see Figure 2F; Table S3). Conversely, 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 were performing 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 to previous investigations on blind individuals; all groups were properly matched by several main variables (size, age, and gender), and all participants had no documented auditory deficits, cognitive impairments, or neurological disorders (blindness was associated only to peripheral damage; **Table S1**; See STAR*METHODS). Moreover, LB participants' performance was hampered only in one of the two experiments, while the observed results in Texture Discrimination were indistinguishable from the ones of the remaining groups, for most of the selected durations. If anything, LB group performance was significantly more efficient at 40ms as compared to CB and partially SC ($p < 0.04$, 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.

In order 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 specific disposition bias towards 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 towards stimulus intervals (see STAR*METHODS; **Figure S4A**).

Another possible confounding effect could have been a between-group difference associated with learning and/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 across all groups emerged for each duration (see STAR*METHODS; **Figure S4B**).

In light of these clarifications, results can genuinely be attributed to late-onset sight loss hampering the use of local features

DISCUSSION

The present study aimed at addressing whether visual experience at different phases of the lifespan exerts a cross-modal influence over specific auditory statistics computations. No difference was observed between CB and SC groups for both modes of representation. Conversely, we found that LB's performance was selectively impaired compared to both SC and CB, for those conditions in which optimal performance relied on local features. On the one hand, these results provided evidence that visual experience is not necessary for the development of auditory statistics. On the other hand, the lack of vision can negatively impact on specific peripheral computations only when vision was available in early development and went missing afterward.

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 our study investigated on one hand 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 due 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 to 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; Gori, Amadeo and Campus, 2020b, 2020a). Second, coherently with previous evidence of enhanced auditory skills in congenitally and early blinds (e.g., Doucet et al., 2005; Muchnik et al., 1991; Röder et al., 1999), lack of vision since birth 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 to natural speech (Moos and Trouvain, 2007; Trouvain, 2007; Dietrich, Hertrich and Ackermann, 2013), an ability perhaps supported by a higher auditory sampling rate capacity. Thus, 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 long durations in Exemplar Discrimination. On the other hand, 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. Thus, our results strongly suggest relative independence of the development of auditory statistic processing from early visual experience. Moreover, the absence of compensatory mechanisms -arising due 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 which 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); for functions in the tactile domain see also (Barolo, Masini and Antonietti, 1990; Röder, Spence and Rösler, 2000). 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, while functions that showed visual input dependencies are mostly the result of higher-order operations at the sensory cortex level. Evidence suggests that multisensory functions performed at subcortical levels seem to exist from early developmental stages (Lewkowicz, 2000), and thus, rely less on experience compared to 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 the full development of the auditory statistics processing of Sound Textures is less experience-dependent from visual inputs than other high-level cortical computations underpinning previously documented behavioral adaptations. Further studies are required to resolve this open-ended issue, including statistics resulting from higher-order computations (Chi, Ru and Shamma, 2005; Norman-Haignere and McDermott, 2018) and involving other types of naturalistic stimuli.

Functional interplay between selective auditory computations and vision

By comparing CB, SC, and LB groups, we can affirm that while early vision is not necessary for the typical development of the auditory statistics processing, the lack of vision can still exert a great influence over the processing of local features

The absence of associations between the accuracy in both tasks and duration or onset of blindness was found and (see STAR*Methods; **Figure S5**) suggested that the performance (i) 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) and (ii) 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 test two early blind individuals, whose blindness onset occurred during their third year of life (See STAR*METHODS, **Figure S1**). Remarkably, their results systematically overlapped with the ones of CB and SC groups, providing further support to the evidence that the dampening of local features processing manifests only if sight loss occurs rather late in the development.

One possibility is that a visual influence on auditory computations can only take place after the complete development of the auditory system has occurred. Consistent evidence shows that, even considering basic auditory functions, human performance keeps improving gradually for over nearly a decade. Frequency discrimination, such as perceiving a difference between two tones presented sequentially, does not mature until roughly 10 years of age for low frequencies (Maxon and Hochberg, 1982; Jensen and Neff, 1993; Moore *et al.*, 2011). Similarly, the thresholds for detecting amplitude modulations become adult-like only after 10 years of age (Banai, Sabin and Wright, 2011). In accordance with these observations, after 10 years of age, the plasticity of both peripheral and central auditory systems is greatly reduced (Sharma, Dorman and Spahr, 2002). The use of local features comprised a number of operations (including the measurement of amplitude modulations over time; Dau, Kollmeier and Kohlrausch, 1997). Thus, it is possible they are designed to develop independently from visual input and only after their functional development is completed, interactions across senses are allowed. This would account for the effect of sight loss becoming explicit only at a rather late developmental stage.

Alternatively, the observed effects might depend on the development of multisensory functions. 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 audio-visual (AV) multisensory facilitations have been found immature until the age of 9 years old while 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). The present data might suggest that the

interactions between basic auditory computations tested here and vision occur only after full development of AV multisensory functions. The availability of visual input in early developmental phases would prompt the typical development of AV interactions. As a result, the sudden loss of visual input could in turn alter auditory processing which had typically developed.

Which reason can we suggest for local features processing being selectively impaired, despite unaltered statistical averaging? Noticeably, the deficit observed in the LB group does not exclude that limiting local features accessibility can provide ecological advantages. Identifying a Sound Texture often equals to recognize 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 -and/or at the expense of temporal details- prompts sound-object recognition in an everyday environment. The results in our LB group could suggest a form of adaptive perceptual learning relying on the implementation of strategical changes to face a remarkable loss in the overall available sensory input. Adult perceptual learning (Fiorentini & Berardi, 1997; Watanabe and Sasaki, 2015) has often been associated with the remapping of the sensory cortex (Gilbert, Sigman and Crist, 2001; Li, Piëch and Gilbert, 2004), is driven by feedback regulations (Schäfer, Vasilaki and Senn, 2007; Gilbert, Li and Piech, 2009), and does not reflect only exposure to sensory inputs, but the salience of stimulus and task to fulfill ecological purposes (Polley, Steinberg and Merzenich, 2006).

The local features processing must involve different steps, specifically features encoding, integration over time, and retention (McDermott, Schemitsch and Simoncelli, 2013). As statistics are averages of temporal features over time (McDermott and Simoncelli, 2011; McDermott, Schemitsch and Simoncelli, 2013), a basic encoding of local features is fundamental for statistical averaging to emerge. Participants in the LB group did not show any impairment in Texture Discrimination, thus they appeared to be capable to encode features and compute averages out of them. On the other hand, results for all groups in Exemplar Discrimination showed that as the system proceeds to integrate features over time, their retention goes missing due to storage capacity limitations (see also McDermott et al., 2013). Thus, we can postulate that statistical averaging could in principle occur normally, despite the temporal dynamics of local details integration and/or retention being altered. Specifically, two alternative interpretations can be proposed. First, the effect observed in the LB group in Exemplar Discrimination hinders principally short intervals (< 500ms) and thus could reflect changes in the temporal window of features integration. This would suggest that LB individuals might start to discriminate sounds only after a consistent amount of information is provided. Alternatively, the deficit observed in our LB sample could be specific for the retention of local features. In other words, for LB participants, local features might become unavailable earlier as compared to the other groups. Further investigations, focused on

breaking down the process into its subcomponents and tackling them separately, are required to disentangle which aspect within the processing of local features is specifically altered after late-sight loss. This will promote further understanding of interdependencies across modalities and the tracking of the developmental trajectory of sensory computations in a multimodal framework.

For example, by testing other participants who became blind within 3 and 10 years of age, it will be possible to better estimate the shape of this auditory function across life cycles in atypical development.

CONCLUSION

Here, we showed that the full development of auditory statistics processing does not require early visual input and may principally rely on both intrinsic biological properties of the auditory system and exposure to statistics included in sounds. However, interactions with vision can emerge when sight goes missing at later developmental stages.

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 form of adaptation to a late-onset sight loss third, it has the potential to expand our approaches in several fields, including auditory, visual, multisensory, and research on brain disorders; fourth, it provides novel ground for applied science, in the context of sensory substitution devices and auditory rehabilitation strategies for people who lost vision compared to those who were born blind. Finally, we prove that combining computational methods with human models of sensory deprivation provides the context to assess the degree of plasticity of specific computations performed by the sensory systems. While 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.

Acknowledgments

We are grateful to Antonio Quatraro for supporting the study at the Istituto per la Ricerca la Formazione e la Riabilitazione (IRIFOR) in Florence and to the Italian associations for the blinds (Unione Italiana Ciechi, UIC) in Lucca, Pisa, Pistoia, Prato, Trento, and the Istituto Cavazza in Bologna for their help during participants recruitment and testing. Finally, we would like to thank Dr. Peter Neri for providing important insights on the analyses and interpretation of the results.

Author Contributions

Conceptualization, D.B. and M.B.; Methodology, M.B. and D.B.; Investigation M.B. and D.B.; Writing – Original Draft, M.B. and D.B.; Writing – Review & Editing, M.B., D.B., E.R., P.P.; Visualization, M.B. and D.B.; Funding D.B. (PRIN 2017 research grant. Prot. 20177894ZH).

Declaration of Interests

The authors declare no competing interests.

REFERENCES

Bacon, S. P., & Grantham, D. W. (1989). Modulation masking: Effects of modulation frequency, depth, and phase. *The Journal of the Acoustical Society of America*, *85*(6), 2575-2580.

Banai, K., Sabin, A. T., & Wright, B. A. (2011). Separable developmental trajectories for the abilities to detect auditory amplitude and frequency modulation. *Hearing research*, *280*(1-2), 219-227.

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

Barolo, E., Masini, R., & Antonietti, A. (1990). Mental rotation of solid objects and problem-solving in sighted and blind subjects. *Journal of Mental Imagery*.

Baumann, S., Griffiths, T. D., Sun, L., Petkov, C. I., Thiele, A., & Rees, A. (2011). Orthogonal representation of sound dimensions in the primate midbrain. *Nature neuroscience*, *14*(4), 423-425.

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cerebral Cortex*, *21*(5), 1042-1055.

Castaldi, E., Lunghi, C., & Morrone, M. C. (2020). Neuroplasticity in adult human visual cortex. *Neuroscience & Biobehavioral Reviews*.

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

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

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, *102*(5), 2892-2905.

de Villers-Sidani, E., & Merzenich, M. M. (2011). Lifelong plasticity in the rat auditory cortex: basic mechanisms and role of sensory experience. In *Progress in brain research* (Vol. 191, pp. 119-131). Elsevier.

Dietrich, S., Hertrich, I., & 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 neuroscience*, *14*(1), 74.

Doucet, M. E., Guillemot, J. P., Lassonde, M., Gagné, J. P., Leclerc, C., & Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Experimental brain research*, *160*(2), 194-202.

Fieger, A., Röder, B., Teder-Sälejärvi, W., Hillyard, S. A., & Neville, H. J. (2006). Auditory spatial tuning in late-onset blindness in humans. *Journal of Cognitive Neuroscience*, *18*(2), 149-157.

Foecker, J., Hoelig, C., Best, A., & Roeder, B. (2015). Neural plasticity of voice processing: Evidence from event-related potentials in late-onset blind and sighted individuals. *Restorative neurology and neuroscience*, *33*(1), 15-30.

Gilbert, C. D., Li, W., & Piech, V. (2009). Perceptual learning and adult cortical plasticity. *The Journal of physiology*, *587*(12), 2743-2751.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, *31*(5), 681-697.

Gori, M., Amadeo, M. B., & Campus, C. (2020). Spatial metric in blindness: behavioural and cortical processing. *Neuroscience & Biobehavioral Reviews*, *109*, 54-62.

Gori, M., Amadeo, M. B., & Campus, C. (2020). Temporal cues trick the visual and auditory cortices mimicking spatial cues in blind individuals. *Human Brain Mapping*, *41*(8), 2077-2091.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, *18*(9), 694-698.

Gori, M., Sandini, G., Martinoli, C., & Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain*, *137*(1), 288-293

Gygi, B., Kidd, G. R., & Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *The Journal of the Acoustical Society of America*, *115*(3), 1252-1265.

Huber, E., Chang, K., Alvarez, I., Hundle, A., Bridge, H., & Fine, I. (2019). Early blindness shapes cortical representations of auditory frequency within auditory cortex. *Journal of Neuroscience*, *39*(26), 5143-5152.

Jensen, J. K., & Neff, D. L. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science*, *4*(2), 104-107.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological reviews*, *84*(2), 541-577.

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychological bulletin*, *126*(2), 281.

Li, W., Piëch, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature neuroscience*, *7*(6), 651-657.

Lorenzi, C., Berthommier, F., Apoux, F., & Bacri, N. (1999). Effects of envelope expansion on speech recognition. *Hearing research*, *136*(1-2), 131-138.

Maxon, A. B., & Hochberg, I. (1982). Development of psychoacoustic behavior: Sensitivity and discrimination. *Ear and Hearing*, *3*(6), 301-308.

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature neuroscience*, *16*(4), 493-498.

McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*, *71*(5), 926-940.

McWalter, R., & McDermott, J. H. (2018). Adaptive and selective time averaging of auditory scenes. *Current Biology*, *28*(9), 1405-1418.

Moore, D. R., Cowan, J. A., Riley, A., Edmondson-Jones, A. M., & Ferguson, M. A. (2011). Development of auditory processing in 6-to 11-yr-old children. *Ear and hearing*, *32*(3), 269-285.

Moos, A., & Trouvain, J. (2007). Comprehension of Ultra-Fast Speech–Blind vs.'Normally Hearing'Persons. In *Proceedings of the 16th International Congress of Phonetic Sciences* (Vol. 1, pp. 677-680). Saarland University Saarbrücken, Germany.

Mowery, T. M., Kotak, V. C., & Sanes, D. H. (2016). The onset of visual experience gates auditory cortex critical periods. *Nature communications*, *7*(1), 1-11.

Muchnik, C., Efrati, M., Nemeth, E., Malin, M., & Hildesheimer, M. (1991). Central auditory skills in blind and sighted subjects. *Scandinavian audiology*, *20*(1), 19-23.

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

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

Picard, D., Dacremont, C., Valentin, D., & Giboreau, A. (2003). Perceptual dimensions of tactile textures. *Acta psychologica*, *114*(2), 165-184.

Plomp, R. (1964). Rate of decay of auditory sensation. *The Journal of the Acoustical Society of America*, *36*(2), 277-282.

Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *Journal of neuroscience*, *26*(18), 4970-4982.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, *40*(1), 49-70.

Putzar, L., Hötting, K., & Röder, B. (2010). Early visual deprivation affects the development of face recognition and of audio-visual speech perception. *Restorative neurology and neuroscience*, *28*(2), 251-257.

Röder, B., & Rösler, F. (2003). Memory for environmental sounds in sighted, congenitally blind and late blind adults: evidence for cross-modal compensation. *International Journal of Psychophysiology*, *50*(1-2), 27-39.

Röder, B., Spence, C., & Rösler, F. (2000). Inhibition of return and oculomotor control in the blind. *Neuroreport*, *11*(13), 3043-3045.

Röder, B., Teder-SaÈlejaÈrvi, W., Sterr, A., Rösler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature*, *400*(6740), 162-166.

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, *33*(12), 2329-2337.

Ruggero, M. A. (1992). Responses to sound of the basilar membrane of the mammalian cochlea. *Current opinion in neurobiology*, *2*(4), 449-456.

Saint-Arnaud, N., & Popat, K. (1995). Analysis and synthesis of sound textures. In *in Readings in Computational Auditory Scene Analysis*.

Schäfer, R., Vasilaki, E., & Senn, W. (2007). Perceptual learning via modification of cortical top-down signals. *PLoS Comput Biol*, *3*(8), e165.

Schwarz, D. (2011, September). State of the art in sound texture synthesis.

Sharma, A., Dorman, M. F., & Spahr, A. J. (2002). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear and hearing*, *23*(6), 532-539.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(6876), 87-90.

Starlinger, I., & Niemeyer, W. (1981). Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness I. Peripheral functions. *Audiology*, *20*(6), 503-509.

Trouvain, J. (2007). On the comprehension of extremely fast synthetic speech.

Voss, P., Gougoux, F., Zatorre, R. J., Lassonde, M., & Lepore, F. (2008). Differential occipital responses in early-and late-blind individuals during a sound-source discrimination task. *Neuroimage*, *40*(2), 746-758.

Weaver, K. E., & Stevens, A. A. (2006). Auditory gap detection in the early blind. *Hearing research*, *211*(1-2), 1-6.

Weber, A. I., Saal, H. P., Lieber, J. D., Cheng, J. W., Manfredi, L. R., Dammann, J. F., & Bensmaia, S. J. (2013). Spatial and temporal codes mediate the tactile perception of natural textures. *Proceedings of the National Academy of Sciences*, *110*(42), 17107-17112.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, *35*(5), 615-619.