Role of the prefrontal cortex in musical and verbal short-term memory: A functional near-infrared spectroscopy study

Jérémie Ginzburg[a], Anne Cheylus[a], Elise Collard[a], Laura Ferreri[b,c], Barbara Tillmann[a,d], Annie Moulin[a], Anne Caclin[a]

[a]Université Claude Bernard Lyon 1, INSERM, CNRS, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, Bron, France

[b]Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy

[c]Laboratoire d'Étude des Mécanismes Cognitifs, Université Lumière Lyon 2, Lyon, France

[d]Laboratory for Research on Learning and Development, LEAD — CNRS UMR5022, Université de Bourgogne, Dijon, France

Corresponding Author: Jérémie Ginzburg (jeremie.ginzburg@inserm.fr)

ABSTRACT

Auditory short-term memory (STM) is a key process in auditory cognition, with evidence for partly distinct networks subtending musical and verbal STM. The delayed matching-to-sample task (DMST) paradigm has been found suitable for comparing musical and verbal STM and for manipulating memory load. In this study, musical and verbal DMSTs were investigated with measures of activity in frontal areas with functional near-infrared spectroscopy (fNIRS): Experiment 1 compared musical and verbal DMSTs with a low-level perception task (that does not entail encoding, retention, or retrieval of information), to identify frontal regions involved in memory processes. Experiment 2 manipulated memory load for musical and verbal materials to uncover frontal brain regions showing parametric changes in activity with load and their potential differences between musical and verbal materials. A FIR model was used to deconvolute fNIRS signals across successive trials without making assumptions with respect to the shape of the hemodynamic response in a DMST. Results revealed the involvement of the dorsolateral prefrontal cortex (dlPFC) and inferior frontal gyri (IFG), but not of the superior frontal gyri (SFG) in both experiments, in keeping with previously reported neuroimaging data (including fMRI). Experiment 2 demonstrated a parametric variation of activity with memory load in bilateral IFGs during the maintenance period, with opposite directions for musical and verbal materials. Activity in the IFGs increased with memory load for verbal sound sequences, in keeping with previous results with n-back tasks. The decreased activity with memory load observed with musical sequences is discussed in relation to previous research on auditory STM rehearsal strategies. This study highlights fNIRS as a promising tool for investigating musical and verbal STM not only for typical populations, but also for populations with developmental language disorders associated with functional alterations in auditory STM.

Keywords: working memory, fNIRS, DMST, auditory, music, language, memory load

1. INTRODUCTION

Auditory short-term memory (STM) plays a crucial role in the processing of auditory information and enables understanding of a dynamic acoustic environment. STM is the active maintenance of information for a brief duration, typically lasting a few seconds (Baddeley & Hitch, 1974; Cowan, 2008). In the present study, "auditory short-term memory" refers to the maintenance of auditory information without manipulation of said information, while any reference to "working memory" refers to paradigms involving both information maintenance and manipulation (as classically defined by Baddeley & Hitch, 1974; Cowan, 2008). Auditory STM has been extensively studied for verbal material, but a growing body of research has also

studied STM for musical material, revealing both shared and specific mechanisms depending on the type of sounds to maintain (Caclin & Tillmann, 2018). Domain-general mechanisms for processing time-based serial order in auditory sequences have been proposed (Gorin et al., 2016), while domain-specific systems have been suggested for encoding and maintenance of distinct materials (Schulze & Tillmann, 2013). Impairment of verbal STM has been shown to be a hallmark of language-related learning disorders, such as dyslexia (Majerus & Cowan, 2016; Roodenrys & Stokes, 2001; Ziegler et al., 2009, 2011) and developmental language disorder (DLD, Archibald & Gathercole, 2006; Nickisch & Von Kries, 2009; Nithart et al., 2009). A few studies have also observed impairment for musical STM in dyslexic children (Couvignou et al., 2023; Forgeard et al., 2008; Ziegler et al., 2012). Moreover, recent studies have shown a sizeable comorbidity (~30%) between dyslexia and congenital amusia (in both adults: Couvignou et al., 2019, and children: Couvignou & Kolinsky, 2021; Couvignou et al., 2023). Congenital amusia is a lifelong disorder characterized by a specific impairment in musical STM (for reviews, see Peretz, 2016; Tillmann et al., 2015, 2023). Overall, these findings suggest impairment of auditory STM in learning disorders, which could either be domain-general or domain-specific. Together with the various patterns of auditory STM impairments observed in neurological diseases (review in Caclin & Tillmann, 2018), they stress the importance of exploring auditory STM for both materials (music and speech).

1.1. Frontotemporal networks for musical and verbal STM

Neuroimaging studies of auditory STM have revealed the involvement of distributed brain networks including auditory areas in the superior temporal lobe, frontal areas, such as the dorsolateral prefrontal cortex (dlPFC) and the inferior frontal gyrus (IFG), parietal areas including the supramarginal gyrus, as well as other brain regions, notably the cerebellum, basal ganglia, and premotor areas (for a review, see Caclin & Tillmann, 2018). Only a few studies using functional magnetic resonance imaging (fMRI) investigated the brain networks involved in STM for both musical and verbal materials. A first study reported overlapping brain networks in the superior temporal, inferior parietal, and frontal areas when participants rehearsed out loud novel melodies and nonsense sentences (Hickok et al., 2003). These findings were supported by a second study that examined sung syllables (Koelsch et al., 2009) in which the authors found similar networks activated during the silent rehearsal of verbal information (syllables) and tonal information (pitches), including mostly lateral

frontal regions: left IFG and bilateral premotor areas, and a small cluster in the left planum temporale. A subsequent fMRI study with sung syllables compared musicians with nonmusicians and revealed that during encoding, parietal and temporal (auditory) cortices showed significant activation in both groups (Schulze et al., 2011). During rehearsal of verbal or tonal information, activation was observed in lateral prefrontal regions (including the dlPFC and IFG) and subcortical structures. While overlapping for both materials, activations were more extended in frontal areas for musical material in musicians. In a recent fMRI study, Albouy et al. (2019) investigated STM for musical and verbal material using a delayed-matching-to-sample task (DMST) and compared activations with control perceptual tasks in amusic participants and matched control participants. During the maintenance of tonal and verbal information, activations emerged for control participants in left lateral frontal regions including the dlPFC and the IFG when comparing the memory task with the simple perception task. However, no activation was observed in temporal auditory regions, suggesting the specific involvement of lateral frontal regions for the maintenance of musical and verbal materials, in line with previous findings (Hickok et al., 2003; Koelsch et al., 2009; Schulze et al., 2011). In contrast to controls, amusics' brain activation patterns revealed several functional alterations during encoding and maintenance of musical information including in the right auditory cortex, the right dlPFC, and the right IFG. These results suggest that networks subtending musical and verbal STM are dissociable when studying a population with a specific deficit for the musical material. Overall, neuroimaging data support the view that temporal regions and lateral frontal regions are involved in the encoding of musical and verbal information, while lateral frontal regions are involved in the maintenance of these information (Ahveninen et al., 2023), along with other parietal and subcortical areas in some studies. Moreover, these networks seem to be largely overlapping for both materials (Schulze & Koelsch, 2012), and evidence for specialized brain networks for musical and verbal STM arose in fMRI data only from the comparison of impaired (amusics) or expert (musicians) populations with controls.

1.2. Memory load manipulation: n-back tasks

Investigating cognitive effort by manipulating memory load is valuable for understanding the underlying mechanisms of auditory STM. Most neuroimaging research on this topic has done so with the verbal identity n-back task that entails maintenance and manipulation of information, hence WM (León-Domínguez et al., 2015). In this task, participants are presented with a series of stimuli

(e.g., letters, digits) and are required to indicate when the current item matches the one presented *n* trials earlier. The task has the advantage of being applicable in the visual or auditory modality. In a meta-analysis of neuroimaging studies with verbal n-back tasks, Owen et al. (2005) found that, regardless of the modality, activity increases with memory load in frontal regions, including dlPFC, IFG, premotor cortex (PMC), and supplementary motor area (SMA). In a meta-analysis with n-back tasks, Sternberg tasks (Sternberg, 1966), and DMSTs, Rottschy et al. (2012) identified bilateral activation patterns where activity increased with memory load, encompassing frontal regions (dlPFC, IFG, PMC, SMA), the middle cingulate cortex, and temporo-occipital areas, regardless of the task, modality, or material (words, pseudowords, pictures…). Note that in this meta-analysis, all Sternberg tasks and DMSTs used visual material. Although the n-back task is frequently employed in neuroimaging studies to manipulate memory load, it is primarily designed to assess WM rather than STM. Indeed, the n-back task involves actively manipulating and updating information, which is specific of WM processes. Furthermore, neuroimaging studies have demonstrated that distinct lateral frontal regions are engaged in forward digit spans (STM) and backward digit spans (WM) (Owen, 2000; Tian et al., 2014). The choice of paradigm thus plays a pivotal role, as it can disentangle whether a study focuses on STM or working memory (WM). Therefore, the n-back task may not be the most suitable for specifically exploring the effects of memory load on STM per se, as it involves additional cognitive operations beyond simple storage, maintenance, and retrieval of information. Alternatively, recall tasks can be used to study STM and WM, but verbal serial recall tasks involve specific cognitive operation, as they depend on several phonological, sublexical, lexical, and semantic factors (e.g., Allen & Hulme, 2006). Furthermore, serial recall tasks are not easytouse for studying musical material due to their reliance on production (Caclin & Tillmann, 2018; but see Gorin et al., 2018; Williamson et al., 2010, for mixed recall/recognition task adaptations to music). Therefore**,** DMSTs seem to be specifically suited to investigate memory load in auditory STM. However, they have only rarely been employed in neuroimaging studies manipulating memory load in auditory STM.

1.3.   Memory load manipulation: DMST

Auditory DMST consists in making a same/different judgment between two sequences separated by a silent delay (Albouy et al., 2019; Caclin & Tillmann, 2018; Ginzburg et al., 2022; Gosselin et al., 2009; Talamini et al., 2022; Tillmann et al., 2009). This paradigm lends itself to memory load manipulation for musical and verbal material, to study both STM and WM (with forward and backward instructions, respectively), and has been used with three types of auditory material in behavioral experiments (words, pitch, and timbre sequences: Schulze & Tillmann, 2013; Schulze et al., 2012). Research using electroencephalography (EEG) with auditory DMST has consistently demonstrated that a sustained anterior negativity (SAN) in the electrophysiological signal serves as a reliable index of memory load. As memory load increases, the amplitude of SAN at frontocentral electrodes increases for both pitch and timbre materials (Alunni-Menichini et al., 2014; Guimond et al., 2011; Lefebvre & Jolicœur, 2016; Lefebvre et al., 2013; Nolden et al., 2013). However, due to the limited spatial resolution of EEG, the precise engagement of frontal regions could not be fully explored. Only one neuroimaging study, using magnetoencephalography (MEG), has thus far investigated the specific brain regions showing parametric activation with memory load increase during DMST with musical material (Grimault et al., 2014). By examining MEG data at the end of the silent retention delay, significant correlations between the memory load and the amplitude of the evoked response were reported for bilateral IFGs, bilateral dlPFCs, bilateral auditory cortices, and the right parahippocampal gyrus. These results indicate that lateral frontal regions (IFG and dlPFC), previously identified to be associated with musical and verbal STM (Albouy et al., 2019; Koelsch et al., 2009; Schulze et al., 2011), show a parametric increase of activity with musical memory load, in keeping with neuroimaging studies using verbal n-back tasks (Owen et al., 2005; Rottschy et al., 2012).

1.4.   fNIRS studies with memory load manipulation in n-back tasks

In recent years, there has been a growing interest for using functional near-infrared spectroscopy (fNIRS) to study brain functioning. fNIRS is a neuroimaging technique that utilizes light sources and detectors at near-infrared wavelengths to measure changes in cerebral metabolism, which serve as an indirect measure of neuronal activity. When neuronal activity increases in a particular cortical area, a rise in oxygenated hemoglobin (HbO) and a decrease in deoxygenated hemoglobin (HbR) occur concurrently, which can both be recorded with fNIRS. This hemodynamic response peaks around 5 seconds after stimulus onset (Fantini et al., 2018; Huppert et al., 2006; Scholkmann et al., 2014). The temporal dynamics of the fNIRS signal is thus similar to the bloodoxygen level-dependent (BOLD) signal in fMRI. Although having a physiological delay typical of hemodynamic signals and a limited spatial resolution, including the difficulty to record deep brain areas (such as the

auditory cortex), fNIRS offers several advantages over other imaging techniques, such as fMRI (Pinti et al., 2020). It provides a quiet and less restrictive environment as it is noninvasive and portable, making it more comfortable and suitable for capturing real-world behavior and associated neural responses in children and clinical populations who may find the fMRI environment distressing (Ferreri et al., 2014). Additionally, fNIRS has a higher tolerance to movement compared with fMRI and electroencephalography (Aslin & Mehler, 2005). Furthermore, whereas the loud sounds of fMRI can present challenges for studying the auditory modality (and in particular for individuals with language processing difficulties as it requires them to listen in a noisy environment), fNIRS is fully silent (Butler et al., 2020; Hancock et al., 2023).

A number of fNIRS studies have employed verbal n-back tasks to investigate memory load.In n-back tasks with auditorily or visually presented consonants, activity in bilateral lateral prefrontal cortices (LPFCs, including dlPFC and IFG) increased with increasing memory load (Rovetti et al., 2021), while bilateral medial prefrontal cortices (MPFCs) do not seem to respond to memory load manipulation. These results are in accordance with previous neuroimaging studies showing the involvement of LPFC structures in WM, with activity increasing parametrically with load (Owen et al., 2005; Rottschy et al., 2012). Using major, minor, and dissonant chords as stimuli in an n-back task, Tseng et al. (2018) showed that activity increased parametrically with load in bilateral orbital PFC and IFG. However, the use of complex chords (more than three tones simultaneously) and major/minor/dissonant chords possibly entailed that participants categorized these stimuli based on the abstract representation of consonance/dissonance, rather than pure pitch information. Recording the same brain regions as in Rovetti et al. (2021), hearing-aid users and normal-hearing participants displayed an increase of activity with load, leveling at 2-back difficulty in ventral and medial PFCs (Rovetti et al., 2019). This parametric activation was observed for verbal memory load in the auditory and visual modality. However, such parametric modulation was not observed in an fNIRS study of WM using a serial recall task with various materials (words, syllables, pitches, timbre, Jeong & Ryu, 2016).

To the best of our knowledge, only one study used fNIRS to investigate auditory WM load in children with developmental language disorder (DLD) and typically developing children (TD) using an auditory n-back task with consonants (Hancock et al., 2023). There was an increase of activity with increasing memory load in the left dlPFC (the right dlPFC was not recorded) and a decrease of activity with increasing memory load in bilat-

eral inferior parietal lobules (IPLs) in the TD group, but not in the DLD group. These results suggest a relationship between DLD and difficulties in engaging neural activity for different auditory WM load in dlPFC and parietal regions. In summary, similarly to fMRI investigations, fNIRS studies using auditory n-back tasks have consistently shown the involvement of lateral frontal regions (IFG and dlPFC) in STM, with activity increasing parametrically with load. In contrast, medial frontal areas appear to have a limited involvement in STM.

1.5. Studying auditory STM with Delayed-Matching-to-Sample Tasks in fNIRS

One methodological study has shown that DMST is a suitable paradigm to investigate auditory STM using fNIRS, using only verbal material and a single sequence length (Yamazaki et al., 2020). After listening to (or watching at) a nine-syllable sequence, participants had to maintain the information during a 9-second retention delay and compare it with a second nine-syllable sequence that could be either identical or different by one syllable. Within a large array of recording channels over the left frontal and temporal areas, significant activation during the encoding and maintenance phase was observed in the auditory modality in the left IFG and dlPFC, respectively, along with other premotor and temporal areas.

The primary objective of the current study was to investigate auditory STM, and more specifically the effect of memory load, for musical and verbal material using a DMST paradigm. Two experiments were conducted to achieve this goal. For both experiments, our focus was on the frontal brain regions consistently reported to be involved in both musical and verbal STM (i.e., IFG and dlPFC). Auditory cortices were not targeted with fNIRS due to their depth in location. Additionally, we recorded medial frontal regions (SFG) as control, since these areas are not typically activated with the fronto-temporal network involved in auditory STM. Experiment 1 adapted the experimental design of Albouy et al. (2019) comparing musical and verbal DMST with a low-level perception task of equivalent duration. Our aim was to replicate findings obtained with fMRI, namely that the lateral prefrontal regions exhibited stronger activation during the memory task compared with the low-level perception task. In Experiment 2, memory load for musical and verbal material was manipulated by varying sequence length. Overall, the research questions were as follows: (1) Is fNIRS suited to explore frontal activations during auditory DMST? (2) What are the frontal brain regions where activity varies parametrically with memory load? (3) Are these regions similar for musical and verbal material?

2. EXPERIMENT 1

2.1. Methods

2.1.1. Participants

Nineteen healthy adults were recruited for Experiment 1. Data from three participants were excluded because of technical problems in the fNIRS signal acquisition. This led to a final sample of 16 right-handed participants (mean age = 39.2 years, sd = 15.2 years, min = 21 years, max = 62 years, 12 females, mean education level = 14.95 years). They all gave written informed consent to participate in the experiment. Prior to the main experiment, all participants were tested with pure tone audiometry (separately for the two ears, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 6000 Hz, 8000 Hz), the Montreal Battery of Evaluation of Amusia (MBEA, Peretz et al., 2003) and a pitch discrimination threshold (PDT) test (Tillmann et al., 2009). All participants presented a normal audiometry (hearing threshold lower than 30 dB at any frequency in both ears). No participant presented any pitch perception or pitch memory impairment (MBEA > 25 (maximum score =30) and PDT < 1 semitone) and they had no or little musical education (mean musical education =0.1 years sd = 0.5 years). No participant presented any neurological or psychiatric history and none reported any past diagnosis of neurodevelopmental disorder. All study procedures were approved by a national ethics committee (CPP Ile de France VI, ID RDC 2018-A02670-55) and participants received a compensation for their participation. During the first session (~1h30), participants performed the audiometry, MBEA, and PDT. During the second session (~1h15) on a different day, participants underwent the fNIRS testing (see below).

2.1.2. Stimuli construction and task design

2.1.2.1. Musical and verbal stimuli.
Musical and verbal stimuli were the same as in Ginzburg et al. (2022). For musical tasks, six musical tones (created with the software Cubase 5.1 (Steinberg) and a Halion Sampler (Steinberg) using an acoustic piano timbre) belonging to the C major scale were used (C2, E2, G2, B2, D3, F3) with frequencies ranging from 131 to 349 Hz (thus encompassing the fundamental frequency range of the vowel recordings: 202-212 Hz). For verbal tasks, the items were consonant-vowel syllables that were selected to show the greatest perceptual distance with each other. Six consonants and 6 vowels were selected: /f/ /t/ /z/ /g/ /m/ /l/ and /i/ /e/ /a/ /y/ /ø/ /u/, thus resulting into 36 syllables that were recorded by a professional mezzo-soprano singer (for details about syllables construction, see Supplementary Figure S1 in Ginzburg et al., 2022). Six syllables were selected: /fi/ /gu/ /ly/ /mø/ /te/ /za/.

Trial examples can be found at https://github.com/jeremieginzburg/supp_mat_STM_adults_fNIRS.

2.1.2.2. Perception and memory tasks.
For each trial of perception and memory tasks, participants were asked to listen to two 5-item auditory sequences (S1 and S2; verbal or musical) separated by a silent retention delay of 6000 ms. Each item lasted 500 ms, the silent interstimulus interval (ISI) between two items lasted 100 ms. Overall, there was a 600 ms stimulus onset asynchrony, leading to a duration of 11800 ms for S1-delay-S2. Participants were given 3000 ms to provide a same/different response after the end of S2. The next trial started after a 5000- to 9000-ms randomly jittered silent interval after the response window. Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) was used to present stimuli and record responses. Additionally, sixteen 22800- to 26800-ms randomlyjittered silent trials were generated to intersperse within each testing block (see *Procedure*).

For the memory task, S1 and S2 sequences could be either identical or different. All items were different within a given sequence, and all musical sequences included at least one ascending interval and one descending interval. When S2 was different, a new item could appear equiprobably at the 2nd, 3rd, 4th position (positions 1 and 5 were not used for changes to minimize primacy and recency effects). For each material, six S1 sequences were used to create "same" trials, with S2 identical to S1. Six other S1 sequences were used to create "different" trials, with S2 sequences differing from S1 by one item. For the musical material, the new item in S2 always changed the contour of the sequence (the contour is the up-and-down scheme of a melody). So, if S1 had a down-up-**down**-up contour (e.g., E2-C2-**B2-G2**-F3), S2 could have a down-up-**up**-up contour (e.g., E2-C2-**B2-D3**-F3).

For the perception task, new sequences were created and S1 and S2 were always different. In the S2 sequence, the last two items could either be identical or different. Except when S2 contained two identical last items, all items were different within a given sequence, and for the musical material, all sequences included at least one ascending interval and one descending interval. For the musical material, when the last two items were different, these two items could not differ by more than three tones. A total of 12 S1-S2 sequences were created for each material, half of them with S2 sequences having 2 identical last items and half of them with S2 sequences having 2 different last items. For this perception task, the participants were asked to ignore S1 and to answer whether the two last items of S2 were the same or different.

2.1.3. fNIRS montage and data acquisition

The absorption of near-infrared light was measured at 760- and 850-nm wavelengths at a sampling frequency of 7.81 Hz using a continuous-wave NIRScout device (NIRx Medical Technologies, LLC). The data were collected using the NIRStar 15.3 acquisition software. Eight light sources and 12 light detectors were attached to a cap with a 10-20-system marking for probe placement. Additionally, eight 8-mm short-distance channels (one for each source) recorded systemic signal.

The montage was created using fOLD (fNIRS Optodes' Location Decider, Zimeo Morais et al., 2018), which allows placement of optodes in the international 10-20 system to maximize coverage of chosen anatomical regions as defined by one of five segmentation atlases. For the segmentation atlas, we chose the AAL2 (Automated Anatomical Labeling, Rolls et al., 2015) to generate a montage (Fig. 1) covering the inferior frontal gyri (IFG) and dorsolateral prefrontal cortex (dlPFC) as they were shown with fMRI to be involved in the encoding and retention phase in memory recognition tasks (Albouy et al., 2019). As a control, our montage additionally covered medially the superior frontal gyri (SFG), as they have not been shown to be involved in memory recognition tasks.

The final montage included 22 measurement channels with distances ranging from 2.9 to 4.1 centimeters (the montage is represented in Fig. 1). Data from two channels that had the highest specificity for the precentral gyrus were not analyzed as they were not in the scope of the current study. We discarded two midline channels because they presented noisy signal. Overall, we retained 18 channels for analysis. Among those channels and as calculated by the fOLD software, in each hemisphere, three of them had the highest level of recording specificity for IFG and four of them for dlPFC. Four channels had the highest specificity for bilateral SFGs. All channels had a maximum specificity of 38% or more for their respective ROIs, and for transparency, we report in Supplementary Table 1 the brain area with the second higher specificity for each channel.

2.1.4. Procedure

The experimenter measured the participant's head circumference to determine cap size. Cap alignment was verified and adjusted if needed so that the probe at Cz was located halfway in the nasion-to-inion and the tragus-to-tragus measurements. The participant was then led into a dimly lit, sound-attenuated booth where, if needed, the participant's hair was moved around the optode locations using a thin wooden stick to provide clear access to the scalp. Optodes were then placed according to the pre-established montage. fNIRS signal was calibrated, checked for quality, and optode placement on the scalp was readjusted until a satisfying signal quality was obtained before proceeding. Right before testing, the experimenter gave the task instructions to the participant.

Presentation of stimuli and recording of behavioral responses were controlled by Presentation© software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Triggers were sent from Presentation© to the fNIRS acquisition system using a parallel port. Participants sat approximately 50 cm from the computer screen, auditory stimuli were played through Sennheiser HD-250Pro headphones at a comfortable intensity (80 dB SPL A-weighted measured with a Larson-Davis System 824 with an AEC101 IEC 318 Artificial Ear Coupler). Headphones were positioned behind the optodes (primarily located on frontal area) and were sufficiently flexible to avoid any interference with the optical fibers. Participants gave their response with a computer mouse with their right hand (leftclick for "same" response and rightclick for "different" response).

All participants underwent four blocks of 16 trials, one block per task and per material: musical perception,
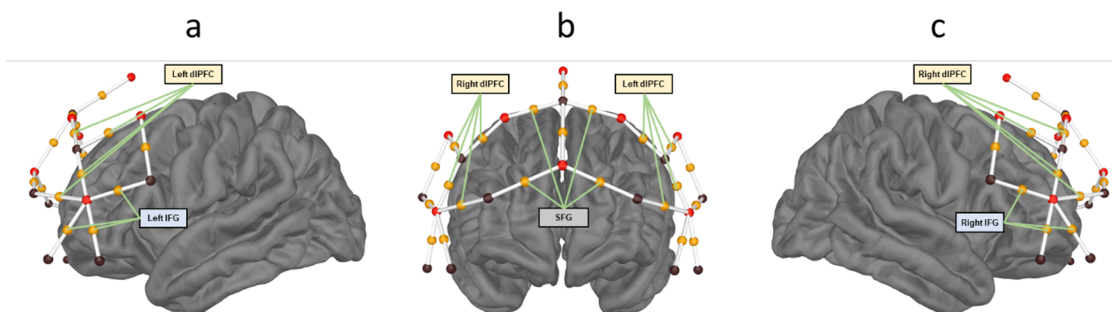

Fig. 1.   Location of sources (black spheres) and detectors (red spheres) and their corresponding midpoint channels (white lines with orange spheres) for (a) left; (b) front; and (c) right views of the brain. Green lines illustrate the ROIs used for analysis: three channels for each IFG (lateral views), four channels for each dlPFC (front and lateral views), four channels covering bilateral SFGs (front view).

musical memory, verbal perception, verbal memory. Among the 16 trials, 12 of them were stimulation trials as described above. Half of the trials were "same" trials and half of the trials were "different" trials. Stimulation trials were pseudo-randomly presented with the constraint that no more than three consecutive "same" (or "different") trials could be presented. Four "silent" trials were played on the 1st, 6th, 11th, and 16th trial position of a block in order to allow for the hemodynamic signal to return to baseline (Balters et al., 2021) and to serve as an implicit baseline in the finite impulse response (FIR, see below) analysis (Cairo et al., 2004; Kharitonova et al., 2015). Each of the four blocks lasted around 6 minutes and was preceded by a four-trial training block with feedback using the same task and material. For the test blocks, no feedback was given. The order of test blocks was counterbalanced across participants using a Latin square balancing for first-order carryover effects using the *crossdes* R package (Sailer, 2022).

2.1.5.　Behavioral data analysis

Measures of d-prime (d') and criterion (c) were obtained according to signal detection theory (SDT) for each task, material, and participant (Macmillan & Creelman, 2005). Hit corresponded to a correct answer for a different trial. False alarm corresponded to an incorrect answer for a same trial. *d'*, or sensitivity, was calculated using the *psycho* package (Makowski, 2018) as the z-score of false alarms subtracted from the z-score of hits. The criterion was calculated as the mean z-score of hits and false-alarm rates multiplied by minus one and reflects an observer's bias to say yes (in our case "different") or no ("same"), an unbiased observer having a value around 0. A liberal bias (tendency to say "different") results in a negative c, a conservative one results in positive c. Correction of extreme values was made following Hautus (1995) who recommends the use of a log-linear rule that consists of increasing each cell frequency of the contingency table by 0.5, irrespective of the content of each cell. Furthermore, we analyzed the response times (RTs) of participants after the end of S2 for correct trials only, averaged separately for each task, material, and each type of trial (same/different). Note that participants had 3 seconds to answer, otherwise the response was counted as irrelevant and thus not considered in the behavioral analysis. Those missed trials represented 0.01 % of all trials.

Analyses were conducted using Bayesian statistics that allow the direct comparison of the predictions of several hypotheses (including the null model) and to estimate a degree of logical support or belief regarding effects of interest and their interactions (Wagenmakers et al., 2018). We report Bayes Factor (BF_{10}) as a relative measure of evidence of an effect compared with the null model. Traditionally, a BF_{10} between 1 and 3 is considered as weak evidence for the tested model, between 3 and 10 as positive evidence, between 10 and 100 as strong evidence, and higher than 100 as decisive evidence. Similarly, to interpret the strength of evidence in favor of the null model, a BF_{10} between 0.33 and 1 is considered as weak evidence, a BF between 0.01 and 0.33 as positive evidence, a BF between 0.001 and 0.01 as strong evidence, and a BF lower than 0.001 as decisive evidence (Lee & Wagenmakers, 2014). For clarity purposes, we report information about the best model only.

Using the R *BayesFactor* package (Morey & Rouder, 2022), d' and c were submitted to a Bayesian repeated-measure ANOVA including task (two levels: perception and memory), material (two levels: musical and verbal), and their interaction as fixed factors. Overall, four models were tested (task, material, task + material, task + material + task:material) and compared with the null model. As recommended by Van Den Bergh et al. (2022), participants were added to all models as random factors using the *lmBF* function of the *BayesFactor* package. Paired Bayesian *t*-tests were performed as post-hoc tests if the best model included the interaction. Correct RTs were submitted to the same Bayesian ANOVA, with the addition of the type of trial (same/different) factor.

Additionally, one-sample Bayesian *t*-tests against 0 were performed on the criterion for each task and material.

Furthermore, we report the results of the analysis of effects using the *bayesfactor_inclusion* function from the R from the *bayestestR* package (Makowski et al., 2019) that compares between models that do or do not incorporate a specific effect, such as a factor or an interaction. The resulting measure, $BF_{inclusion}$, serves as a relative indicator of the evidence favoring the inclusion of a factor.

2.1.6.　fNIRS data preprocessing and signal deconvolution

fNIRS data were preprocessed using the NIRS Brain-AnalyzIR toolbox (Santosa et al., 2018) and custom-written scripts. Raw intensity signals were first converted to changes in optical density. To correct for motion artifacts from excessive head movements, we applied Temporal Derivative Distribution Repair (TDDR) (Fishburn et al., 2019), a robust regression approach to remove large fluctuations in the optical density signal (motion artifacts), while keeping smaller fluctuations (hemodynamic activity). Corrected optical density was then bandpass filtered between 0.01 and 0.2 Hz to remove cardiac (~1.2 Hz) and respiratory activity (~0.25 Hz). Finally, corrected and filtered optical densities were transformed

into (de)oxygenated hemoglobin concentrations using the modified Beer-Lambert Law.

Data were then processed with a General Linear Model (GLM) using a Finite Impulse Response (FIR) model to deconvolute the signals from successive trials for each channel. This method does not make assumptions regarding the shape of the hemodynamic response and allows for an unconstrained estimation of the full hemo-dynamic response during stimulation and maintenance. To do so, 36 one-second boxcar regressors were fitted around S1 onsets to encompass the total duration of stimulation (-5 seconds to 30 seconds around S1 onset) for each task and material. A boxcar regressor per block (encompassing the entire 12 stimulation trials and 4 silence trials) was added to account for possible HbO/HbR signal changes across blocks of recordings. Finally, data from all short-channels (eight HbO and eight HbR measures) were orthogonalized and added as regressors of no interest in the GLM in order to further clean the sig-nal from systemic components (Luke et al., 2021). Over-all, for each chromophore (HbO/HbR), each recording channel of each participant was regressed using an Ordi-nary Least Squares (OLS) GLM with a design matrix including block regressors, 36 one-second boxcar regressors for each task and material (thus 144 regres-sors for the FIR models), and all short-channel signals (16 regressors). Note that no baseline correction is needed in the FIR GLM approach as silent trials, for which no regressors are fitted in the GLM, act as implicit baseline (Cairo et al., 2004; Kharitonova et al., 2015).

2.1.7.   fNIRS data analyses

Deconvoluted data were then analyzed using a Bayesian ANOVA on beta coefficients for each 1-second time-window of the FIR models and each ROI (see below). The use of Bayes factors to analyze time-course data shows great promise with robustness to type I errors without the need for corrections (see Teichmann et al., 2021 for an empirical comparison between cluster-based corrected time-course data and the Bayes Factor approach; and Kruschke, 2010; Neath et al., 2018 for a thorough assess-ment of Bayesian statistics robustness to type I errors). Moreover, as compared with the traditional frequentist approach to time-course data that usually only allows for the comparison of two conditions when applying correc-tions for multiple testing, Bayesian statistics allow testing for multiple factors and their interaction at each time sample.

Five ROIs (see Fig. 1 and Supplementary Table 1) were created with the channels showing the largest specificity for the left or right IFG (three channels in each hemi-sphere), the left or right dlPFC (four channels in each

hemisphere), and the bilateral SFGs (four channels). For each participant, time point, task, and material, betas were averaged across channels making up each ROI. These ROI-averaged betas were then tested across par-ticipants, for each time point from -5 to 18 seconds around S1 onset (i.e., until 6 seconds after the end of S2, as the deconvoluted signal would then include the motor response which we did not intend to analyze) for each ROI with a repeated measure Bayesian ANOVA, as for behavioral data, including task (two levels: perception and memory), material (two levels: musical and verbal), and their interaction as fixed factors. Participants and their interaction with the task and material factors (ran-dom slopes) were added to all models as random factors. We report here the best model (as compared with the null model, $BF_{10} > 1$) for each time sample. Paired Bayesian *t*-tests were performed as post-hoc tests if the best model included the interaction.

We report in the main text only HbO results, as HbO tends to show higher amplitude changes and a higher SNR (Pinti et al., 2020) than HbR. However, HbR results are available in Supplementary Materials and commented in the discussion.

2.2.   Results

2.2.1.   Behavioral results

For d' (Fig. 2b), the best model explaining the data included both fixed factors (task and material) and their interaction ($BF_{10} = 3.5e+5$). The analysis of effects across matched models revealed decisive evidence for the task ($BF_{inclusion} = 9.76e+03$) and the material ($BF_{inclusion} = 278.57$) effects and weak evidence for the task:material interac-tion ($BF_{inclusion} = 1.06$). d' was higher for the verbal material as compared with the musical material and higher for the perception task as compared with the memory task (Fig. 2b). Post-hoc Bayesian *t*-tests revealed strong evi-dence for the task effect in the musical material ($BF_{10} = 39.46$) and positive evidence for the task effect in the verbal material ($BF_{10} = 8.33$).

For correct RTs (Fig. 2c), the best model was the one including only the task effect, but with only weak evi-dence relative to the null model ($BF_{10} = 2.8$). RTs were slightly longer in the perception task compared with the memory task.

For c, results are reported in Supplementary Figure S1.

2.2.2.   fNIRS results

HbO deconvoluted signals within the targeted ROIs are represented with corresponding statistics in Figure 3 (HbR results are represented in Supplementary Fig. S3).
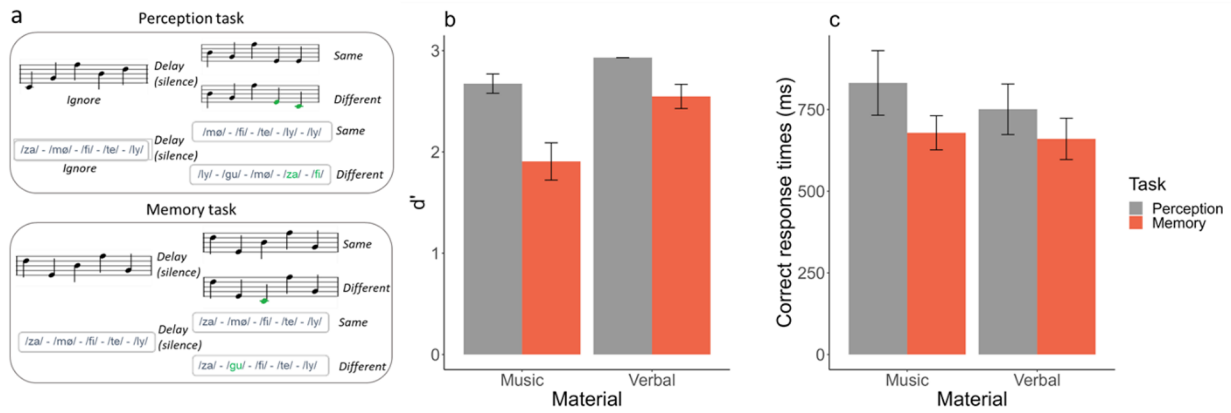
Fig. 2. Tasks and average performance as a function of task (perception in gray /memory in orange) and material (musical/verbal). (a) Examples of trials for the perception and memory tasks for musical and verbal materials. (b) Mean and standard error of participant's sensitivity (d'). (c) Mean and standard error of averaged response times for correct trials (time in millisecond that participants spent after the end of S2 before giving a "same" or "different" answer).

Fig. 3. (a) Average beta (plain line) and standard error (shaded area) from the FIR deconvolution performed on HbO data for all participants (n = 16), across five ROIs (left and right IFG, left and right dlPFC, bilateral SFGs), in a time-window ranging from -5 to 30 seconds relative to S1 onset (gray dotted vertical line), for the perception task (gray) and the memory task (red). Top panel: musical material; bottom panel: verbal material. For clarity purposes, S1, D (silent retention delay), and S2 durations are indicated with double-headed arrows. (b) Time-course representation of Bayesian analysis, each blue shade represents the best model as compared with the null model ($BF_{10} > 1$). No statistics were performed after 18 seconds (end of S2 + peak of the hemodynamic response, ~6 s), because the deconvoluted signal would then include the motor response which we did not intend to analyze. (c) Time-course representation of Bayesian post-hoc analysis performed only if the best model included the interaction. Purple shades indicate if there was evidence in favor of the task effect ($BF_{10} > 1$) for the musical material, the verbal material, or both. Dotted areas in all panels represent portions of the signal that were not statistically analyzed.

Topographic representations and individual responses around the effects of interest are represented Figure 4 (topographic representations during the whole time-course are available in Supplementary Fig. S2).

For the left IFG, we found weak evidence from -1 to 1 seconds around S1 onset for the model including the task effect ($1.47 < BF_{10} < 1.84$) and weak evidence at 8 seconds for the model with the task effect ($BF_{10} = 1.03$), with in both cases higher betas for the memory task than for the perception task.

For the right IFG, we found weak to positive evidence from -3 to 4 seconds around S1 onset for the model including the task effect ($1.11 < BF_{10} < 7$) with higher betas for the memory task than for the perception task.

For the left dlPFC, we only found weak evidence at 5 and 6 seconds for the model including the material effect ($1 < BF_{10} < 1.08$), with higher betas for the musical material than for the verbal material.

For the right dlPFC, we found weak to decisive evidence for the model including the task effect from 8 to 17 seconds around S1 onset ($1.07 < BF_{10} < 2.9e+03$) with higher betas for the memory task than the perception task. At the 14-second post-S1 time sample, we found strong evidence for the model including all effects and their interaction ($BF_{10} = 56.5$). Post-hoc tests at this time sample revealed strong evidence for the task effect only for the verbal material ($BF_{10} = 27$) with higher betas for the memory task than the perception task.

For the SFG, no model was better than the null model to explain the data (all $BF_{10} < 0.76$).

2.3. Discussion

Experiment 1 showed better behavioral performance and faster response times for the verbal material, as compared with the musical material for both perception and memory
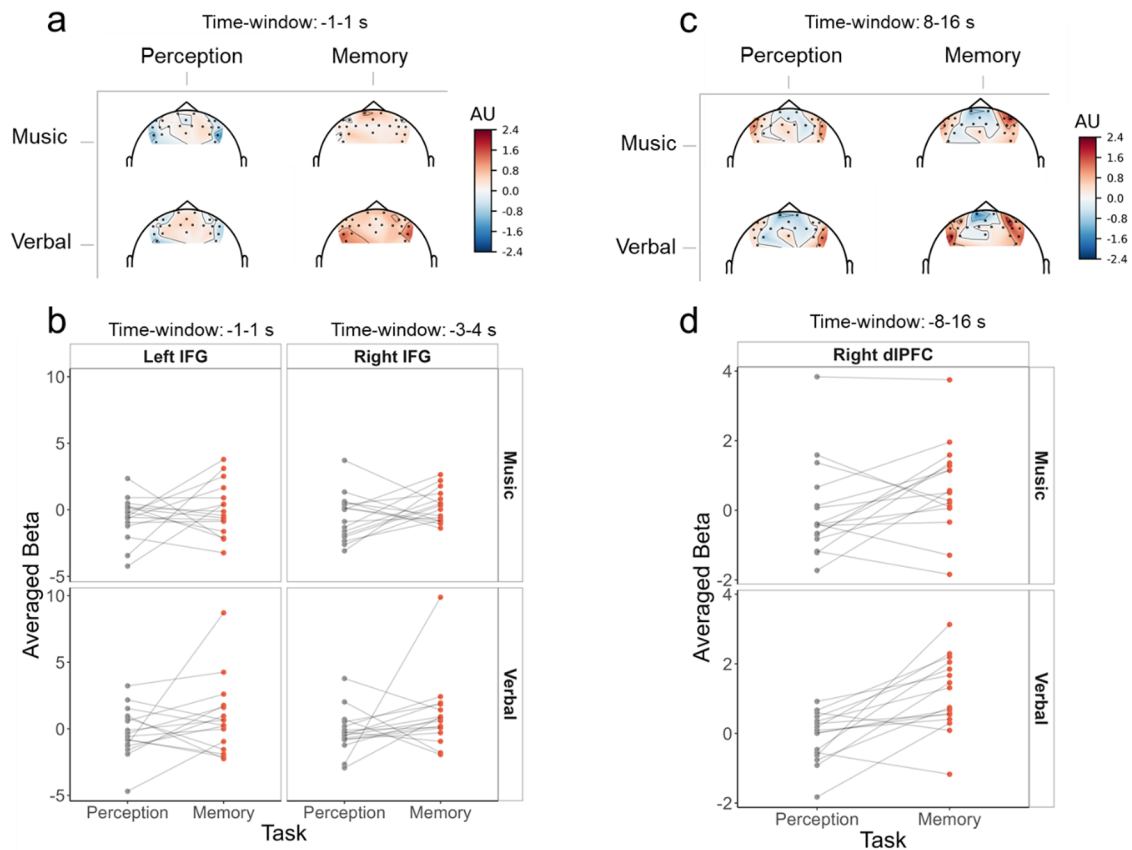

Fig. 4.   (a) Topographical representation of averaged beta -1 to 1 seconds relative to S1 onset (corresponding to the time-window where an effect of task was observed in the main analysis in both IFGs) for each material (music/verbal) and each task (perception/memory). (b) Individual averaged beta -1 to 1 seconds relative to S1 onset for the left IFG and -3 to 4 seconds relative to S1 onset for the right IFG (corresponding to the time-windows where an effect of the task was observed in each ROI) for each material (music/verbal, upper/lower rows, respectively) and each task (perception/memory in gray/orange). (c) Topographical representation of averaged beta 8 to 16 seconds after S1 onset (corresponding to the time-window where an effect of task was observed in the main analysis in the right dlPFC) for each material (music/verbal) and each task (perception/memory). (d) Individual averaged beta 8 to 16 seconds after S1 onset for the right dlPFC (corresponding to the time-window where an effect of the task was observed) for each material (music/verbal) and each task (perception/memory).

tasks. These results are in accordance with a previous study testing children with the same stimuli and the same memory paradigm that reported higher performance for the verbal material than for the musical material (Ginzburg et al., 2022). Other studies using different verbal stimuli, but the same memory task, reported a material effect in the opposite direction as observed here (verbal performance < musical performance, Albouy et al., 2019; Tillmann et al., 2009). This is due to the use of other verbal stimuli than the current study: monosyllabic words, resynthesized to obtain a constant fundamental frequency and differing only by consonants in order to exploit the phonological similarity effect (Baddeley, 1966). In contrast, our verbal material aimed for phonologically distinct items as they differ in vowel as well (hence, easier perception and more contrasting material to be memorized). As expected and as in Albouy et al.(2019), we also observed higher performance in the perception task compared with the memory task for both materials. Analysis of criterion revealed a weak tendency for a conservative bias for the musical material (more errors in "different" trials) and a more liberal bias for the verbal material (more errors in "same" trials). Based on these results, we increased the sequence length of the verbal material as compared with the musical material for Experiment 2, in order to equalize performance between materials.

HbO deconvoluted fNIRS signals over frontal areas showed a clear pattern of activation in lateral frontal channels over the time-course of the trial for both materials and both tasks (see Supplementary Fig. S2), with a higher increase in activity for the memory tasks as compared with the perception tasks. Conversely, a deactivation takes place in channels placed above medial frontal areas but does not seem to be different for the memory and the perception tasks. Analyses of activation in frontal ROIs showed evidence for higher activation in the memory task as compared with the perception task in the right dlPFC for both materials between 8 and 17 seconds after S1 onset ($BF_{10} > 1$) and in particular 10 to 13 seconds after S1 onset (all $BF_{10} > 107$). Considering that the hemodynamic response peaks ~5-6 seconds after stimulus onset (Pinti et al., 2020), signals for these time samples should mainly correspond to the cortical activity during the silent retention delay (which occurred 2.9— 8.9 seconds relative to S1 onset). In bilateral IFGs, we also observed higher activation in the memory task compared with the perception task between -1 and 1 seconds relative to S1 onset for the left hemisphere ($1.5 < BF_{10} < 1.8$) and between -3 to 4 seconds in the right hemisphere ($3.9 < BF_{10} < 7$). Given the hemodynamic delay, these responses in IFGs correspond to an anticipation period (before 0 second). The activation observed during the anticipation period in this experiment is likely attributable to the block design employed, where

each block comprised trials of the same task and material. As in-depth encoding of S1 is only needed in memory blocks, participants seemed to have adopted different attending strategies depending on the blocks. To avoid such block-related anticipatory effects, conditions were randomized within blocks in Experiment 2.

The analysis on HbR deconvoluted signal (Supplementary Fig. S3) mirrored HbO results in lateral frontal regions: we observed in the right dlPFC evidence for a stronger decrease of HbR for the memory task as compared with the perception task between 9 and 17 seconds after S1 onset ($BF_{10} > 1$) and in particular, 12 to 14 seconds after S1 onset (all $BF_{10} > 13.4$). We also observed a stronger decrease of HbR for the memory task compared with the perception task in the right IFG between -1 to 6 seconds ($1 < BF_{10} < 114$). Finally, no effect of task was observed for the SFG. Overall, these results are in accordance with previously reported findings that observed the involvement of lateral PFC areas in auditory memory processes, but not of medial frontal areas (Owen et al., 2005; Rottschy et al., 2012; Rovetti et al., 2021). They thus confirm that fNIRS is a suitable technique to explore the involvement of frontal areas in musical and verbal STM processes using a DMST paradigm.

3. EXPERIMENT 2

3.1. Methods

3.1.1. Participants

Twenty-four healthy adults were recruited for Experiment 2 (mean age =28.9 years, sd = 9.3 years, min = 21 years, max = 66 years, 6 left handed, 16 females, mean education level = 15.44 years, mean musical education = 0.48 years, sd = 1 year). Participants underwent the same inclusion procedure as for Experiment 1. They all gave written informed consent to participate in the experiment (ethical authorization: CPP Ile de France VI, ID RDC 2018-A02670-55) and were given a compensation for their participation. The first session (~1h30 minutes) was the same as for Experiment 1 and the second session during which participants underwent fNIRS recordings lasted around 1hour 30 minutes.

3.1.2. Stimuli and task design

In Experiment 2, only the STM task was used. There were three memory load (ML) levels that differed in sequence length: ML1, ML2, and ML3. For the musical material, the three MLs consisted in, respectively, four-, five-, and six-item sequences; for the verbal material, they consisted in respectively six-, seven-, and eight-item sequences. Within a trial, S1 and S2 always had the same number of

items. Item duration, interstimulus interval, delay duration, response time-window, and jittered intertrial interval were as described for Experiment 1. Due to the ML manipulation, the durations of S1-delay-S2 were, respectively, 10600 ms, 11800 ms, and 13000 ms for ML1, ML2, and ML3 for the musical material, and, respectively, 13000 ms, 14200 ms, and 15400 ms for ML1, ML2, and ML3 for the verbal material.

Twelve S1 sequences were created for each ML level and for each material: six S1 sequences were used to create "same" trials, with S2 identical to S1. Six other S1 sequences were used to create "different" trials, with S2 sequences differing from S1. Only six items per material were available in order to maximize phonological discriminability and as the verbal stimuli are intended to be used in children with language disorders as well as in adults. Since there were sequences with more than six items, we allowed for item repetitions within a given sequence with the following constraints: the first item could not be repeated within a sequence, items could be repeated only with at least two items inbetween, one item could not be repeated more than three times, and three- or four-item patterns could not be repeated. When S2 was different, two adjacent items were switched (instead of introducing a new item), thus systematically changing S2 contour for the musical material. Any item could be switched with the next one, except for the first item. When it was possible, there was an equiprobable number of sequences with each position of item switch (e.g., for musical ML1 "different" S2 sequences, there were three trials with a switch between the 2nd and the 3rd item and three trials with a switch between the 3rd and the 4th item). When it was not possible, the remaining number of trials was randomly assigned to an item position switch (e.g., for verbal ML1 "different" S2 sequences, there were five trials with an item switch, respectively, between the 2nd, the 3rd, the 4th, and the 5th item with the adjacent one and one randomly assigned trial with a switch between the 3rd and the 4th items). Additionally, twenty-four 26400- to 30400-ms randomly jittered silent trials were generated to intersperse within each testing block.

3.1.3.   fNIRS data acquisition and montage

Data acquisition and fNIRS montage were the same as described in Experiment 1.

3.1.4.   Procedure

Participants underwent 6 blocks of 16 trials each. Each block could randomly contain any stimulation trial as there were only memory trials for Experiment 2 and we wanted to avoid condition-dependent anticipation effects (as observed in Experiment 1 with a block design). There

could be no more than three consecutive "same" (or "different") trials, no more than three consecutive trials of the same memory load, and no more than four consecutive trials of the same material. Four silent trials were displayed on the 1st, 6th, 11th, and 16th trial position of each block. There was a six-trial training block at the beginning of the experiment, with one trial from each ML condition, half of them "same" and half of them "different". Then, participants underwent the six consecutive test blocks. Each of the six blocks lasted around 6 minutes. The order of test blocks was counterbalanced across participants using a Latin square balanced for first-order carryover effects using the *crossdes* R package (Sailer, 2022).

3.1.5.   Behavioral data analysis

Behavioral analysis was performed, as for Experiment 1, on d', c, and correct RTs. For d' and c, a Bayesian repeated-measure ANOVA was performed with Memory Load (three levels: ML1, ML2, ML3), Material (two levels: musical and verbal), their interaction as fixed factors and participants as random factor. Overall, four models were tested (memory load, material, memory load + material, memory load + material + memory load:material) and compared with the null model. Paired Bayesian *t*-tests were performed as post-hoc tests if the best model included the memory load effect or the interaction. Correct RTs were submitted to the same Bayesian ANOVA, with the addition of the type of trial (same/different) as factor. Additionally, one-sample Bayesian *t*-tests against 0 were performed on the criterion (c) for each memory load and material.

3.1.6.   fNIRS data preprocessing and deconvolution

The raw data were preprocessed using the same pipeline as for Experiment 1.

Data were then analyzed with the same FIR model as in Experiment 1. However, as S1 sequences had different durations according to memory load, we centered the deconvolution around the onset of the delay. Hence, 36 one-second boxcar regressors were fitted around delay onset (-9 seconds to 27 seconds). Overall, each recording channel of each participant was regressed using an Ordinary Least Squares (OLS) GLM with a design matrix including block regressors, 36 boxcar regressors around delay onsets for each memory load and material for the FIR deconvolution (216 regressors in total), and all short-channel signals (16 regressors).

3.1.7.   fNIRS data analysis

The same ROIs as in Experiment 1 were used, with betas averaged across channels making up each ROI. These

ROI-averaged betas were then tested across participants every second from -9 to 12 seconds around delay onset (i.e., until 6 seconds after the end of S2 for the highest ML) in each ROI with a repeated-measure Bayesian ANOVA including Memory Load (three levels: ML1, ML2, and ML3), material (two levels: musical and verbal), and their interaction as fixed factors, with the exact same procedure as for the analysis of the behavioral data. Participants were added to all models as random factors with their random slopes for task and material. If the interaction effect was included in the best model, a Bayesian repeated-measure ANOVA with memory load as fixed factor and participants as random factor was performed separately for each material. Then for time samples for which the best model included the memory load effect for a given material, paired Bayesian t-tests were performed as post-hoc tests between each memory load levels (ML1 vs. ML2, ML1 vs. ML3, and ML2 vs. ML3).

As for Experiment 1, we report in the main text only results for HbO, and HbR results are available in Supplementary Figure S4 and commented in the discussion.

3.2. Results

3.2.1. Behavioral results

For d' (Fig. 5a), the best model explaining the data included both fixed factors (memory load and material, decisive evidence, $BF_{10} = 6.1e+15$) but not the interaction between both. The analysis of effects across matched models revealed decisive evidence for memory load ($BF_{inclusion} = 2.83e+08$) and material ($BF_{inclusion} = 1.43e+10$) factors. d' was lower for the verbal material as compared with the musical material, and decreased as load increased (Fig. 5a). Post-hoc Bayesian t-tests for the memory load factor revealed decisive evidence for a dif-

ference between ML1 and ML2 ($BF_{10} = 3e+04$) with lower d' for ML2 as compared with ML1, decisive evidence for a difference between ML1 and ML3 ($BF_{10} = 2.2e+05$) with lower d' for ML3 as compared with ML1, and weak evidence for a difference between ML2 and ML3 ($BF_{10} = 2$) with lower d' for ML3 as compared with ML2.

For correct RTs (Fig. 5b), the best model was, as for d', the one including the fixed factors memory load and material (decisive evidence, $BF10 = 1.4e+03$). The analysis of effects across matched models revealed decisive evidence for the memory load factor ($BF_{inclusion} = 2.3e+05$) and strong evidence for the material factor ($BF_{inclusion} = 65$). RTs were longer for the verbal material as compared with the musical material, and increased with load (Fig. 5b). Post-hoc Bayesian t-tests for the memory load factor revealed positive evidence for a difference between ML1 and ML2 ($BF_{10} = 8.8$) with longer RTs for ML2 as compared with ML1, decisive evidence for a difference between ML1 and ML3 ($BF_{10} = 1.3e+04$) with longer RTs for ML3 as compared with ML1, and weak evidence for a difference between ML2 and ML3 ($BF_{10} = 1$) with slightly longer RTs for ML3 as compared with ML2.

For c, results are reported in Supplementary Figure S4.

3.2.2. fNIRS results

HbO deconvoluted signals within the targeted ROIs are shown with corresponding statistics in Figure 6 (HbR results are represented in Supplementary Fig. S6). Topographic representations and individual responses around the effects of interest are represented Figure 7 (topographic representations during the whole time-course are available in Supplementary Fig. S5). For completeness we report all effects below, but our interest was only in models including the memory load factor or its interaction with the material factor.
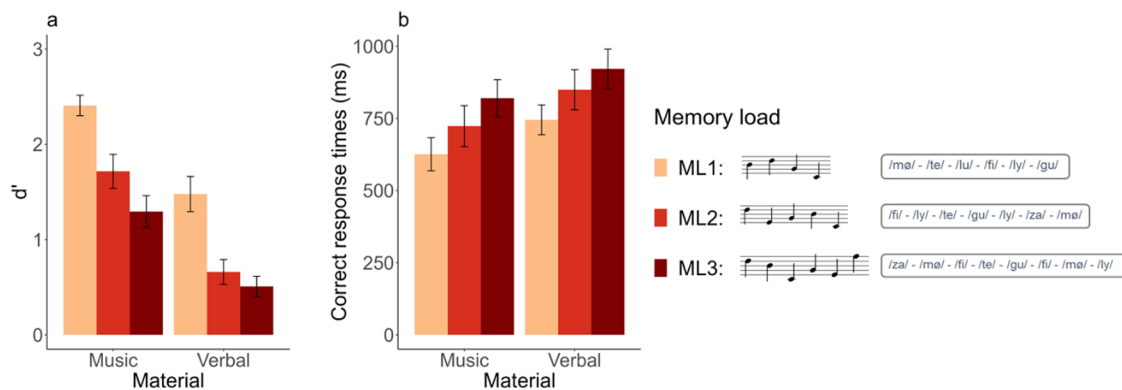

Fig. 5. Average performance as a function of memory load (ML) (ML1 in orange/ML2 in red/ML3 in dark red) and material (musical/verbal). (a) Mean and standard error of participants' sensitivity (d'). (b) Mean and standard error of averaged response times for correct trials (time in millisecond that participants spent after the end of S2 before giving the "same" or "different" answer). Examples of S1 sequences for each ML are shown in the legend.
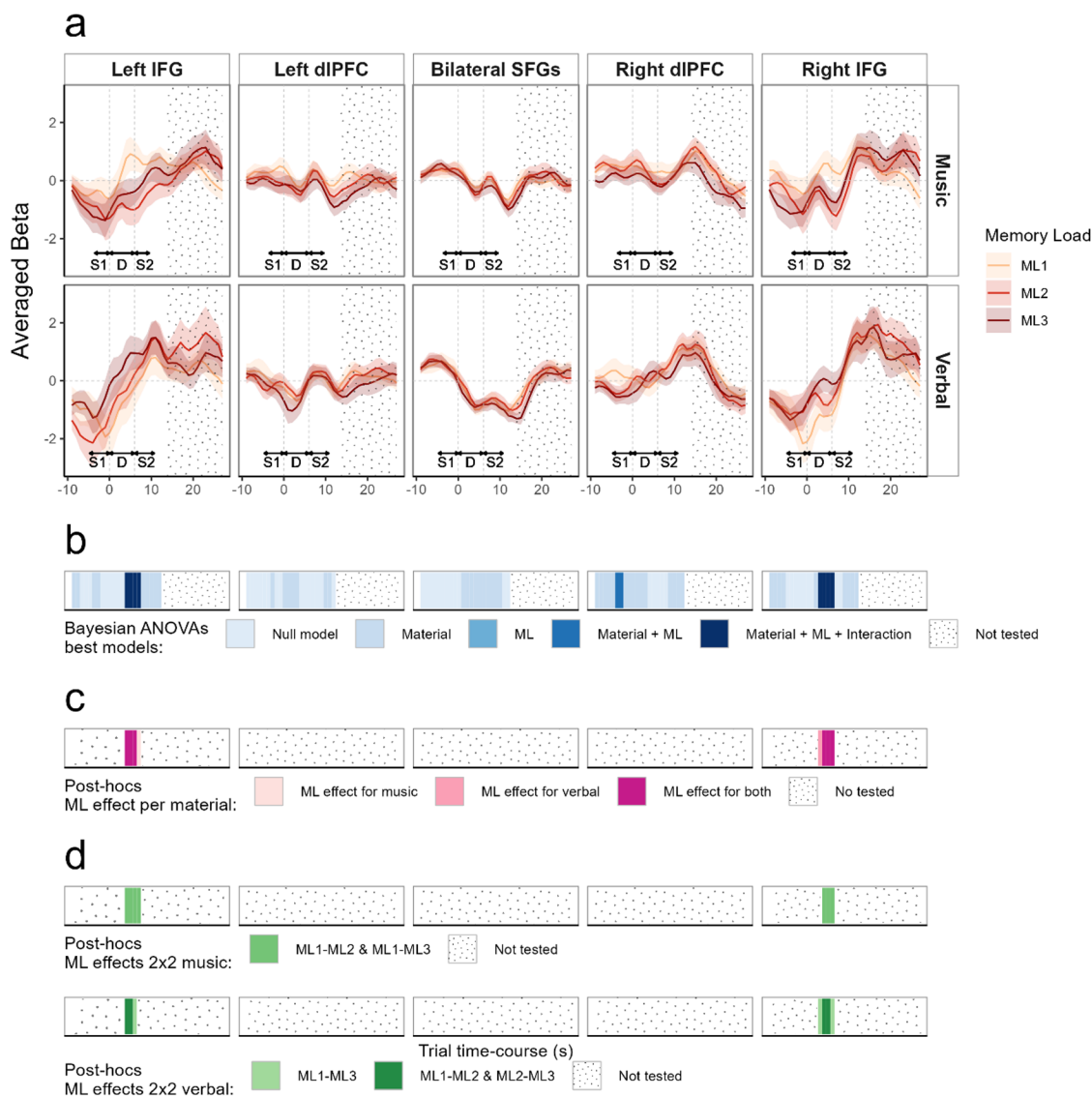
Fig. 6. (a) Average beta (plain line) and standard error (shaded area) from the FIR deconvolution performed on HbO data for all participants (n = 24), across five ROIs (left and right IFG, left and right dlPFC, bilateral SFGs), in a time-window ranging from -9 to 27 seconds around delay onset (gray dotted vertical line), for the three memory load levels (ML1/ML2/ML3), for the musical material (top panel) and verbal material (bottom panel). For clarity purposes, S1, silent retention delay (D), and S2 durations are indicated with double-headed arrows, S1 and S2 arrows are dotted to indicate their variable duration according to the memory load. The delay onset and offset are represented by dotted vertical line. (b) Time-course representation of Bayesian analysis, each blue shade represents the best model as compared with the null model ($BF_{10} > 1$). No statistics were performed beyond 12 seconds after delay onset (end of S2 for the longest sequence + peak of the hemodynamic response, ~6 s). (c) Time-course representation of post-hoc Bayesian ANOVAs per material with memory load (ML) as factor performed only if the best model included the interaction. Purple shades indicate if there was evidence in favor of a memory load effect ($BF_{10} > 1$) for the musical material, the verbal material, or both. (d) Post-hoc Bayesian t-tests comparing memory load levels two by two for the musical material (top panel) and verbal material (bottom panel) performed only if the previous post-hoc analysis showed evidence for a memory load effect for the corresponding material. Dotted areas in all panels represent portions of the signal that were not statistically analyzed.

For the left IFG, we found weak to strong evidence for the model with the material factor at time samples -9, -8, -4, and -3 seconds before delay onset ($1.2 < BF_{10} < 18.6$) with higher betas for the musical material as compared with the verbal material, and weak to strong evidence for the model with the material factor 8 to 12 seconds after delay onset ($2 < BF_{10} < 39.4$) with higher betas for the verbal material as compared with the musical material. Importantly, we found positive to strong evidence for the model including the interaction between material and memory load 4 to 7 seconds after delay onset ($3.4 < BF_{10} < 46.4$). In the musical material, post-hoc tests for the memory
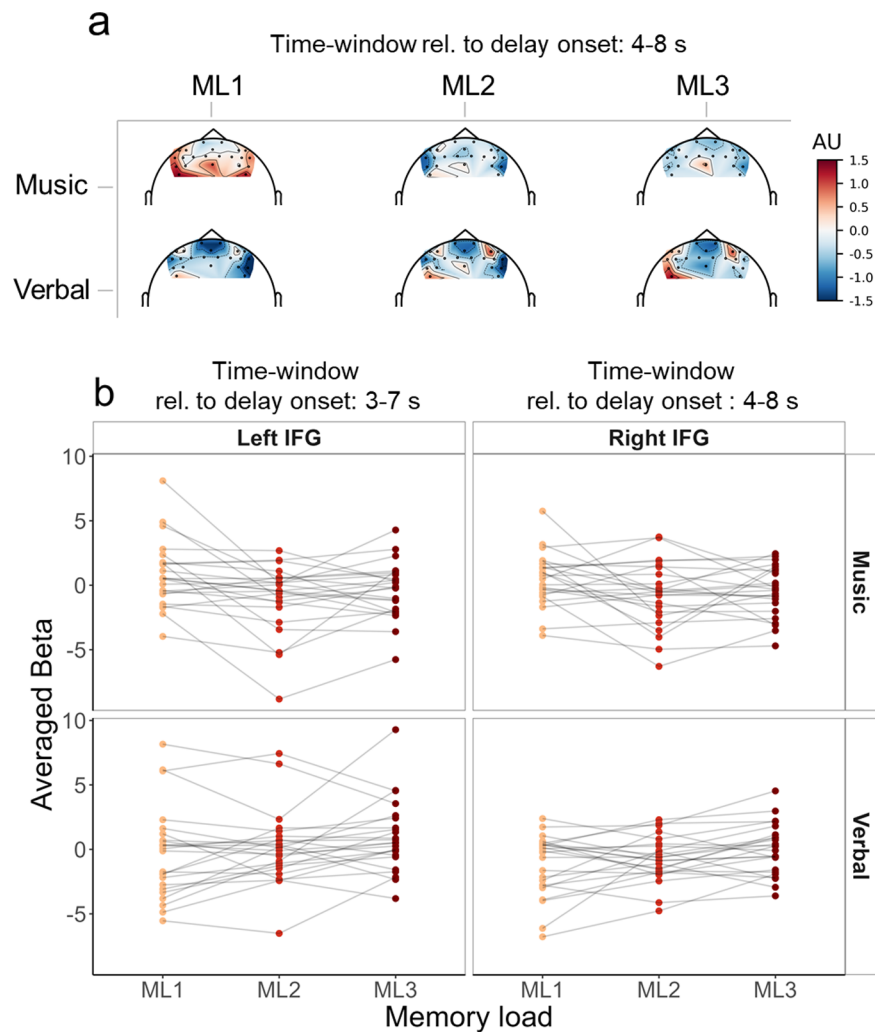
Fig. 7. (a) Topographical representation of averaged beta 4 to 8 seconds after delay onset (corresponding to the time-window where an effect of memory load was observed in the main analysis in both IFGs) for each material (music/verbal) and each memory load (ML1/ML2/ML3). (b) Individual averaged beta 3 to 7 seconds after delay onset for the left IFG and 4 to 8 seconds after delay onset for the right IFG (corresponding to the time-windows where an effect of the memory load was observed in each ROI) for each material (music/verbal, upper/lower rows, respectively) and each memory load (ML1/ML2/ML3, yellow/red/dark-red, respectively).

load effect revealed a positive to strong evidence for the memory load effect for the four tested time samples ($10.3 < BF_{10} < 49.7$). Pairwise post-hoc t-tests in the musical material revealed positive to strong evidence for a difference between ML1 and ML2 ($4.4 < BF_{10} < 10.4$) and weak to positive evidence for a difference between ML1 and ML3 ($1.8 < BF_{10} < 4.9$) for the four tested time samples. Unexpectedly, in the musical material, betas were higher for the ML1 condition as compared with the ML2 and ML3 conditions. In the verbal material, post-hoc tests for the memory load effect revealed weak to positive evidence for the memory load effect at time samples 4, 5, and 6 seconds after delay onset ($1.3 < BF_{10} < 5.3$). Pairwise post-hoc tests in the verbal material revealed weak evidence for a difference between ML1 and ML3 ($1.2 < BF_{10} < 2.4$) for the three tested time samples and

weak evidence for a difference between ML2 and ML3 ($1.2 < BF_{10} < 1.5$) for time samples 4 and 5. As expected, for the verbal material, betas were higher for the ML3 condition as compared with the ML1 and ML2 conditions. In other words, higher activations were found for the ML1 condition as compared with ML2 and ML3 in the musical material, and higher activations were found for the ML3 condition as compared with ML2 and ML1 in the verbal material.

For the right IFG, we found weak to positive evidence for the model including the material factor -9 to -6 seconds and at 2 seconds around delay onset ($1.1 < BF_{10} < 10$) with higher betas for the musical material as compared with the verbal material. Importantly, we found weak to strong evidence for the model including the interaction 3 to 6 seconds after delay onset ($1 < BF_{10} < 19.8$). Post-hoc

tests for the memory load effect in the musical material revealed a weak to positive evidence for the memory load effect for time samples 4, 5, and 6 seconds after delay onset ($2.1 < BF_{10} < 7.7$). Pairwise post-hoc Bayesian *t*-tests in the musical material and for these three time samples revealed weak to positive evidence for a difference between ML1 and ML2 ($1.2 < BF_{10} < 3.5$) and weak to positive evidence for a difference between ML1 and ML3 ($2.8 < BF_{10} < 5$). As for the left IFG for the musical material, betas were higher for the ML1 condition as compared with the ML2 and ML3 conditions. Post-hoc tests for the memory load effect in the verbal material revealed weak to strong evidence for the memory load effect for the four tested time samples ($1.5 < BF_{10} < 15.7$). Pairwise post-hoc *t*-tests in the verbal material revealed weak to positive evidence for a difference between ML1 and ML3 ($1.6 < BF_{10} < 7.8$) for the four tested time samples, and positive evidence for a difference between ML2 and ML3 ($3.9 < BF_{10} < 4.9$) at time samples 4 and 5 seconds after delay onset. As expected, for the verbal material, betas were higher for the ML3 condition as compared with the ML1 and ML2 conditions.

For the left dlPFC, we found weak to strong evidence for the model including the material factor at -3 seconds and from 0 to 3 seconds around delay onset ($1 < BF_{10} < 11.2$) with higher betas for the musical material as compared with the verbal material and weak evidence for the model with the material factor 10 and 11 seconds after delay onset ($1.2 < BF_{10} < 1.4$) with higher betas for the verbal material as compared with the musical material.

For the right dlPFC, we found weak to strong evidence for the model including the material factor at -9 to -5 seconds and -2 to 3 seconds around delay onset ($1.2 < BF_{10} < 19.1$) with higher betas for the musical material as compared with the verbal material. We also found weak to strong evidence for the model with the material factor 9 to 12 seconds after delay onset ($1.3 < BF_{10} < 15.2$) with higher betas for the verbal material as compared with the musical material. We found strong evidence for the model including material and memory load as factors -4 and -3 seconds before delay onset ($20.4 < BF_{10} < 44.2$) with higher betas for the musical material as compared with the verbal material. Post-hoc tests averaged over materials for the memory load effect (not shown in Fig. 6) revealed weak evidence for a difference between ML1 and ML2 ($1.4 < BF_{10} < 1.9$) and positive to strong evidence for a difference between ML1 and ML3 ($9.8 < BF_{10} < 10.1$), with higher betas for ML1 as compared with ML2 and ML3.

For the SFG, we found weak to strong evidence for the model including the material factor from 1 to 10 seconds after delay onset ($1.1 < BF_{10} < 22.8$) with higher betas for the musical material as compared with the verbal material.

3.3. Discussion

The aim of Experiment 2 was to manipulate memory load for musical and verbal STM in order to identify the lateral frontal regions showing parametric involvement with memory load and to explore whether their involvement differs for the two materials.

Behavioral data showed a parametric decrease of performance and an increase of RTs for higher memory loads with strong to decisive evidence for differences between ML1 and the other two memory loads (Fig. 5), as previously observed in studies manipulating memory load with DMST (Grimault et al., 2014; Schulze & Tillmann, 2013; Schulze et al., 2012). Analysis of the criterion (c) revealed weak evidence for a liberal bias for low memory loads for the musical material. The performance was overall lower and RTs longer for the verbal material as a consequence of (1) increasing sequence length by two items in the verbal material compared with the musical material and (2) switching from a change of item to a change of order when S2 sequences were different from S1. In future studies using these stimuli, similar performance between musical and verbal material should be reached by increasing sequence length for the verbal material by one item compared with the musical material. Nevertheless, we managed to manipulate memory load in a similar fashion for both materials, as revealed by the absence of interaction between material and memory load for both performance and RTs. These results provide evidence that the DMST paradigm is highly effective in manipulating memory load, in line with previous behavioral studies with auditory material (Schulze & Tillmann, 2013; Schulze et al., 2012) and the few neuroimaging studies employing DMST for this purpose in the visual (Habeck et al., 2005; Kaiser et al., 2010; Robitaille et al., 2010) and auditory (Grimault et al., 2014) modalities.

As S1 duration was different for each memory load (different sequence lengths), rendering hemodynamic time-course comparison in the response to S1 difficult, we analyzed the hemodynamic responses from the start of the retention delay. Note that material effects in the time ranges corresponding to S1 encoding or S2 processing are likely reflecting the differences in sequence lengths for the two materials. Hence, the interpretation of the data focuses on the delay period. The topographical representation of HbO deconvoluted signal (Supplementary Fig. S5) showed a parametric modulation of activity in lateral frontal channels for both materials after the delay onset with higher activation for lower memory loads for the musical material and, conversely, higher activation for higher memory loads for the verbal material. In the ROI analysis (Fig. 6), the parametric effects in the IFGs were present 3 to 6 seconds after delay onset but not

later, confirming that the parametric effect specifically concerned processes at the beginning of the silent retention delay. For the verbal material, we observed the expected activation increase with higher memory loads but intriguingly, for the musical material, the effect of memory load was not the expected one, with higher activation for ML1 as compared with ML2 and ML3. We discuss these results, along with those from Experiment 1, in the following general discussion.

Channels placed along the medial part of the PFC (SFGs) were not sensitive to memory load, as expected. They showed a global deactivation after delay onset regardless of the memory load, but more strongly for verbal stimuli. In the ROI analysis (Fig. 6), we observed that the dlPFC and SFG did not show any memory load effect, and that the parametric memory load effects only involved the bilateral IFGs. Furthermore, in the SFG, the material effect revealed greater deactivation for the verbal material, likely reflecting the increased difficulty of the task, as revealed by the lower performance for the verbal material than for the musical material.

4.   GENERAL DISCUSSION

The aims of the present study were three-fold. (1) We aimed at validating the application of fNIRS in investigating frontal activations during auditory DMST. In Experiment 1, the IFG and dlPFC exhibited greater activation for the memory task as compared with the perception task for both materials while the SFG did not exhibit any task effect. (2) Our second goal was to identify the specific frontal brain regions where activity demonstrates parametric variation with memory load. In Experiment 2, only the IFG showed parametric activations with memory load while the dlPFC and the SFG did not respond to memory load manipulation. (3) We wanted to determine whether these regions exhibit similarities when processing musical and verbal materials. Both materials showed similar patterns in Experiment 1 (memory > perception in the IFG and the dlPFC) and memory load processing involved the same region (IFG) in Experiment 2. However, different profiles were observed between materials with increasing activations for increasing memory loads in the verbal material and decreasing activation for increasing memory loads in the musical material.

4.1.   Involvement of frontal regions in auditory STM

Previous research has consistently indicated the involvement of lateral prefrontal regions during maintenance of musical and verbal material in auditory STM, specifically the IFG and dlPFC. None of the previous studies has reported the involvement of medial prefrontal regions

(Grimault et al., 2014; Rottschy et al., 2012). In the current study, both experiments provide converging evidence with these observations. In particular, Experiment 1 provided (1) evidence for the involvement of bilateral IFGs during the anticipation period (-3 to 2 seconds around S1 onset), only for the memory task. These results are an addition to previous fMRI studies that did not investigate anticipation effects. (2) During the maintenance of information (i.e., the silent retention delay), we found that the dlPFC showed higher activation for the memory task as compared with the perception task, while this pattern was not observed in the IFG. The findings in previous studies present inconsistencies regarding the specific involvement of the dlPFC or IFG in the maintenance of musical and verbal STM. While Koelsch et al. (2009) observed activations in the IFG, but not the dlPFC, during the maintenance phase for musical and verbal information, Schulze et al. (2011) and Albouy et al. (2019) reported involvement of both regions during the maintenance period. Furthermore, the inconsistency observed in the varying engagement of the dlPFC and IFG could arise from the ambiguity surrounding the definition of processes in auditory STM and WM. In our current study, we have adhered to a specific definition of STM, involving the maintenance of auditory information without manipulation. Given this definition, studies mentioned earlier can be categorized as investigating STM (Albouy et al., 2019; Schulze et al., 2011), while others, e.g., using suppression conditions (Koelsch et al., 2009) or direct information manipulation such as in n-back tasks (Rottschy et al., 2012) and modified serial order recall tasks (Jeong & Ryu, 2016), investigate WM. Given that our study is the first to investigate STM under the previously outlined definition by directly comparing materials using a neuroimaging technique that differs from fMRI, and considering the scarcity of existing research in this area, further studies are necessary to draw conclusive insights regarding whether the IFG and/or the dlPFC is preferentially involved in maintenance in auditory STM tasks (vs. WM tasks). In Experiment 1, our findings revealed, in line with previous fMRI studies, that when comparing memory processes with perception processes in control participants, no difference between musical and verbal material was observed. In addition, we did not observe differences in the medial frontal regions activity for memory processes compared with perception processes. Overall, by demonstrating differential activation patterns between memory and perception tasks in lateral PFC, our study contributes to the growing body of evidence linking lateral prefrontal regions to the encoding and maintenance of auditory information in STM, a link that was further explored in Experiment 2 by manipulating memory load.

4.2. Memory load manipulation for musical and verbal STM

In Experiment 2, our results demonstrated that, as the verbal memory load increased, there was a corresponding increase in activation in bilateral IFGs during the maintenance of information. In contrast, the verbal memory load did not impact bilateral dlPFC and SFG activations. Previous studies in fMRI report a linear increase in activity with visual verbal memory load using n-back tasks (assessing WM) in the left IFG and in bilateral dlPFC (Braver et al., 1997) and in the left IFG and right dlPFC (Cohen et al., 1997). fNIRS studies have also yielded similar results for visual-verbal memory load using n-back tasks (Fishburn et al., 2014; Khaksari et al., 2019). Fishburn et al. (2014) showed an increase in dlPFC with-load but not in the IFG, medial PFC, or parietal areas. Khaksari et al.(2019) found an increase of activity in bilateral dlPFC memory load (they did not record bilateral IFGs). Interestingly, they also found a parametric increase in medial PFC. However, these results can be mitigated as most of the fNIRS studies analyzed HbO and HbR data, whereas in this study, the authors only analyzed total hemoglobin concentration changes, making the comparison difficult. In the auditory modality, fNIRS studies have found an increase of activity with verbal WM load in bilateral IFGs and dlPFCs (without differentiating them in the fNIRS montage, Rovetti et al., 2021) and in bilateral orbital PFC and IFG, but without recording bilateral dlPFCs (Tseng et al., 2018), see, however, Jeong & Ryu (2016) for negative findings.

Overall, for verbal material, WM studies consistently report a linear bilateral increase in lateral prefrontal regions with memory load, but most of the time not in medial frontal regions. The relative involvement of the dlPFC and the IFG varies across studies. One potential explanation for the present findings implicating mostly the IFG is that our DMST paradigm targeted STM rather than WM processes. With fMRI, using forward (STM) and backward (WM) digit span tasks, the orbital part of the IFG showed significant activation for both tasks while the dlPFC showed significant activation only for the WM task. With fNIRS, the contrast between forward and backward digit span tasks shows that only the dlPFC is involved in WM processes (Tian et al., 2014). The relative involvement of IFG, dlPFC, and SFG might thus depend on task requirements, neuroimaging technique, modality, and analysis pipelines.

For musical STM, our study revealed that during the maintenance of information, activity varied parametrically with load in bilateral IFGs but not in bilateral dlPFCs nor SFGs. However, a striking feature of the current results is that despite similar behavioral changes (decrease in accuracy and increase in RT) when memory load increases for the two types of sound sequences (music and verbal), the modulation of activity in the IFG goes in opposite directions for musical and verbal sequences. Specifically, we observed higher activations for the 1st level of memory load (four-item sequences) as compared with levels 2 and 3 (five- and six-item sequences). These results are not in line with the only study that investigated memory load effects using a DMST for musical material using MEG: Grimault et al. (2014) reported an increase of activity with memory load in bilateral IFGs, dlPFCs, and temporal regions. The main difference between the latter study and the present study, beyond the use of different neuroimaging techniques, is that they used four-item sequences as their maximum memory load, when in the present study a four-item sequence corresponded to the minimum memory load, six-item sequences being the maximum one. Furthermore, their silent retention interval lasted 2 seconds when ours lasted 6 seconds. Therefore, the task used in the present study was far more challenging than the task used in Grimault et al. (2014). While studies employing verbal n-back tasks have reported deactivation in lateral prefrontal regions when memory loads exceed participants' WM capacity (Nyberg et al., 2009), we can exclude this explanation for the present study. Indeed, our behavioral results demonstrate that participants were able to perform the musical task even at the highest memory load, and we observed a parametric increase of IFG activity with load in the more difficult verbal task. A possible explanation for these findings could be that when processing musical sequences in the lower memory load condition, participants employ the same strategy as for the verbal material and thus engage the same frontal network, as evidenced by the increased activity in the IFGs in the ML1 condition. However, for higher memory loads, participants might change strategy. It has been shown that for the encoding of verbal information, when the acoustic context facilitates chunking strategies, activity decreases in lateral frontal regions (Ferreri et al., 2015). Such activity decreases in lateral PFC are also observed when participants are given specific instructions to use a chunking strategy (Matsui et al., 2007). For musical material, participants might have relied on different strategies to handle higher memory loads. A previous study comparing musicians and non-musicians has evidenced the existence of different (and more or less efficient) strategies for musical STM, notably using or not contour information (Talamini et al., 2022). Some of these alternative strategies (chunking, contour-based, etc.) could involve different brain regions than the ones recorded here. Interestingly, when testing the impact of memory load for musical material with MEG, significant clusters of activation were observed not only

in frontal areas, but also in temporal and parietal areas (Grimault et al., 2014). Future studies using musical material and a DMST should gather data about the strategy employed by participants to perform the task and record activity in temporal and parietal areas. Moreover, these maintenance strategies could be studied for other auditory material, such as timbre. Indeed, it has been suggested that strategies used to maintain for timbre information in WM differs from tonal and verbal material and that they would rely on sensory imagery rather than internal rehearsal or moto-related processes (Schulze & Tillmann, 2013).

Previous neuroimaging studies comparing musical and verbal STM consistently found overlapping regions involved for both materials, except in impaired or expert populations (Caclin & Tillmann, 2018). Our results support the observation that the same brain regions are recruited for musical and verbal STM, but by manipulating memory load, we found that their involvement differed between the two materials. These results yield great promise for the identification of specific neurophysiological markers of auditory STM for specific materials. Here, by increasing the length of the verbal sequences by two items compared with the musical sequences, we managed to have a similar influence of increases in memory loads on behavior, but with an overall decrease of performance for verbal sequences compared with musical sequences. Considering that most neuroimaging studies investigating the effect of memory load employ paradigms involving not only maintenance but also manipulation of information, it would be particularly valuable for future studies to compare the engagement of prefrontal regions depending on memory load, for both materials, and both with and without information manipulation. For instance, incorporating a suppression task during the maintenance period of a DMST, or requesting the participant to reverse the sequence (backward task), could provide insightful comparisons with the classic "forward DMST without suppression" (as used here). To go further, future studies should record other regions that have been shown to be involved in memory processes (e.g., temporal and parietal regions). One potential limitation of fNIRS is its limited ability to record deeper brain regions, such as the primary auditory cortex and surrounding areas, which could also be involved in musical memory. While some studies have reported success in detecting auditory cortex activity using fNIRS (Plichta et al., 2011; Santosa et al., 2014), it remains a methodological challenge. Additionally, it would be valuable to investigate the activity of parietal regions, such as the superior and inferior parietal lobule, which activity has been shown to be modulated by memory load for musical and verbal material (Grimault et al., 2014; Rovetti et al., 2021).

4.3. Perspectives for studies in children with learning disorders

fNIRS, a portable and non-invasive neuroimaging tool, holds promise for studying auditory cognition in children, including those with learning disorders like dyslexia and developmental language disorder (DLD). This study demonstrates fNIRS' effectiveness in measuring neurophysiological markers of auditory STM and its ability to differentiate memory load effects across materials. Previous research using fNIRS in DLD children revealed decreased activation in specific brain regions during language comprehension and verbal WM tasks compared with typically developing (TD) children (Fu et al., 2016; Hancock et al., 2023). These findings, along with prior studies, underscore fNIRS' potential for exploring functional differences between DLD and TD children. Future investigations could delve into the engagement of frontal regions during auditory STM for various materials in children with dyslexia and DLD, contributing to a better understanding of neural mechanisms and potentially informing targeted interventions for this population.

DATA AND CODE AVAILABILITY

Access to the data requires the completion of a Data Use Agreement (DUA), available upon request to the corresponding author. The analytic code necessary to reproduce the analyses presented in this paper is publicly accessible at the following URL: https://github.com/jeremieginzburg/fNIRS_STM_adults.

AUTHOR CONTRIBUTIONS

Jérémie Ginzburg: Conceptualization, Investigation, Methodology, Visualization, Resources, Software, Writing–original draft, Writing–review & editing. Anne Cheylus: Methodology, Software, Writing–review & editing. Elise Collard: Investigation, Writing–review & editing. Laura Ferreri: Conceptualization, Writing–review & editing, Funding acquisition. Barbara Tillmann: Conceptualization, Resources, Writing–review & editing, Funding acquisition. Annie Moulin: Conceptualization, Resources, Writing–review & editing, Funding acquisition. Anne caclin: Conceptualization, Supervision, Methodology, Resources, Software, Writing–review & editing, Funding acquisition.

FUNDING

DECLARATION OF COMPETING INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: https://doi.org/10.1162/imag_a_00168.

REFERENCES

Ahveninen, J., Uluç, I., Raij, T., Nummenmaa, A., & Mamashli, F. (2023). Spectrotemporal content of human auditory working memory represented in functional connectivity patterns. *Communications Biology*, 6(1), 294. https://doi.org/10.1038/s42003-023-04675-8

Albouy, P., Peretz, I., Bermudez, P., Zatorre, R. J., Tillmann, B., & Caclin, A. (2019). Specialized neural dynamics for verbal and tonal memory: fMRI evidence in congenital amusia. *Human Brain Mapping*, 40(3), 855–867. https://doi.org/10.1002/hbm.24416

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, 55(1), 64–88. https://doi.org/10.1016/j.jml.2006.02.002

Alunni-Menichini, K., Guimond, S., Bermudez, P., Nolden, S., Lefebvre, C., & Jolicoeur, P. (2014). Saturation of auditory short-term memory causes a plateau in the sustained anterior negativity event-related potential. *Brain Research*, 1592, 55–64. https://doi.org/10.1016/j.brainres.2014.09.047

Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, 41(6), 675–693. https://doi.org/10.1080/13682820500442602

Aslin, R. N., & Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants: Promise, prospects, and challenges. *Journal of Biomedical Optics*, 10(1), 011009. https://doi.org/10.1117/1.1854672

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, 18(4), 302–309. https://doi.org/10.1080/14640746608400047

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The Psychology of Learning and Motivation*, 8, 47–90. https://doi.org/10.1016/S0079-7421(08)60452-1

Balters, S., Baker, J. M., Geeseman, J. W., & Reiss, A. L. (2021). A methodological review of fNIRS in driving research: Relevance to the future of autonomous vehicles. *Frontiers in Human Neuroscience*, 15, 637589. https://doi.org/10.3389/fnhum.2021.637589

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, 5(1), 49–62. https://doi.org/10.1006/nimg.1996.0247

Butler, L. K., Kiran, S., & Tager-Flusberg, H. (2020). Functional near-infrared spectroscopy in the study of speech and language impairment across the life span: A systematic review. *American Journal of Speech-Language Pathology*, 29(3), 1674–1701. https://doi.org/10.1044/2020_AJSLP-19-00050

Caclin, A., & Tillmann, B. (2018). Musical and verbal short-term memory: Insights from neurodevelopmental and neurological disorders: Musical and verbal short-term memory. *Annals of the New York Academy of Sciences*, 1423(1), 155–165. https://doi.org/10.1111/nyas.13733

Cairo, T. A., Liddle, P. F., Woodward, T. S., & Ngan, E. T. C. (2004). The influence of working memory load on phase specific patterns of cortical activity. *Cognitive Brain Research*, 21(3), 377–387. https://doi.org/10.1016/j.cogbrainres.2004.06.014

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, 386(6625), 604–608. https://doi.org/10.1038/386604a0

Couvignou, M., & Kolinsky, R. (2021). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia in children. *Neuropsychologia*, 155, 107811. https://doi.org/10.1016/j.neuropsychologia.2021.107811

Couvignou, M., Peretz, I., & Ramus, F. (2019). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia. *Cognitive Neuropsychology*, 36(1–2), 1–17. https://doi.org/10.1080/02643294.2019.1578205

Couvignou, M., Tillmann, B., Caclin, A., & Kolinsky, R. (2023). Do developmental dyslexia and congenital amusia share underlying impairments? *Child Neuropsychology*, 29(8), 1294–1340. https://doi.org/10.1080/09297049.2022.2162031

Cowan, N. (2008). Chapter 20 What are the differences between long-term, short-term, and working memory? In *Progress in brain research* (Vol. 169, pp. 323–338). Elsevier. https://doi.org/10.1016/S0079-6123(07)00020-9

Fantini, S., Frederick, B., & Sassaroli, A. (2018). Perspective: Prospects of non-invasive sensing of the human brain with diffuse optical imaging. *APL Photonics*, 3(11), 110901. https://doi.org/10.1063/1.5038571

Ferreri, L., Bigand, E., Bard, P., & Bugaiska, A. (2015). The influence of music on prefrontal cortex during episodic encoding and retrieval of verbal information: A multichannel fNIRS study. *Behavioural Neurology*, 2015, 1–11. https://doi.org/10.1155/2015/707625

Ferreri, L., Bigand, E., Perrey, S., & Bugaiska, A. (2014). The promise of near-infrared spectroscopy (NIRS) for psychological research: A brief review. *L'Année Psychologique*, 114(03), 537–569. https://doi.org/10.4074/S0003503314003054

Fishburn, F. A., Ludlum, R. S., Vaidya, C. J., & Medvedev, A. V. (2019). Temporal derivative distribution repair

(TDDR): A motion correction method for fNIRS. *NeuroImage*, *184*, 171–179. https://doi.org/10.1016/j.neuroimage.2018.09.025

Fishburn, F. A., Norr, M. E., Medvedev, A. V., & Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00076

Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., & Winner, E. (2008). The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 383–390. https://doi.org/10.1525/mp.2008.25.4.383

Fu, G., Wan, N. J. A., Baker, J. M., Montgomery, J. W., Evans, J. L., & Gillam, R. B. (2016). A proof of concept study of function-based statistical analysis of fNIRS data: Syntax comprehension in children with specific language impairment compared to typically-developing controls. *Frontiers in Behavioral Neuroscience*, *10*. https://doi.org/10.3389/fnbeh.2016.00108

Ginzburg, J., Moulin, A., Fornoni, L., Talamini, F., Tillmann, B., & Caclin, A. (2022). Development of auditory cognition in 5- to 10-year-old children: Focus on musical and verbal short-term memory. *Developmental Science*, *25*(3), e13188. https://onlinelibrary.wiley.com/action/showCitFormats?doi=10.1111%2Fdesc.13188

Gorin, S., Kowialiewski, B., & Majerus, S. (2016). Domain-generality of timing-based serial order processes in short-term memory: New insights from musical and verbal domains. *PLOS ONE*, *11*(12), e0168699. https://doi.org/10.1371/journal.pone.0168699

Gorin, S., Mengal, P., & Majerus, S. (2018). A comparison of serial order short-term memory effects across verbal and musical domains. *Memory & Cognition*, *46*(3), 464–481. https://doi.org/10.3758/s13421-017-0778-0

Gosselin, N., Jolicœur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*(1), 270–272. https://doi.org/10.1111/j.1749-6632.2009.04762.x

Grimault, S., Nolden, S., Lefebvre, C., Vachon, F., Hyde, K., Peretz, I., Zatorre, R. J., Robitaille, N., & Jolicoeur, P. (2014). Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch: Evidence from magnetoencephalography. *NeuroImage*, *94*, 96–106. https://doi.org/10.1016/j.neuroimage.2014.03.020

Guimond, S., Vachon, F., Nolden, S., Lefebvre, C., Grimault, S., & Jolicoeur, P. (2011). Electrophysiological correlates of the maintenance of the representation of pitch objects in acoustic short-term memory: Pitch and ASTM. *Psychophysiology*, *48*(11), 1500–1509. https://doi.org/10.1111/j.1469-8986.2011.01234.x

Habeck, C., Rakitin, B. C., Moeller, J., Scarmeas, N., Zarahn, E., Brown, T., & Stern, Y. (2005). An event-related fMRI study of the neural networks underlying the encoding, maintenance, and retrieval phase in a delayed-match-to-sample task. *Cognitive Brain Research*, *23*(2–3), 207–220. https://doi.org/10.1016/j.cogbrainres.2004.10.010

Hancock, A. S., Warren, C. M., Barrett, T. S., Bolton, D. A. E., & Gillam, R. B. (2023). Functional near-infrared spectroscopy measures of neural activity in children with and without developmental language disorder during a working memory task. *Brain and Behavior*, *13*(2), e2895. https://doi.org/10.1002/brb3.2895

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–682. https://doi.org/10.1162/089892903322307393

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, *29*(2), 368–382. https://doi.org/10.1016/j.neuroimage.2005.08.065

Jeong, E., & Ryu, H. (2016). Nonverbal auditory working memory: Can music indicate the capacity? *Brain and Cognition*, *105*, 9–21. https://doi.org/10.1016/j.bandc.2016.03.003

Kaiser, S., Kopka, M.-L., Rentrop, M., Walther, S., Kronmüller, K., Olbrich, R., Weisbrod, M., & Stippich, C. (2010). Maintenance of real objects and their verbal designations in working memory. *Neuroscience Letters*, *469*(1), 65–69. https://doi.org/10.1016/j.neulet.2009.11.045

Khaksari, K., Condy, E., Millerhagen, J., Anderson, A., Dashtestani, H., & Gandjbakhche, A. (2019). Effects of performance and task duration on mental workload during working memory task. *Photonics*, *6*(3), 94. https://doi.org/10.3390/photonics6030094

Kharitonova, M., Winter, W., & Sheridan, M. A. (2015). As working memory grows: A developmental account of neural bases of working memory capacity in 5- to 8-year old children and adults. *Journal of Cognitive Neuroscience*, *27*(9), 1775–1788. https://doi.org/10.1162/jocn_a_00824

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An FMRI study. *Human Brain Mapping*, *30*(3), 859–873. https://doi.org/10.1002/hbm.20550

Kruschke, J. K. (2010). Bayesian data analysis. *WIREs Cognitive Science*, *1*(5), 658–676. https://doi.org/10.1002/wcs.72

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lefebvre, C., & Jolicœur, P. (2016). Memory for pure tone sequences without contour. *Brain Research*, *1640*, 222–231. https://doi.org/10.1016/j.brainres.2016.02.025

Lefebvre, C., Vachon, F., Grimault, S., Thibault, J., Guimond, S., Peretz, I., Zatorre, R. J., & Jolicœur, P. (2013). Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. *Neuropsychologia*, *51*(13), 2939–2952. https://doi.org/10.1016/j.neuropsychologia.2013.08.003

León-Domínguez, U., Martín-Rodríguez, J. F., & León-Carrión, J. (2015). Executive n-back tasks for the neuropsychological assessment of working memory. *Behavioural Brain Research*, *292*, 167–173. https://doi.org/10.1016/j.bbr.2015.06.002

Luke, R., Larson, E., Shader, M. J., Innes-Brown, H., Van Yper, L., Lee, A. K. C., Sowman, P. F., & McAlpine, D. (2021). Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm. *Neurophotonics*, *8*(02). https://doi.org/10.1117/1.NPh.8.2.025008

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide, 2nd ed*. (p. xix, 492). Lawrence Erlbaum Associates Publishers.

Majerus, S., & Cowan, N. (2016). The nature of verbal short-term impairment in dyslexia: The importance of serial order. *Frontiers in Psychology*, *7*, 1522. https://doi.org/10.3389/fpsyg.2016.01522

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *The Journal of Open Source Software*, 3(22), 470. https://doi.org/10.21105/joss.00470

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, 4(40), 1541. https://doi.org/10.21105/joss.01541

Matsui, M., Tanaka, K., Yonezawa, M., & Kurachi, M. (2007). Activation of the prefrontal cortex during memory learning: Near-infrared spectroscopy study. *Psychiatry and Clinical Neurosciences*, 61(1), 31–38. https://doi.org/10.1111/j.1440-1819.2007.01607.x

Morey, R. D., & Rouder, J. N. (2022). *BayesFactor: Computation of Bayes Factors for Common Designs*. https://CRAN.R-project.org/package=BayesFactor

Neath, A. A., Flores, J. E., & Cavanaugh, J. E. (2018). Bayesian multiple comparisons and model selection. *WIREs Computational Statistics*, 10(2), e1420. https://wires.onlinelibrary.wiley.com/action/showCitFormats?doi=10.1002%2Fwics.1420

Nickisch, A., & Von Kries, R. (2009). Short-term memory (STM) constraints in children with specific language impairment (SLI): Are there differences between receptive and expressive SLI? *Journal of Speech, Language, and Hearing Research*, 52(3), 578–595. https://doi.org/10.1044/1092-4388(2008/07-0150)

Nithart, C., Demont, E., Majerus, S., Leybaert, J., Poncelet, M., & Metz-Lutz, M.-N. (2009). Reading disabilities in SLI and dyslexia result from distinct phonological impairments. *Developmental Neuropsychology*, 34(3), 296–311. https://doi.org/10.1080/87565640902801841

Nolden, S., Bermudez, P., Alunni-Menichini, K., Lefebvre, C., Grimault, S., & Jolicoeur, P. (2013). Electrophysiological correlates of the retention of tones differing in timbre in auditory short-term memory. *Neuropsychologia*, 51(13), 2740–2746. https://doi.org/10.1016/j.neuropsychologia.2013.09.010

Nyberg, L., Dahlin, E., Stigsdotter Neely, A., & Bäckman, L. (2009). Neural correlates of variable working memory load across adult age and skill: Dissociative patterns within the fronto-parietal network. *Scandinavian Journal of Psychology*, 50(1), 41–46. https://doi.org/10.1111/j.1467-9450.2008.00678.x

Owen, A. M. (2000). The role of the lateral frontal cortex in mnemonic processing: The contribution of functional neuroimaging. *Experimental Brain Research*, 133(1), 33–43. https://doi.org/10.1007/s002210000398

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1), 46–59. https://doi.org/10.1002/hbm.20131

Peretz, I. (2016). Neurobiology of congenital amusia. *Trends in Cognitive Sciences*, 20(11), 857–867. https://doi.org/10.1016/j.tics.2016.09.002

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The montreal battery of evaluation of amusia. *Annals of the New York Academy of Sciences*, 999(1), 58–75. https://doi.org/10.1196/annals.1284.006

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, 1464(1), 5–29. https://doi.org/10.1111/nyas.13948

Plichta, M. M., Gerdes, A. B. M., Alpers, G. W., Harnisch, W., Brill, S., Wieser, M. J., & Fallgatter, A. J. (2011). Auditory cortex activation is modulated by emotion: A functional near-infrared spectroscopy (fNIRS) study. *NeuroImage*, 55(3), 1200–1207. https://doi.org/10.1016/j.neuroimage.2011.01.011

Robitaille, N., Marois, R., Todd, J., Grimault, S., Cheyne, D., & Jolicoeur, P. (2010). Distinguishing between lateralized and nonlateralized brain activity associated with visual short-term memory: fMRI, MEG, and EEG evidence from the same observers. *NeuroImage*, 53(4), 1334–1345. https://doi.org/10.1016/j.neuroimage.2010.07.027

Rolls, E. T., Joliot, M., & Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, 122, 1–5. https://doi.org/10.1016/j.neuroimage.2015.07.075

Roodenrys, S., & Stokes, J. (2001). Serial recall and nonword repetition in reading disabled children. *Reading and Writing*, 14, 379–394. https://doi.org/10.1023/A:1011123406884

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, 60(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050

Rovetti, J., Goy, H., Nurgitz, R., & Russo, F. A. (2021). Comparing verbal working memory load in auditory and visual modalities using functional near-infrared spectroscopy. *Behavioural Brain Research*, 402, 113102. https://doi.org/10.1016/j.bbr.2020.113102

Rovetti, J., Goy, H., Pichora-Fuller, M. K., & Russo, F. A. (2019). Functional near-infrared spectroscopy as a measure of listening effort in older adults who use hearing aids. *Trends in Hearing*, 23, 233121651988672. https://doi.org/10.1177/2331216519886722

Sailer, M. O. (2022). *crossdes: Construction of Crossover Designs*. https://CRAN.R-project.org/package=crossdes

Santosa, H., Hong, M. J., & Hong, K.-S. (2014). Lateralization of music processing with noises in the auditory cortex: An fNIRS study. *Frontiers in Behavioral Neuroscience*, 8. https://doi.org/10.3389/fnbeh.2014.00418

Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS brain AnalyzIR toolbox. *Algorithms*, 11(5), 73. https://doi.org/10.3390/a11050073

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, 85, 6–27. https://doi.org/10.1016/j.neuroimage.2013.05.004

Schulze, K., Jay Dowling, W., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception*, 29(3), 255–267. https://doi.org/10.1525/mp.2012.29.3.255

Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, 1252(1), 229–236. https://doi.org/10.1111/j.1749-6632.2012.06447.x

Schulze, K., & Tillmann, B. (2013). Working memory for pitch, timbre, and words. *Memory*, 21(3), 377–395. https://doi.org/10.1080/09658211.2012.731070

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, 32(5), 771–783. https://doi.org/10.1002/hbm.21060

Sternberg, S. (1966). High-speed scanning in human memory. *Science (New York, N.Y.)*, *153*(3736), 652–654. https://doi.org/10.1126/science.153.3736.652

Talamini, F., Blain, S., Ginzburg, J., Houix, O., Bouchet, P., Grassi, M., Tillmann, B., & Caclin, A. (2022). Auditory and visual short-term memory: Influence of material type, contour, and musical expertise. *Psychological Research*. https://doi.org/10.1007/s00426-021-01519-0

Teichmann, L., Moerel, D., Baker, C., & Grootswagers, T. (2021). An empirically-driven guide on using Bayes factors for M/EEG decoding [Preprint]. Neuroscience. https://doi.org/10.1101/2021.06.23.449663

Tian, F., Yennu, A., Smith-Osborne, A., Gonzalez-Lima, F., North, C. S., & Liu, H. (2014). Prefrontal responses to digit span memory phases in patients with post-traumatic stress disorder (PTSD): A functional near infrared spectroscopy study. *NeuroImage: Clinical*, *4*, 808–819. https://doi.org/10.1016/j.nicl.2014.05.005

Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital amusias. In *Handbook of clinical neurology* (Vol. *129*, pp. 589–605). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00033-0

Tillmann, B., Graves, J. E., Talamini, F., Leveque, Y., Fornoni, L., Hoarau, C., Pralus, A., Ginzburg, J., Albouy, P., & Caclin, A. (2023). Auditory cortex and beyond: Deficits in congenital amusia. *Hearing Research*, *437*, 108855. https://doi.org/10.1016/j.heares.2023.108855

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition*, *71*(3), 259–264. https://doi.org/10.1016/j.bandc.2009.08.003

Tseng, Y.-L., Lu, C.-F., Wu, S.-M., Shimada, S., Huang, T., & Lu, G.-Y. (2018). A functional near-infrared spectroscopy study of state anxiety and auditory working memory load. *Frontiers in Human Neuroscience*, *12*, 313. https://doi.org/10.3389/fnhum.2018.00313

Van Den Bergh, D., Wagenmakers, E.-J., & Aust, F. (2022). Bayesian repeated-measures ANOVA: An updated methodology implemented in JASP [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fb8zn

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory & Cognition*, *38*(2), 163–175. https://doi.org/10.3758/MC.38.2.163

Yamazaki, H., Kanazawa, Y., & Omori, K. (2020). Advantages of double density alignment of fNIRSoptodes to evaluate cortical activities related to phonological short-term memory using NIRS-SPM. *Hearing Research*, *395*, 108024. https://doi.org/10.1016/j.heares.2020.108024

Ziegler, J. C., Pech-Georgel, C., George, F., & Foxton, J. M. (2012). Global and local pitch perception in children with developmental dyslexia. *Brain and Language*, *120*(3), 265–270. https://doi.org/10.1016/j.bandl.2011.12.002

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, *12*(5), 732–745. https://doi.org/10.1111/j.1467-7687.2009.00817.x

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2011). Noise on, voicing off: Speech perception deficits in children with specific language impairment. *Journal of Experimental Child Psychology*, *110*(3), 362–372. https://doi.org/10.1016/j.jecp.2011.05.001

Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). fNIRSOptodes' location decider (fOLD): A toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports*, *8*(1), 3341. https://doi.org/10.1038/s41598-018-21716-z