Received: 16 October 2020 | Revised: 20 October 2021 | Accepted: 22 October 2021

DOI: 10.1111/desc.13188

PAPER

Developmental Science WILEY

Development of auditory cognition in 5- to 10-year-old children: Focus on musical and verbal short-term memory

Jérémie Ginzburg[1,2] | Annie Moulin[1,2] | Lesly Fornoni[1,2] | Francesca Talamini[3] | Barbara Tillmann[1,2] | Anne Caclin[1,2]

[1] Lyon Neuroscience Research Center, UMR5292, INSERM, U1028, CNRS, Lyon, France

[2] University Lyon 1, Lyon, France

[3] Institute für Psychologie, Universität Innsbruck, Innsbruck, Austria

Correspondence
Jérémie Ginzburg, Lyon Neuroscience Research Center, CNRS, UMR5292, INSERM, U1028, Lyon, F-69000, France.
Email: jeremie.ginzburg@inserm.fr

Abstract

Developmental aspects of auditory cognition were investigated in 5-to-10-year-old children (n = 100). Musical and verbal short-term memory (STM) were assessed by means of delayed matching-to-sample tasks (DMST) (comparison of two four-item sequences separated by a silent retention delay), with two levels of difficulty. For musical and verbal materials, children's performance increased from 5 years to about 7 years of age, then remained stable up to 10 years of age, with performance remaining inferior to performance of young adults. Children and adults performed better with verbal material than with musical material. To investigate auditory cognition beyond STM, we assessed speech-in-noise perception with a four-alternative forced-choice task with two conditions of phonological difficulty and two levels of cocktail-party noise intensity. Partial correlations, factoring out the effect of age, showed a significant link between musical STM and speech-in-noise perception in the condition with increased noise intensity. Our findings reveal that auditory STM improves over development with a critical phase around 6–7 years of age, yet these abilities appear to be still immature at 10 years. Musical and verbal STM might in particular share procedural and serial order processes. Furthermore, musical STM and the ability to perceive relevant speech signals in cocktail-party noise might rely on shared cognitive resources, possibly related to pitch encoding. To the best of our knowledge, this is the first time that auditory STM is assessed with the same paradigm for musical and verbal material during childhood, providing perspectives regarding diagnosis and remediation in developmental learning disorders.

KEYWORDS
auditory perception, music, recognition task, speech, speech-in-noise, working memory

1 | INTRODUCTION

Auditory Short-Term Memory (STM) allows for encoding, storage, and retrieval of auditory information during a short amount of time (within several seconds, Baddeley & Hitch, 1974; Cowan, 2008).[1] Along with auditory scene analysis and auditory attention, it is a key component of central auditory processing that subtends auditory cognition and allows making sense of the ever-changing acoustic environment.

[1] Note that "Working memory" is sometimes used in a larger sense encompassing this STM definition. Here we will refer to working memory only when the paradigm entails a manipulation of information.

Numerous relationships have been described between central auditory processing disorders (CAPD) and learning disabilities in children (Goswami, 2011; Iliadou & Iakovides, 2003; Medwetsky, 2011; Moore et al., 2010). Recent research suggests a link between verbal STM deficits and learning disorders (Männel et al., 2015; Nithart et al., 2009; Perez et al., 2012) and a link between impaired speech-in-noise processing and learning disorders (Bradlow et al., 2003; Sperling et al., 2005; Ziegler et al., 2009, 2011). Specifically, verbal STM impairment has been observed in dyslexic children with reduced digit spans or poor non-word repetition in recall tasks (Adlard & Hazan, 1998; Kramer et al., 2000; Majerus & Cowan, 2016; Nithart et al., 2009; Plaza et al., 2002; Roodenrys & Stokes, 2001; Tijms, 2004). The link between impaired verbal STM or working memory (WM) and learning disorders has also been reported for specific language impairment (SLI, Nithart et al., 2009) and dyscalculia (Attout & Majerus, 2015). It appears that both verbal STM and speech-in-noise perception are consistently reported to be impaired in learning disorders and CAPD (Moore et al., 2010; Perez et al., 2012; Ziegler et al., 2011).

Similarly to verbal STM deficits observed in several language-related disorders (Adlard & Hazan, 1998; Kramer et al., 2000; Majerus & Cowan, 2016; Nithart et al., 2009; Plaza et al., 2002; Roodenrys & Stokes, 2001; Tijms, 2004), musical STM deficits have also been reported in children with neurodevelopmental learning disorders. Dyslexic children display lower performance than typically-developing children in recognition STM tasks for pitch (Ziegler et al., 2012) and in tonal recognition tasks from the Primary Measure of Music Audiation (PMMA, Gordon, 1986) as shown by Atterbury (1985) and Forgeard et al. (2008). Furthermore, a sizeable comorbidity between dyslexia and congenital amusia has been observed in adults and children (Couvignou & Kolinsky, 2021; Couvignou et al., 2019). Congenital amusia is characterized by a deficit in music processing and in particular musical STM (Tillmann et al., 2009, 2016). These findings stress the importance of characterizing the development of central auditory processing in typically developing children, in particular to improve diagnosis and rehabilitation of central auditory processing deficits associated with learning disorders.

It has been well documented that verbal short-term storage capacity increases during childhood (Alloway et al., 2006; Chuah & Maybery, 1999; Cowan et al., 1999; Dempster, 1981; Gathercole, 1999; Gathercole et al., 2004; Orsini et al., 1987). Gathercole (1999) considers age as the most powerful factor influencing verbal STM capacity. These verbal STM abilities, tested with digit, word, and non-word spans (i.e., recall tasks), increase linearly with age between 4 and 14 years of age and appear to level-off by the age of 14–15 years (Gathercole et al., 2004). The development of STM over childhood was thus mostly investigated using serial recall paradigms that are reliant on verbal production, hence specific to verbal and/or phonological material (Gathercole, 1999). Only a few studies have investigated children's STM for other types of auditory information (e.g., pitch, timbre, or rhythm and temporal processing), but not yet over development nor in comparison to verbal material. Reviewing previous research reveals that only few studies investigated musical STM over children development. Pitch memory for single tones arise as early as 6 months of age (Plantinga & Trainor, 2008). Later in development, STM for tone sequences appears to mature from early childhood (6 years old) to pre-teenage years (13 years old, Clark et al., 2018). Indeed, Clark et al. (2018) found a developmental increase in the memory capacity for single tones as well as tone sequences similar to the developmental trajectory of visuo-spatial memory in children from 6- to 13-years of age, using an adapted part-set cueing task. In addition to an increase of capacity, Keller and Cowan (1994) found a decrease in STM trace decay for pitch over time in children from 4- to-12 years old using a two-tone comparison task. This latter study found an increase in the persistence of memory for pitch between the ages of 6 to 7 years. To our knowledge, no study has compared directly the precise development of musical and verbal STM. Our study aimed at doing so systematically in children from 5- to 10-years old by using a paradigm that allows for direct comparison between musical and verbal STM.

RESEARCH HIGHLIGHTS

- Auditory short-term memory was assessed in 5- to 10-year-old children and young adults using musical and verbal materials in a delayed matching-to-sample task.
- Musical and verbal short-term memory shared a similar developmental trajectory and are still under development at 10 years of age.
- Correlations with speech perception in cocktail-party noise suggest shared cognitive resources between musical short-term-memory and speech in cocktail-party noise perception capacities.
- Testing both musical and verbal short-term memory provides perspectives for diagnosis and training in developmental learning disorders.

Musical and verbal STM are difficult to compare if standard recall paradigms (most frequently used to evaluate verbal STM) are used. Even if some studies compared musical and verbal STM using mixed recall/reconstruction paradigms (Gorin et al., 2016, 2018; Williamson et al., 2010), they nonetheless required production processes that are difficult to adapt for children. The classical delayed matching-to-sample task (DMST) allows us to circumvent the need of oral or motor production that is required in a recall paradigm. In a DMST, participants have to memorize a first (S1) sequence of sounds (or an isolated sound). After a delay, a second stimulus is presented (S2) and participants have to report whether S1 and S2 are identical or different. This task has the advantage of allowing for the use of different kinds of materials (verbal, musical, environmental sound…, e.g., Talamini et al., 2021) and entails the three memorization steps of encoding (during S1), retention (during the delay), and retrieval (during S2) without relying on a production phase (as required in recall tasks). Hence, the DMST appears to be a well-suited paradigm to assess the development of auditory STM (for both musical and verbal materials) in children.

In adults, it has been suggested that auditory STM, rather than being a unitary phenomenon, could be based on partly separate subsystems for different types of material, in particular musical and verbal information (for a review, see Caclin & Tillmann, 2018). Berz (1995) and Pechmann and Mohr (1992) proposed a musical/tonal loop that would account for a storage component specific to the representational features of tonal information, based on Baddeley and Hitch (1974) multicomponent model of WM. Ockelford (2007), based on the same multicomponent model, has also suggested to add a musical central executive component that would entail attentional processes specific to musical information. These models might predict different maturation patterns for each material type, leading to domain-specific patterns of developmental trajectory (i.e., different patterns of recency effects and absence of correlations between musical and verbal STM). In another line of research, other models have postulated more general attentional processes involved in the maintaining of information in a short-term storage and more specific item-related processes concerning the encoding of information (Barrouillet & Camos, 2007; Cowan, 1998). These latter models would predict a greater involvement of domain-general attentional processes in STM for both materials and would probably predict a similar developmental trajectory for musical and verbal STM as well as similitudes of domain-general processes between them (similar recency effects and correlations between both materials).

In addition, to further our understanding of the distinct mechanisms between musical and verbal STM, our study aimed at testing the development of auditory cognition beyond STM. It has been proposed that STM and speech-in-noise perception might share encoding processes (Murphy et al., 2000; Pichora-Fuller et al., 1995; Sarampalis et al., 2009), and both are deficient in learning disorders (Perez et al., 2012; Ziegler et al., 2009). In order to confirm that STM and speech-in-noise perception share encoding processes and to shed light on material-specific perceptual processes shared by auditory STM and speech-in-noise, we compared musical and verbal STM with speech-in-noise perception abilities. The two types of WM models described in the previous paragraph would both predict different encoding processes for musical and verbal material. Consequently, if STM and speech-in-noise share encoding processes, we predict differential links between musical and verbal STM, and speech-in-noise (e.g., only musical STM is linked to speech-in-noise perception or only verbal STM). If only musical STM shares processes with speech-in-noise, these results would be in line with the already observed reliance of sound segregation on pitch processing (Oxenham, 2008) and bring evidence for the observed musician-advantage for speech-in-noise processing (Chandrasekaran & Kraus, 2010; Parbery-Clark et al., 2009; Strait et al., 2012; Zendel et al., 2015). If, however, only verbal STM is linked to speech-in-noise, the evidence would be in favor of top-down lexical influences or fine temporal structure encoding shared by speech-in-noise perception and verbal STM (Zekveld et al., 2013).

The aims of the present work are thus threefold. First, this study aims at describing the developmental trajectory of musical and verbal STM as, until now, only the trajectory of verbal STM using recall tasks has been described. The use of a DMST provides the potential to investigate specific patterns of development that could be task-related and/or material-specific, as it allows us to directly compare musical and verbal STM. Second, the present study aims at bringing insights about the shared and distinct mechanisms between musical and verbal STM, as domain-general and domain-specific processes have been described for musical and verbal STM in adults (Gorin et al., 2016, 2018). Third, in order to go beyond auditory STM and to scrutinize domain-specific processes in auditory STM, dependences between auditory STM and speech-in-noise perception were explored using cocktail party noise. As speech-in-noise and auditory STM would share encoding processes (Murphy et al., 2000; Pichora-Fuller et al., 1995; Sarampalis et al., 2009), finding out if both materials are linked to speech-in-noise or only one of them would bring insights into the existence of shared (e.g., both materials linked to speech-in-noise) or distinct (e.g., one of the materials linked to speech-in-noise) encoding processes.

In the present study, we created a child-adapted DMST with four-tone sequences for the musical material and with sequences of four consonant-vowel syllables for the verbal material. We implemented two levels of difficulty and tested 100 children ranging from 5 to 10 years of age (from kindergarten to fifth grade). The same 100 children were tested with a four-alternative forced-choice speech-in-noise perception task.

2 | METHODS

2.1 | Participants

One hundred children (mean age = 7.6 years old; min = 54-month-old, max = 127 month-old, six left-handed), attending a public primary school in South-East of France, participated in this study. The children were tested during school hours. Participation in the study was proposed to a total of 114 children and they were included in the study only if both parents or legal tutors provided written informed consent. The study was approved by the relevant services of the French public education services (Inspection de l'Education Nationale [IEN] and Direction Académique des Services de l'Education Nationale [DASEN] of the Isere department). Children from Kindergarten (n = 12), first (n = 17), second (n = 18), third (n = 19), fourth (n = 16), and fifth (n = 18) grade underwent the experiment (Table 1). Among the 100 children, parents' responses to questionnaires revealed that five children had a diagnosed learning disability (dyslexia, dysphasia, dyscalculia, dysorthographia, or dysgraphia), 30 had seen at least once a speech-therapist and six had already worn grommets. Nine out of the hundred children had specific musical training for more than 2 months (mean = 2.01 years, SD = 1.29), other children had never had any musical training apart from normal school curriculum. As we aimed here to explore the cognitive abilities of children among a representative set of the population, we present the results including all children.

Twelve adults (mean age = 26.5 years, SD = 9.6, one left-handed) were also included in the study. None reported any neurological or psychiatric troubles. Six of them had had a few years of musical education (mean = 4 years, SD = 2.8) but none of them were practicing any

TABLE 1 Number of participants and mean age (SD in parentheses) for each grade and matching English and French label for educational level

English label	Kindergarten (KG)	First grade	Second grade	Third grade	Fourth grade	Fifth grade
French label	Grande Section de Maternelle (GS)	Cours préparatoire (CP)	Cours élémentaire 1 (CE1)	Cours élémentaire 2 (CE2)	Cours Moyen 1 (CM1)	Cours Moyen 2 (CM2)
N	12	17	18	19	16	18
Mean age in years (SD)	5.01 (0.31)	6.03 (0.29)	6.93 (0.30)	8.01 (0.34)	8.92 (0.26)	9.94 (0.29)

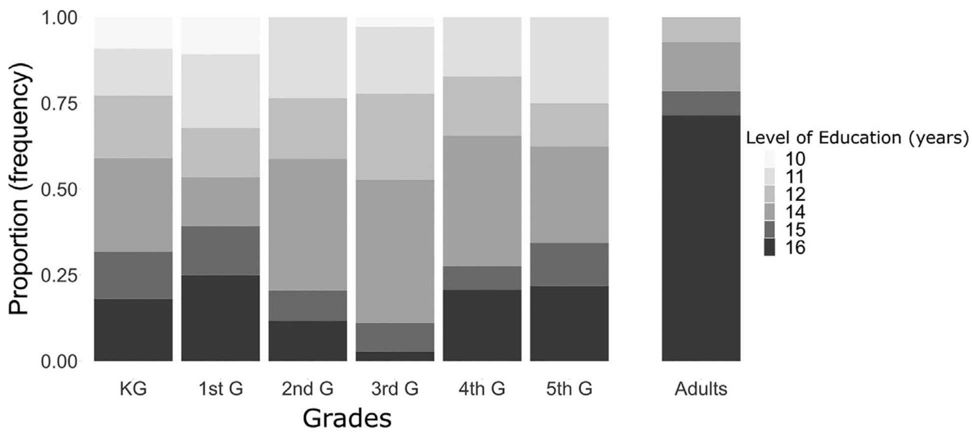

FIGURE 1 Distribution of children's parents and adult participants' level of education. Level of education categories correspond to the number of years spent in scholar institutions, from first grade on. A Pearson's Chi-squared test between all children's grades and the number of parents per category of level of education revealed that the proportion of parental level of education did not significantly differ by grade X^2 (25, $N = 181$) = 25.97, $p = .41$. A Chi-squared was also performed between adult's distribution of level of education and the mean proportion of children's parental level of education. The proportion of parental level of education differed by group (parents/adults) X^2 (5, $N = 195$) = 25.6, $p = .00011$[2]

instrument at the time of the experiment and this for at least the last 10 years. Their level of education, along with children's parental level of education, are shown in Figure 1.

2.2 | Stimuli construction and task design

2.2.1 | Short-term memory task

As shown in Figure 2, each trial of the STM recognition task consisted in listening to a four-item auditory sequence (S1), then after a silent retention delay of 2000 ms, to another four-item sequence (S2) that could be identical or different. When S2 was different, a new item could appear equiprobably at the 2nd, 3rd, or 4th position of the sequence. Each item lasted 500 ms, the silent inter-stimulus interval (ISI) between two items lasted 100 ms so overall there was a 600 ms stimulus onset asynchrony (SOA), leading to a duration of 6600 ms for S1 and S2 sequences. Children were given unlimited time to give a response. The next trial started after a 1500 ms delay. Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) was used to present stimuli and record responses in the STM task.

[2] We aimed at reaching homogeneity between children's parental level of education and adult participants' level of education but recruitment of adult participants fell during the Covid-19 pandemic, preventing us to fully achieve that goal. In the sample tested, all adult participants had very high level of performance irrespective of their education level.

Based on previous data-points in the literature (Gathercole et al., 2001; Jarrold et al., 2009; Majerus et al., 2006, 2007) we found that the use of four-item sequences was an optimal trade-off for younger children to be able to do the task and for older children not being at ceiling.

There were two material conditions for the STM task: musical piano tones (music condition) or syllables (verbal condition).

For the musical STM task, six musical tones (piano timbre, Cubase database, Steinberg) belonging to the C major scale were used (C2, D2, E2, F2, G2, A2) with frequencies ranging from 131 to 220 Hz (thus encompassing the fundamental frequency range of the vowel recordings: 202–212 Hz). A total of 48 four-item sequences were generated, all tones were different within a given sequence and they all contained at least one ascending interval and one descending interval (to avoid simple, constantly rising or falling patterns that could facilitate memorization). Twenty-four S1 sequences were used as S2 sequences for the "same" trials, half for the easy condition and half for the difficult one. There was no difference between the difficulty conditions for the "same" trials. Twenty-four S2 sequences were generated for the "different" trials, which can be of two types depending on whether the change violated contour or not. Previous research has shown that a contour-violation leads to better performance in melody discrimination tasks than contour preservation (Dyson & Watkins, 1984; Monahan et al., 1987; Peretz & Babaï, 1992; Ziegler et al., 2012). For the easy condition, twelve S2 sequences were generated with the new item in S2 changing the contour of the sequence (the contour is the

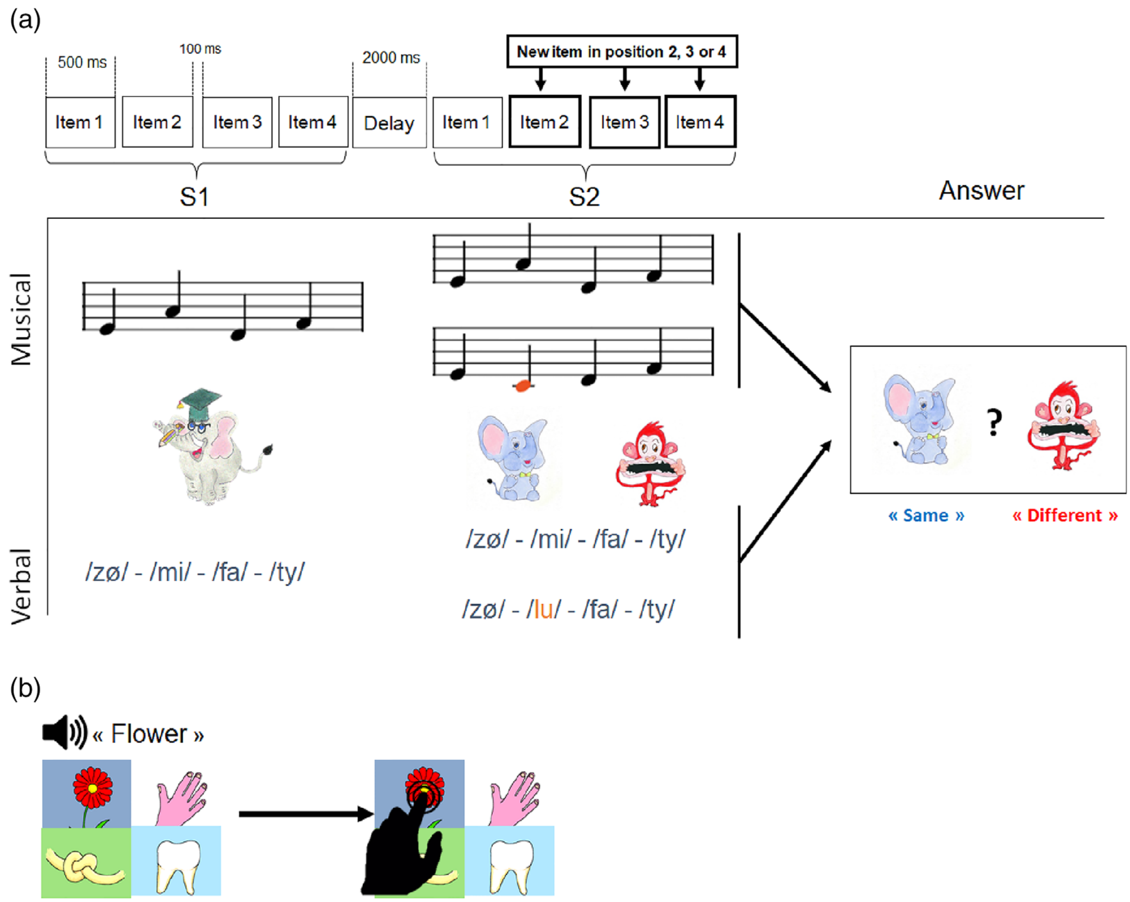

FIGURE 2 (a) DMST paradigm. In each trial, the child hears the first sequence S1 and after a 2000 ms delay, the second sequence S2. In a block, in half of the trials, the S2 sequence was identical to S1, in the other half S2 was different. A new item could equiprobably appear in the second, third, or fourth position of S2. We represent here only the easy condition (Supplementary material, Sounds 1 & 3). In the difficult musical condition, the new item in the S2 sequence did not violate the contour of the melody (Supplementary material, Sound 2); in the difficult verbal condition, the new item in the S2 sequence changed only by the vowel (Supplementary material, Sound 4). Visual stimuli were part of the playful story children were told to understand the task, and were inspired by melodic and rhythmic discrimination task from Ireland et al. (2018) and Wieland et al. (2015). (b) Speech-in-noise paradigm. Children heard a word and four pictures appeared simultaneously on the touchpad. They had to tap on the corresponding picture as fast as possible. A cocktail-party noise was presented at either −3 or ± 3 dB SNR in the background and two conditions of phonological proximity were possible between the words corresponding to the four images: phonologically distant (e.g., fleur-main-noeud-dent, \flœʁ\ \mɛ̃\ \nø\ \dɑ̃\, as shown in the figure) or phonologically close (fleur-coeur-beurre-l'heure, \flœʁ\ \kœʁ\ \bœʁ\ \lœʁ\)

up-and-down scheme of a melody,). So, if S1 had an up-down-up contour (e.g., E-A-D-F), S2 could have a down-up-up contour (e.g., E-C-D-F, Supplementary Material, Sound 1). In the difficult condition, 12 S2 sequences were generated, the new item in S2 did not change the contour (Supplementary Materials, Sound 2). It should be noted that here, greater difficulty of the musical memory task is induced by not only the absence of contour change but also by smaller pitch differences between the new item and the original one in the difficult condition compared to the easy condition.

For the verbal STM task, the items were Consonant-Vowel syllables. To avoid difficulties due to children's phonological skills, the consonant-vowel stimuli were selected to show the greatest perceptual distance with each other, within a S1 (or S2) sequence Six consonants and six vowels were thus selected: /f/ /t/ /z/ /g/ /m/ /l/ and /i/ /e/ /a/ /y/ /ø/ /u/ resulting into 36 syllables that were then recorded by a professional mezzo-soprano singer (for details about syllables construction, see Figure S1). A total of 48 four-item sequences were generated to be used as S1. All vowels and all consonants were different in a given sequence. Twenty-four S1 sequences were used as S2 sequences for the "same" trials, half for the easy condition and half for the difficult one. For the other 24 S1 sequences, 24 S2 sequences were created for the "different" trials. For the easy condition, 12 S2 sequences were created with the new item in S2 differing from the item in S1 by both consonant and vowel (e.g., /lu/ instead of /mi/, Supplementary Material, Sound 3). For the difficult condition, 12 S2 sequences were created with the new item in S2 differing by the vowel only (e.g., /lu/ instead of /la/, Supplementary Material, Sound 4). When S2 was different, no substitution was made between /i/, /e/, and /y/ because of their shorter distance on the vowel triangle compared to other possible substitutions.

2.2.2 | Speech-in-noise task

The speech-in-noise task was specifically designed for children. It was a French language adaptation (Moulin et al., 2013) of Foster and Haggard (1987)'s British Four Alternative Auditory Feature test. The adaptation for children (Bourgeois–Vionnet et al., 2020; Ginzburg et al., 2019) used pictures instead of written words and was implemented on a touch-tablet (iPad). For each trial, children had to match an aurally-presented word with its corresponding image among four pictures, by tapping on the corresponding picture on the tablet. The material consisted of 24 spoken words selected as a function of their concreteness, their frequency of occurrence in the French language and their age of acquisition (New et al., 2004). The 24 selected words were recorded several times by a French-native female speaker in order to have at least nine different sound exemplars of each word. All the exemplars were equalized in RMS amplitude and two listeners chose independently the best sounding exemplars for each word that was included in the final test. From this list of 24 words, six four-images arrays were created so that each array contained images for which the words denominating them were phonological neighbors (difficult condition, e.g., [flœʁ][bœːʀ][lœʁ][kœʁ]: fleur, beurre, l'heure, coeur – flower, butter, hour, heart). With the same words, six four-images arrays were created with phonologically distant words that had at least two different phonemes (easy condition, e.g., [flœʁ][mɛ̃][nø][dɑ̃]: fleur, main, nœud, dent – flower, hand, node, tooth). For each trial, an array of four-images was displayed on the touchpad's screen followed by the sound stimuli: a target word denominating one of the four images after a 600 ms delay. The child had to tap with its finger on the matching image as quickly as possible but without any time limit. The next trial was triggered by the child's click on the screen. Each word was presented twice: once in the difficult condition (phonological neighbors) and once in the easy condition (phonologically distant words) so that a total of 48 sets of four-images-array were presented. During the task, a continuous Cocktail-Party noise, made of 16 unintelligible French male and female voices, was presented binaurally via headphones. The target-word was systematically presented at 66 dB SPL. The Cocktail-party noise was presented at either 63 dB SPL (signal/noise ratio = +3 dB), or 69 dB SPL (signal/noise ratio = −3 dB).

2.3 | Procedure

Before testing, parents filled a questionnaire about their child's level of education and their own level of education, the child's laterality, possible vision or auditory impairments, musical activities, bilingualism, learning disabilities, and 11 questions adapted from adults' musical listening questionnaires (Lévêque et al., 2018; Tillmann et al., 2014).

Children were tested by groups of five or six in the gym of their school. Third to fifth graders were tested first and KG to second graders were tested a week after. Before testing, each child sat in front of a desk, listening to the experimenter's instructions. Before the STM tasks, the experimenter told a story about the elephant-professor and his two pupils, corresponding to the visual stimuli displayed on the computer screen during the task. These stimuli and the overall cover story for the child-friendly STM implementation were adapted from the melodic and rhythmic discrimination tasks of Ireland et al. (2018) and Wieland et al. (2015). After the six children were settled in front of their table, the experimenter sat in front of them and started telling the cover-story for the task instructions using cardboard panels on which the task visual stimuli were printed. During the first sequence, a cartoon picturing an elephant teacher would appear on the computer screen and during the second sequence, a cartoon of a nice baby-elephant would appear on the left and a cartoon of a grimacing baby-monkey would appear on the right. After the second sequence, a question mark would appear on the screen between the two cartoons. Children were given the instruction that the baby-elephant was always repeating correctly the sequence produced by the elephant-professor and that the grimacing-monkey was always repeating incorrectly. They had to give their answer by clicking on the left button of the laptop trackpad for the baby-elephant ("same") and on the right button for the grimacing-monkey ("different"). After the STM task, the experimenter explained the speech-in-noise task with cartoons printed on cardboard panels: children saw 4 images appear on the touchpad and at the same time, they heard a word through their headphones. They had to find the spoken word in one of the four images and tap on it as quickly as possible. They had to ignore the people talking in the background (cocktail-party noise). A training block of four trials was given first. The entire testing session lasted around 30 min, including instructions and breaks.

All children underwent:

-The STM recognition task with both material (verbal/musical) and difficulty levels (easy/difficult). To avoid having a testing session of more than 30 min, younger children (KG, first and second grade) were not tested with the music difficult block since they were slower than the older children and the results of older children tested during the first day revealed that this condition was the most difficult. Auditory stimuli were presented with AKG-142-HD headphones and visual stimuli via laptop computer screens.

For each material, the STM task was divided into four blocks of 24 trials (two materials and two conditions of difficulty). In each block, half of the trials were "same" trials (identical S1 and S2) and the other half were "different" trials. These 24 trials were pseudorandomly presented during a block, with the constraint that no more than three "same" or three "different" trials could appear consecutively. Children always began with the two easy blocks. Half of the children began with the verbal material and the other half with the musical material. Each of the four STM blocks lasted around 4 min and was preceded by a training with corresponding material and difficulty conditions. The training consisted of two "same" and two "different" trials with a smiley-shaped error feedback for each trial. At the end of the training blocks and the test blocks, a feedback specified the number of correct answers. For the test blocks, no trial-based error feedback was given. Overall, the STM tasks lasted around 20 min.

-The speech-in-noise test with two levels of noise (−3 dB SNR and +3 dB SNR) and two levels of phonological proximity (Bourgeois–Vionnet et al., 2020; Ginzburg et al., 2019). Auditory stimuli were displayed through Sennheiser HD-250-Pro headphones and the task was performed on touchpads. Children underwent a training comprising eight sets of images with a signal/noise ratio (SNR) of +3 dB, then a 48-trial block (24 words in each of the two phonological conditions, in pseudo-random order, so that the same word was not presented one after the other) with a SNR of +3 dB and finally a second 48-trial block with a SNR of −3 dB. Each block lasted around 3 min, so the speech-in-noise test lasted around 10 min overall with the instructions.

Adult participants were tested in the lab with the four STM tests, as for the older children.

2.4 | STM data preprocessing

To assess whether instructions were well understood, we computed the STM task's answers at inappropriate time points, the correct response would be any answer given after S2. Responses during the S2 sequence were coded as "anticipation", responses between the child's response and the next trial were coded as "responses between trials" and responses during S1 or the retention delay between S1 and S2 as "within-trial error ". Distribution of responses between trials and anticipations showed no particular information about a possible indicator of misunderstanding and/or problematic impulsivity of children during the task. However, within-trial errors displayed interesting information about children's ability to understand the task. A high rate of errors during the S1 sequence corresponded to the experimenters' observation of children's inability to understand the task correctly during the experimental session. On the account that they displayed more than five within-trial errors out of 24 trials during at least one block, we excluded eight children among the 100 from the analysis (two of them were in kindergarten, three were in first grade, one in second grade, one in fourth grade and one in fifth grade).

2.5 | Data analyses

In all analyses, the developmental aspects of children's performance[3] were tested as a function of children's school grades, allowing for homogeneous groups with children being equally scholarly educated regarding reading and other abilities.

For the STM task, measures of d' and bias (c parameter) were obtained according to Signal Detection Theory (SDT) for each material (verbal/musical), for each difficulty level (easy/difficult) and for each participant (Macmillan & Creelman, 1991). Hit corresponded to a correct answer for different trials. False alarm corresponded to an incorrect answer for same trials. *d'* and criterion (*c*) were calculated using Dominique Makowski's "dprime" R function from the *psycho* package (Makowski, 2018). *d'* or sensitivity, was calculated as the *z*-score of False Alarms subtracted from the *z*-score of Hits. The criterion, *c*, is 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 are made following the recommendation of Hautus (1995). Furthermore, we analyzed the response times of participants after the end of S2, on correct trials. Recency effects were examined by analyzing the percentage of correct responses as a function of the position of the new item in "different" trials (second, third, or fourth position).

For the speech-in-noise task, we obtained percentages of correct answers for each child, for each condition of phonological proximity (close/distant) and for each signal/noise ratio (−3 dB and +3 dB).

Statistical analyses were performed on R (3.5.5 version). Mixed-design ANOVAs were performed to analyze data from the STM tasks using the *rstatix* R package (Kassambara, 2020). Due to the unbalanced experimental design (younger children did not undergo the difficult musical STM task), three sets of analyses were performed to assess all aspects of the data (factors for each analysis are detailed in the results section). Greenhouse-Geisser correction was applied when the sphericity assumption was violated, assessed by Mauchly's test. Significant effects and interactions were analyzed using pairwise *t*-tests as post-hoc tests adjusted with the Tukey method. We also performed one-sampled *t*-tests comparing *c* values to 0, the value of an unbiased observer. Spearman correlations were performed between age and performance in the DMST and the speech-in-noise task. Spearman partial correlation analyses with age as control variable were performed between performance in the STM task and the speech-in-noise task using the *ppcor* R package (Kim, 2015).

Two sets of ANOVAs are presented in the results section: the first one concerns children of all six grades (KG to fifth grade) on the easy conditions of the STM task in order to examine the developmental trajectory of STM for both materials. The second one concerns older children (from third to fifth grade) and examines the differential effects of material and its interaction with the task difficulty. A third ANOVA has been performed with children of all six grades on the verbal conditions in order to assess difficulty effects and is provided as a supplementary figure (Figure S2 and Table S1).

3 | RESULTS

3.1 | Musical and verbal STM: Easy conditions

We computed two 6 × 2 mixed-design analyses of variance (ANOVA) for the easy conditions on *d'* (Figure 3) and *c* (Table 2) with Grade as a between-subjects factor (kindergarten to fifth grade) and with Material as within-subject factor (musical/verbal), as well as a 6 × 2 × 2

[3] it should be noted that from this point on, the term "performance" refers to the sensitivity measure (d') described in this section

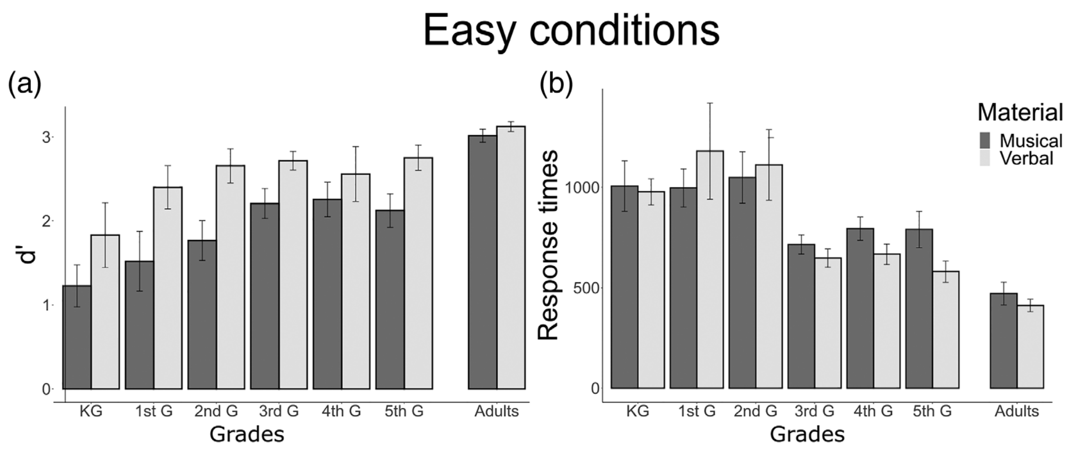

FIGURE 3 Performance for the easy conditions as a function of material (musical or verbal) for all children's grades (KG, kindergarten; G, grade) and for adults. (a) Mean and standard error of children and adult's sensitivity (calculated as the *d'*). (b) Mean and standard error of children and adult's correct response times (time in millisecond that subjects spent after the end of S2 before giving a "same" or "different" answer)

TABLE 2 Mean *c* values and standard deviation as a function of grade (columns) and condition (rows)

Condition	KG	First grade	Second grade	Third grade	Fourth grade	Fifth grade	Adults
Music easy	0.26 (0.62)	0.22 (0.30)	0.34* (0.33)	0.36* (0.29)	0.19* (0.19)	0.30* (0.26)	0.16 (0.19)
Music difficult	NA	NA	NA	0.69* (0.28)	0.70* (0.28)	0.55* (0.36)	0.54* (0.34)
Verbal easy	0.079 (0.63)	0.039 (0.14)	0.02 (0.25)	0.023 (0.28)	0.17 (0.21)	0.04 (0.21)	0.14* (0.13)
Verbal difficult	0.25 (0.62)	−0.11 (0.29)	0.0074 (0.27)	0.06 (0.26)	0.18* (0.20)	0.099 (0.20)	0.14* (0.13)

Note: As the mean value of the *c* parameter of an unbiased observer is 0, we performed one-sample *t*-tests (Bonferroni corrected per condition) comparing *c* mean values per grade and condition to 0. All mean *c* values (except first grades in the verbal difficult condition) were numerically higher than 0 (conservative bias) as is usually observed in DMST tasks. Asterisks indicate a *c* value significantly different from 0.

mixed-design ANOVA for response times analysis (Figure 3) with the aforementioned factors and with type of trial as within-factor (same/different). We also performed a 6 × 2 × 3 ANOVA on the percentage of correct responses in different trials to assess recency effects, with Grade as between-subject factor, and Material and Position (second, third, and fourth position) as within-subject factors. The position factor corresponded to the position of the item that changed during the S2 of different trials. Complete ANOVA results are shown in Table 3, and we present significant effects and interactions below.

3.1.1 | *d'*

The main effect of Grade was significant, $F(5,86) = 3.030$, $p = 0.014$, $\eta_p^2 = 0.257$, revealing better performance for older children. Post-hoc tests showed a significantly lower performance of kindergarten children compared to all other grades (all $p < 0.02$) except first and second grade (both $p > 0.13$). The main effect of Material was significant $F(1,86) = 32.33$, $p < 0.001$, $\eta_p^2 = 0.273$, showing better performance with the verbal material than with musical material. The interaction between grade and material was not significant ($p = 0.644$) even though the observation of the results suggests that third graders performed better for the musical material than second graders whereas for the verbal material second graders seemed to reach a similar performance level as did third graders (Figure 3).

3.1.2 | Correct response times

There were no significant effects in the response time analysis. We expected a grade effect which was marginally significant, $F(5,85) = 2.238$, $p = 0.058$, $\eta_p^2 = 0.116$.

3.1.3 | Bias

Comparison of the mean *c* parameter per grade and condition to zero showed that second, third, fourth, and fifth grade children displayed a positive response bias *c* that was significantly above zero in the easy musical condition (Table 2). The ANOVA revealed a significant effect of Material $F(1,86) = 40.89$, $p < 0.001$, $\eta_p^2 = 0.322$, with a higher *c* for the musical material than the verbal material.

TABLE 3 For the easy conditions, results of the repeated measures ANOVAs on each measure (d', c, response times, recency effects)

Measure	Factor(s)	df1	df2	*F*-value	*p*-value	η_p^2	*W*	ε
d'	**Grade**	**5**	**86**	**3.03**	**0.0144**	**0.257**	–	–
	Material	**1**	**86**	**32.329**	**1.75e – 07**	**0.273**	–	–
	Grade: Material	5	86	0.675	0.644	0.038	–	–
c	Grade	5	86	0.085	0.994	0.013	–	–
	Material	**1**	**86**	**40.890**	**8.05e - 09**	**0.322**	–	–
	Grade: Material	5	86	1.845	0.113	0.097	–	–
Response times	Grade	5	85	2.338	0.058	0.116	–	–
	Material	1	85	0.307	0.581	0.004	–	–
	Type of trial	1	85	0.907	0.344	0.011		
	Grade: Material	5	85	0.982	0.434	0.055	–	–
	Grade: Type of trial	5	85	1.554	0.182	0.084		
	Material: Type of trial	1	85	2.210	0.141	0.025		
	Grade: Material: Type of trial	5	85	1.675	0.149	0.090		
Recency effects	Position	2	261	0.682	0.507	0.008	0.953	–
	Grade	5	261	1.814	0.118	0.094	–	–
	Material	**1**	**261**	**54.850**	**7.84e – 11**	**0.387**	–	–
	Position: Grade	10	261	1.030	0.420	0.056	0.953	–
	Position: Material	2 (1.84)	261 (159.83)	1.551	0.217	0.018	0.911*	0.919
	Grade: Material	5	261	1.669	0.151	0.088	–	–
	Position: Grade: Material	10 (9.19)	261 (159.83)	0.872	0.553	0.048	0.911*	0.919

Note: For d' and c, grade was used as between-subject factor (six levels: KG, first, second, third, fourth, fifth grade) and material (two levels: verbal, musical) as within-subject factor. For the response times analysis, the type of trial was taken into account as within-subject factor (same/different). The analysis of recency effects was done on the percentage of correct responses for different trials with position (three levels: second, third, fourth position) and grade as between-subject factors and material as within-subject factor. Significant effects are in bold font. df, degrees of freedom; η_p^2, partial eta-squared; *W*, Mauchly's statistic. An asterisk indicates a significant *W*. If so, corrected degrees of freedom are reported in parenthesis in the df1 and df2 columns and the corresponding statistics are corrected with ε: Greenhouse-Geisser estimates of sphericity.

3.1.4 | Recency effects

Only the main effect of Material was significant, $F(1,87) = 54.850$, $p < 0.001$, $\eta_p^2 = 0.387$, with better performance for verbal than musical material, thus mirroring the *d'* analysis.

When children were gathered in three groups of age (KG-1st, 2nd-3rd, 4th-5th), the ANOVA on recency effect showed a main effect of Grade, albeit with a small effect size $F(2,89) = 3.591$, $p = 0.032$, $\eta_p^2 = 0.075$.

3.2 | Musical and verbal STM: Effect of difficulty and material

Two 3 × 2 × 2 mixed-design ANOVA were performed on *d'* (Figure 4) and *c* (Table 4) with Grade as a between-subjects factor (third to fifth grade), and Material (verbal/musical) and Difficulty (easy/difficult) as within-subject factors. A 3 × 2 × 2 × 2 mixed-design ANOVA was performed on response times (Figure 4) with the aforementioned factors and with type of trial as within-subject factor (same/different). A 3 × 2 × 2 × 3 mixed design ANOVA was performed on the percentage of correct response for different trials, adding the Position factor (second, third, and fourth position) as within-subject factor. These analyses were only possible with data from children between third and fifth grade ($n = 50$).

3.2.1 | d'

No significant effect of Grade was observed in these groups spanning a more restricted age range, $F(2,47) = 0.325$, $p = 0.7$, $\eta_p^2 = 0.038$. We observed a main effect of Material, $F(1,47) = 94.8$, $p < 0.001$, $\eta_p^2 = 0.787$, with lower performance for musical material than for verbal material. The main effect of Difficulty was significant, $F(1,47) = 31.24$, $p < 0.001$, $\eta_p^2 = 0.544$ and interacted with Material, $F(1,47) = 42.4$, $p < 0.001$, $\eta_p^2 = 0.474$: performance was lower in the difficult condition than in the easy condition for the musical material ($p < 0.001$), but not for the verbal material ($p = 0.96$).

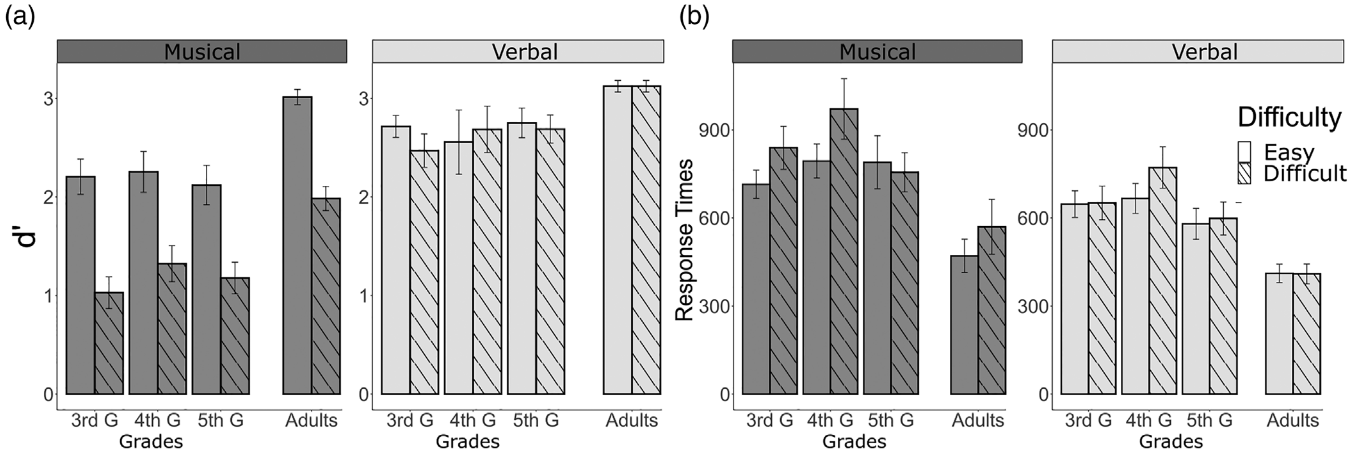

FIGURE 4 Performance for all conditions of Material (musical or verbal) and Difficulty (easy or difficult) for children of third, fourth, and fifth grades and for adults. (a) Mean and standard error of children and adult's sensitivity (calculated as the *d′*). (b) Mean and standard error of children and adult's response time (time in millisecond that subjects spent before giving a "same" or "different" answer)

TABLE 4 With older children, results of the repeated measures ANOVAs on each measure (*d′*, *c*, response times and recency effects)

Measure	Factor(s)	df1	df2	*F*-value	*p*-value	η_p^2	*W*	ε
d'	Grade	2	47	0.174	0.841	0.029	–	–
	Material	**1**	**47**	**94.803**	**7.52e –13**	**0.787**	–	–
	Difficulty	**1**	**47**	**31.237**	**1.12e – 06**	**0.544**	–	–
	Grade: Material	2	47	0.462	0.633	0.035	–	–
	Grade: Difficulty	2	47	0.872	0.425	0.062	–	–
	Material: Difficulty	**1**	**47**	**42.426**	**4.48e – 08**	0.474	–	–
	Grade: Material: Difficulty	2	47	0.125	0.883	0.005	–	–
c	Grade	2	47	0.616	0.545	0.043	–	–
	Material	**1**	**47**	**92.585**	**1.09e – 12**	0.734	–	–
	Difficulty	**1**	**47**	**40.558**	**7.45e – 08**	**0.432**	–	–
	Grade: Material	2	47	2.509	0.0922	0.130	–	–
	Grade: Difficulty	2	47	1.064	0.353	0.038	–	–
	Material: Difficulty	**1**	**47**	**23.118**	**1.6e – 05**	**0.330**	–	–
	Grade: Material: Difficulty	2	47	1.775	0.181	0.070	–	–
Response times	Grade	2	47	0.760	0.473	0.031	–	–
	Material	**1**	**47**	**27.73**	**3.4e – 06**	**0.371**	–	–
	Difficulty	1	47	3.381	0.072	0.067	–	–
	Type of trial	1	47	1.331	0.254	0.028		
	Grade: Material	2	47	0.314	0.732	0.013	–	–
	Grade: Difficulty	2	47	1.397	0.257	0.056	–	–
	Grade: Type of trial	2	47	0.087	0.916	0.004		
	Material: Difficulty	1	47	0.917	0.343	0.019	–	–
	Material: Type of trial	**1**	**47**	**7.874**	**0.007**	**0.143**		
	Difficulty: Type of trial	1	47	0.139	0.711	0.003		
	Grade: Material: Difficulty	2	47	0.168	0.32	0.047	–	–
	Grade: Material: Type of trial	2	47	0.853	0.433	0.035		
	Grade: Difficulty: Type of trial	2	47	1.611	0.211	0.064		

(Continues)

TABLE 4 (Continued)

Measure	Factor(s)	df1	df2	*F*-value	*p*-value	η_p^2	*W*	ε
	Material: Difficulty: Type of trial	1	47	0.051	0.822	0.001		
	Grade: Material: Difficulty: Type of trial	2	47	0.019	0.981	0		
Recency effects	**Position**	**2**	**144**	**32.381**	**1.79e – 11**	**0.403**	**0.922**	–
	Grade	2	144	0.355	0.703	0.015	–	–
	Material	**1**	**144**	**161.142**	**5.92e – 17**	**0.770**	–	–
	Difficulty	**1**	**144**	**73.882**	**2.82e – 11**	**0.606**	–	–
	Position: Grade	4	144	0.571	0.684	0.023	0.922	–
	Position: Material	**2 (1.68)**	**144 (80.68)**	**11.053**	**1.49e – 04**	**0.187**	**0.810***	**0.840**
	Position: Difficulty	**2**	**144**	**42.613**	**5.68e –14**	**0.470**	**0.995**	–
	Grade: Material	2	144	2.274	0.114	0.087	–	–
	Grade: Difficulty	2	144	0.435	0.65	0.018	–	–
	Material: Difficulty	**1**	**144**	**99**	**2.98 e –13**	**0.673**	–	–
	Position: Grade: Material	4 (3.36)	144 (80.68)	0.862	0.475	0.035	0.810*	0.840
	Position: Grade: Difficulty	4	144	1.081	0.37	0.043	0.880	–
	Position: Material: Difficulty	**2**	**144**	**38.196**	**6.26e – 13**	**0.443**	**0.960**	–
	Grade: Material: Difficulty	2	144	1.494	0.235	0.059	–	–
	Position: Grade: Material: Difficulty	4	144	0.882	0.478	0.035	0.960	–

Note: For d' and *c, grade was used as between-subject factor (three levels: third, fourth, fifth grade) and material (two levels: Verbal, musical) and difficulty (two levels: Easy, difficult) as within-subject factors.* For the response times analysis, the type of trial was taken into account as within-subject factor (same/different). The analysis of recency effects was done on the percentage of correct responses for different trials with position (three levels: second, third, fourth position) and grade as between-subject factors and material and difficulty as within-subject factors. Significant effects are in bold font. df, degrees of freedom; η_p^2, partial eta-squared; *W*, Mauchly's statistic. An asterisk indicates a significant *W*. If so, corrected degrees of freedom are reported in parenthesis in the df1 and df2 columns and the corresponding statistics are corrected with ε: Greenhouse-Geisser estimates of sphericity.

3.2.2 | Correct response times

We found a significant effect of Material $F(1,47) = 22.373$, $p < 0.001$, $\eta_p^2 = 0.428$ with longer response times for the musical material than for the verbal material. The material effect interacted significantly with the type of trial factor $F(1,47) = 7.874$, $p = 0.007$, $\eta_p^2 = 0.143$ with a difference between musical and verbal materials only for different trials ($p < 0.001$).

3.2.3 | Bias

Comparison of the mean *c* parameter per grade and condition to zero showed that all children from third to fifth grade displayed positive response bias *c* that was significantly above 0 in the difficult musical condition (Table 2). The ANOVA revealed a significant effect of material $F(1,47) = 92.585$, $p < 0.001$, $\eta_p^2 = 0.734$ with a higher positive *c* for the musical material than the verbal material. The main effect of difficulty was significant, $F(1,47) = 40.558$, $p < 0.001$, $\eta_p^2 = 0.432$, children displaying a higher *c* for the difficult condition. The material and difficulty interaction was significant $F(1,47) = 23.118$, $p < 0.001$, $\eta_p^2 = 0.330$. Post-hoc tests revealed a significantly higher *c* in the difficult condition compared to the easy condition, only for the musical material ($p < 0.001$) but not for the verbal material ($p = 0.46$).

3.2.4 | Recency effects

We found a Position effect $F(2,144) = 32.381$ $p < 0.001$, $\eta_p^2 = 0.403$ with performance on the second position lower than for third position ($p = 0.035$) and fourth position ($p < 0.001$) and third position lower than fourth ($p = 0.028$). The Position, Material, and Difficulty interaction $F(2,144) = 38.196$, $p < 0.001$, $\eta_p^2 = 0.442$ showed a lower performance on the second position compared to the third and fourth and between third and fourth only for the difficult musical material ($p < 0.001$).

3.3 | Partial correlations between STM and speech-in-noise performance

For the speech-in-noise task, as expected, effects of age, phonological proximity, and noise intensity were observed (Ginzburg et al., 2019). The correlation analysis shown here-after include 87 children: data from five children were excluded because of technical difficulties when recording speech-in-noise data.

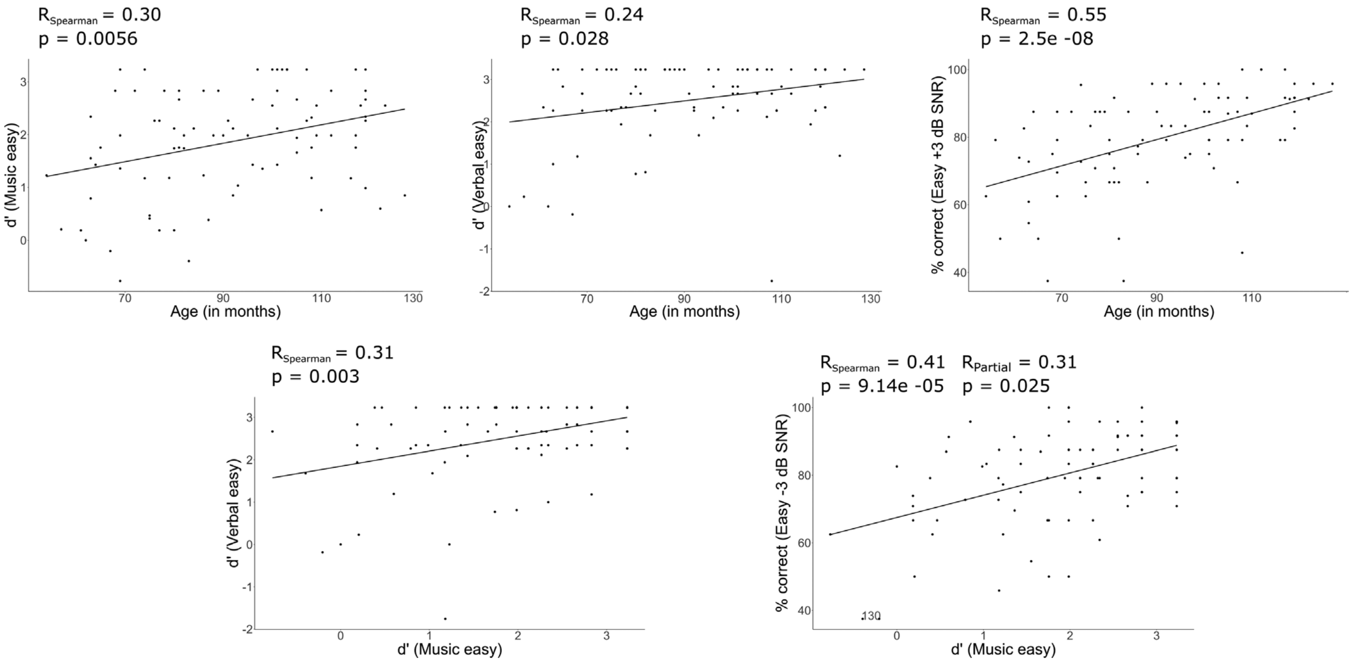

FIGURE 5 Top panels: Spearman correlation between children's age (in number of months) and the performance of children for the STM task (calculated as the *d'*) in the musical easy condition (top left), the verbal easy condition (top middle) and the performance of children in the speech-in-noise task (in percentage of correct response) for the easy condition and a −3 dB SNR (top right). Lower left panel: correlation between the musical easy condition and the verbal easy condition in the STM task. Lower right panel: partial correlation between the musical easy condition in the STM task and the speech-in-noise task in the easy phonological condition and −3 dB SNR. $R_{Spearman}$: Spearman correlation coefficient and its corresponding *p*-value. $R_{partial}$: partial correlation coefficient after factoring out the effect of age and its corresponding *p*-value. Partial correlations were Bonferroni corrected for eight comparisons (two STM conditions and four speech-in-noise conditions): the p-values were multiplied by eight. The lower right panel correlation was the only correlation between STM performance and Speech-in-Noise performance that showed significance when a partial correlation was computed with age as a control variable

We first performed correlations between musical and verbal STM performance (*d'*, in the easy condition only, as data for the difficult condition was not available for children from KG to second grade) and the age of children (in months). Both conditions were significantly correlated with age $r(85) = 0.30$, $p = 0.0056$ for the musical material and $r(85) = 0.24$, $p = 0.028$ for the verbal material and these two correlation coefficient were not significantly different ($p = 0.80$). We then performed partial correlation accounting for the age variable between musical and verbal STM performance (*d'*), in the easy condition only, as data for the difficult condition was not available for children from KG to second grade. Musical and verbal STM were significantly correlated even after factoring out the effect of age, $r(85) = 0.26$, $p = 0.014$. Next, we performed partial correlations between performance in the STM tasks (easy conditions) and in the speech-in-noise tasks accounting for the age variable. The analysis revealed that only the correlation between music STM performance with performance in the phonologically distant condition with a −3 dB SNR in the speech-in-noise task was significant, $r(85) = 0.32$, $p = 0.025$, after Bonferroni correction for multiple comparisons (Figure 5). No other significant correlations were observed between performance in STM tasks and speech-in-noise tasks when age was considered ($p > 0.53$).

3.4 | Effect of learning disorders or musical training

Five children, out of the 92 included in the analyses above, had various learning disorders diagnosis (dyslexia, dysphasia, dyscalculia, dysorthographia, dysgraphia), according to parental reports. No statistical analysis was conducted on their performance because of their small number. Nonetheless, for the STM task, we compared each child's performance in every condition to the corresponding median and their placement in the quartiles of performance of other children of the same grade and in the same condition. This analysis revealed that in the musical easy condition, all five children with learning disorders scored below the median. Among those five children, four of them performed in the first quartile. In the other conditions, at most two out of these five children performed above the median. All the ANOVAs described before were performed excluding these five children and, aside from a small decrease of statistical power, the effects remained consistent.

For the speech-in-noise task, we compared these children's performance in each condition of phonological difficulty (easy/hard) and noise intensity (SNR-3/SNR+3) to the median of the other children for the corresponding grade, phonological difficulty, and noise intensity. Noticeably in the easy condition (phonologically distant) with a +3 dB

SNR, all of the five children with learning disorders performed below the median, in the lowest quartile. In both difficulty conditions for the −3 dB SNR, at most one child performed above the median, but never the same child. In the difficult condition for +3 dB SNR, only two children performed below the median.

Furthermore, nine children reported musical training of at least 1 year (conservatory or communal music school). Given their small number, we did not perform any statistical analysis but when we removed them from the analysis, the effects remained consistent. Seven out of these nine children performed above the median and five out of these seven performed in the upper quartile in the music easy condition and in the verbal difficult condition. Five out of these nine children underwent the musical difficult condition (the others being too young) and all of them performed above the median for this condition. For the speech-in-noise task, a majority of children with musical training (at least seven out of nine in each condition) scored above the median in all conditions.

3.5 | Comparison between children and adults

To compare performance of the oldest children with performance of adults, three 2 × 2 × 2 mixed-design ANOVA on *d*′, RTs and *c* were performed with Group (adults [$n = 12$] and children [fifth graders, $n = 17$]) as between-subject factor, and Material (verbal/musical) and Difficulty (easy/difficult) as within-subject factors and type of trials as within-subject factor (same/different) for the response times analysis. A 2 × 2 × 2 × 3 mixed-design ANOVA was performed on the percentage of correct responses for different trials adding the Position factor (seconf, third, and fourth position) as within-subject factor. Complete results are shown in Table 5.

3.5.1 | d′

The main effect of Group was significant, $F(1,27) = 13.25$ $p = 0.00114$, $\eta_p{}^2 = 0.743$ with children having poorer performance than adults. A significant Group and Material interaction was found $F(1,27) = 4.981$, $p = 0.0341$, $\eta_p{}^2 = 0.269$. Post-hoc tests showed a significant difference between all comparisons ($p < 0.003$) except for a marginally significant difference between children and adults for the verbal material ($p = 0.061$). Adults displayed a significantly smaller difference between musical and verbal material than children ($p = 0.02$, *t*-test comparing the subtraction of musical and verbal performance between adults and children). All other effects mirrored the former *d*′ analysis on children's data.

3.5.2 | Correct response times

The main effect of Group was significant $F(1,27) = 4.825$, $p = 0.037$ $\eta_p{}^2 = 0.152$ with longer response times for children than for adults. The main effect of Material was significant $F(1,27) = 9.911$, $p = 0.004$ $\eta_p{}^2 = 0.269$ with longer response times for the musical material. There was also a significant interaction between Material and Type of trial $F(1,27) = 8.625$, $p = 0.007$, $\eta_p{}^2 = 0.242$ with a difference between musical and verbal material only in different trials ($p < 0.001$).

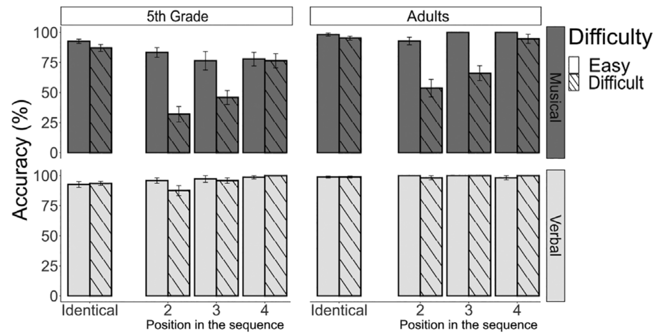

FIGURE 6 Recency effects for the older children (fifth grade, $n = 17$) and adults (% of correct responses for different trials). Results are presented as a function of the material musical/verbal), difficulty (easy/difficult) and the position of the item change in the S2 sequence (second, third, or fourth position). Percentage of correct response for identical S1-S2 are also represented for comparison. These effects were similar for the other younger children

3.5.3 | Bias

Comparison of the mean *c* parameter per condition to 0 showed that adults displayed a positive bias parameter *c* for all conditions and all of them except the musical easy condition were significantly different from zero (Table 2). No effect of group was found in the ANOVA.

3.5.4 | Recency effects

We observed an effect of the Group factor $F(1,84) = 10.99$, $p = 0.00254$, $\eta_p{}^2 = 0.177$ with lower performance for children. None of the other significant effects or interactions involved the Group or the Position factor (Figure 6).

We emphasize the fact that ceiling performance (e.g., 100% correct responses) was observed for the majority of adults in the easy and difficult conditions for the verbal material and for the easy musical condition. However less than half of fifth graders reached ceiling in all conditions.

4 | DISCUSSION

The aim of this study was to investigate the development of auditory STM in 5 to 10-year-old children for musical and verbal material. A DMST was created in a child-friendly version in highly comparable ways for both materials, each one with two levels of difficulty. Results showed that overall, younger children (KG and first grade, see Table 1

TABLE 5 Comparison of children and adult data, results of the repeated measures ANOVAs on each measure (*d′*, c, response times, recency effects)

Measure	Factor(s)	df1	df2	*F*-value	*p*-value	η_p^2	*W*	ε
d′	**Group**	**1**	**27**	**13.25**	**0.00114**	**0.743**	–	–
	Material	**1**	**27**	**89.526**	**4.59e – 10**	**0.869**	–	–
	Difficulty	**1**	**27**	**78.887**	**1.7e – 09**	**0.693**	–	–
	Group: Material	**1**	**27**	**4.981**	**0.0341**	**0.269**	–	–
	Group: Difficulty	1	27	0.142	0.709	0.004	–	–
	Material: Difficulty	**1**	**27**	**52.691**	**8.3e – 08**	**0.661**	–	–
	Group: Material: Difficulty	1	27	0.631	0.434	0.023	–	–
c	Group	1	27	0.003	0.957	0	–	–
	Material	**1**	**27**	**29.615**	**9.31e – 06**	**0.719**	–	–
	Difficulty	**1**	**27**	**28.635**	**1.19e – 05**	**0.454**	–	–
	Group: Material	1	27	1.752	0.197	0.131	–	–
	Group: Difficulty	1	27	0.322	0.575	0.009	–	–
	Material: Difficulty	**1**	**27**	**14.159**	**0.000826**	**0.344**	–	–
	Group: Material: Difficulty	1	27	1.752	0.197	0.061	–	–
Response times	**Group**	**1**	**27**	**4.825**	**0.037**	**0.152**	–	–
	Material	**1**	**27**	**9.911**	**0.004**	**0.269**	–	–
	Difficulty	1	27	0.196	0.661	0.007	–	–
	Type of trial	1	27	0.874	0.358	0.031	–	–
	Group: Material	1	27	0.618	0.439	0.022	–	–
	Group: Difficulty	1	27	0.383	0.541	0.014	–	–
	Group: Type of trial	1	27	0.120	0.731	0.004	–	–
	Material: Difficulty	1	27	0.130	0.721	0.005	–	–
	Material: Type of trial	**1**	**27**	**8.625**	**0.007**	**0.242**	–	–
	Difficulty: Type of trial	1	27	0.002	0.969	0	–	–
	Group: Material: Difficulty	1	27	1.316	0.261	0.046	–	–
	Group: Material: Type of trial	1	27	0.086	0.772	0.003	–	–
	Group: Difficulty: Type of trial	1	27	2.019	0.167	0.07	–	–
	Material: Difficulty: Type of trial	1	27	0.670	0.420	0.024	–	–
	Group: Material: Difficulty: Type of trial	1	27	0.137	0.714	0.005	–	–
Recency effects	**Position**	**2**	**84**	**25.711**	**8.87e – 11**	**0.187**	**0.862**	–
	Group	**1**	**84**	**10.99**	**0.00254**	**0.177**	–	–
	Material	**1**	**84**	**124.42**	**8.18e – 12**	**0.542**	–	–
	Difficulty	**1**	**84**	**112.231**	**2.65e – 11**	**0.304**	–	–
	Position: Group	2	84	0.151	0.860	0.001	0.862	–
	Position: Material	**2**	**84**	**12.332**	**8.30e – 06**	**0.099**	**0.847**	–
	Position: Difficulty	**2**	**84**	**25.426**	**1.12e – 10**	**0.185**	**0.985**	–
	Group: Material	**1**	**84**	**10.39**	**0.00321**	**0.090**	–	–
	Group: Difficulty	1	84	0.058	0.812	0	–	–
	Material: Difficulty	**1**	**84**	**57.729**	**2.82e – 08**	**0.256**	–	–
	Position: Group: Material	2	84	1.626	0.199	0.014	0.847	–
	Position: Group: Difficulty	2	84	1.023	0.361	0.009	0.985	–

(Continues)

TABLE 5 (Continued)

Measure	Factor(s)	df1	df2	F-value	p-value	η_p^2	W	ε
	Position: Material: Difficulty	**2**	**84**	**13.656**	**2.53 – 06**	**0.109**	**0.973**	–
	Group: Material: Difficulty	1	84	0.338	0.565	0.002	–	–
	Position: Group: Material: Difficulty	2	84	0.218	0.804	0.002	0.973	–

Note: For d' and *c*, *group was used as between-subject factor (two levels: fifth grade and adults) and material (two levels: Verbal, musical) and difficulty (two levels: Easy, difficult) as within-subject factors.* For the response times analysis, the type of trial was taken into account as within-subject factor (same/different). The analysis of recency effects was done on the percentage of correct responses for different trials with position (three levels: second, third, fourth position) and group as between-subject factors and material and difficulty as within-subject factors. Significant effects are in bold font. df, degrees of freedom; η_p^2, partial eta-squared; *W*, Mauchly's statistic. An asterisk indicates a significant *W*: Mauchly's statistic. An asterisk indicates a significant *W*. If so, corrected degrees of freedom are reported in parenthesis in the df1 and df2 columns and the corresponding statistics are corrected with ε: Greenhouse-Geisser estimates of sphericity.

for age equivalences) displayed poorer memory performance than older children (second, third, fourth, and fifth grade). This was observed for both materials: children's performance increased from KG to second grade, and then remained stable until fifth grade. Overall, children's performance was lower for musical material (sequences of four tones) than verbal material (sequences of four syllables). Adults showed the same pattern of performance, but with ceiling performance for verbal material. For both musical and verbal materials, children did not reach the same level of performance as adults, thus revealing the immaturity of auditory STM even in 10-year-old children. Response times analysis showed the same pattern as performance analysis: younger children were slower than older children thus confirming the absence of an impulsivity issue or a potential difference in speed-accuracy trade-off between age-groups. An effect of difficulty level was found for the musical material, revealing a facilitatory effect of contour violation and large interval changes, in both children and adults. Manipulating the difficulty in the verbal material (i.e., changing only a vowel in the second sequence in the difficult task instead of both consonant and vowel in the easy version) elicited only minor decrease in performance, if any. As for the recency effects expected in that kind of task, they were more apparent in the difficult conditions. Children displayed a recency effect for both materials, but older children and adults mostly for the difficult musical material.

4.1 | Developmental trajectory of auditory STM

The present study, based on a recognition paradigm with a fixed number of items, suggests the following pattern of development for musical and verbal STM: increasing performance from KG to second grade (5–7 years-old) followed by a standstill until fifth grade (10 years old). However, neither Gathercole et al. (2004) nor Alloway et al. (2006) observed such a pattern for verbal STM (for a comparison between the former two studies and the present one, see Figure 7), they rather observed a linear increase in capacity. Regarding musical STM, Clark et al. (2018) found a developmental increase in children from 6 to 13 years-old similar to the one of visuo-spatial STM, with a linear increase in storage capacity as observed in former studies (Alloway et al., 2006; Gathercole et al., 2004).

The difference of developmental pattern observed between our study and the former ones may rely on the difference of cognitive processes used in recall tasks and DMST. Recall tasks might rely more strongly on memory capacity and long-term knowledge (if it is used with digits or words), whereas the DMST task might involve different procedural processes as it involves a comparison process between S1 and S2 and rely more on sensory memory traces as they can be used in stimulus recognition (Cowan, 1984). Regarding the weight of long-term knowledge, lexical knowledge has indeed been found to have a significant influence on verbal STM in children when recall tasks are used (Gathercole et al., 2001; Messer et al., 2015) and also when serial-order STM reconstruction tasks are used (Gorin et al., 2018; Leclercq & Majerus, 2010; Majerus et al., 2006). The reconstruction task is supposed to minimize phonological and lexical demands; however, these studies are using verbal materials (like animal names in Leclercq & Majerus, 2010, e.g.). Forward recall tasks also rely on phonological production and thus present constraints linked to the phonological loop and specific mechanisms of motor production.

Several factors have been proposed to account for the increase of performance in studies investigating the development of STM: the decrease of memory trace decay throughout childhood, the increase of memory capacity, the increasing involvement of executive functions as they mature, and the influence of long-term knowledge that would support short-term storage (Messer et al., 2015). The developmental trajectory observed in the present study could thus arise from the use of a DMST that reduces lexical demands and necessitates different executive demands during the comparison between the two sequences. Furthermore, DMST has the significant advantage of allowing the same implementation for different auditory material (verbal, musical, timbre), which allows direct comparisons.

4.2 | Development of musical and verbal STM: Shared and distinct mechanisms

As mentioned in the introduction, several WM model and in particular their account for short-term storage could be divided into two groups regarding their prediction for the developmental trajectory of musical and verbal STM. On one hand, theoretical framework based on

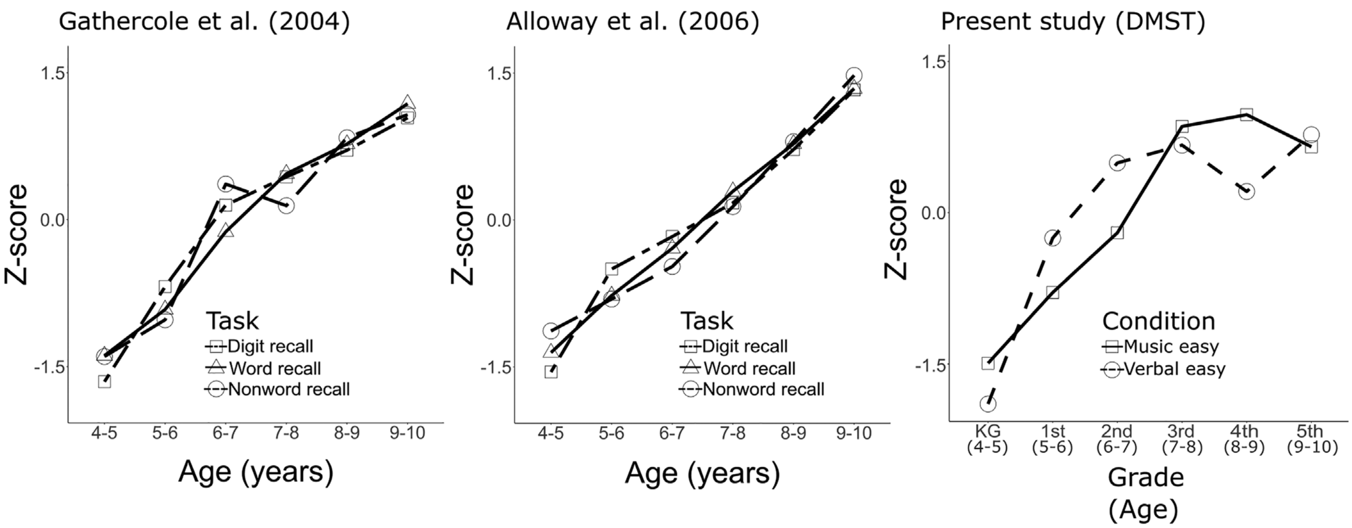

FIGURE 7 Left and middle panels: *z*-scores of mean performance in verbal forward recall tasks with digits, words, and nonwords in children from 4 to 10 years of age, adapted from Gathercole et al. (2004) and Alloway et al. (2006), respectively. Right panel: *z*-scores of mean musical and verbal recognition tasks of the present study in the easy condition

Baddeley and Hitch (1974) WM model postulate the existence of a separate short-term storage for musical material (Pechmann & Mohr, 1992) and even a specific musical central executive (Ockelford, 2007). These frameworks might predict two different developmental trajectories for musical and verbal STM as they would rely on different attentional processes. On the other hand, other models consider that attentional processes involved in the maintenance of items in STM do not present such a modular structure (Barrouillet & Camos, 2007; Cowan et al., 1998). A common consideration between these models is that maintenance processes involving attentional resources used to refresh memory traces in short-term storage are non-modular and can be allocated to different materials. Such models account for the distinction between domain-general and domain-specific processes in musical and verbal STM (Gorin et al., 2018) and they would probably account for a similar developmental trajectory between musical and verbal STM as it was observed in this study. In addition to a similar trajectory, we found recency effects for both materials. Some models posit that recency effect would represent behavioral signatures of serial order constructs in recall and recognition tasks (Hurlstone et al., 2014; Lewandowsky & Farrell, 2008). Indeed, it has been argued that musical and verbal STM systems may rely on similar sequential processes. These similar sequential processes were characterized in adults by Gorin et al (2018) who found similar selective sensitivity to time-based interference in musical and verbal STM (with a mixed recall/recognition paradigm) and similar transposition gradients. They also found similar error patterns, sequence length effects, and recency and primacy effects for both materials as well as similar limited capacity and an effect of pitch proximity, comparable to phonological proximity (Williamson et al., 2010). Another finding in favor of shared mechanisms between the two materials is the significant correlation between musical and verbal STM performance, even when the effect of age was factored out.

We observed similarities in the developmental trajectory for both materials, but we also observed better performance for the verbal material, despite using the same number of items for both tasks, both for children and adults (Figure 4). Although this latter finding could be related to discriminability differences for musical and verbal items used in the DMST, we cannot rule out the possibility that modality-specific STM systems could treat the two materials differently, as domain-specific systems have been suggested for encoding, storage, and maintenance of musical and verbal information (Gorin et al., 2018; Schulze & Tillmann, 2013). In addition, the fact that only musical STM performance correlated with the speech-in-noise performance in the present study suggests differentiated encoding processes between musical and verbal STM. A study involving comparison between STM for words, tones, and timbres using a recognition paradigm in adults, suggested similar storage of musical and verbal information and different internal sensorimotor codes used to maintain musical and verbal information (Schulze & Tillmann, 2013). In our present study, the similarity of the developmental trajectory for both materials could be the consequence of the involvement of shared domain-general systems in serial order coding, given in particular the finding of a recency effect for both materials, shared maintenance processes, and similar rehearsal mechanisms. Conversely, the difference of performance that we found in children between materials, might arise from different sensorimotor codes used for musical and verbal information.

4.3 | Development of auditory cognition

The secondary aim of this study was to investigate auditory cognition beyond auditory STM. Indeed, the links between CAPDs and learning disorder, as well as the advantage of musicians in speech-in-noise abilities (Parbery-Clark et al., 2009; Slater et al., 2015; Zendel et al., 2015), suggest that the joint investigation of speech-in-noise and verbal and non-verbal STM would be informative. Here, children underwent a speech-in-noise task (Ginzburg et al., 2019) and it was found that only performance in the musical DMST correlated with speech-in-noise performance, specifically in the phonologically easy condition with a −3 dB SNR (Figure 6). This finding suggests that children might rely on perceptual processes used in pitch encoding to process speech in cocktail-party noise, these processes being shared with musical STM but not with verbal STM. These results verify the already observed link between sound segregation and pitch information processing (Oxenham, 2008). Overall, these results seem to corroborate the hypothesis that musical and verbal STM have distinct mechanisms regarding item-information processing. Indeed, we showed that domain-general procedural and attentional processing seem to be involved in both musical and verbal STM given their similar developmental trajectory. The distinct mechanisms involved in musical and verbal STM appear to lie in item-specific encoding processes that only speech-in-noise and musical STM share in the present study.

4.4 | Auditory cognition and learning disorders

These two components of auditory cognition (STM and speech-in-noise) might thus share processes that are of particular interest to understand the underpinnings of central auditory processing. We found, in the present study, that children who presented learning disorders all performed below the median in the musical easy condition of the DMST and in the easy condition at +3 dB SNR of the speech-in-noise task. These results indicate that the musical easy condition in the STM task and the easy +3 dB SNR condition in the speech-in-noise task might allow discriminating children with learning disorders. These results are in agreement with the hypothesis of a general deficit in auditory perception in learning disorders (Nithart et al., 2009) encompassing verbal STM deficits (Perez et al., 2012) and speech-in-noise deficits (Ziegler et al., 2005). Therefore, it seems that STM for pitch and speech in cocktail party noise could be of great interest to identify early CAPD and, as suggested by these studies, early learning disorders. As comorbidity between CAPD and learning disorders has been reported, the use of early identification of CAPD would facilitate the adaptation of a child's school and home environment earlier in development. Indeed, the diagnosis of learning disorders is currently highly reliant on reading abilities and thus cannot be done before reading acquisition. The use of auditory child-adapted tasks, such as the ones used here, allows overriding the reading acquisition problematic as no reading abilities are required. Furthermore, remediation and clinical treatment could benefit from these insights as pitch encoding seems to be closely related to learning disorders. Indeed, clinical population could take advantage of the indirect effect of pitch encoding enhancement that arise from musical training (Forgeard et al., 2008).

5 | CONCLUSION

This study investigated the developmental trajectory of auditory STM in 5–10 years-old children for musical and verbal material. Results suggest that auditory STM is developing throughout childhood and is still under maturation at the age of 10. With the use of a recognition task, we observed that STM performance increases until second grade and levels-off until fifth grade for musical and verbal material. Using the DMST allowed discussing shared processes between the two materials and the different processes at stake in recognition tasks compared to forward recall tasks. Children, as well as adults, also showed poorer performance for the musical material compared to the verbal material, providing some evidence for specific mechanisms for the processing of the two materials in STM. We also observed that similar processes might be at stake between musical STM and speech-in-noise perception revealing the relevance of investigating the specific processes involved in musical information processing. Future studies should further investigate the development of auditory cognition in a systematic way that relies less on verbal production and LTM knowledge as serial recall tasks do. Future investigation should also assess jointly STM and speech-in-noise processing in children with typical development and learning disorders, to pave the way for new diagnosis and rehabilitation tools of central auditory processing deficits.

ACKNOWLEDGMENTS

We thank all the children who took part in the study, along with their parents, and their welcoming teachers. We also thank Nathalie Bedoin and Emmanuel Ferragne for their help for the creation of the stimuli, as well as Céline Hidalgo and Daniele Schön for their help with "p-center" calculation. This work was conducted within the framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique", ANR-10-LABX-0060) and of the LabEx Cortex ("Construction, Cognitive Function and Rehabilitation of the Cortex", ANR-11-LABX-0042) of Université de Lyon, within the program "Investissements d'avenir" (ANR-16-IDEX-0005) operated by the French National Research Agency (ANR). This work was funded by a Pack Ambition Recherche (COGAUDYS project) from the Auvergne-Rhône-Alpes Region to AC, AM, and BT.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Jérémie Ginzburg https://orcid.org/0000-0003-4851-146X

REFERENCES

Adlard, A., & Hazan, V. (1998). Speech perception in children with specific reading difficulties (Dyslexia). *The Quarterly Journal of Experimental Psychology Section A*, *51*(1), 153–177. https://doi.org/10.1080/713755750

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*(6), 1698–1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Atterbury, B. W. (1985). Musical differences in learning-disabled and normal-achieving readers, aged seven, eight and nine. *Psychology of Music*, *13*(2), 114–123. https://doi.org/10.1177/0305735685132005

Attout, L., & Majerus, S. (2015). Working memory deficits in developmental dyscalculia: The importance of serial order. *Child Neuropsychology*, *21*(4), 432–450. https://doi.org/10.1080/09297049.2014.922170

Baddeley, A., & 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

Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. In N. Osaka, R. H. Logie, & M. D'Esposito (Éds.), *The cognitive neuroscience of working memory* (pp. 59–80). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198570394.003.0004

Berz, W. L. (1995). Working memory in music: A theoretical model. *Music Perception: An Interdisciplinary Journal*, *12*(3), 353–364. https://doi.org/10.2307/40286188

Bourgeois–Vionnet, J., Moulin, A., Hermier, M., Pralus, A., Tillmann, B., Caclin, A., & Nighoghossian, N. (2020). A case of verbal and emotional prosody processing dissociation after a right temporal venous infarct. *Neurological Sciences*, *41*(6), 1615–1618. https://doi.org/10.1007/s10072-020-04259-y

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). *Speaking clearly for children with learning disabilitiessentence perception in noise*. 18.

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

Chandrasekaran, B., & Kraus, N. (2010). Music, noise-exclusion, and learning. *Music Perception*, *27*(4), 297–306. https://doi.org/10.1525/mp.2010.27.4.297

Chuah, Y. M. L., & Maybery, M. T. (1999). Verbal and spatial short-term memory: Common sources of developmental change? *Journal of Experimental Child Psychology*, *73*(1), 7–44. https://doi.org/10.1006/jecp.1999.2493

Clark, K. M., Hardman, K. O., Schachtman, T. R., Saults, J. S., Glass, B. A., & Cowan, N. (2018). Tone series and the nature of working memory capacity development. *Developmental Psychology*, *54*(4), 663–676. https://doi.org/10.1037/dev0000466

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

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, *96*(2), 341–370. https://doi.org/10.1037/0033-2909.96.2.341

Cowan, N. (1998). *Attention and memory: An integrated framework*.

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

Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development*, *70*(5), 1082–1097. https://doi.org/10.1111/1467-8624.00080

Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). *Two separate verbal processing rates contributing to short-term memory span*. 20.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bullelin*, *89*(1), 63–100. https://doi.org/10.1037/0033-2909.89.1.63

Dyson, M. C., & Watkins, A. J. (1984). A figural approach to the role of melodic contour in melody recognition. *Perception & Psychophysics*, *35*(5), 477–488. https://doi.org/10.3758/BF03203924

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

Foster, J., & Haggard, M. (1987). The Four Alternative Auditory Feature test (FAAF)—Linguistic and psychometric properties of the material with normative data in noise. *British Journal of Audiology*, *21*(3), 165–174. https://doi.org/10.3109/03005368709076402

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410–419. https://doi.org/10.1016/S1364-6613(99)01388-1

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology Section A*, *54*(1), 1–30. https://doi.org/10.1080/02724980042000002

Gathercole, S. E., Pickering, S. J., Ambridge, B., Wearing, H. (2004). The Structure of Working Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Ginzburg, J., Fornoni, L., Popieul, J., Pierre, C., Caclin, A., & Moulin, A. (2019). Neurodevelopmental aspects of hearing-in-noise in 5 to 11 year-old children. Oral communication at the 14th congress of the European Federation of Audiological Societies, 22–25 may 2019, Lisbon (Portugal). *Journal of Hearing Science*, *9(1)*(1), Supplement 41.

Gordon, E. E. (1986). *Manual for the primary measures of music audiation and the intermediate measures of music audiation*. G.I.A Publications, Inc.

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

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. https://doi.org/10.1016/j.tics.2010.10.001

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

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. https://doi.org/10.1037/a0034221

Iliadou, V., & Iakovides, S. (2003). Contribution of psychoacoustics and neuroaudiology in revealing correlation of mental disorders with central auditory processing disorders. *Annals of General Hospital Psychiatry*, *2*(1), 5. https://doi.org/10.1186/1475-2832-2-5

Ireland, K., Parker, A., Foster, N., & Penhune, V. (2018). Rhythm and melody tasks for school-aged children with and without musical training: Age-equivalent scores and reliability. *Frontiers in Psychology*, *9*, 426. https://doi.org/10.3389/fpsyg.2018.00426

Jarrold, C., Thorn, A. S.C., Stephens, E. (2009). The relationships among verbal short-term memory, phonological awareness, and new word learning: Evidence from typical development and Down syndrome. *Journal of Experimental Child Psychology*, *102*(2), 196–218. https://doi.org/10.1016/j.jecp.2008.07.001

Kassambara, A. (2020). rstatix: Pipe-Friendly framework for basic statistical tests. *R version 0.5.0*, https://CRAN.R-project.org/package=rstatix

Keller, T. A., & Cowan, N. (1994). Developmental increase in the duration of memory for tone pitch. *Developmental Psychology*, *30*(6), 855–863. https://doi.org/10.1037/0012-1649.30.6.855

Kim, S. (2015). ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665–674. https://doi.org/10.5351/CSAM.2015.22.6.665

Kramer, J. H., Knee, K., & Delis, D. C. (2000). Verbal memory impairments in dyslexia. *Archives of Clinical Neuropsychology*, *15*(1), 83–93. https://doi.org/10.1093/arclin/15.1.83

Leclercq, A.-L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development: Evidence from a longitudinal study. *Developmental Psychology*, *46*(2), 417–427. https://doi.org/10.1037/a0018540

Lévêque, Y., Teyssier, P., Bouchet, P., Bigand, E., Caclin, A., & Tillmann, B. (2018). Musical emotions in congenital amusia: Impaired recognition, but preserved emotional intensity. *Neuropsychology*, *32*(7), 880–894. https://doi.org/10.1037/neu0000461

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In *Psychology of Learning and Motivation*, (Vol., *49*, p. 1–48). Elsevier. https://doi.org/10.1016/S0079-7421(08)00001-7

Macmillan, N. A., & Creelman, C. Douglas(1991). *Detection theory: A user's guide*. Cambridge University Press.

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

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology*, *93*(2), 95–119. https://doi.org/10.1016/j.jecp.2005.07.005

Majerus, S., Van der Linden, M., Braissand, V., Eliez, S. (2007). Verbal Short-Term Memory in Individuals With Chromosome 22q11.2 Deletion: Specific Deficit in Serial Order Retention Capacities?. *American Journal on Mental Retardation*, *112*(2), 79. https://doi.org/10.1352/0895-8017(2007)112[79:vsmiiw]2.0.co;2

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

Männel, C., Meyer, L., Wilcke, A., Boltze, J., Kirsten, H., & Friederici, A. D. (2015). Working-memory endophenotype and dyslexia-associated genetic variant predict dyslexia phenotype. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, *71*, 291–305. https://doi.org/10.1016/j.cortex.2015.06.029

Medwetsky, L. (2011). Spoken language processing model: Bridging auditory and language processing to guide assessment and intervention. *Language, Speech, and Hearing Services in School*, *42*, 286–296. https://doi.org/10.1044/0161-1461(2011/10-0036

Messer, M. H., Verhagen, J., Boom, J., Mayo, A. Y., & Leseman, P. P. M. (2015). Growth of verbal short-term memory of nonwords varying in phonotactic probability: A longitudinal study with monolingual and bilingual children. *Journal of Memory and Language*, *84*, 24–36. https://doi.org/10.1016/j.jml.2015.05.001

Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics*, *41*(6), 576–600. https://doi.org/10.3758/BF03210491

Moore, D. R., Ferguson, M. A., Edmondson-Jones, A. M., Ratib, S., & Riley, A. (2010). Nature of auditory processing disorder in children. *Pediatrics*, *126*(2), e382--e390. https://doi.org/10.1542/peds.2009-2826

Moulin, A., Garcia, S., Jeanvoine, A., & Richard, C. (2013). French "Audimots": Development of a French version of the Four Alternative Auditory Features Test. 11th European Federation of Audiology Societies (EFAS) Congress. *Otorhinolaryngologia Hungarica*, *59*(2), 98–99.

Murphy, D. R., Craik, F. I. M., Li, K. Z. H., & Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychology and Aging*, *15*(2), 323–334. https://doi.org/10.1037/0882-7974.15.2.323

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524. https://doi.org/10.3758/BF03195598

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

Ockelford, A. (2007). A music module in working memory? Evidence from the performance of a prodigious musical savant. *Musicae Scientiae*, *11*(2_suppl), 5–36. https://doi.org/10.1177/10298649070110S202

Orsini, A., Grossi, D., Capitani, E., Laiacona, M., Papagno, C., & Vallar, G. (1987). Verbal and spatial immediate memory span: Normative data from 1355 adults and 1112 children. *The Italian Journal of Neurological Sciences*, *8*(6), 537–548. https://doi.org/10.1007/BF02333660

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification*, *12*(4), 316–331. https://doi.org/10.1177/1084713808325881

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661. https://doi.org/10.1097/AUD.0b013e3181b412e9

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition*, *20*(3), 314–320. https://doi.org/10.3758/BF03199668

Peretz, I., & Babaï, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia*, *30*(3), 277–292. https://doi.org/10.1016/0028-3932(92)90005-7

Perez, T. M., Majerus, S., Mahot, A., & Poncelet, M. (2012). Evidence for a specific impairment of serial order short-term memory in dyslexic children: Serial order STM deficits in dyslexia. *Dyslexia*, *18*(2), 94–109. https://doi.org/10.1002/dys.1438

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608. https://doi.org/10.1121/1.412282

Plantinga, J., & Trainor, L. J. (2008). Infants' memory for isolated tones and the effects of interference. *Music Perception: An Interdisciplinary Journal*, *26*(2), 121–127. https://doi.org/10.1525/mp.2008.26.2.121

Plaza, M., Cohen, H., & Chevrie-Muller, C. (2002). Oral language deficits in dyslexic children: Weaknesses in working memory and verbal planning. *Brain and Cognition*, *48*(2–3), 505–512. https://doi.org/10.1006/brcg.2001.1407

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

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1230–1240. https://doi.org/10.1044/1092-4388(2009/08-0111

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

Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural Brain Research*, *291*, 244–252. https://doi.org/10.1016/j.bbr.2015.05.026

Sperling, A. J., Lu, Z.-L., Manis, F. R., & Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience*, *8*(7), 862–863. https://doi.org/10.1038/nn1474

Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, *123*(3), 191–201. https://doi.org/10.1016/j.bandl.2012.09.001

Talamini, F., Blain, S., Ginzburg, J., Houix, O., Bouchet, P., Grassi, M., Tillmann, B., & Caclin, A. (2021). 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

Tijms, J. (2004). Verbal memory and phonological processing in dyslexia. *Journal of Research in Reading*, *27*(3), 300–310. https://doi.org/10.1111/j.1467-9817.2004.00233.x

Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia: Evidence from a gating paradigm. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, *59*, 84–94. https://doi.org/10.1016/j.cortex.2014.07.012

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, *1640*, 251–263. https://doi.org/10.1016/j.brainres.2015.10.035

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

Wieland, E. A., McAuley, J. D., Dilley, L. C., & Chang, S.-E. (2015). Evidence for a rhythm perception deficit in children who stutter. *Brain and Language*, *144*, 26–34. https://doi.org/10.1016/j.bandl.2015.03.008

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and non-musicians' 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

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, *134*(3), 2225–2234. https://doi.org/10.1121/1.4817926

Zendel, B. R., Tremblay, C.-D., Belleville, S., & Peretz, I. (2015). The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *Journal of Cognitive Neuroscience*, *27*(5), 1044–1059. https://doi.org/10.1162/jocn_a_00758

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences*, *102*(39), 14110–14115. https://doi.org/10.1073/pnas.0504446102

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

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: 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*, e13188. https://doi.org/10.1111/desc.13188