Cross-modal Implicit Learning of Random Time Patterns

HiJee Kang[1,2]*, Ryszard Auksztulewicz[1,3]*, Chi Hong Chan[1], Drew Cappotto[1,4], Vani Gurusamy Rajendran[1,5], Jan W. H. Schnupp[1]

[1] Department of Neuroscience, City University of Hong Kong, Hong Kong S.A.R.
[2] Department of Biomedical Engineering , Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
[3] Center for Cognitive Neuroscience Berlin, Free University Berlin, Berlin, Germany
[4] UCL Ear Institute, University College London, London, United Kingdom
[5] Department of Cognitive Neuroscience, Instituto de Neurobiología, Universidad Nacional Autónoma de México, Querétaro, Mexico.

* Equal contribution

Abbreviated title: Cross-modal Implicit Learning of Random Time Patterns

Corresponding author email address: wschnupp@cityu.edu.hk

 https://doi.org/10.1016/j.heares.2023.108857

Highlights

- Neural signatures of learning transfer across modalities were observed.
- Neural correlates of initial learning rely on modality-specific brain regions.
- Neural correlates of learning transfer across modalities rely on frontal regions.
- In frontal regions, the cross-modal effect could be linked to beta-band activity.

https://doi.org/10.1016/j.heares.2023.108857

Abstract

Perception is sensitive to statistical regularities in the environment, including temporal characteristics of sensory inputs. Interestingly, implicit learning of temporal patterns in one modality can also improve their processing in another modality. However, it is unclear how cross-modal learning transfer affects neural responses to sensory stimuli. Here, we recorded neural activity of human volunteers using electroencephalography (EEG), while participants were exposed to brief sequences of randomly timed auditory or visual pulses. Some trials consisted of a repetition of the temporal pattern within the sequence, and subjects were tasked with detecting these trials. Unknown to the participants, some trials reappeared throughout the experiment across both modalities (Transfer) or only within a modality (Control), enabling implicit learning in one modality and its transfer. Using a novel method of analysis of single-trial EEG responses, we showed that learning temporal structures within and across modalities is reflected in neural learning curves. These putative neural correlates of learning transfer were similar both when temporal information learned in audition was transferred to visual stimuli and vice versa. The modality-specific mechanisms for learning of temporal information, and general mechanisms which mediate learning transfer across modalities, had distinct physiological signatures: temporal learning within modalities relied on modality-specific brain regions, while learning transfer affected beta-band activity in frontal regions.

 https://doi.org/10.1016/j.heares.2023.108857

Introduction

Sensory signals in each modality contain complex feature-based information. Unlike modality-specific features - such as sound frequency or colour spectrum - timing information is universal across modalities (Hardy and Buonomano, 2016; Lim et al., 2016). Extracting statistical regularities based on temporal stimulus features is essential for efficient sensory perception (Nobre and van Ede, 2018; Sohoglu and Chait, 2016). However, it is unclear whether this ability is mediated by global mechanisms or differs across sensory modalities. For instance, the auditory modality is highly sensitive to timing information, enabling temporal discrimination of sensory inputs within a millisecond range (Goodfellow, 1934; Grondin, 1993; Grondin and McAuley, 2009; Grondin and Rousseau, 1991; Lechelt, 1975) and facilitating the processing of rhythmic patterns (Grahn et al., 2011; Grondin and McAuley, 2009; Patel et al., 2005). More efficient temporal processing in the auditory modality has also been claimed when sensory information is presented in multiple modalities concurrently, with participants' task performance being more biased by auditory inputs when multimodal sensory inputs are not synchronised (Ball et al., 2018a; Ball et al., 2018b; Ball et al., 2022; Guttman et al., 2005; Recanzone, 2003; Repp and Penel, 2002). Furthermore, neuroimaging studies reported the involvement of auditory regions during visual timing processing, further implying the dominance of the auditory system for processing timing regardless of stimulus modality (Grahn et al., 2011; Kanai et al., 2011).

However, there is no consensus on whether temporal processing follows modality-specific or modality-general mechanisms, or both. Some studies reported that auditory dominance in temporal processing could only be driven by the experimental settings requiring timing sensitivity as a crucial factor (Grahn, 2012; Rammsayer, 2014). In contrast, human neuroimaging studies reported similar EEG responses in temporal processing across modalities as evidence of

 https://doi.org/10.1016/j.heares.2023.108857

generalised mechanisms (Merchant et al., 2008; Pasinski et al., 2016). Therefore, while it is clear that there is a quantitative temporal sensitivity difference across modalities, further investigation is required to test whether temporal processing is modality-specific when the effect of temporal acuity is minimised. Importantly, mechanisms of temporal processing identified for very simple stimuli might not generalise to more naturalistic stimuli (Schutz and Gillard, 2020), whose processing likely involves higher-order sensory abilities such as memory and learning.

The learning of complex sequential sensory inputs is often studied using statistical learning paradigms in which sensory information is learned based on stimulus regularity patterns. It has been suggested that statistical learning of sequences is subserved by both modality-specific and modality-independent mechanisms (Conway and Christiansen, 2006; Mitchel and Weiss, 2011). Sequence learning in both auditory and visual modalities has been linked to modality-specific cortices, as well as to shared regions, including the inferior frontal gyrus as well as subcortical regions such as the hippocampus, caudate, and thalamus (Milne et al., 2018). However, stimuli used in statistical learning often include additional modality-specific features such as auditory frequency or visual spatial cues, and a traditional training-and-test experimental setting reflects an explicit learning scheme. Furthermore, while most statistical learning paradigms typically rely on learning contingencies between particular stimuli, which may not easily transfer across modalities, learning more abstract patterns (e.g., based on stimulus repetition) can be more easily transferred (Milne et al., 2018). Thus, it remains unclear how unsupervised learning of abstract, randomly generated sensory input is processed across modalities.

One recent human psychophysics study attempted to investigate implicit learning of timing information, including audition, touch, and vision (Kang et al., 2018). Random time patterns, a sensory feature that exists in all sensory modalities, were used as stimuli to minimise any modality-

 https://doi.org/10.1016/j.heares.2023.108857

driven effects on learning. While generating patterns with temporal irregularity could reduce the learning effect (Dauer et al., 2022), it reduces a chance for participants to rely on temporal predictability that could be used as another cue for learning. In a task relying on repetition detection within trials (Agus et al., 2010; Ringer et al., 2022; 2023), participants showed behavioural improvement only for a specific complex time pattern that also re-appeared across trials, even in an unsupervised learning setting. These improvements were observed for all three tested modalities, suggesting rapid and implicit learning of arbitrary temporal structures. Regardless of better overall task performance in audition than vision, the learning effect was qualitatively similar across modalities. Interestingly, participants also showed successful learning transfer from audition to other modalities, suggesting that the learning of random time patterns may be mediated by crosstalk between modalities. For sensory processing and implicit learning in audition, frontal and parietal cortices have been suggested as candidates where relevant neural signatures emerge (e.g., Hermann et al., 2022; Teki et al., 2016; Gutschalk et al., 2008). However, it is unclear which brain regions and neural activity patterns are involved in implicit learning transfer across modalities, and whether learning transfer can occur from other modalities into audition.

In an attempt to identify brain activity that may underlie these forms of transfer of implicit temporal pattern learning between sensory modalities, we measured brain responses of human volunteers using EEG, during a repetition detection task very similar to that used by Kang et al. (2018) in a study that demonstrated clear transfer learning. Participants were exposed to sequences of randomly timed pulses either in audition or vision, and the modality switched halfway through the test session. Some temporal patterns were repeated throughout the session, and we examined whether the EEG responses evoked by a given temporal stimulus patterns in one modality differed systematically depending on whether or not the same temporal pattern had already been

 https://doi.org/10.1016/j.heares.2023.108857

encountered previously in the other modality. To this end, we developed a novel method for analysing EEG-based learning curves in sensor and source space, in order to investigate which brain regions mediate learning transfer across modalities. In addition, we analysed EEG data in the time-frequency domain, given recent reports that neural sensitivity to repeated patterns is specifically subserved by beta-band activity (Kang et al., 2021).

Materials and methods

Participants

A total of 24 participants enrolled in the study (12 females, 12 males; mean age = 20.63 years, std = 3.24 years). The sample size of 24 participants was based on a power analysis (G*Power; Faul et al., 2007) for a two-way interaction in a repeated-measures ANOVA, under an assumed power of 0.8, taking an approximation of the effect sizes from the previous literature (Andrillon et al., 2015), in which a similar paradigm yielded moderate effect sizes (Cohen's d ≈ 0.5) for differences in ERP amplitudes between repeated and non-repeated stimuli. All participants self-reported as having normal hearing and normal or corrected-to-normal vision, with 3 self-reported as left-handed and the remaining as right-handed. Participants gave informed consent to taking part in the experiments and received cash reimbursement for their time after participating in the study. All study protocols were approved by the Human Subjects Ethics Sub-Committee of the City University of Hong Kong.

Experimental design and statistical analyses

Block design

Participants were randomly assigned to one of two groups, under the constraint of maintaining gender balance in each group. One group performed the task described below in

 https://doi.org/10.1016/j.heares.2023.108857

Audition first and then moved on to Vision (AV). The other group performed the tasks in the opposite order (VA). Each participant completed a total of 4 test blocks. Each block contained 160 trials - 80 trials for the first modality and 80 trials for the second modality. Halfway through each block, the subjects were advised of the switch in modality. In each trial, the subject was presented with a temporal pattern ("sequence") of simple clicks or flashes described in more detail below, and they had to decide whether a sequence contained a repeated temporal pattern or not.

Stimuli

Auditory or visual stimulus "sequences" (brief temporal patterns of clicks or visual flashes) were presented in training or testing blocks which are described in detail below. Each sequence consisted of a set of 0.2 ms rectangular stimulus pulses delivered at an average rate of 7 Hz. To generate temporal sequences following a Poisson distribution (with a 10 ms refractory period), inter-pulse time intervals were drawn from an exponential distribution, and intervals shorter than 10 ms were discarded. The onset of the temporal sequences (i.e., the first pulse) was always fixed. The length of each sequence was limited to 2.4 s. Sequences either had random time intervals for the full duration (random patterns, P), or a half-length (1.2 sec.) of random intervals seamlessly presented twice in a row (repeated patterns, RP). While both P and RP stimuli were generated afresh each time and occurred only once within the test block, one specific RP sequence (reference repeated patterns, RefRP) randomly re-occurred repeatedly within the test block. Each block of trials contained a randomized set of 40 P, 20 RP and 20 RefRP trials for each modality. Each block contained a newly generated RefRP. As in Agus et al. (2010), the rationale is that repeats in RefRP trials creates a familiarity which makes repeats in the sequence easier to detect, even though the

 https://doi.org/10.1016/j.heares.2023.108857

subjects are generally unaware that they had encountered the RefRP pattern before in the same block.

Test blocks were divided into two conditions: in the Transfer condition, RefRPs had an identical temporal structure for both modalities, while in the Control condition, the RefRP temporal structure differed between modalities (Fig. 1). Each participant performed two Transfer blocks and two Control blocks, in a randomised order.

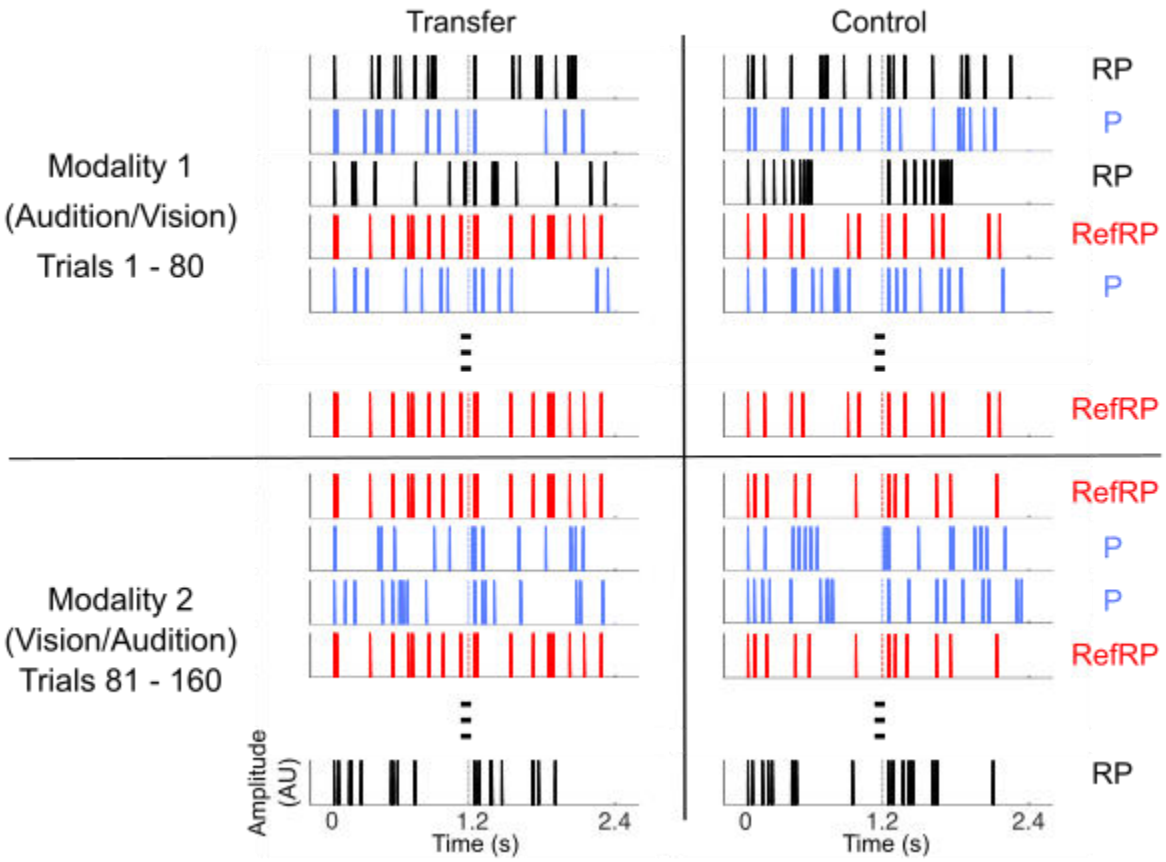

Figure 1. Example schematic of the stimulus patterns. Each stimulus example refers to one trial of either RP (black), P (blue), or RefRP (red). Different trial types were presented in a randomized order. Dotted lines within each stimulus indicate the midpoint of the pattern where the repetition begins for RP and RefRP trials. While stimuli are newly generated for RP and P, the RefRP is identical across trials in a block. The black solid horizontal line indicates the middle of the test block, at which point in time the stimuli are presented with a new modality. In the Transfer condition, the same RefRP is presented between two modalities, while in the Control condition, a different RefRP is presented after the modality switch.

In the auditory modality (Audition), the click trains were high-pass filtered at 2 kHz to avoid the possibility of spectral cues arising from filter ringing at a lower frequency range (Kang

https://doi.org/10.1016/j.heares.2023.108857

et al., 2017; Pollack, 1968). The low frequency range below 2 kHz was filled with band-passed pink noise between 50 Hz and 2 kHz at -2 dB SNR relative to sequences. The sound level for the auditory stimulus sequences was kept at 72 dB SPL, and they were presented at a 44.1 kHz sampling rate. In the visual modality (Vision) no filtering was applied to sequences, and identical filtered pink noise to the auditory stimulus was presented acoustically during the visual stimuli via a separate channel for consistency between modalities. All stimulus sequence waveforms were generated using custom Matlab scripts (R2017a, MathWorks Inc.), and played via MOTU UltraLite-mk3 Hybrid soundcard. Auditory stimuli were delivered to the participants through Brainwavz B100 earphones.

For visual stimuli, a single green LED light was housed in a custom-made 7x4x3 cm box covered with translucent plastic film to diffuse dimmed light. It was attached on top of the screen at the subjects' eye level. The LED lights were driven by a RM2 mobile processor (Tucker-Davis Technologies, TDT) running at a sample rate of 44.1 kHz, which was triggered by the MOTU soundcard.

Apparatus and Procedure

Participants were seated in a sound-attenuated and electrically shielded room in front of a computer screen. The screen was used to deliver instructions and display a white fixation cross on a grey background during the auditory task. During the visual task, the screen was black, and participants were instead asked to fixate on the LED delivering the visual sequences. After each trial (auditory or visual), participants were shown an instruction to indicate whether the stimulus of the trial contained a within-sequence repetition or not. Their responses were acquired by a computer keyboard button press (1 for a repetition; 2 for no repetition). Before performing the test

 https://doi.org/10.1016/j.heares.2023.108857

session, all subjects received a short training session to confirm that they correctly understood the task and performed above chance. There were three training blocks for each modality. The first training block for each modality had 10 trials of either RPs with 10 repetitions of a short 1.2 sec. segment of sequences, or the equivalent length of a random non-repeating pattern (P). Under these conditions, the repetition in RP trials was very salient so that participants could familiarise themselves with the task. The second and third training blocks consisted of 20 trials each, with the number of repetitions within each RP trial first reduced to 3, and then to 2. During the training blocks, participants received visual feedback after each trial on whether they made a correct response, only to ensure that participants understood the task. No feedback was given during the experimental test blocks. Since the training session was solely designed to ensure that participants understood the task requirements, no RefRP stimuli were included in the training. The training was done first on auditory, then on visual sequences. Note that the task performed after training is designed to only trigger implicit learning behaviour. To achieve this, the participants are simply asked to focus only on within-sequence repetition detection, and are not told that some patterns (the RefRP) will re-occur.

Prior to performing the test session, the participants were fitted with an EEG cap, and a conducting gel was applied on the scalp under each of the EEG electrode contacts. Participants were instructed to minimise blinking and body movement during stimulus presentation. EEG signals were recorded using ANT Neuro EEGo Sports 64 channel device at a sampling rate of 1000 Hz, grounded at the nasion and referenced to the Cpz electrode. No online filtering was used.

Two seconds after the onset of the stimulus sequence in each trial, a visual prompt emerged on the screen asking the subject to be ready to respond whether they had perceived a within-sequence repetition after the sequence ended.

 https://doi.org/10.1016/j.heares.2023.108857

Behavioural data analysis

First, based on signal detection theory, the behavioural *d'* sensitivity index was calculated per participant, separately for RefRP and RP stimuli following the equation:

$$d' = z(\text{Hit}) - z(\text{False Alarms}) \quad (1)$$

Correctly reported "Repeated" responses for either RefRP or RP stimuli were considered as hits and wrongly reported "repeated" responses for P stimuli were considered as false alarms. These values were computed per each block. To test whether participants were more accurate in their detection of repetition in RefRP rather than RP trials, and how this repetition detection might depend on condition or modality, a mixed-design analysis of variance (ANOVA) was performed on *d'* values, with stimulus type (RefRP or RP), condition (Transfer or Control), and modality (Audition or Vision) as within-subject factors, order (AV or VA) as a between-subject factor, and participant ID as a random factor. Post-hoc repeated-measures ANOVAs were conducted separately for each level of the modality and order factors. Performance increase in this study refers to the performance difference between RefRP and RP (RefRP - RP), which was expected to be observed in all cases regardless of condition and modality order. However, if there is any transfer effect, a significant interaction between stimulus type and condition only for the second modality should be observed.

Lastly, in an attempt to quantify the rate of performance improvements as subjects become familiar with RefRP stimuli, we subjected the dataset to the extra sum-of-squares F test (Agus et al., 2010; Motulsky and Christopoulos, 2004), which fits several nested models to the data and allows for an inference which parameters provide a significant contribution to the fit. To do so, trial-by-trial hit-rate performance of RefRP per each modality (A or V) and each condition

 https://doi.org/10.1016/j.heares.2023.108857

(Transfer or Control) respectively was fitted with either a constant line (one-parameter, intercept, equation 2) model or a three-parameter learning model (equation 3), used in a previous behavioural study in a similar paradigm (Agus et al., 2010):

$$y = mean(hitrate) \quad (2)$$

$$y = a + (i - a) * t^{\frac{(x-1)}{c}} \quad (3)$$

where x is trial index (1 – 20), i is initial score, a is asymptote, c is time constant, t is learning threshold, and y is the average performance at each trial x. An F-test was performed to investigate whether the three-parameter model fit the data significantly better than the one-parameter model. We computed the test statistic:

$$F = \frac{\left(\frac{ss_1 - ss_3}{df_1 - df_3}\right)}{\frac{ss_3}{df_3}} \quad (4)$$

where ss_1, ss_3 refer to sum of squares for the one and three parameter models respectively, and df_1, df_3 refer to the corresponding degrees of freedom. This test statistic was then compared against the critical values of the F distribution ($p < 0.05$) with 2 degrees of freedom.

Additionally, to compare the model used for fitting behavioural data (equations 2-3) and a logarithmic model selected for fitting EEG signals (equation 8 in the section "EEG analysis: learning curve estimation" below), we also applied the logarithmic model to behavioural data, which yielded essentially identical outcomes. For the sake of brevity we will omit reporting the logarithmic model fits to behavioural data and only report results based on fits to equations 2 and 3 above, which is an approach identical to that used in previous studies (e.g., Agus et al., 2010; Kang et al., 2018).

EEG analysis: preprocessing

 https://doi.org/10.1016/j.heares.2023.108857

The recorded EEG multichannel signals were preprocessed using the SPM12 Toolbox (Wellcome Trust Centre for Neuroimaging, University College London) for Matlab. The EEG data were high-pass filtered at a cut-off frequency of 0.01 Hz using a 5th order two-pass Butterworth filter to remove low frequency noise. A 5th order two-pass Butterworth notch filter (48-52 Hz) was applied to remove line artefacts. Then, a 5th order two-pass Butterworth low-pass filter at 90 Hz was applied to reduce any high-frequency environmental or physiological noise. Eyeblink artefacts were removed by subtracting the first two spatiotemporal components associated with each eyeblink (Ille et al., 2002), as implemented in SPM12. Specifically, the first two principal components were extracted from the time course and topography of the average eyeblink-evoked response, and removed from the raw EEG data at the time of each blink using spatial filtering.

After these preprocessing steps, the data were further denoised based on Dynamic Separation of Sources (de Cheveigné and Simon, 2008). This denoising procedure is commonly used to increase the signal-to-noise ratio of event-related potentials (ERPs) by maximizing the reproducibility of stimulus-evoked responses across trials. As part of the denoising procedure, data were re-referenced offline to the average of all channels. To calculate ERPs, continuous data were epoched between 200 ms before stimulus onset to 2600 ms after stimulus offset. Each epoch (segment) was baseline-corrected to the mean of the pre-stimulus period (i.e., from -200 ms to 0 ms relative to stimulus onset). In order to exclude epochs contaminated by transient artefacts, we removed epochs that had an average RMS (root mean square) amplitude exceeding the median by 2 standard deviations from further analysis. This procedure resulted in rejecting 5.71% (± 0.29%, SEM across participants) trials on average.

EEG analysis: learning curve estimation

 https://doi.org/10.1016/j.heares.2023.108857

The remaining epochs were subjected to single-trial learning-curve fitting. First, to determine the learning curve equation, we fitted four models to determine which best described any observed changes in RMS ERP amplitude as a function of trial. ERP RMS values (over the entire post-stimulus time window), rather than peak voltage amplitude values, were included in this analysis as a measure of "signal power" of the neural responses which is independent of signal polarity. To compute the ERP RMS values used to fit learning curves first - "relative RMS response amplitudes" were calculated for each of the 20 trials of each stimulus type as RMS(stimulus)-RMS(baseline). Then, these single-trial relative RMS values were averaged across participants, sessions, stimulus types, and channels, but not over trial number. The rationale for averaging these single-trial RMS values across stimulus types was to avoid biasing the subsequent analysis by any specific stimulus type (e.g., RefRP). The four models used to fit the grand-average data were based on previous studies in which learning curves were quantified in the context of statistical learning (Choi et al., 2020; Kepinska et al., 2017; Siegelman et al., 2018) and sensory adaptation (Ulanovsky et al., 2004). They included linear, quadratic, exponential, and logarithmic models, described by the following four equations respectively:

$$y = a\,xax + b \quad (5)$$
$$y = aexp(-bx) + c \quad (6)$$
$$y = ax^2 - bx + c \quad (7)$$
$$y = a(logx) + b \quad (8)$$

In equations (5)-(8), *x* is the trial index number (ranging from 1 to 20); *a, b* and *c* are model coefficients; log denotes the natural logarithm; and *y* is the grand-averaged RMS amplitude of all data at trial *x*. To select the most appropriate among these four candidate models, the models were compared using the adjusted R^2 metric, which quantifies the goodness of fit of each model, while penalising for the number of model coefficients (Miles, 2005). The adjusted R^2 values were: linear

 https://doi.org/10.1016/j.heares.2023.108857

(eqn 5): 0.309; exponential (eqn 6): 0.340; quadratic (eqn 7): 0.348; and logarithmic (eqn 8): 0.384. Since the logarithmic model yielded the highest adjusted R^2 for the averaged data, that model was selected for subsequent learning curve estimation, and this was performed separately for each individual participant and condition (Fig. 4A).

To estimate learning curves, epoched EEG data from each session were sorted according to the 6 trial types defined by modality (Audition; Vision) and stimulus type (P; RP; RefRP), and by condition (Transfer; Control). For each participant, channel, time point, and stimulus/trial type (design cell), single-trial EEG amplitude values corresponding to the first 20 trials of a particular stimulus/trial type were fitted with the logarithmic fit, yielding the logarithmic coefficient *a* (quantifying the "learning rate") and a constant *b* (quantifying the "average" ERP amplitude, after regressing out the learning curve). The logarithmic coefficient *a* was positive when EEG amplitudes increased over trials; negative when they decreased over trials; and equal to zero when there was no consistent change over trials. Finally, to allow a statistical examination of the learning curve parameters, these were grouped, across participants, into a between-subjects factor of modality order (AV; VA) to yield a 2×3×2×2 design with three within-subjects factors (modality, stimulus, condition) and one between-subjects factor (order). We performed further control analyses (see Supplementary Information) to rule out the possibility that learning rate parameters are biased by the fact that RefRP stimuli are repeated while RP stimuli are unique for each trial, as well as to quantify test-retest reliability of the coefficient estimates.

EEG analysis: channel-by-time statistical inference

The channel-by-time topography maps of ERPs (constant parameters *b*) and learning curves (logarithmic coefficients *a*) resulting from the curve fitting procedure described above were

 https://doi.org/10.1016/j.heares.2023.108857

analysed separately, in a series of repeated-measures ANOVAs. We aimed at identifying spatiotemporal (channel-by-time) clusters in which ERPs and/or learning curves were modulated by experimental factors. Specifically, we designed separate "flexible factorial designs" (repeated-measures ANOVAs) in SPM12 for the following factors: dependent variable (ERP amplitude or learning rate, separately for each channel and time point); order (AV; VA), and modality (Audition; Vision). Each ANOVA had the within-subjects factors "stimulus" (P; RP; RefRP) and "condition" (Transfer; Control), as well as a random "participant ID" factor. ERP and learning curve data were converted into 3D images (2D: scalp topography; 1D: time), and smoothed with an 8 mm × 8 mm × 50 ms Gaussian kernel to ensure that data conform to random field theory assumptions, used in SPM12 to correct for multiple comparisons across time points and channels. The smoothed images were then entered into the flexible factorial designs (GLMs equivalent to repeated-measures ANOVAs, a standard implementation of statistical inference in SPM12). Statistical parametric maps (SPMs) were thresholded at $p < 0.005$ (uncorrected) and significant effects were inferred at a cluster-level family-wise error (FWE) of $p < 0.05$, correcting for multiple comparisons across time points and channels under random field theory assumptions (Kilner et al., 2005).

For each ANOVA, the following contrasts were examined: First, we looked at the main effects of "stimulus" (RefRP vs. RP) to test for the effect of stimulus repetition across trials. Second, we examined the main effect of "condition" (Transfer vs. Control). This was a sanity check, as we hypothesised that this main effect should not be significant, given that the effect of transfer should be limited to RefRP stimuli. Finally, we tested the interaction effect of "stimulus" (RefRP vs. RP) and "condition" (Transfer vs. Control), as evidence for cross-modal learning transfer effects. We reasoned that, if there is indeed a transfer of pattern memory across modalities, then in the second modality (but not in the first), the neural processing of RefRP stimuli in the

 https://doi.org/10.1016/j.heares.2023.108857

Transfer condition would be different from that in the Control condition, given that in the Transfer condition the subjects had just experienced the same temporal pattern in the first modality, while in the Control condition they had not. Meanwhile, simple RP stimuli, which were always generated afresh and thus allowed no possibility of a memory transfer across modality, should be processed identically in the Transfer and Control conditions.

Any significant interaction effects observed between the factors of "stimulus" (RefRP vs. RP) and "condition" (Transfer vs. Control) were subjected to an additional post-hoc analysis, verifying whether the effects of "learning" a RefRP pattern in the previous modality significantly affected the initial responses to RefRP stimuli in the Transfer condition, in the second modality. This would indicate that prior experience of the temporal pattern of a given RefRP that had been learned in one modality could alter the responses to the same pattern in another modality. To set up the post-hoc analysis, we investigated the differences in trial-by-trial changes in EEG signals between four trial types (stimulus: RefRP vs. RP; condition: Transfer vs. Control). Per participant and trial, we extracted data from EEG channels corresponding to the significant SPM cluster (using the 2D coordinates of each EEG channel mapped onto the 2D cluster coordinates), calculated the RMS of the significant time window (averaged across channels), and subtracted the RMS of the pre-stimulus baseline. Based on previous studies (Kang et al., 2017; Kang et al., 2018), we expected rapid learning to occur, requiring perhaps no more than the first five RefRP trials. Any true transfer effect ought to manifest early during this rapid learning phase. The RMS data for the first three trials of each type were therefore averaged, normalised per participant by z-scoring each data point (pooling across trial types), and subjected to pairwise comparisons (paired *t*-tests) to confirm whether RefRP Transfer condition did indeed differ from the others, as hypothesized.

 https://doi.org/10.1016/j.heares.2023.108857

EEG analysis: source reconstruction

Having identified significant channel-by-time clusters in which ERPs and/or learning curves differed between conditions, we sought to reconstruct the most likely cortical sources of the observed sensor-level effects. To this end, the spatiotemporal clusters showing significant effects were source-localised using the multiple-sparse-priors approach under group constraints (Litvak and Friston, 2008). For each significant main effect, sensor-level data were subject to source localisation for a time window in which the significant effect was observed, plus an additional 100 ms of signal on either side, given that, in multiple-sparse-priors source reconstruction, rise and fall times of neural signals should be included, as source activity estimates are based on signal variance across time (López et al., 2014). For each participant, the source estimates were converted into 3D images (in MNI space), smoothed with a 5×5×5 mm Gaussian kernel, and entered into flexible factorial designs with one within-subjects "stimulus" factor and one between-subjects "participant" factor. The resulting statistical parametric maps were thresholded at $p < 0.05$ (two-tailed, uncorrected) and significant effects were inferred at a cluster-level $p < 0.05$ family-wise error (FWE, small-volume corrected), correcting for multiple comparisons across voxels under random field theory assumptions (Kilner et al., 2005). Sources were labelled using the Neuromorphometrics atlas, as implemented in SPM12. As a sanity check, we also source-localised early evoked responses (0-100 ms relative to stimulus onset), confirming auditory and visual cortical sources of auditory and visual sequence onset responses respectively, in addition to a less prominent right prefrontal source found for both modalities (Fig. S3, Table S1).

In case of significant interaction effects, the source reconstruction procedure was identical as for the main effects, except that sensor-level data were used to calculate contrast time-series

 https://doi.org/10.1016/j.heares.2023.108857

(RefRP-RP). Flexible factorial designs had one within-subjects factor ("condition": Transfer vs. Control) and two between-subjects factors ("participant", and "order": AV vs. VA). Statistical thresholds were set at $p < 0.05$ (uncorrected), correcting for multiple comparisons across time points at a cluster-level $p < 0.05$ family-wise error (FWE).

EEG analysis: time-frequency statistical inference

We also investigated whether pattern learning transfer was reflected in changes in activity in specific EEG frequency bands. Since we observed significant interaction effects between stimulus and condition for both AV and VA participant groups, which in both cases source-localised to the right inferior frontal gyrus (rIFG; see Results), in this post-hoc analysis we focused on the time-frequency patterns of activity localized to the rIFG. To this end, continuous EEG data were re-epoched from 800 ms before stimulus onset to 800 ms after stimulus offset. Dipole waveforms were estimated for rIFG coordinates (dipole location: peak MNI coordinates of the source reconstruction in each AV and VA group; dipole orientations: three orthogonal dipole moments), resulting in three dipole time-series (corresponding to three orthogonal dipole orientations) per trial. The dipole time-series were subject to a time-frequency decomposition using multi-taper convolution with a frequency range of 8-48 Hz in 1 Hz steps, and a time window of 400 ms (time step: 50 ms). The fixed time window length was chosen to yield a relatively high temporal resolution and a sufficient number of cycles per frequency, while at the same time avoiding the smearing of evoked response from the (speeded) button presses. Thus, frequencies below 8 Hz were not analysed a priori. This yielded three time-frequency maps per trial (one per orthogonal dipole moment). For each trial and time-frequency point, maximum power was selected across the three orthogonal dipole orientations, and log-rescaled to the pre-stimulus baseline (from

 https://doi.org/10.1016/j.heares.2023.108857

-600 to -200 ms relative to stimulus onset, to avoid baseline contamination by the time window used for time-frequency estimation). These single-trial rescaled time-frequency estimates were subject to logarithmic fitting, as described above.

The resulting time-frequency maps of model coefficients (learning curves) were converted into 2D (time × frequency) images for each trial type, smoothed with a 5 Hz × 50 ms Gaussian kernel (to ensure that data conform to random field theory assumptions), and entered into a factorial design with two within-subjects factors ("stimulus": RefRP vs. RP; "condition": Transfer vs. Control) and one between-subjects factor ("participant"). Statistical parametric maps were thresholded at $p < 0.05$ (two-tailed, uncorrected) and significant effects were inferred at $p < 0.05$ family-wise error (FWE), correcting for multiple comparisons across time and frequency under random field theory assumptions (Kilner et al., 2005). As in the case of EEG amplitudes, the significant interaction effects were subject to an additional post-hoc analysis to test if the interaction between "stimulus" and "condition" modulates time-frequency responses for the initial trials. Per participant and trial, we pooled power estimates from the significant cluster, averaged them across the first three trials of each type, and normalised the resulting estimates of initial power per participant by dividing each data point by the standard deviation (pooling across trial types). These data were subject to pairwise comparisons (paired *t*-tests).

Correlation between EEG and behaviour

Finally, we tested whether behavioural sensitivity correlates with neural signatures of learning transfer (i.e., the effects of learning transfer on EEG amplitude and time-frequency activity) across participants. As the behavioural measure, per participant and condition (Transfer vs. Control) we calculated the difference in *d'* between RefRP and RP stimuli presented in the

 https://doi.org/10.1016/j.heares.2023.108857

second modality. As the neural signature of learning transfer on EEG amplitude, we calculated the differences in EEG-based logarithmic fits (pooling from significant clusters shown in Fig. 4BC and 3EF; see Results) between RefRP and RP stimuli, separately for each participant and condition (Transfer vs. Control). Similarly, as the neural signature of learning transfer on time-frequency activity, we calculated the corresponding differences in beta-band logarithmic fits (pooling from significant clusters shown in Fig. 5B). The individual estimates of differences in behavioural sensitivity to RefRP and RP stimuli were entered as a dependent variable into an analysis of covariance (ANCOVA) with two continuous covariates (EEG signature of learning transfer (as a continuous regressor; time-frequency signature of learning transfer) and two categorical independent variables (condition: Transfer vs. Control); order: AV vs. VA), as well as their interactions. Significant interactions were inspected using post-hoc ANCOVAs.

Results

Behavioural effects

Figure 2A shows behavioural performances for detection of repetition in RefRP and RP stimuli respectively, for each modality and each condition. A mixed-design ANOVA on the *d'* performance scores combining both groups of different modality orders confirmed significant effects of stimulus type ($F_{1,22} = 19.33$, $p < 0.001$) and modality ($F_{1,22} = 26.39$, $p < 0.001$). If the effect of stimulus type and condition depended on modality, a significant interaction for the modality order × modality × stimulus type × condition should be observed, which was the case in our data ($F_{1,22} = 6.88$, $p = 0.016$). Thus, we ran separate post-hoc repeated-measures ANOVAs on each modality for the AV and the VA participant groups respectively to study whether effects of stimulus type, modality, and condition were observed in both groups. Based on our hypotheses,

 https://doi.org/10.1016/j.heares.2023.108857

we expected 1) an easier detection of within-sequence repetition for RefRP compared to RP, which would manifest as a significant effect of stimulus type on performance; 2) no qualitative performance difference between the two conditions (Transfer vs. Control), and therefore a significant effect of stimulus type on performance for all cases; 3) a significant interaction between stimulus type and condition only for the second modality (V in the AV group, A in the VA group) as an indication of a transfer effect resulting from the same RefRP temporal pattern being presented across the two modalities only for Transfer condition. Note that no significant interaction would be expected in the first modality, given that the Transfer and Control conditions only begin to differ once RefRP stimuli are presented in the second modality.

In the AV group, ANOVA on the first modality (Audition) confirmed that there was a significant performance difference between RP and RefRP ($F_{1,11} = 6.53$, $p = 0.027$) and no overall performance difference between Transfer and Control conditions ($F_{1,11} = 0.003$, $p = 0.96$) nor an interaction between the stimulus type and condition ($F_{1,11} = 0.07$, $p = 0.79$). In contrast, an ANOVA on the second modality (Vision) revealed no significant performance difference between RP and RefRP ($F_{1,11} = 0.20$, $p = 0.65$) or between Transfer and Control ($F_{1,11} = 0.04$, $p = 0.85$). The interaction between the stimulus type and condition approached significance ($F_{1,11} = 4.33$, $p = 0.06$). For the VA group, a post-hoc ANOVA on each modality confirmed that both modalities showed an effect of stimulus type, reflecting significantly greater performance for RefRP than RP (V: $F_{1,11} = 14.05$, $p = 0.004$; A: $F_{1,11} = 6.38$, $p = 0.028$). No effect of condition (V: $F_{1,11} = 1.62$, $p = 0.23$; A: $F_{1,11} = 1.08$, $p = 0.32$) or interaction between condition and stimulus type were observed (V: $F_{1,11} = 3.43$, $p = 0.09$; A: $F_{1,11} = 0.02$, $p = 0.89$; Fig. 1A), which may be due to greater general task performance in the second (A) modality with higher temporal sensitivity (Gault & Goodfellow, 1938; Goodfellow, 1934; Lechelt, 1975). In summary, significant performance differences

 https://doi.org/10.1016/j.heares.2023.108857

between RP and RefRP were observed in all cases except for the second modality of the AV group, and, as expected, no significant main effects of condition were observed in all cases. However, contrary to our hypotheses, interactions between stimulus type and condition in the second modality failed to reach our significance level (p = 0.05) in both AV and VA groups, and came close to the significance criterion only in the AV group (p = 0.06).

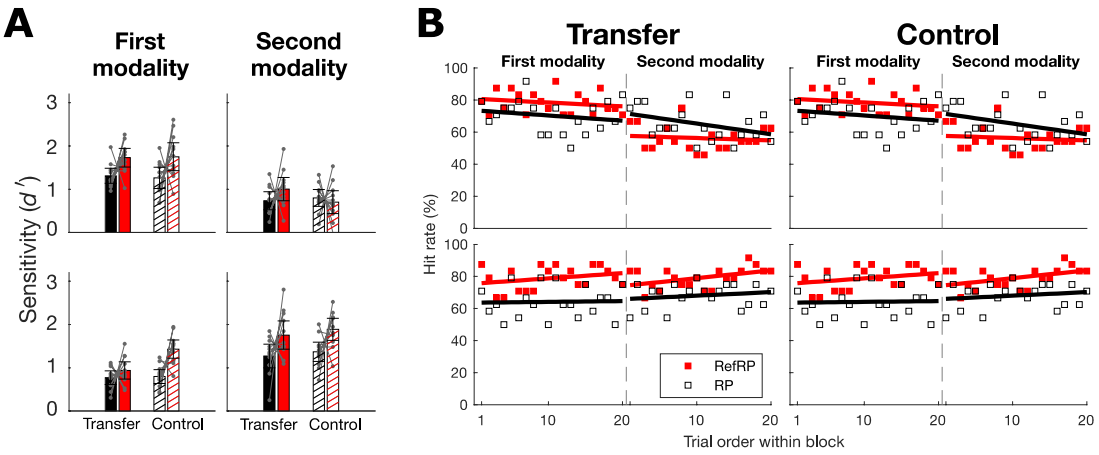

Figure 2. Behavioural results. **(A)** Sensitivity index *d'* values averaged across subjects for RP (black) and RefRP (red) for each modality (top: AV group, bottom: VA group) and each condition (filled: Transfer, hatched: Control). *d'* values for individual participants were rescaled relative to the average performance across participants only for the visualization. Gray lines on the bar graphs indicate individual subject's *d'* values per stimulus type and condition. Error bars represent 95% confidence intervals. **(B)** The time course of RefRP (red) and RP (black) performance for Transfer and Control conditions (top: AV group, bottom: VA group). Dashed lines indicate the midpoint of the test block where the modality switch occurred.

To investigate whether observed performance differences between stimulus type and modality type were due to a performance improvement for RefRP, and whether there is a performance difference between the Transfer and Control conditions, we computed a hit rate between RefRP and RP for each trial, for each modality order group. Unlike previous studies (e.g., Agus et al., 2010; Kang et al., 2018), we did not observe a clear, rapid performance improvement, nor a difference between conditions (F-statistic of three-parameter fitted model compared to a constant-line model, all p > 0.05), which may be due to a limited number of test blocks and the difficulty of the task.

https://doi.org/10.1016/j.heares.2023.108857

Stimulus repetition modulates ERP amplitude

In the analysis of the effects of stimulus repetition across trials (RefRP vs. RP) on ERP amplitude (i.e., the constant coefficients of the single-trial model; see Methods), two significant spatiotemporal clusters were identified (Fig. 3AD). In the first modality of the AV blocks (i.e., for Auditory stimuli), RefRP stimuli differed in amplitude from RP stimuli over posterior channels, during the second, repeated stimulus segment (Table 1), while for the first modality of the VA blocks (i.e., for Visual stimuli), responses to RefRP stimuli differed in amplitude from those to RP stimuli over central channels during both stimulus segments. In the source reconstruction of the two significant clusters (Fig. 3CF), the difference in auditory RefRP and RP stimuli was attributed to the right middle/inferior temporal gyrus (MNI peak: [54 -24 -22], cluster-level $p_{FWE} = 0.047$, small-volume corrected, $T_{max} = 2.42$, $Z_{max} = 2.37$), in the immediate vicinity of secondary auditory regions, while the difference in visual RefRP and RP stimuli was attributed to the left inferior occipital gyrus (MNI peak: [-22 -98 0], cluster-level $p_{FWE} = 0.047$, small-volume corrected, $T_{max} = 2.30$, $Z_{max} = 2.26$), in the early visual cortex.

Table 1. Statistical results of the ERP analysis (constant and logarithmic coefficients).

Dependent variable	Effect	Group	Block	Latency range (ms)	Cluster-level pFWE	Fmax	Zmax
ERP amplitude (constant coefficient)	RefRP vs. RP	AV	Audition	1110-1900	< 0.001	34.27	5.37
		VA	Vision	372-1442	< 0.001	29.21	4.99
ERP	Interactio	AV	Vision	704-942	0.042	15.41	3.63

 https://doi.org/10.1016/j.heares.2023.108857

learning curve (logarithmic coefficient)	n: (RefRP vs. RP) x (Transfer vs. Control)	VA	Audition	476-700	0.003	23.30	4.47

Prior to identifying these two significant clusters (corresponding to post-hoc comparisons), we tested for the corresponding interaction effects. Given our 2 × 2 × 2 × 2 design (order: AV vs. VA; modality: Auditory vs. Visual; stimulus type: RefRP vs. RP; condition: Transfer vs. Control), but also our priori interest in contextual effects on ERPs within each modality (rather than in trivial amplitude differences between modalities), we first tested for two 3-way interactions (order, condition, stimulus type) for the two modalities separately. In both modalities separately, the interaction tests yielded significant clusters while correcting for multiple comparisons across channels and time points ($pFWE < 0.05$). Based on these 3-way interactions, we then tested for main effects, as well as 2-way interactions (condition, stimulus type) for the two blocks separately (first vs. second modality block). These tests showed that the only significant effects were found in the first modality, and corresponded to the main effect of stimulus type, as described above. The remaining effects (main effects of condition and/or interaction in the first modality; main effects and/or interactions in the second modality) were not significant ($pFWE > 0.05$).

 https://doi.org/10.1016/j.heares.2023.108857

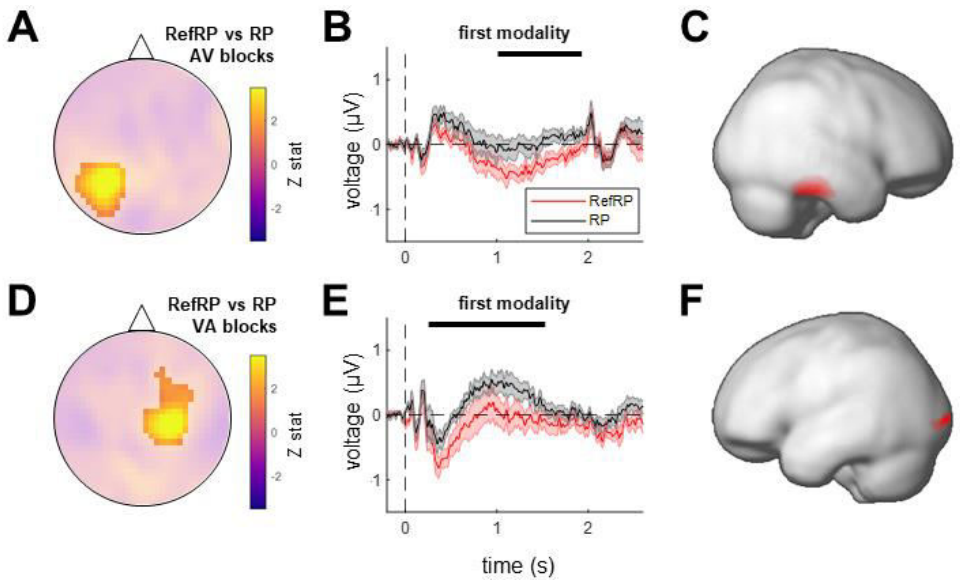

Figure 3. ERP effects of stimulus repetition. **(A)** Scalp topographies of repetition effects across trials (RefRP vs. RP stimuli), in the first modality of AV blocks. Highlighted significant clusters ($p_{FWE} < 0.05$). **(B)** Time courses of the effects of repetition across trials, in the first modality of AV blocks. Red: RefRP stimuli; black: RP stimuli. Shaded areas denote SEM across participants. Black bars denote the significant effect time windows. A peak at 2 s is where the visual prompt appeared to guide subjects to make responses. **(C)** Source estimates of the significant clusters in the AV blocks: right middle/inferior temporal gyrus (peak MNI [54 -24 -22]). **(D)** Scalp topographies of repetition effects across trials (RefRP vs. RP stimuli), in the first modality of VA blocks. Highlights as in (A). **(E)** Time courses of the effects of repetition across trials, in the first modality of VA blocks. Legend as in (B). **(F)** Source estimates of the significant clusters in the VA blocks: left inferior occipital gyrus (peak MNI [-22 -98 0]).

Cross-modal transfer modulates EEG learning curves

In the analysis of the effects of cross-modal transfer (interaction: [RefRP vs. RP] × [Transfer vs. Control]) on EEG learning curves (i.e., the logarithmic coefficients of the single-trial model; see Methods), two significant clusters were identified. In the second modality of the AV blocks (i.e., for visual stimuli), there was a significant interaction between stimulus (RefRP vs. RP) and condition (Transfer vs. Control) over right-lateralized central electrodes during the first stimulus segment (Table 1; Fig. 4BC). Similarly, for the second modality of the VA blocks (i.e., for auditory stimuli), the same interaction also yielded a significant cluster over right-lateralized frontal electrodes, also during the first stimulus segment (Table 1; Fig. 4EF). The source reconstruction of these clusters, attributed both significant interaction effects to source activity in

 https://doi.org/10.1016/j.heares.2023.108857

the right inferior frontal gyrus (rIFG; AV blocks, MNI peak: [46 40 -4], cluster-level $p_{FWE} = 0.049$, small-volume corrected across voxels, $T_{max} = 3.17$, $Z_{max} = 3.05$; Fig. 4D; VA blocks, MNI peak: [52 14 20], cluster-level $p_{FWE} = 0.049$, small-volume corrected across voxels, $T_{max} = 3.67$, $Z_{max} = 3.50$; Fig. 4G).

As in the analysis of ERP amplitudes, prior to identifying these two significant clusters, we tested for two 3-way interactions (order, condition, stimulus type) for the two modalities separately, which showed significant clusters in both modalities (pFWE < 0.05). Then, we tested for the main effects and 2-way interactions (condition, stimulus type), separately for the first vs. second modality block. These tests showed that the only significant effects were found in the second modality, and corresponded to the interaction effect of stimulus type and condition. The remaining effects (main effects in the second modality; main effects and/or interactions in the first modality) were not significant (pFWE > 0.05).

A visual inspection of the EEG learning curve interaction between stimulus type (RefRP, RP) and Condition (Transfer, Control; Fig. 4CF) revealed that, in both AV and VA blocks, the learning curves were positive for the RefRP stimuli presented in the Control condition, reflecting a gradual increase of RMS values over trials, while they were closer to zero for the RefRP stimuli presented in the Transfer condition, reflecting a relatively flat trajectory of RMS values over trials. For the RP stimuli, the opposite pattern of results was found. However, a further formal inspection of the interaction effect (Fig. 4HI) revealed that, for the initial three trials of the second modality, the interaction was driven only by the significant difference in ERP amplitude between RefRP stimuli presented in Transfer vs. Control conditions ($t_{23} = 2.38$, $p = 0.026$), with no significant differences between the remaining stimulus/condition pairs (all other $p > 0.2$), including between RP in Transfer vs. Control conditions. Furthermore, the difference between RefRP and RP was

 https://doi.org/10.1016/j.heares.2023.108857

greater in the Transfer than in the Control condition ($t_{23} = 2.15$, $p = 0.043$). Thus, RefRP stimuli were characterised by significantly higher ERP amplitudes from the onset of the Transfer blocks, while in the Control blocks RefRP as well as RP stimuli were associated with initially relatively low ERP amplitudes that gradually increased over trials, likely reflecting generic improvements in stimulus processing.

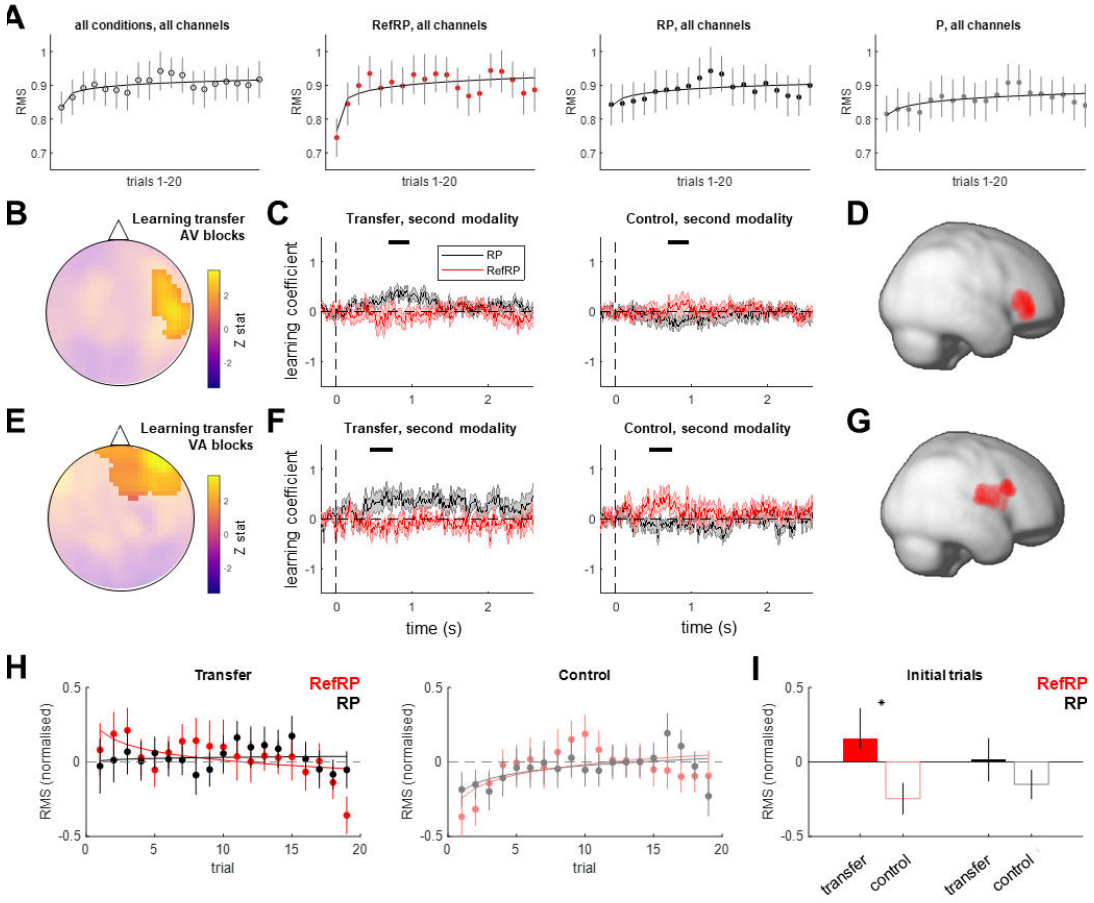

Figure 4. EEG learning curves reflect cross-modal transfer. **(A)** Learning curve estimation by logarithmic fitting. Plots show RMS of ERP responses (relative to baseline), averaged across channels and participants for all conditions (left), and for RefRP, RP, and P stimuli separately (middle/right). For display purposes, data were smoothed by a 3 point moving average. Error bars denote SEM across participants. **(B)** Scalp topography of the cross-modal transfer effect (interaction: [RefRP vs. RP] × [Transfer vs. Control]) on EEG learning curves, in the second modality of AV blocks. Highlighted significant cluster ($p_{FWE} < 0.05$). **(C)** Time course of AV cross-modal transfer effect. Left: transfer condition, right: control condition. Red: RefRP stimuli; black: RP stimuli. Shaded areas denote SEM across participants. Black bars denote the significant interaction effect time windows. **(D)** Source estimates of the significant cluster: right inferior frontal gyrus (peak MNI [46 40 -4]). **(E)** Scalp topography of the cross-modal transfer effect in

https://doi.org/10.1016/j.heares.2023.108857

the second modality of VA blocks. Highlighted significant cluster ($p_{FWE} < 0.05$). **(F)** Time course of VA cross-modal transfer effect. Legend as in (B). **(G)** Source estimates of the significant cluster: right inferior frontal gyrus (peak MNI [52 14 20]). **(H)** Single-trial EEG amplitude (RMS). Left: transfer condition, right: control condition. Red: RefRP stimuli; black: RP stimuli. Solid curves denote logarithmic fits to the group average. Error bars denote SEM across participants. **(I)** Post-hoc tests of the EEG amplitude (RMS) averaged across the first three trials. Only the difference between RefRP EEG amplitudes (RMS) in transfer vs. control conditions was significant ($p < 0.05$).

Cross-modal transfer modulates frontal beta power

For the post-hoc analysis of time-frequency activity, we focused on source activity localized to the rIFG region, and tested whether any time-frequency clusters were sensitive to cross-modal transfer (interaction effect) as identified above for ERPs. This analysis revealed that, across AV and VA blocks, beta-band power was modulated by cross-modal transfer during the first stimulus segment (latency range 500-800 ms, frequency range 15-20 Hz, cluster-level p_{FWE} = 0.016, T_{max} = 3.25, Z_{max} = 3.19; Fig. 5AB). The main effect of block order (AV vs. VA) was not significant. A closer inspection of this interaction effect for the first three trials (Fig. 5CD) revealed that, as in the case of the ERPs, the effect was driven by the significant difference beta power between RefRP stimuli presented in Transfer vs. Control conditions (t_{23} = -2.59, p = 0.016), with no significant differences between the remaining stimulus/condition pairs (all other $p > 0.1$). In particular, the visually robust difference between RefRP and RP initial responses in the control condition was not statistically significant. Taken together, the beta-band effects revealed a strong initial beta desynchronisation for RefRP stimuli presented in the Transfer condition, which then gradually returned towards baseline with consecutive trials. Conversely, RefRP stimuli presented in the Control condition showed a relative decrease in beta power with consecutive trials. On the other hand, RP stimuli presented in both conditions showed a relatively flat beta power trajectory across trials.

 https://doi.org/10.1016/j.heares.2023.108857

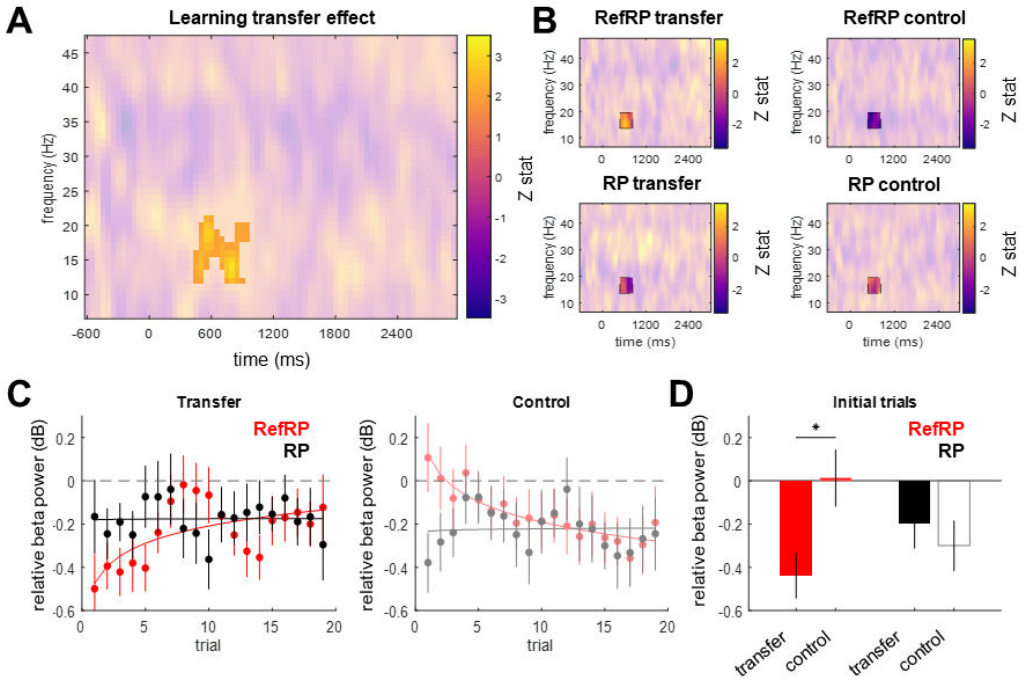

Figure 5. Frontal beta power reflects cross-modal transfer. **(A)** Time-frequency map of the cross-modal transfer effect (interaction: [RefRP vs. RP] × [Transfer vs. Control]) on single-trial changes (logarithmic coefficients) in rIFG activity, in the second modality of both AV and VA blocks. The significant interaction cluster ($p_{FWE} < 0.05$) is highlighted with saturated colours. **(B)** Time-frequency maps for each condition. The highlighted cluster denotes the time and frequency window (corresponding to the maximum range of the interaction cluster depicted in A) used to obtain single-trial estimates in (C,D). **(C)** Single-trial frontal beta power relative to the baseline. Left: transfer condition, right: control condition. Red: RefRP stimuli; black: RP stimuli. Solid curves denote logarithmic fits to the group average. Error bars denote SEM across participants. **(D)** Post-hoc tests of the frontal beta power, averaged across the first three trials. Only the difference between RefRP frontal beta power in transfer vs. control conditions was significant ($p < 0.05$).

Neural signatures of learning transfer covary with behavioural advantages of RefRP over RP processing

While we did not observe robust behavioural effects of learning transfer at the group level (see above), we also investigated the intra-individual variability of behavioural benefits to test whether individual participants' behaviour covaries with their neural signatures of learning transfer. Across participants, in an ANCOVA with *d'* benefit between RefRP and RP as the dependent variable, we found these behavioural advantages to be linked to a significant three-way interaction between the continuous covariate representing the EEG signatures of learning transfer, and the

 https://doi.org/10.1016/j.heares.2023.108857

categorical factors of "condition" and "order" ($F_{1,31}$ = 4.36; p = 0.045). Further, we also found a significant four-way interaction between the continuous covariates representing the EEG and beta-band activity signatures of learning transfer, and the categorical factors of "condition" and "order" ($F_{1,31}$ = 5.02, p = 0.032). The remaining main and interaction effects were not significant (all p > 0.1). A further inspection of the significant interaction effects revealed that, in the Transfer condition of the AV group, behavioural sensitivity benefits were subject to a significant main effect of EEG signatures of learning transfer ($F_{1,8}$ = 5.52, p = 0.047, β = 0.425, reflecting a positive correlation), and a significant interaction effect between EEG and beta-band signatures of learning transfer ($F_{1,8}$ = 7.67, p = 0.024, β = -0.262, reflecting a negative correlation with the interaction term). These effects were not significant in the Control condition or in the VA group (all p > 0.05). In other words, at an individual participant level, behavioural advantages of RefRP vs. RP processing covaried with neural signatures of learning transfer only for the AV group and in the Transfer condition.

Discussion

Learning of temporal patterns within modalities

In this study we investigated characteristics of neural changes for implicit learning of temporal patterns in audition and vision, as well as transfer across modalities of the learnt pattern. First, we observed a significant ERP difference between RefRP and RP in both audition and vision, complementing previous studies using similar paradigms which reported a modulation of neural responses to stimulus repetition in audition (Andrillon et al., 2015; Luo et al., 2013; Barascud et al., 2016; Teki et al., 2016; Herrmann et al., 2021). Based on previous studies using similar

 https://doi.org/10.1016/j.heares.2023.108857

paradigms (Andrillon et al., 2015; Luo et al., 2013; Herrmann et al., 2021), this study focused a priori on comparing responses to RefRP and RP stimuli.

A source localisation of the ERP differences between RefRP and RP stimuli showed that, in audition, these differences were primarily due to activity in or near the auditory cortex, while in vision, these differences originated from the visual cortex. This is consistent with observations made by other authors who have studied the encoding of temporal information (Buonomano and Maass, 2009; Ivry et al., 1988; Ivry and Schlerf, 2008), that parts of modality specific sensory cortical regions may play a key role in learning temporal patterns presented in a their preferred modality but our finding also contrasts with previous studies that have suggested an involvement of auditory regions in temporal processing of both auditory and visual stimuli (Guttman et al., 2005; Kanai et al., 2011), or an involvement of visual regions in mnemonic processing of both auditory and visual stimuli (Wolff et al., 2020).

Additionally, we replicated that the temporal patterns of occasionally repeated stimuli (RefRP) are learned and exploited when these stimuli are presented within a particular modality (in both audition and vision). During the learning of temporal patterns in the first modality of presentation (either audition or vision), participants had better behavioural performance (within-trial repetition detection) for RefRP stimuli, presented multiple times during the experiment, than for RP stimuli which were only presented once. These performance benefits were observed in both audition and vision as the first modality, consistent with previous studies using the same paradigm (Bale et al., 2017; Gold et al., 2014; Kang et al., 2017; Kang et al., 2018). These improvements can serve as an index of learning for RefRP.

Learning transfer between modalities

 https://doi.org/10.1016/j.heares.2023.108857

Previous studies of cross modal transfer of temporal pattern learning have reported somewhat inconsistent findings. For example, while some studies reported that training in audition improves visual temporal processing but not vice versa (Barakat et al., 2015; Bratzke et al., 2012), others did report a generalization of temporal perceptual learning from vision to audition (Bueti and Buonomano, 2014). Yet other studies reported no transfer effect at all (Ball et al., 2018b; Lapid et al., 2008).

The analysis of EEG responses showed robust neural signatures of potential learning transfer across modalities, in both directions (from audition to vision and from vision to audition). Thus, even if the behavioural evidence for transfer was weak in the present study, we did observe robust physiological effects which were also fully compatible with previously reported, significant transfer effects across modality using a very similar paradigm (Kang et al., 2018). Similarly, neural signatures for pattern learning have been observed even in a passive manner (Andrillon et al., 2015). Previous auditory imaging studies in humans under the same paradigm focused on the inter-trial coherence as an index of neural activity modulated by the re-occurrence of reference sequences (Andrillon et al., 2015; Luo et al., 2013). However, the inter-trial coherence by definition averages information across trials. In the present study, since we were primarily interested in learning transfer rather than merely pattern learning, we estimated a neural learning curve to repeated presentations of RefRP stimuli across trials, based on a logarithmic fit of trial-by-trial responses. Our model-based choice of logarithmic fitting, as opposed to other (e.g., linear and exponential) fits analysed here, was consistent with the results of previous work on statistical learning in adults (Siegelman et al., 2018) and infants (Choi et al., 2020) which has shown that logarithmic fits efficiently describe learning curves. This allowed us to quantify the gradual changes in neural responses, tapping more directly into the learning mechanisms. In addition to

 https://doi.org/10.1016/j.heares.2023.108857

analysing the logarithmic fits (learning curves), we also analysed responses in the initial trials after the modality switch, based on the assumption that they provide complementary information. Logarithmic fits quantify changes in response amplitude across trials. In contrast, analysis of initial trials can characterise the time scale of these changes. Combining these two analyses can discriminate between alternative hypotheses regarding transfer learning: (1) cross-modal pattern transfer may result in faster learning for previously experienced stimuli, which would be reflected in steep learning curves for RefRP stimuli in the transfer condition, without necessarily affecting the response amplitude to the initial trials; (2) cross-modal transfer modulates the processing of the initial stimuli, without necessarily affecting the response amplitudes to the later trials (which may be e.g. masked by ceiling effects). Using these methods, we found that when RefRP temporal patterns presented in one modality later reoccur in another modality, the ERP learning curves estimated for RefRP stimuli in the second modality (for both AV and VA groups) are significantly different in the transfer vs. control conditions. Importantly, these effects were driven by the initial few trials, corresponding to a negative learning coefficient for RefRP in the transfer condition and a positive learning coefficient for RefRP in the control condition. Unlike the first modality, in which we observed overall differences between ERP amplitudes evoked by RefRP and RP stimuli, here the average differences in ERP amplitudes were not significant. Furthermore, in the AV group, the intra-individual differences in measured behavioural transfer benefits (RefRP vs. RP) positively covaried with the ERP signatures of learning transfer (i.e., EEG-based learning curves predicted the size of the behavioural benefit in the Transfer condition, but not in the Control condition). We speculate that the greater temporal sensitivity in audition compared to vision may explain why no such effect was observed in the VA group. In other words, sensitivity to temporal patterns in audition may be so high, and learning so fast and robust (Agus, 2010), that it

 https://doi.org/10.1016/j.heares.2023.108857

may be very difficult to further significantly improve on it through cross-modal transfer. Taken together, these results indicate that transfer of learned temporal patterns across modalities manifests in distinct difference in neural responses, and this is largely limited to the initial stimulus presentations of the learned pattern, as would be expected given that repeated presentations will facilitate intra-modality learning which in turn will erase any modality transfer advantage.

Interestingly, for both AV and VA groups, source localization revealed that the learning transfer effect was associated with activity modulations in the right IFG, an area typically associated with working memory, attention, and detection of relevant targets, regardless of modality (Corbetta and Shulman, 2002; Hampshire et al., 2010; Linden et al., 1999). However, the IFG is also considered a multimodal hub region in the context of sequence learning, with previous studies consistently identifying the IFG as mediating sequence learning in both auditory and visual modalities (Milne et al., 2018). (In addition to IFG, subcortical regions such as the hippocampus and thalamus are also thought to be involved, but these are less accessible with EEG recordings.) Our study thus suggests that the IFG may also mediate cross-modal transfer, in addition to modality-general learning effects. Furthermore, in a post-hoc analysis of oscillatory activity within the rIFG cluster, significant signal modulation due to the learning transfer effect was observed specifically in the beta-band oscillations. While our previous study in rodents identified beta-band correlates of auditory pattern learning (Kang et al., 2021), those results were based on direct recordings limited to the auditory cortex, which in humans constitutes a relatively deep cortical source. However, prefrontal beta-band activity is routinely observed in human EEG studies and linked to a range of phenomena linked to learning and memory, including working memory (Gelastopoulos et al., 2019; Spitzer and Blankenburg, 2012; von Lautz et al., 2017), memory formation (Hanslmayr et al., 2014), implicit learning (Pahi et al., 2020), and stimulus familiarity

 https://doi.org/10.1016/j.heares.2023.108857

(Ketz et al., 2015). Previous work has identified right prefrontal beta-band activity as a neural correlate of working memory maintenance independent of stimulus modality or type (Spitzer et al., 2014; Wimmer et al., 2016). Furthermore, there is mounting evidence linking prefrontal beta rhythms to the reactivation of working memory (Spitzer and Haegens, 2017). Our findings are thus consistent with the notion that learning transfer may require a reactivation of a previously learned temporal pattern from supra-modal working memory, so that temporal patterns experienced in a new sensory format can be matched to previously learned patterns irrespective of stimulus modality. An interesting direction for future research would be to use multivariate methods such as representational similarity analysis (Cappotto et al., 2021), to test whether the initial neural responses after the modality switch share representational contents with the final trials before the modality switch, potentially reflecting a shared mnemonic code across modalities (Wolff et al., 2020; but see Wu et al., 2018).

In terms of the functional relevance of the beta-band signature of learning transfer, the interaction of beta-band and ERP-based learning curves could be related to the behavioural benefits (in the AV group), suggesting that the interplay between both kinds of neural signatures has behavioural relevance. Interestingly, the negative slope of the interaction effect between the two neural signatures on the behavioural benefits suggests that participants for whom the EEG and beta-band signatures of learning transfer are similar (e.g., both consistently low or consistently high) may show relatively weaker behavioural benefits than those participants for whom the EEG and beta-band signatures are dissimilar. This may suggest a number of non-exclusive possibilities: (1) interference between EEG- and beta-band signatures of transfer learning; (2) heterogeneity in the neural correlates of transfer learning across participants (possibly due to different strategies in approaching the task). However, given the nuanced pattern of results, the post-hoc character of the

 https://doi.org/10.1016/j.heares.2023.108857

time-frequency analysis, and a relatively low sample size (N=12 for the AV group), the latter finding should be replicated in a follow-up study.

The present behavioural data showed a trend suggestive of learning transfer across modalities, in that repetition in visually presented RefRP stimuli was detected at a higher rate if stimuli with the same temporal characteristics were first presented in audition but not vice versa. Note that, while this transfer effect from audition to vision failed to reach statistical significance in the current study ($p = 0.06$), a previous study using a very similar behavioural paradigm (Kang et al., 2018) observed a robust audition to vision transfer effect, making the overall evidence that behaviourally measurable modality transfer effects of temporal pattern learning can occur reasonably strong.

A likely explanation for why the behavioural effects in the current study failed to reach statistical significance, in contrast to previous work, is that this study used fewer test blocks than previous studies (Agus et al., 2010; Kang et al., 2018). Since we recorded neural responses concurrently, the test blocks needed to be slower to allow a sufficient inter-trial-interval for EEG signals. This reduced the number of test blocks that could be performed within feasible experiment durations, and this in turn would be expected to decrease the sensitivity of the experiment to learning of RefRP, as well as for transfer. This factor may have been particularly pronounced in the case of visual stimuli, given that vision has lower temporal sensitivity than audition, and participants found it generally more difficult to perform the task in the visual modality. In contrast, the participants' general task performance was greater for the second modality (audition) in the VA group. The greater temporal sensitivity in audition than vision made the auditory task relatively easier than the visual task, which may have obscured cross-modal memory transfer effects as no transfer benefits were necessary to perform at a very high level at this part of the behavioural

 https://doi.org/10.1016/j.heares.2023.108857

paradigm. Nonetheless, neural signatures for sensory processing could still be captured, while behavioural results could have been affected by other external factors such as task difficulty. Thus, neural signatures observed for implicit learning from EEG data still suggest shared mechanisms for temporal pattern learning.

Taken together, the present study did not reveal convincing behavioural evidence for a cross-modality transfer effect of implicit learning. However, neural data point to both modality-specific and modality-general mechanisms for temporal information processing and transfer, at least from audition to vision. Familiarization with temporal sequences recruits both modality specific sensory cortex (as suggested also in the "intrinsic model" by Buonomano and Maas (Buonomano and Maass, 2009), as well as supra-modal working memory circuits (Spitzer and Haegens, 2017), which may make familiarity with the learned temporal patterns accessible across sensory modalities.

Conflict of interest

The authors declare no competing financial interests.

Acknowledgments

This work has been supported by the European Commission's Marie Skłodowska-Curie Global Fellowship (750459 to R.A.), the Hong Kong General Research Fund (11100518 to R.A. and J.S.), a grant from European Community/Hong Kong Research Grants Council Joint Research Scheme (9051402 to R.A. and J.S.), the German Science Foundation (507693881 to R.A.), and Fyssen Foundation (study grant to H.K.). We would like to thank Reuben Chaudhuri for help with data acquisition, and Sir Colin Blakemore for helpful discussions.

 https://doi.org/10.1016/j.heares.2023.108857

https://doi.org/10.1016/j.heares.2023.108857

References

Agus, T., Thorpe, S., Pressnitzer, D., 2010. Rapid Formation of Robust Auditory Memories: Insights from Noise. Neuron 66(4), 610-618. https://doi.org/10.1016/j.neuron.2010.04.014.

Andrillon, T., Kouider, S., Agus, T., Pressnitzer, D., 2015. Perceptual Learning of Acoustic Noise Generates Memory-Evoked Potentials. Current Biology 25(21), 2823-2829. https://doi.org/10.1016/j.cub.2015.09.027.

Bale, M.R., Bitzidou, M., Pitas, A., Brebner, L.S., Khazim, L., Anagnou, S.T., Stevenson, C.D., Maravall, M., 2017. Learning and recognition of tactile temporal sequences by mice and humans. Elife 6, e27333.

Ball, F., Fuehrmann, F., Stratil, F., Noesselt, T., 2018a. Phasic and sustained interactions of multisensory interplay and temporal expectation. Sci Rep 8(1), 10208. https://doi.org/10.1038/s41598-018-28495-7.

Ball, F., Michels, L.E., Thiele, C., Noesselt, T., 2018b. The role of multisensory interplay in enabling temporal expectations. Cognition 170, 130-146. https://doi.org/10.1016/j.cognition.2017.09.015.

Ball, F., Nentwich, A., Noesselt, T., 2022. Cross-modal perceptual enhancement of unisensory targets is uni-directional and does not affect temporal expectations. Vision Research 190, 107962. https://doi.org/https://doi.org/10.1016/j.visres.2021.107962.

Barakat, B., Seitz, A.R., Shams, L., 2015. Visual rhythm perception improves through auditory but not visual training. Current Biology 25(2), R60-R61. https://doi.org/https://doi.org/10.1016/j.cub.2014.12.011.

 https://doi.org/10.1016/j.heares.2023.108857

Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., Chait, M. 2016. Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proceedings of the National Academy of Sciences, 113(5), E616-E625.

Bratzke, D., Seifried, T., Ulrich, R., 2012. Perceptual learning in temporal discrimination: asymmetric cross-modal transfer from audition to vision. Experimental Brain Research 221(2), 205-210. https://doi.org/10.1007/s00221-012-3162-0.

Bueti, D., Buonomano, D.V., 2014. Temporal Perceptual Learning. Timing & Time Perception 2(3), 261-289. https://doi.org/https://doi.org/10.1163/22134468-00002023.

Buonomano, D.V., Maass, W., 2009. State-dependent computations: spatiotemporal processing in cortical networks. Nat Rev Neurosci 10(2), 113-125. https://doi.org/10.1038/nrn2558.

Cappotto, D., Auksztulewicz, R., Kang, H., Poeppel, D., Melloni, L., Schnupp, J. 2021. Decoding the content of auditory sensory memory across species. *Cerebral Cortex*, *31*(7), 3226-3236.

Choi, D., Batterink, L.J., Black, A.K., Paller, K.A., Werker, J.F., 2020. Preverbal Infants Discover Statistical Word Patterns at Similar Rates as Adults: Evidence From Neural Entrainment. Psychol Sci 31(9), 1161-1173. https://doi.org/10.1177/0956797620933237.

Conway, C.M., Christiansen, M.H., 2006. Statistical Learning Within and Between Modalities: Pitting Abstract Against Stimulus-Specific Representations. Psychological Science 17(10), 905-912. https://doi.org/10.1111/j.1467-9280.2006.01801.x.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3(3), 201-215. https://doi.org/10.1038/nrn755.

de Cheveigné, A., Simon, J.Z., 2008. Denoising based on spatial filtering. J Neurosci Methods 171(2), 331-339. https://doi.org/10.1016/j.jneumeth.2008.03.015.

 https://doi.org/10.1016/j.heares.2023.108857

Faul, F., Erdfelder, E., Lang, A. G., Buchner, A. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods, 39(2), 175-191.

Gault, R. H., Goodfellow, L. D. 1938. An empirical comparison of audition, vision, and touch in the discrimination of temporal patterns and ability to reproduce them. The Journal of General Psychology, 18(1), 41-47.

Gelastopoulos, A., Whittington, M.A., Kopell, N.J., 2019. Parietal low beta rhythm provides a dynamical substrate for a working memory buffer. Proc Natl Acad Sci U S A 116(33), 16613-16620. https://doi.org/10.1073/pnas.1902305116.

Gold, J.M., Aizenman, A., Bond, S.M., Sekuler, R., 2014. Memory and incidental learning for visual frozen noise sequences. Vision research 99, 19-36.

Goodfellow, L., 1934. An empirical comparison of audition, vision, and touch in the discrimination of short intervals of time. American Journal of Psychology 46, 243-258. https://doi.org/10.2307/1416558.

Grahn, J., 2012. Neural Mechanisms of Rhythm Perception: Current Findings and Future Perspectives. Topics in Cognitive Science 4(4), 585-606. https://doi.org/10.1111/j.1756-8765.2012.01213.x.

Grahn, J.A., Henry, M.J., McAuley, J.D., 2011. FMRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa. NeuroImage 54(2), 1231-1243. https://doi.org/https://doi.org/10.1016/j.neuroimage.2010.09.033.

Grondin, S., 1993. Duration discrimination of empty and filled intervals marked by auditory and visual signals. Percept Psychophys 54(3), 383-394. https://doi.org/10.3758/bf03205274.

 https://doi.org/10.1016/j.heares.2023.108857

Grondin, S., McAuley, D., 2009. Duration discrimination in crossmodal sequences. Perception 38(10), 1542-1559. https://doi.org/10.1068/p6359.

Grondin, S., Rousseau, R., 1991. Judging the relative duration of multimodal short empty time intervals. Percept Psychophys 49(3), 245-256. https://doi.org/10.3758/bf03214309.

Guttman, S.E., Gilroy, L.A., Blake, R., 2005. Hearing what the eyes see: auditory encoding of visual temporal sequences. Psychol Sci 16(3), 228-235. https://doi.org/10.1111/j.0956-7976.2005.00808.x.

Hampshire, A., Chamberlain, S.R., Monti, M.M., Duncan, J., Owen, A.M., 2010. The role of the right inferior frontal gyrus: inhibition and attentional control. Neuroimage 50(3), 1313-1319. https://doi.org/10.1016/j.neuroimage.2009.12.109.

Hanslmayr, S., Matuschek, J., Fellner, M. C. 2014. Entrainment of prefrontal beta oscillations induces an endogenous echo and impairs memory formation. Current biology, 24(8), 904-909.

Hardy, N.F., Buonomano, D.V., 2016. Neurocomputational Models of Interval and Pattern Timing. Curr Opin Behav Sci 8, 250-257. https://doi.org/10.1016/j.cobeha.2016.01.012.

Herrmann, B., Araz, K., Johnsrude, I. S. 2021. Sustained neural activity correlates with rapid perceptual learning of auditory patterns. *NeuroImage*, *238*, 118238.

Herrmann, B., Maess, B., Johnsrude, I. S. 2022. A neural signature of regularity in sound is reduced in older adults. *Neurobiology of Aging*, *109*, 1-10.

Ille, N., Berg, P., Scherg, M., 2002. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. J Clin Neurophysiol 19(2), 113-124. https://doi.org/10.1097/00004691-200203000-00002.

 https://doi.org/10.1016/j.heares.2023.108857