

## Original Articles

# Cognitive representation of “musical fractals”: Processing hierarchy and recursion in the auditory domain



Mauricio Dias Martins<sup>a,b,c,\*</sup>, Bruno Gingras<sup>c,d</sup>, Estela Puig-Waldmueller<sup>c</sup>, W. Tecumseh Fitch<sup>c</sup>

<sup>a</sup> Berlin School of Mind and Brain, Humboldt University, Berlin, Germany

<sup>b</sup> Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

<sup>c</sup> Department of Cognitive Biology, University of Vienna, Vienna, Austria

<sup>d</sup> Institute of Psychology, University of Innsbruck, Austria

## ARTICLE INFO

## Article history:

Received 30 November 2015

Revised 1 October 2016

Accepted 3 January 2017

Available online 16 January 2017

## Keywords:

Hierarchy

Recursion

Fractals

Music

Auditory

## ABSTRACT

The human ability to process hierarchical structures has been a longstanding research topic. However, the nature of the cognitive machinery underlying this faculty remains controversial. Recursion, the ability to embed structures within structures of the same kind, has been proposed as a key component of our ability to parse and generate complex hierarchies. Here, we investigated the cognitive representation of both recursive and iterative processes in the auditory domain. The experiment used a two-alternative forced-choice paradigm: participants were exposed to three-step processes in which pure-tone sequences were built either through recursive or iterative processes, and had to choose the correct completion. Foils were constructed according to generative processes that did not match the previous steps. Both musicians and non-musicians were able to represent recursion in the auditory domain, although musicians performed better. We also observed that general ‘musical’ aptitudes played a role in both recursion and iteration, although the influence of musical training was somehow independent from melodic memory. Moreover, unlike iteration, recursion in audition was well correlated with its non-auditory (recursive) analogues in the visual and action sequencing domains. These results suggest that the cognitive machinery involved in establishing recursive representations is domain-general, even though this machinery requires access to information resulting from domain-specific processes.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The capacity to represent and generate hierarchical structures is a fundamental human trait that is used in virtually every domain of activity. Even though this trait is to some extent present in other species (Bergman, Beehner, Cheney, & Seyfarth, 2003; Massen, Pašukonis, Schmidt, & Bugnyar, 2014; Seyfarth & Cheney, 2008; Seyfarth, Silk, & Cheney, 2014), it seems to be especially developed in humans (Conway & Christiansen, 2001; Fitch & Friederici, 2012; Hauser, Chomsky, & Fitch, 2002; ten Cate & Okanoya, 2012; Vasconcelos, 2008). Not only are human-generated hierarchies more complex, but they are also more general, since the average person can generate visual, social, linguistic and action hierarchies (Altmann, Bülthoff, & Kourtzi, 2003; Badre, 2008; Badre & D’Esposito, 2009; Badre, Hoffman, Cooney, & D’Esposito, 2009; Bahlmann, Schubotz, & Friederici, 2008; Chomsky, 1995; Eglash,

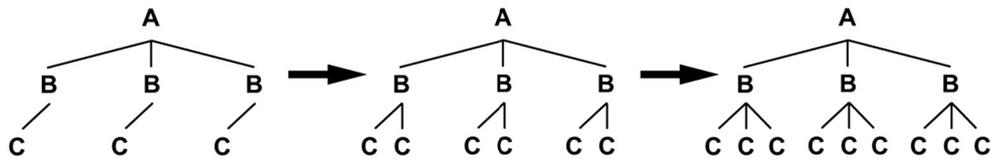
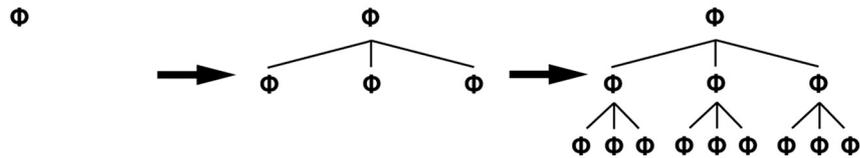
1998; Fitch & Martins, 2014; Friederici, 2011; Hunyady, 2010; Jackendoff, 2003; Jackendoff, 2009; Kravitz, Saleem, Baker, & Mishkin, 2011; Picard et al., 2010; Zink et al., 2008). This complexity and generality could be explained either by a general increase in processing power, due to a larger brain, or by the existence of additional specialized processes or abilities in human cognitive architecture.

One such ability, which could potentially explain the human cognitive exceptionality, is recursion (Hauser et al., 2002). Recursion can be understood as the ability to embed elements within elements of the same kind (Fitch, 2010; Hulst, 2010; Martins, 2012). Recursion is a particular principle to represent and generate hierarchies which allows the generation of multiple levels with a single rule (Fig. 1B).

This increased generative power of recursion in comparison with other kinds of hierarchical principles is thought of as being instantiated by cognitive representations of a higher level of abstraction (Martins, 2012). For instance, instead of representing each hierarchical relation with its own rule, of the kind  $A \rightarrow B$  and  $B \rightarrow C$  (Fig. 1A), humans are able to understand that these

\* Corresponding author at: Berlin School of Mind and Brain, Humboldt University, Luisenstrasse 56, 10117 Berlin, Germany.

E-mail address: [diasmarm@hu-berlin.de](mailto:diasmarm@hu-berlin.de) (M.D. Martins).

**(A) Within-level addition rule: Add another C to existing level under each B.****(B) Cross-level recursive rule: Add three  $\Phi$  to new level under each  $\Phi$ .**

**Fig. 1.** Recursive and non-recursive procedures to generate hierarchies. (A) Simple iterative procedures add elements to a hierarchy, within fixed levels. In order to generate a level under 'C', another rule would be required – add another 'D' under each 'C'. (B) Recursive rules are more abstract. They can be used to characterize hierarchical relations across many different levels of the hierarchy. With the same rule (B), an infinite number of hierarchical levels could be added to the structure.

relations have commonalities, which allows the induction of a more general rule  $\Phi \rightarrow \Phi$  (Fig. 1B). To our knowledge, this level of hierarchical abstraction seems to be specifically available to human cognition (Berwick, Friederici, Chomsky, & Bolhuis, 2013; Fitch, Hauser, & Chomsky, 2005; Hauser et al., 2002; ten Cate & Okanoya, 2012).

### 1.1. Cognitive assessment of recursion

Formally, structures that can be understood as recursive have been suggested to exist in visual art (Eglash, 1997), visuo-spatial processing (Martins, 2012; Martins & Fitch, 2012), music (Jackendoff & Lerdahl, 2006; Lerdahl & Jackendoff, 1996), architecture (Eglash, 1998), humour (Eisenberg, 2008), theory of mind (Miller, 2009; Tomasello, 2008), problem solving (Schiemanz, 2002), action sequencing (Pulvermüller & Fadiga, 2010), syntax (Chomsky, 1995; Karlsson, 2010; Mithun, 2010; Roeper, 2009), phonology (Hulst, 2010; Hunyady, 2010; Schreuder, Gilbers, & Quené, 2009; Wagner, 2010), pragmatics (Levinson, 2013), conceptual structure (Hofstadter, 2000; Picard et al., 2010), mathematical proofs (Odifreddi, 1999), natural numbers (Hauser et al., 2002), and arithmetic operations (Friederici, Bahlmann, Friedrich, & Makuuchi, 2011). In all these domains it is possible to build recursive algorithms that generate hierarchical structures. However, it is not clear that in all these domains humans actually represent the recursive character of these structures, and use these representations productively. Such demonstrations require empirical rather than theoretical tools.

To our knowledge, the ability to induce recursive rules was empirically demonstrated first in the linguistic domain (Alegre & Gordon, 1996; Roeper, 2011). In this domain, recursion seems to be universally used (Reboul, 2012), and although some researchers argue that it is rare in common speech (Laury & Ono, 2010), most language users in the world are likely to have generated multiple recursive sentences in their lifetimes (for instance, compound noun phrases such as “[[[student] film]] committee”). Furthermore, the ability to extract the correct meaning from recursive sentences seems to be available early during ontogeny (Alegre & Gordon, 1996; Roeper, 2009). This interesting relationship between language and recursion, yet undemonstrated in other domains, has led some authors to propose that recursion is a domain-specific “linguistic computational system [...], independent of the other systems with which it interacts and interfaces” (Fitch et al., 2005; Hauser et al., 2002). This hypothesis implies that

the use of recursion in other domains is dependent on verbal resources. Coincidentally, the ability to perform second-order theory of mind tasks (e.g., [I think that [she thinks that [John thinks something]]]) correlates with language abilities (Miller, 2009, for a review), and verbal interference tasks block the ability to use natural numbers (Gordon, 2004). These results were taken as strong evidence that recursion is a linguistic domain-specific ability.

Recently, in a series of experiments, human adults and children have also been shown to represent recursion in the visuo-spatial domain (Martins, Fischmeister, et al., 2014; Martins, Laaha, Freiburger, Choi, & Fitch, 2014; Martins, Mursic, Oh, & Fitch, 2015). In this domain, subjects were able to induce recursive rules generating visual fractals, and to use these rules productively. Crucially, this ability was not specifically related with grammar comprehension (Martins, Laaha, et al., 2014), and it neither required verbal resources (Martins et al., 2015), nor generated activation in classical language brain areas (Martins, Fischmeister, et al., 2014). However, performance correlated with an action sequencing task, the Tower of Hanoi, which is best solved using recursive strategies (Martins, Fischmeister, et al., 2014).

These findings suggest that recursion does not necessarily require linguistic resources (arguing against the primacy of language). However, it is still possible that the same cognitive machinery is used to implement recursion in both domains (language and vision), if a domain-general (abstract) code were used instead of a linguistic or verbal one (Martins et al., 2015).

In this paper, we aim to expand our understanding of recursion in human cognition by focusing on another non-linguistic domain – using acoustic stimuli – and measure how recursive capacities in this domain correlate with recursion in other domains. Towards this goal, we will assess how humans represent “musical fractals”.

### 1.2. Hierarchical processing of music

Like language, music is a domain known to require the processing of hierarchical relations (Jackendoff, 2009; Jackendoff & Lerdahl, 2006; Koelsch, Rohrmeier, Torrecuso, & Jentschke, 2013; Rohrmeier, 2011). These relations involve the embedding of discrete acoustic events into higher-order structures, according to their rhythmical and pitch relationships. For instance, in tonal structures, there are precise relations between tones and the context in which these tones are embedded. Thus, in Western music, the same tone (or chord) can be perceived as decreasing or increasing the tension of a musical sequence according to the context,

creating emotional feelings of completeness versus incompleteness (Jackendoff & Lerdahl, 2006). The rules that govern these dynamics of tension and attraction are usually acquired during the process of enculturation with the musical canon (Tillmann, 2012; Tillmann & Bigand, 2004), and are commonly referred to as musical syntax, in analogy to linguistic rules (e.g., Fadiga, Craighero, & D'Ausilio, 2009; Jeon, 2014; Koelsch, 2006; Koelsch et al., 2013; Patel, 2003; Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Sammler et al., 2009).

But unlike in language, the understanding of long-distance hierarchical relations in music seems to be particularly difficult, especially for non-musicians (Besson, Faïta, & Requin, 1994; Tillmann & Bigand, 2004). The effect of expertise seems to be more influential in the ability to make explicit decisions, rather than in perceptual aspects (Besson & Faïta, 1995).

To our knowledge, only one study has suggested that non-musicians might be sensitive to long-distance relations in music (Koelsch et al., 2013). However, since many other studies have failed to detect this sensitivity (e.g., Cook, 1987; Tillmann, Bigand, & Madurell, 1988), it remains unclear to which extent non-musicians can represent complex hierarchical relations in music, even though some capacity can be acquired through passive listening (Bigand & Poulin-Charronnat, 2006). Interestingly, some studies suggest that the processing of syntactic violations in music seems to elicit neural signals analogous to the processing of linguistic syntax (Koelsch, 2006; Maess, Koelsch, Gunter, & Friederici, 2001; Patel et al., 1998; Sammler et al., 2009). This provides evidence for an influential view that music and linguistic syntax have common resources, and perhaps a common evolutionary history (Brown, 2000; Fitch, 2006; Patel, 2003).

Finally, hierarchical relations of tonal structures have also been formalized as recursive (Jackendoff & Lerdahl, 2006; Lerdahl & Jackendoff, 1983; Rohrmeier, 2011). However, as in many other domains, it remains an open question whether human adults, and especially non-musicians, can induce these recursive rules and use them productively. In this article we describe an attempt to experimentally answer this question.

### 1.3. Aims

The current study has two main aims: The first is to assess whether human adults, in particular non-musicians, are able to represent hierarchical relations in the auditory domain. In particular, we will test whether they are able to induce and apply within-level iterative rules (Fig. 1A) and cross-level recursive rules (Fig. 1B) in the domain of structured tonal sequences. The second goal is to assess whether the ability to represent recursion in the musical domain correlates with analogous tasks in the visual and action domain. This will address the question of whether the ability to form recursive representations relies on domain-specific resources or a domain-general cognitive machinery.

## 2. Experiment 1: Can humans represent recursion in the auditory domain?

Previous research by our group has shown that the ability to represent recursion is present in the general population in the visuo-spatial domain (Martins, 2012; Martins, Fischmeister, et al., 2014; Martins, Laaha, et al., 2014; Martins et al., 2015). In the current experiment, our goal was to determine whether this ability is also present in the auditory domain. Analogous to our work in vision, we applied recursive rules over sequences of tones, and tested whether participants were able to extract these rules. We employed a two-alternative forced-choice paradigm: participants were exposed to three steps of a recursive process generating auditory fractals (using pure tone sequences), and then asked to

discriminate between a well-formed fourth step of the same process and a foil. In addition to this task, which we named Auditory Recursion Task (ART), we devised a control task, the Auditory Iteration Task (AIT), which shared the hierarchical, iterative and sequential aspects of ART, but which did not use recursive procedures to generate the hierarchical structures. Our previous research - behavioral, developmental and neuroimaging - suggests that hierarchical processing taps into different systems depending on whether an iterative or recursive rule is primed (Martins, 2012; Martins, Fischmeister, et al., 2014; Martins, Laaha, et al., 2014).

In this experiment, we tested both musicians and non-musicians. On the one hand, musicians are known to be able to represent rules governing tone structures (Tillmann, 2012; Tillmann & Bigand, 2004, for reviews). If the ability to represent recursion is available in the auditory domain, then musicians should be able to adequately encode our tone sequences as formed by recursive rules. On the other hand, non-musicians are known to have difficulties in explicitly representing some elementary qualities of musical sounds (Rammsayer & Altenmüller, 2006; Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005). Even though non-musicians might be able to process the dynamics of musical tension and relaxation resulting from hierarchical relations (Koelsch et al., 2013; Maess et al., 2001), their ability to process pitch and rhythm is inferior to that of musicians (Rammsayer & Altenmüller, 2006; Tervaniemi et al., 2005; Tillmann & Bigand, 2004), as is their ability to retain auditory sequences in memory (Cohen, Evans, Horowitz, & Wolfe, 2011). In our task, we used somewhat impoverished stimuli from a musical viewpoint. Although we used rhythmical and tone structures, these comprised only simple pitch intervals and durational ratios. Our goal was to increase the likelihood that non-musicians would be able to parse our stimuli as structured sequences of tones. Nevertheless, a failure of the non-musician population could either mean that the ability to represent recursion is not available, or that non-musicians cannot easily parse tone structure in general. General impairments could be caused by bottlenecks either at the proximal auditory level (Bigand & Poulin-Charronnat, 2006), or at the memory level (Cohen et al., 2011), since music stimuli are fast and transient. We tried to reduce the latter effect by allowing participants to listen to all stimuli as many times as they wished before providing a response.

Finally, we also included an initial training task, in which stimuli were simple structures of three notes, directly utilizing the ability to parse tone contour.

In sum, our goal was to assess whether humans could represent recursive rules in auditory sequences. If this ability is potentially present in human cognition, then at least musicians should perform adequately. To test whether this capacity was present in the non-expert general population, we also tested non-musicians.

### 2.1. Methods

#### 2.1.1. Participants

We tested 30 non-musicians (17 females) aged between 19 and 38 ( $M = 25.1$ ,  $SD = 5$ ) with the Auditory Recursion Task, and 24 different non-musicians (17 Females) aged between 19 and 46 ( $M = 25.7$ ,  $SD = 7$ ) with the Auditory Iteration Task. In addition, we tested 20 musicians (16 females) aged between 19 and 38 ( $M = 24.0$ ,  $SD = 4$ ) with the Auditory Recursion Task. Musicians were defined as participants with more than 8 years of musical training who regularly practiced an instrument or singing (more than once a week). Non-musicians were participants with less than 2 years of music classes who reported not regularly playing any musical instrument. All participants were recruited at the University of Vienna. All were tested using the same experimental apparatus, and all reported normal or corrected-to-normal visual and

auditory acuity. All participants gave their prior written consent, and were paid 10 euros. The research conformed to institutional guidelines of the University of Vienna and Austrian national legislation regarding ethics.

### 2.1.2. Stimuli and procedure

**Stimulus generation.** The stimuli comprising the Auditory Recursion Task were built as an auditory analog of visual fractals (Mandelbrot, 1977). Visual fractals are structures that can be generated from single constituents such as squares or triangles (*the initiators*) by applying a simple transformation rule (*the generator*) a given number of times (iterations). The structures generated by iterating this process are hierarchical and self-similar (see Fig. 2 for an exemplary overview of such a process).

In the visual domain, hierarchical level is denoted by constituent size (larger constituents being dominant, and smaller subordinate), and the transformation rule (*generator*) encodes the spatial position of a set of subordinate elements (three surrounding elements in Fig. 2) in relation to the dominant. Here, we built our auditory fractals using auditory features analogous to these parameters: we took note duration and pitch to denote hierarchical level (longer and lower-pitched notes being dominant over shorter and higher ones); and tone space as the parameter being modulated by the *generator*. Specifically, for each tone present in one iteration (hierarchically dominant), we added three new tones (hierarchically subordinate) higher in pitch (and shorter in duration), with a particular contour (variably ascending or descending), and at a certain pitch distance from the dominant (inter-level interval); these subordinate-note contours constitute the recursive rule operative over multiple hierarchical levels.

A typical example of a music fractal is represented in Fig. 3. The target stimulus was built in four steps or iterations (Fig. 4). The first iteration stimulus consists simply of a low-pitch pure tone with a duration of 7.3 s (*the initiator*). The second iteration retains the initial tone, and adds a set of three new tones, according to a particular rule (*the generator*). This rule manipulates: (i) pitch contour, which can be either ‘ascending’ or ‘descending’; (ii) pitch interval between consecutive tones, which can correspond either to four semitones (major third) or eight semitones (minor sixth); and (iii) pitch interval between successive levels, which can also correspond to either four or eight semitones (Supplementary Materials 1 contain sound files corresponding to the four successive steps involved in the generation of a tonal auditory fractal). The same *generator* was applied to build all new hierarchical levels. Thus, the pitch and rhythmical relations between dominant and subordinate elements were kept constant recursively across hierarchical levels. This produced a hierarchical self-similar structure.

These pitch intervals (major third and minor sixth) were chosen because they are consonant, and because simultaneous combinations of major thirds and minor sixths also usually generate consonant combinations (with the exception of augmented chords). Moreover, an interval of 4 semitones is considered sufficient to ensure auditory stream segregation between the various hierarchical levels (Vliegen & Oxenham, 1999). Furthermore, to maximize

auditory stream segregation, we applied different amplitude modulation rates to the tones at different levels, with no amplitude modulation (0 Hz) for the lowest (first) and highest (last) hierarchical levels and amplitude modulation rates of 6 and 18 Hz for the second and third hierarchical levels (Grimault, Bacon, & Michey, 2002). The pure tone frequencies used in the stimuli ranged between 50.5 Hz and 3846 Hz, a frequency range within which humans can reliably recognize frequency intervals (Semal & Demany, 1990). For comparison, the note frequencies found on a 88-key modern piano range between 27.5 and 4186 Hz.

We produced four successive iterations of 24 different types of auditory fractals, generated using custom code in MATLAB R2012b (Mathworks, Natick, MA). For each of these 24 fractals, we produced (1) a correct fourth continuation of the first three iterative steps and (2) an incorrect continuation as a “foil” stimulus. This incorrect fourth iteration was produced by applying a different *generator* to the third stage. There were three categories of foils (Fig. 5): (i) positional, (ii) odd, and (iii) repetition (sound files corresponding to well-formed fourth iterations and foils can be found in Supplementary Materials II).

**Auditory Recursion Task (ART).** Participants sat in front of a computer screen and used the mouse to interact with a custom GUI interface implemented using Python programming language. Sounds were delivered through Sennheiser HC 520 headphones.

The task was composed of 24 trials. The set of 24 stimuli was a fully balanced distribution of all categories (see stimulus generation): inter-note interval (with 12 stimuli using the major third and 12 the minor sixth), direction (12 ascending and 12 descending), inter-level interval (12 using the major third and 12 the minor sixth), and foil (8 repetition, 8 position, and 8 odd).

On each trial, a set of five plain black square buttons was displayed on the screen, one button corresponding to each different tone sequence (Fig. 6). The squares corresponding to the iterations 1, 2 and 3 were depicted on the top of the screen; and the possible continuations (correct and foil) were displayed at the bottom. The position (LEFT and RIGHT) of the correct and incorrect fourth iterations were counterbalanced.

At the beginning of each trial, iterations 1, 2, and 3 were automatically played in sequence. Each sound had a duration of 7.3 s, and they were presented with a pause of 1 s between them. There was a visual cue (change from black to gray) indicating which iteration (square button) was being played. After the 3 first iterations were played, participants were asked to choose from 2 alternatives (Choice A and Choice B) which corresponded to the correct fourth iteration. Participants could listen to the alternatives (and repeat the other sounds) as many times as they wished by clicking on the corresponding squares. Responses were delivered by clicking on the buttons ‘Choose A’ or ‘Choose B’ displayed under the corresponding black squares (Fig. 6). These buttons became active only after all 5 sounds were played. After a choice was made, the next trial started.

Since participants were allowed to listen to all sounds as many times as they wished, the total duration of the task was variable. The average was around 45 min.

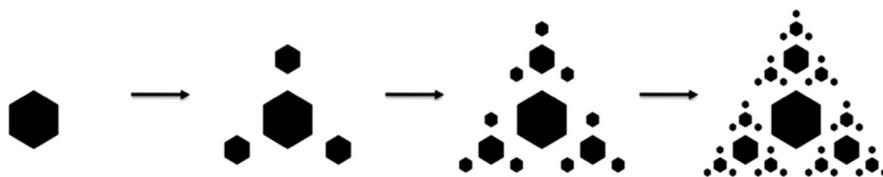
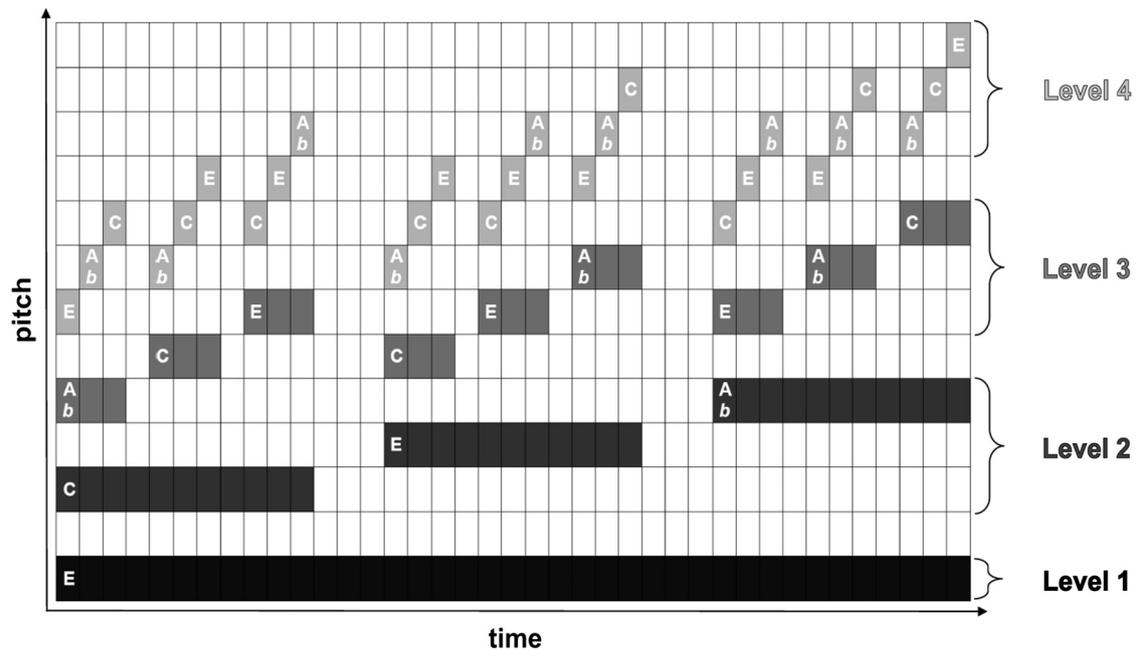
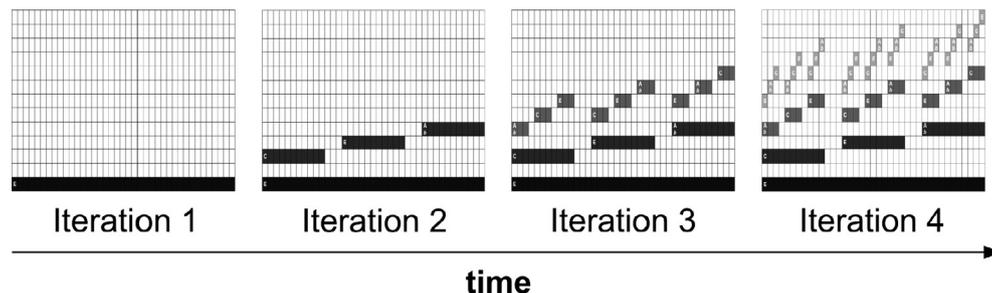


Fig. 2. Example of a recursive process generating a visual fractal.



**Fig. 3.** Example of a tonal auditory fractal. This auditory structure is composed of 4 hierarchical levels, denoted with different shades of gray. The dominant level (Level 1) is a low-pitch note with duration of 7.3 s. The second level (Level 2) is composed of three notes, each with a duration of slightly less than 1/3 the dominant note (Level 1), and with short silent pauses between them. This set of three notes is arranged in an ascending directionality, with a fixed pitch interval between each pair within these three notes. The same principle is used to generate Level 3, in which each set of three notes is added at a certain pitch interval in relation to a dominant note (each note in Level 2).



**Fig. 4.** Recursive process generating a tonal auditory fractal. At each step of the recursive process, a new hierarchical level is added (in the figure, a lighter shade of gray) containing notes of a higher pitch and shorter duration. A certain note is dominant (e.g., Level 1) over a set of three notes (e.g., Level 2) if they occur simultaneously in the sequence (time-wise), and the former is of a lower pitch and longer duration. The pitch and rhythmical relations between dominant and subordinate elements are kept constant recursively across hierarchical levels. Thus, if the dominant level displays an 'ascending' contour, so does the subordinate level. This between-level regularity occurs regarding the distance between levels, and distance within level (between notes of each set of three). In our recursion task, participants were first exposed to the first three iterations of the process, and then asked to discriminate between the fourth iteration and a foil (Fig. 5).

### 2.1.3. Auditory Iteration Task (AIT)

The experimental procedure and stimuli used in AIT were similar to what was described above for ART. The crucial difference between the tasks was the procedure used to generate Iteration 4 (Fig. 7). While in ART each iteration embedded the new tonal elements recursively within new hierarchical levels (different tone durations at each step, Fig. 4), in AIT elements were embedded within a single fixed hierarchical level (always the same tone duration), without adding new levels to the structure (sound files corresponding to AIT sequences and foils can be found in [Supplementary Materials III](#)). Experimentally, this means that Sounds 1, 2 and 3 were different between ART and AIT. However, the choice pairs (Choice A and Choice B) were the same between tasks. Foil stimuli were also identical, except for the 'Repetition' foils, since Iteration 3 in AIT is different from Iteration 3 in ART. All other parameters were kept constant between tasks.

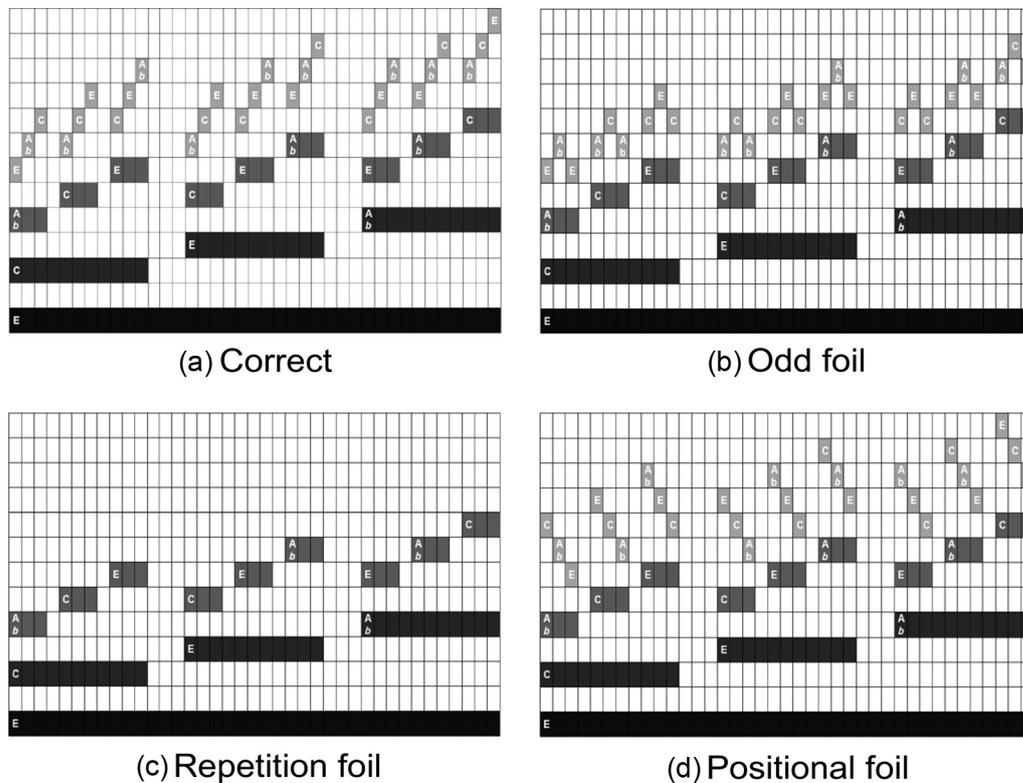
**Training.** The goal of the training procedure was to allow participants to familiarize themselves with the overall framework,

and with the concept of ascending and descending sequences of three tones, a crucial prerequisite to perform both ART and AIT. Each trial presented participants with a sequence of two tones (Sound #1), in which the second tone was either of a higher or lower pitch than the first one (ascending and descending direction, respectively).

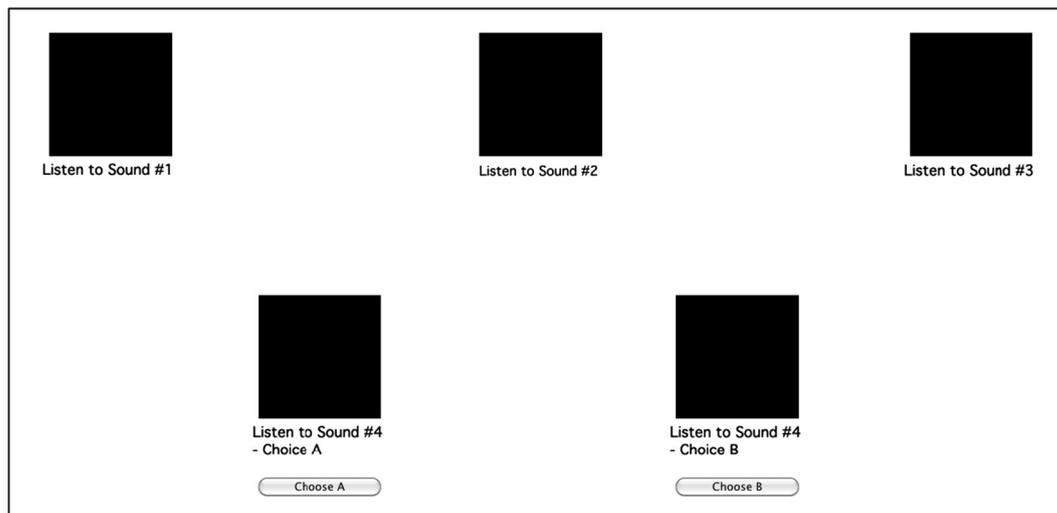
After this sequence was played, we asked participants to decide, from a set of two alternatives (Choice A and Choice B), which one correctly completed 'Sound #1'. The correct item was a sequence of three tones in which the third followed the same direction (ascending or descending) as the first two. Foils and interface were similar to ART, except that only one square was presented in the top center of the screen (corresponding to Sound #1).

### 2.1.3. Procedure

All participants started the procedure by filling in a questionnaire concerning their level of musical education and practice. All performed a training session as described above. Then,



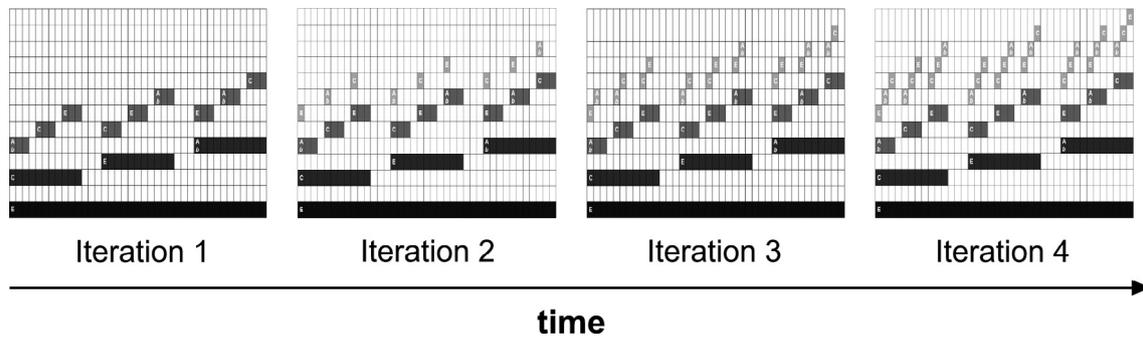
**Fig. 5.** Foil categories. For each music fractal, we generated the correct continuation (a.) of the recursive process in Fig. 4 and three different foils. The 'Repetition' foil (c.) is identical to the third iteration, and fails to continue the process, however it keeps the overall configuration of the auditory sequence ('ascending') constant. In both 'Odd' (b.) and 'Positional' (d.) foils, a new hierarchical level is added to the structure, however, this level is not consistent in contour with the pattern in previous iterations. In 'Odd' (b.) foils, the last note of each set of three is of the same pitch as the first note within that set. This disrupts the expected directionality (ascending or descending). In 'Positional' foils (d.), the directionality is consistent within each set of three, but not consistent with other hierarchical levels. In this example, the higher level has a descending contour, even though the previous levels display an ascending contour.



**Fig. 6.** Screenshot of a typical MRT trial. At the beginning of each trial, sounds (iterations) #1, #2 and #3 were automatically played in sequence. Participants were then asked to choose from 2 alternatives (Choice A and Choice B) which corresponded to the correct Sound #4. Participants could listen to the stimuli as many times as they wished. Responses were presented by clicking on the buttons 'Choose A' or 'Choose B'. These buttons became active only after all 5 sounds had been played at least once. After a choice was presented, the next trial started.

depending on the group they were assigned to, each participant either performed the ART or AIT. For both these tasks, instructions were minimal. Participants were simply asked to listen carefully to the first 3 iterations and try to imagine how the fourth iteration would sound. They were then asked to select the correct fourth iteration among 2 alternatives (A complete

PowerPoint presentation with instructions can be found in [Supplementary Materials IV](#)). Importantly, participants were not explicitly instructed on the concepts of recursion or iteration, and had to extract these regularities by themselves while being exposed to stimuli examples. The overall procedure lasted approximately 45 min.



**Fig. 7.** (Non-recursive) iterative process generating a music fractal. Each iteration of this process adds one subordinate note (Level 4) for each dominant note (Level 3). The pitch interval between the three notes being added to Level 4 (within the duration of each note in Level 3) is kept constant. To solve this task, participants need to discriminate the pitch contour (ascending or descending) between the first two light gray notes (within each set of three) in Iteration 3, and generalize this contour to Iteration 4. Foil categories were similar to ART.

#### 2.1.4. Analysis

To compare the proportion of correct answers across groups and conditions, we used generalized linear mixed model analysis which was implemented in R 3.1.1 (R Core Team, 2015) using the glmer function from package lme4 to build the models (Bates, Maechler, Bolker, & Walker, 2015), the fitLmer function from package LMERConvenienceFunctions to select the best-fitting models (Tremblay & Ransijn, 2015) and the anova function from package car to obtain significance tests (Fox and Weisberg, 2011). Correlational analyses were done using Spearman's rho. To assess whether there were effects of learning, we tested whether accuracy and RT across trial were well explained by power curves (Anderson, 1982). To this effect we performed Curve Estimation analyses using SPSS 22 (IBM).

## 2.2. Results

### 2.2.1. Training

Prior to both ART and AIT, participants performed a simple training in which they had to discriminate the correct continuation of a (either ascending or descending) sequence of two tones (separated by either 4 or 8 semi-tones). Mean accuracy in the training was 77% in the group of musicians ( $SD = 24$ ), and 76% in the group of non-musicians ( $SD = 17$ ). The best fitting model included effects of 'foil' (Positional:  $b = 1.0$ ,  $SE = 0.16$ ; Repetition:  $b = 0.4$ ,  $SE = 0.15$ ,  $\chi^2(2) = 41.9$ ,  $p < 0.001$ ), 'tone distance' (major third:  $b = 0.5$ ,  $SE = 0.16$ ,  $\chi^2(1) = 15.7$ ,  $p < 0.001$ ) and 'directionality' (descending:  $b = -0.5$ ,  $SE = 0.13$ ,  $\chi^2(1) = 18.7$ ,  $p < 0.001$ ). Having musical training did not significantly affected accuracy in the training ( $b = 0.3$ ,  $SE = 0.4$ ,  $\chi^2(1) = 0.7$ ,  $p = 0.4$ ).

### 2.2.2. ART vs AIT: non-musicians

On average, participants scored 70.8% correct ( $SD = 18$ ) in ART and 70.1% in AIT ( $SD = 16\%$ ) (Fig. 8). Response accuracy was above chance in both tasks (binomial tests,  $p < 0.001$ ). Our best accuracy model converged to an overall effect of foil category ( $\chi^2(2) = 7.7$ ,  $p = 0.02$ ). We will probe into this effect in the "Rejection of foils section". Importantly, we found no accuracy differences between ART and AIT ( $b = -0.07$ ,  $SE = 0.25$ ,  $\chi^2(1) = 0.1$ ,  $p = 0.8$ ), and no effects related to other stimuli features (inter-tone, inter-level interval, and direction) (all  $ps > 0.1$ ).

Mean RT, measuring from the end of the last obligatory playback, was 37.5 s ( $SD = 10.7$ ) in ART and 31.1 s ( $SD = 6.1$ ) in AIT. The best-fit model also included effects of foil category ( $\chi^2(2) = 10.1$ ,  $p < 0.01$ ). There were no significant differences between tasks ( $b = -3.8$ ,  $SE = 2.4$ ,  $\chi^2(1) = 2.5$ ,  $p = 0.1$ ), and no other effects related to stimuli features (all  $ps > 0.1$ ).

### 2.2.3. Learning curves

In order to assess whether participants induced recursive and iterative rules, we analyzed performance across trials (1–24) and including all foil categories (repetition, positional and odd).

In ART, participants' proportion of correct answers increased with trial number, significantly fitting a power curve ( $F(1,22) = 8.0$ ,  $p = 0.01$ ,  $R^2 = 0.27$ ), and suggesting a learning effect (Fig. 9). In AIT, this effect was not significant ( $F(1,22) = 2.5$ ,  $p = 0.1$ ,  $R^2 = 0.10$ ). When we analyzed the evolution of RT across trials, we found that in both tasks RT became significantly shorter in later trials, also fitting a power curve ( $p < 0.001$ ,  $R^2 = 0.80$ , for both tasks).

In the musicians group, we found similar learning effects for both accuracy ( $F(1,22) = 7.0$ ,  $p = 0.015$ ,  $R^2 = 0.24$ ) and RT ( $F(1,22) = 177$ ,  $p < 0.001$ ,  $R^2 = 0.89$ ) in ART.

### 2.2.4. Rejection of foils

In ART, the best fit was an intercept model only (null model). Participants were equally likely to correctly reject the incorrect choice across all foil categories (Fig. 10) ( $\chi^2(2) = 0.5$ ,  $p = 0.8$ ). In contrast, in AIT, we found that the best model included an effect of foil category (odd:  $b = -0.38$ ,  $SE = 0.22$ ; repetition:  $b = 0.4$ ,  $SE = 0.24$ ,  $\chi^2(2) = 11.4$ ,  $p < 0.01$ ): participants were more likely to reject 'repetition foils' than 'odd' foils ( $p = 0.03$ ) (Fig. 10). For both tasks, we found that no other factor, such as directionality (ascending vs. descending), inter-note interval (4 vs. 8 semitones) and inter-level interval (4 vs. 8 semitones) was a significant predictor of accuracy in either of the tasks (all  $ps > 0.1$ ).

### 2.2.5. ART: Musicians vs. non-musicians

On average, the percentage of correct answers in ART was 84% ( $SD = 37$ ) in the group of musicians and 70.8% ( $SD = 18$ ) in the group of non-musicians (Fig. 8). The best-fitting model included a significant main effect of music training ( $b = 2.1$ ,  $SE = 0.5$ ,  $\chi^2(1) = 8.4$ ,  $p < 0.001$ ) and a significant interaction between foil and music training. In particular, musicians were worse in rejecting 'Repetition' foils than other kinds of foils ( $b = -2.0$ ,  $SE = 0.4$ ,  $\chi^2(2) = 32.6$ ,  $p < 0.001$ ). There were no other significant stimuli effects (inter-tone, inter-level interval, and direction) (all  $ps > 0.1$ ).

Mean RT was 26.8 s ( $SD = 20$ ) for musicians and 37.5 s ( $SD = 10.7$ ) for non-musicians. The best model included a main effect of music training ( $b = 10.1$ ,  $SE = 1.6$ ,  $\chi^2(1) = 14$ ,  $p < 0.001$ ), and no other significant terms (all  $ps > 0.1$ ).

## 2.3. Discussion

In this experiment we found that participants, both with and without musical training, are able to induce recursive rules

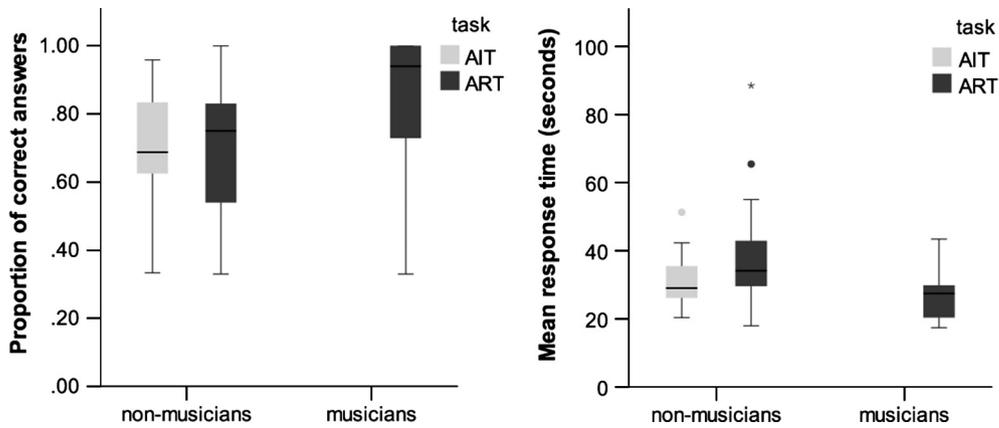


Fig. 8. Accuracy (left) and response times (right) in ART. The boxplot divides the scores into quartiles, the 'box' represents the distance from the 25th percentile to the 75th percentile (interquartile range). The horizontal dark line is the median. ° are outliers deviating from the box between 1.5 and 3 times the interquartile range; \* are outliers deviating from the box more than 3 times the interquartile range.

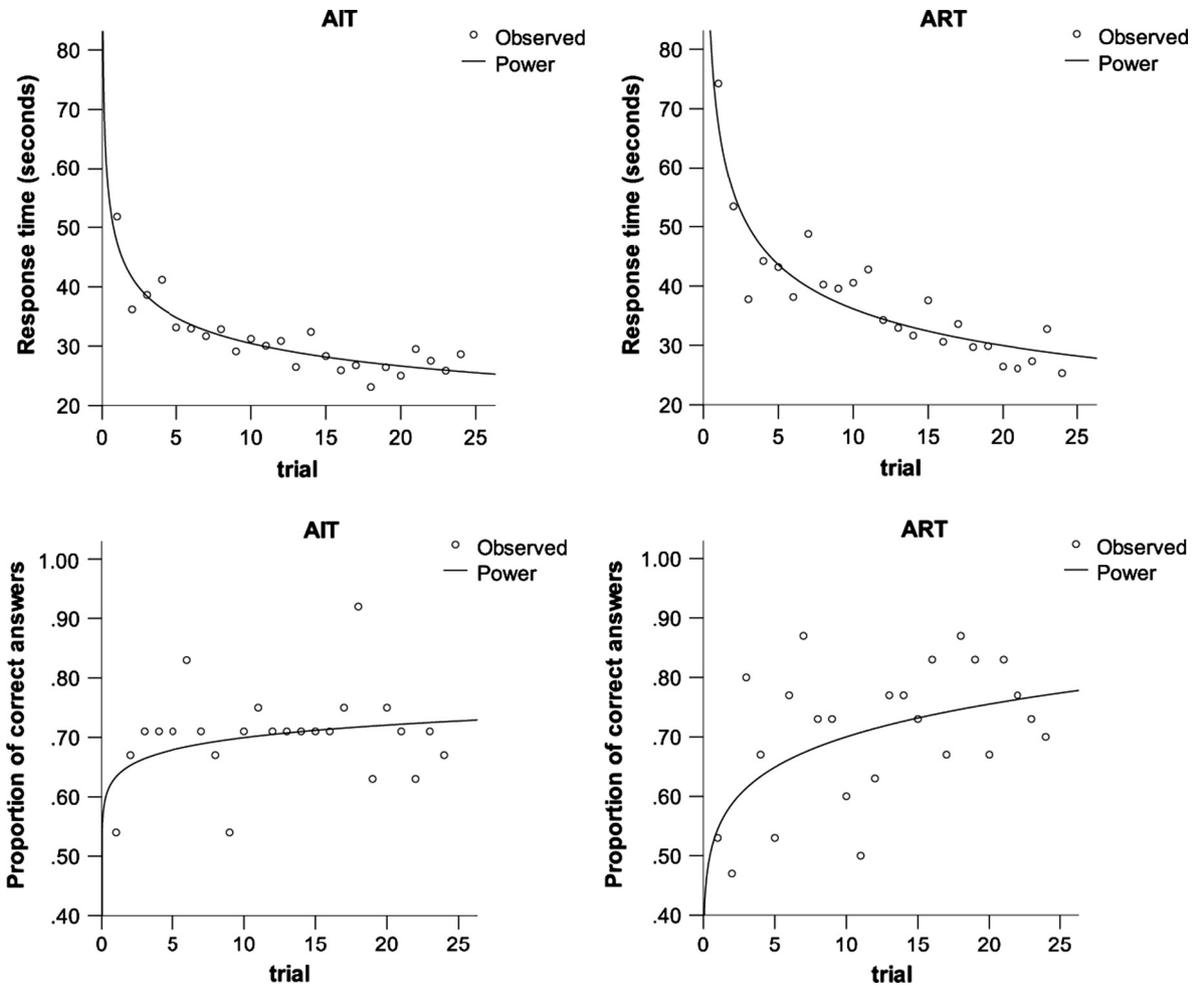
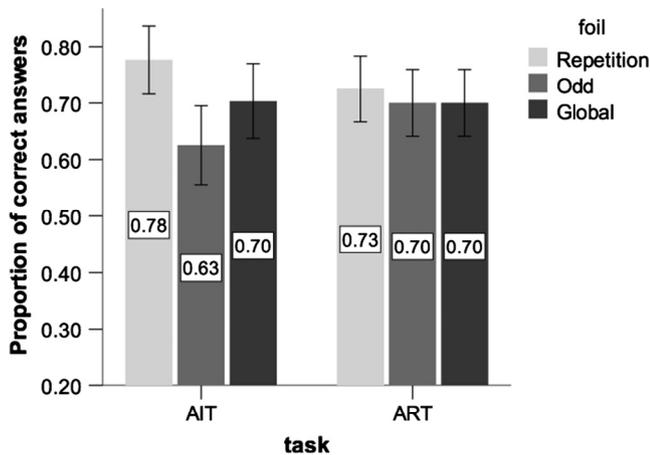


Fig. 9. Response time (top) and response accuracy (bottom) learning curves. Auditory Iteration Task (AIT; left); Auditory Recursion Task (ART; right). In this figure we have included only non-musicians, since performance was at ceiling for most musician participants.

governing the generation of music fractals, and to use these rules productively. We have shown that, without any feedback or explicit instructions, participants were able to learn and use recursive rules in the Auditory Recursion Task (ART). They were able to induce and transfer information from one trial to the next, as shown by the learning curves, and to consistently reject

incorrect continuations of recursive processes across three different foil categories. This suggests that no single simple auditory heuristic was used to solve the task.

We also introduced a control task, the Auditory Iteration Task (AIT), which is similar in many respects to ART. In particular, participants were asked to discriminate between similar pairs of



**Fig. 10.** Response accuracy across foil categories (non-musicians only). AIT (Auditory Iteration Task); ART (Auditory Recursion Task). In this analysis we have included only non-musicians.

choices (correct and foil) in AIT and ART, but the type of rules that were used to generate these choice items was different. In ART the generating rule is recursive, and each step generates a new level hierarchical level, while in AIT the generating rule is iterative, and performs transformations within a fixed level of the hierarchy.

When we compared performance across the two tasks, we found that accuracy was globally similar: participants performed adequately in both AIT and ART. However, we also found differences between the tasks: (1) the accuracy learning curve was somewhat steeper in ART than in AIT; (2) While participants rejected all foil categories equally well in ART, they were particularly bad at rejecting the ‘Odd’ foils in AIT.

These differences all align well with previous findings in the visual domain (Martins, 2012; Martins, Laaha, et al., 2014). The strategy used to acquire the recursive principles might differ from simple iteration (Martins, 2012). This strategy involves the induction of a rule, which seems to play a bigger role in ART than in AIT, as suggested by the accuracy learning curve (Dewar & Xu, 2010). The principle induced in ART allows participants to reject equally well all three different kinds of foils. In AIT, however, performance was less consistent across foil categories, suggesting that heuristic strategies might play a bigger role, although not an exclusive one, in solving this task. Partial heuristics are adequate to reject ‘Repetition’ foils (in which the foil is identical to the third iteration), but inadequate to reject ‘Odd’ foils, which are somewhat more subtle. The difficulty in rejecting ‘Odd’ foils while using iterative, but not recursive, representations of hierarchical structures is also present in the visual domain (Martins, Laaha, et al., 2014) and might be caused by a greater ability to process fine-grained details of hierarchical structures when using more abstract representations (Alvarez, 2011). We suggest that this is the case of recursion vs. iteration, as suggested by our previous behavioral, developmental and neuroimaging findings (Martins, 2012; Martins, Fischmeister, et al., 2014; Martins, Laaha, et al., 2014).

### 2.3.1. Musicians vs. non-musicians

Although non-musicians performed above chance in ART, their performance was not very high (71%), at least in comparison with the ability to perform an analogous task in the visuo-spatial domain (>84%). Interestingly, the accuracy level of musicians (84%) was similar to the accuracy of the general population in the visuo-spatial recursion task (Martins, 2012). This finding is interesting and suggests, in line with previous findings, that

specific practice and expertise effects could play a bigger role in the auditory domain than in the visual domain (Cohen et al., 2011). Even though the capacity to process hierarchies in both domains may be cognitively correlated, the processing of local (vs. global) features in the auditory domain (e.g., ‘direction’ of tone sequences) seems to be more difficult than the processing of local features in the visual domain (e.g., geometric shapes) (Bouvet, Rousset, Valdois, & Donnadieu, 2011). In that regard, musicians generally show smaller pitch interval discrimination thresholds than non-musicians, although the thresholds previously reported for non-musicians were also under 4 semitones, which was the smallest inter-note or inter-level interval used here (McDermott, Keebler, Michéyl, & Oxenham, 2010). Alternatively, these differences in difficulty might be related not to the domain itself (visual vs. auditory), but simply reflect the fact that, as a task, ART is more difficult than VRT. Perhaps if VRT were made more difficult it would similarly depend on practice effects. Future research investigating the capacity of visually sophisticated experts on a difficult version of VRT could address this question.

In experiment 2 we compare ART and AIT with their analog in the visual domain, and test our participants with specific melodic memory tasks. With this procedure we will assess how much of ART performance is explained by recursion itself, and how much is explained by general auditory or musical processing.

### 3. Experiment 2: Is auditory recursion domain-specific?

In Experiment 1, we have shown that human adults can perform adequately in an auditory recursion task and in an auditory iteration task. We have also shown that participants’ behavior was consistent with rule induction, especially in the Auditory Recursion Task (ART), given that there was consistent performance across items using different foil categories, and a consistent learning effect.

In this experiment we will attempt to investigate more closely what kind of rule is being induced in ART. To that effect, we will investigate the relationship between accuracy in ART and in other non-auditory recursive tasks, namely the Visual Recursion Task (VRT) (Martins et al., 2015) and the Tower of Hanoi task (ToH; which requires recursive action sequencing) (Goel & Grafman, 1995a; Halford, Wilson, & Phillips, 1998; Halford, Wilson, & Phillips, 2010). Performance in VRT and ToH was previously shown to be correlated (Martins, Fischmeister, et al., 2014) and they thus offer appropriate tools to tap into the ability to establish cognitive representations of recursion. A close correlation between ART, VRT and ToH would corroborate that ART taps into something specific to recursion.

However, ART is an auditory task which also requires the ability to perceive musical tone structure. To quantify these general musical and auditory expertise effects we included in the test battery our control Auditory Iteration Task, a Melodic Memory Task (MMT), which requires the ability to compare transposed tone contours (Mullensiefen, Gingras, Musil, & Stewart, 2014), and a (non-auditory) Visual Iteration Task (VIT), to control for variance explained by specific visual resources. In addition, we took into account the number of years of musical training of the participants.

In sum, with this test battery, comprising ART, AIT, ToH, VRT, VIT, MMT, combined with the number of years of musical training, we hoped to investigate whether, after controlling for effects specific to auditory and visual processing, ART was specifically correlated with other recursive tasks. If this were the case, this would support the hypothesis that our novel task taps into the ability to represent recursion in the auditory domain, and would be relevant to the question of whether recursion is domain-general or

domain-specific. The general methodology for this investigation involved exploratory correlations between task scores and a Principal Component Analysis.

### 3.1. Methods

#### 3.1.1. Participants

We tested 40 participants (24 females) aged between 18 and 49 years ( $M = 26$ ,  $SD = 7$ ). In contrast to Experiment 1, in this experiment we included participants with various levels of musical expertise, from no musical training to 16 years of musical training ( $M = 3.3$ ,  $SD = 4$ ). All participants were tested with all tasks and were paid 20 euros. All other admission and testing conditions were similar to experiment 1.

#### 3.1.2. Auditory recursion and auditory iteration tasks

We used exactly the same ART and AIT versions as in Experiment 1.

#### 3.1.3. Visual recursion and visual iteration tasks

The Visual Recursion Task (VRT) and Visual Iteration Task (VIT) were adaptations of the tasks used and described in detail elsewhere (Martins, 2012; Martins, Fischmeister, et al., 2014; Martins, Laaha, et al., 2014). These were visual analogs of ART and AIT. In both tasks, participants were exposed to a sequence of three images (Iterations 1, 2, and 3) which depicted a certain process generating a fractal. This process could be iterative or recursive (Fig. 11). After being exposed to the first three iterations, participants were asked to discriminate, from two choices, the image corresponding to the correct continuation of the previous

sequence of three (i.e. the fourth iteration). One of the choices was the correct image, and the other was a foil. Foil categories were identical to the ones used in ART and AIT (Odd, Positional, and Repetition).

Both VRT and VIT were composed of 27 trials, 9 of each foil category. Variability was achieved by varying the number of constituents composing the visual fractal, as described in (Martins, Laaha, et al., 2014).

#### 3.1.4. Tower of Hanoi (ToH)

ToH is a visuo-motor task widely used in clinical and cognitive assessment (Shallice, 1982). ToH requires the hierarchical movement of disks across pegs to solve puzzles, according to well-defined rules, and is best solved using a recursive strategy (Goel & Grafman, 1995a). We used an open-source computer version of ToH retrieved from The Psychological Experiment Building Language (PEBL; <http://pebl.sf.net/battery.html>; Mueller, 2011; Peirce, 2007). In this computerized version, participants were exposed to ten trials with increasing levels of difficulty. Level of difficulty was indexed by the minimum number of movements required to complete the trial. Participants were instructed to solve the trials mentally before starting to move the disks. The final score in this task was the length of the most challenging problem each participant could solve without mistakes. This score reflected the ability to generate and maintain recursive movement sequences before initiating the motor program necessary to solve the task (Halford et al., 1998; Halford et al., 2010).

#### 3.1.5. Melodic Memory Task (MMT)

This task was an authorized adaptation of the MMT to the Psychopy program (Peirce, 2007), originally part of the Goldsmiths Musical Sophistication Index battery (Gold-MSI) (Mullensiefen et al., 2014). The goal of this task is to assess participants' memory for short melodies. Participants were asked to listen to pairs of short melodies (containing between 10 and 17 notes) and to indicate whether the two melodies had an identical pitch interval structure or not (by selecting "same" or "different"). Participants also rated the confidence of their judgment on a 3-point scale ("I'm totally sure", "I think so", "I'm guessing"). In "same" trials, the second melody had the same pitch interval structure as the first one, but was transposed by a semitone or by a fifth. In "different" trials, in addition to being transposed, the second melody was modified by changing two notes by an interval varying between 1 and 4 semitones (for details, see Mullensiefen et al., 2014). The task was composed of 13 trials, including 2 initial training trials, and had a total duration of around 10 min. Accuracy was calculated as  $d'$  scores.

#### 3.1.6. Procedure

All participants performed ART, AIT, VRT, VIT, ToH and MMT. The procedure was divided in two blocks of 50 min each: the 'auditory block' - comprised of ART, AIT, and MMT - and the 'non-auditory block' - comprised of VRT, VIT and ToH. Each participant executed the two blocks with a break of 10 min in between. The order of the blocks was counter-balanced: 20 participants started the procedure with the 'auditory block' and 20 participants started with the 'non-auditory block'. The order of tasks within each block was also counter-balanced: 20 participants started both blocks with recursive tasks first (e.g., ART-MMT-AIT-break-VRT-ToH-VIT) and 20 participants started each block with iterative tasks first (e.g., AIT-MMT-ART, VIT-ToH-VRT). ToH and MMT were always performed in the middle of their blocks.

All tasks were executed in the same room and computer screen, and auditory stimuli were delivered through headphones. The procedure had a total duration of 2 h.

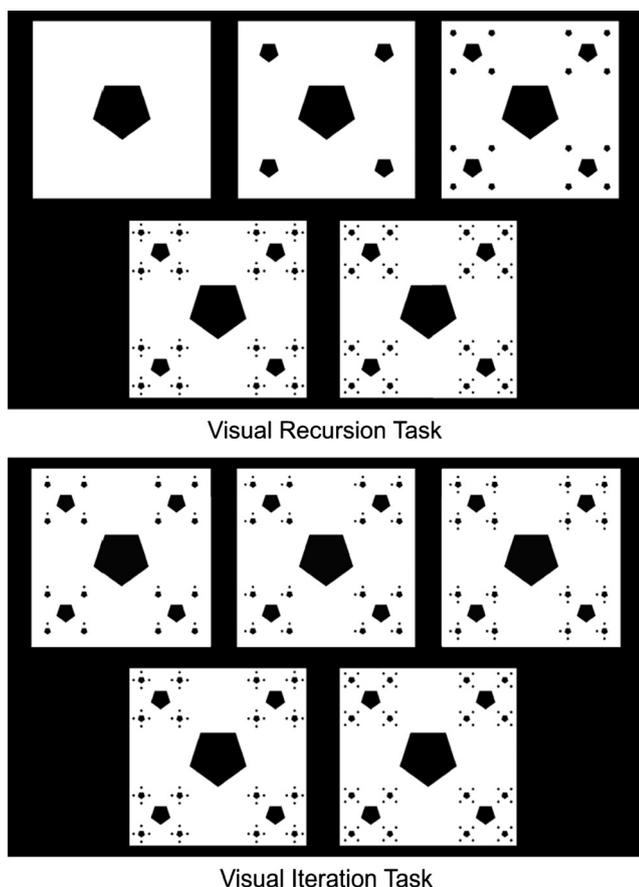


Fig. 11. Screenshots of the Visual Recursion Task (top) and the Visual Iteration task (VIT).

### 3.1.7. Analysis

First, we standardized the data by doing z-transformations. Then, we performed exploratory correlational analysis using Spearman correlations, and stepwise correlational analyses to determine the best model fit to predict both ART and AIT performance. We used the stepAIC function from the R library MASS (Venables & Ripley, 2002). Finally, we submitted all variables to a Principal Component Analysis (PCA) with Varimax rotation. The PCA was performed with SPSS 22.

## 3.2. Results

### 3.2.1. Replication of Experiment 1

In Experiment 2, we did not enforce strict inclusion criteria regarding music training since we were interested in parametric differences between participants. Thus, the sample included participants with a wide range of musical training (from 0 to 16 years). However, when we applied (post hoc) the same criteria to classify musicians and non-musicians as in Experiment 1, we obtained similar results: Musicians (8 years or more of musical training,  $N = 7$ ) were more accurate in ART ( $M = 96\%$ ,  $SD = 8\%$ ) than non-musicians (less than 2 years of musical training,  $N = 23$ ,  $M = 76\%$ ,  $SD = 22\%$ ) (Mann-Whitney U:  $z = -2.53$ ,  $p = 0.012$ ). We also found that musicians were more accurate in AIT ( $M = 90\%$ ,  $SD = 14\%$ ) than non-musicians ( $M = 73\%$ ,  $SD = 20\%$ ) (Mann-Whitney U:  $z = -2.00$ ,  $p = 0.05$ ). This confirms that the effect of musical expertise is present for both the recursive and iterative tasks.

### 3.2.2. Correlational analyses

We z-transformed all variables (raw data is presented in Table 1) to homogenize the scales, and performed exploratory correlation analyses. We found that performance in both ART and AIT was correlated with the number of years of music training ( $r_s = 0.38$ ,  $p = 0.01$ ;  $r_s = 0.34$ ,  $p = 0.04$ , respectively), with performance in the Melodic Memory ( $r_s = 0.36$ ,  $p = 0.02$ ;  $r_s = 0.33$ ,  $p = 0.04$ , respectively) and Visual Recursion tasks ( $r_s = 0.37$ ,  $p = 0.02$ ;  $r_s = 0.32$ ,  $p = 0.05$ , respectively) (see Table 2).

To assess whether there were specific differences between ART and AIT, we performed two stepwise regression analyses, with both AIT and ART as dependent variables. ART was best predicted by a model including AIT ( $b = 0.69$ ,  $t(35) = 8.8$ ,  $p < 0.001$ ), VRT ( $b = 0.15$ ,  $t(35) = 2.1$ ,  $p = 0.04$ ), ToH ( $b = 0.15$ ,  $t(35) = 1.5$ ,  $p = 0.14$ ) and Music training ( $b = 0.13$ ,  $t(35) = 1.4$ ,  $p = 0.18$ ) (model AIC =  $-40.84$ ; significance tests conducted on the predictors entered in the same order as above, using Type 1 sums of squares). In the inverse analysis, AIT was best predicted by a model including only ART ( $b = 0.8$ ,  $F(1,35) = 67.9$ ,  $p < 0.001$ ).

### 3.2.3. Principal component analysis

We then submitted all variables to a Principal Component Analysis. The analysis clustered the data into two Principal Components ( $KMO = 0.68$ , Bartlett's Test for Sphericity:  $\chi^2(21) = 75.6$ ,  $p < 0.001$ ):

the first explaining 41.6% of the variance and the second 15.8%. Results with Varimax rotation are depicted in Table 3.

With the exception of the Tower of Hanoi, all tasks had a loading coefficient higher than 0.35 on Principal Component 1 (PC1), which explained the most variance and may reflect general cognitive capacity. The tasks with higher loading in this component were AIT, ART, as well as music training, which hints of some predominance of auditory perception. On the other hand, Principal Component 2 (PC2) was dominated by ToH, which is known to be a recursive task (Halford et al., 1998; Halford et al., 2010). Both VRT and ART had higher loadings on this component, suggesting that PC2 might have some specificity for recursive tasks. Interestingly, MMT also had a high loading on this component, but not VIT, AIT or musical training. This suggests that in addition to general auditory processing, MMT might also require complex hierarchical processing.

## 3.3. Discussion

In these experiments, we sought to investigate whether the Auditory Recursion Task tapped into cognitive constructs specific to recursion. We did this by testing each participant with a battery of tests including ART, AIT, two non-auditory recursion tasks (VRT and ToH), a non-auditory non-recursive task (VIT), and a melodic memory task (MMT); we then performed several correlational analyses.

The first and core finding of our experiments was a confirmation that humans can represent recursion in the auditory domain. When we tested participants with different levels of musical expertise (from 0 to 16 years of music training) global performance in ART rose to 82%, in more musically-trained participants. We also confirmed that musical training was a major predictor of performance in both VRT and ART, as was the ability to detect changes in melodic contour (MMT). Notably, performance was higher in Experiment 2 than in Experiment 1, both for musicians and non-musicians, and both for ART and AIT. This might have been caused by task order effects, since in Experiment 2 participants performed six tasks instead of one. Consistent with this interpretation, when participants started the procedure with the 'visual block' (VRT, VIT and ToH), performance was on average 82% in AIT and 86% in ART. When participants started the procedure with the 'auditory block' (ART, AIT and MMT), performance was on average 74% in AIT and 78% in ART.

The second finding was that, when we controlled for the variance explained by a control auditory task – AIT – in all respects identical to ART but without recursion, we found that ART was specifically predicted by VRT and ToH (two non-auditory recursive tasks) (Halford et al., 1998; Halford et al., 2010; Martins et al., 2015). This result was confirmed by a PCA. The latter analysis yielded two major components, one relating with general cognitive capacity, strongly influenced by musical training and both AIT and ART, and another component, PC2, explaining 16% of the variance, which was dominated by the three recursive tasks. However, per-

**Table 1**

Descriptive data for all the measures included in the battery. ART (Auditory Recursion Task), VRT (Visual Recursion Task), AIT (Auditory Iteration Task), VIT (Visual Iteration Task), ToH (Tower of Hanoi), MMT (Melodic Memory Task), and Musical training.

	N	Minimum	Maximum	Mean	SD
ART	40	0.42	1.00	0.82	0.20
VRT	40	0.52	1.00	0.89	0.12
AIT	40	0.33	1.00	0.78	0.19
VIT	40	0.70	1.00	0.88	0.09
ToH	40	3.00	7.00	5.82	1.06
MMT (d')	40	-0.84	3.46	1.07	0.97
Music training (years)	40	0.00	16.00	3.27	4.02

**Table 2**  
Correlations between AIT (Auditory Iteration Task), ART (Auditory Recursion Task) and the other variables. MMT (Melodic Memory Task), MTrain (Music training), ToH (Tower of Hanoi), VIT (Visual Iteration Task), and VRT (Visual Recursion Task). \*\* $p < 0.01$  \* $p < 0.05$ . Correlation coefficients are Spearman's rho.

		AIT	MMT	MTrain	ToH	VIT	VRT
AIT	$r_s$		0.33*	0.34*	−0.00	0.17	0.32*
ART	$r_s$	0.83**	0.36*	0.38*	0.13	0.28	0.37*

**Table 3**  
Principal component analysis. AIT (Auditory Iteration Task), ART (Auditory Recursion Task), Music training, VIT (Visual Iteration Task), ToH (Tower of Hanoi), MMT (Melodic Memory Task) and VRT (Visual Recursion Task).

	Component	
	1	2
AIT	0.769	
ART	0.753	0.410
Music training (years)	0.732	
VIT	0.543	
ToH		0.838
MMT (d')	0.356	0.631
VRT	0.386	0.624

formance on the MMT task also loaded strongly on this putative 'domain-general' recursive component. This finding is intriguing and invites the speculation that perhaps the explicit processing of musical structure in melodies invokes some sort of domain-general hierarchical or recursive processing, as suggested by Lerdahl & Jackendoff (Jackendoff & Lerdahl, 2006; Lerdahl & Jackendoff, 1983; Lerdahl & Jackendoff, 1996). Alternatively, the component (PC2) comprising VRT, ART, ToH and MMT could be accounted by general intelligence in some way. However, we think that PC1 is more likely than PC2 to represent 'general capacity', given that all variables (VRT, VIT, ART, AIT, MMT and Music training) had moderate loadings on this component with the exception of ToH. Furthermore, memory for musical phrases (as tapped by MMT) has been shown to be somewhat independent of general intelligence (Pring, Woolf, & Tadic, 2008; Steinke, Cuddy, & Holden, 1997); and IQ seems to be also weakly correlated with VRT (Martins, Fischmeister, et al., 2014) and ToH (Bishop, Aamodt-Leeper, Creswell, McGurk, & Skuse, 2001; Byrnes & Spitz, 1979; Goel & Grafman, 1995b), even though IQ might explain a small proportion of variance in these tasks (Martins, 2012; Zook, Davalos, DeLosh, & Davis, 2004). Taken together, previous research and current findings make PC2 unlikely to reflect a 'general intelligence' component independent of PC1.

Finally, when we accounted for the shared variance with ART, we did not find specific correlations between the AIT and other cognitive measures, including with VIT. Instead, this task was strongly related with other auditory measures. This suggests that the auditory processing of simple iteration relies more on modality-specific resources that do not generalize across domains.

#### 4. General discussion

In this article we introduced two novel tasks – AIT and ART – designed to test participants' ability to represent iterative and recursive processes in the auditory domain. Through a series of two experiments, we have shown that human adults can process and represent the recursive structure underlying the generation of music fractals, and can use this information to discriminate between well-formed hierarchies and multiple categories of foils.

In the first experiment, we found that performance was consistent with the induction of an abstract rule: participants could correctly reject different foil categories and the increase in accuracy across trials was highly suggestive of a learning effect. Second,

we found that musical expertise had a significant impact on the ability to extract a recursive rule from sequences of tones. While musicians displayed an accuracy level similar to levels previously shown by the general population in a visual recursion task (Martins, 2012), non-musicians had a significantly lower score (71%). This specific difficulty in the auditory domain could be caused by proximal factors, involving attention to or perception of tone sequences (Bigand & Poulin-Charronnat, 2006; McDermott et al., 2010; Tervaniemi et al., 2005) and/or distal factors, more related to the extraction of abstract principles from musical stimuli (Besson & Faïta, 1995; Besson et al., 1994; Cook, 1987; Tillmann, 2012; Tillmann et al., 1988). Performance could also be constrained by auditory memory limitations in non-musicians (Cohen et al., 2011). Interestingly, non-musicians' performance was already low in AIT, a task simpler than ART, in which participants only had to discriminate the correct directionality (ascending or descending) of a sequence of three tones. This suggests that the difficulty of non-musicians was partially mediated by proximal factors related with the ability to perceive melodic structure even in very simple stimuli (Bigand & Poulin-Charronnat, 2006) or to discriminate between different pitch intervals (McDermott et al., 2010), possibly related to a difficulty directing attention to the relevant acoustic parameters (Tervaniemi et al., 2005).

In the second experiment, we applied a larger test battery to a new set of participants; the battery included not only auditory recursive and iterative tasks, but also visual recursion, visual iteration, melodic memory and Tower of Hanoi (a visuo-motor recursive task). The results of this experiment strongly suggest that ART correlates well with other recursive abilities, but that it also depends on general capacities to process auditory stimuli. This finding is consistent with previous research showing that global-to-local interference effects in the processing of hierarchies is significantly correlated between visual and music modalities (Bouvet et al., 2011). Here we have extended this result to the representation of recursion.

##### 4.1. Is recursion domain-general?

Recursion as a cognitive capacity allows the generation of multiple hierarchical levels with a single rule (Fitch, 2010; Hulst, 2010; Martins, 2012). This makes recursion especially useful in the processing and generation of complex hierarchies, with a specific potential to build hierarchies of very large depth (Hauser et al., 2002; Koike & Yoshihara, 1993; Schiemenz, 2002). Recursion has been claimed to be present in many domains (Alegre & Gordon, 1996; Arsenijević & Wolfram, 2010; Busi, Gabbriellini, & Zavattaro, 2009; Chomsky, 1995; de Vries, Christiansen, & Petersson, 2011; Eglash, 1997; Eglash, 1998; Eisenberg, 2008; Fitch, 2010; Fitch et al., 2005; Hauser et al., 2002; Hulst, 2010; Hunyady, 2010; Jackendoff & Lerdahl, 2006; Karlsson, 2010; Koike & Yoshihara, 1993; Komara, 2011; Latek, Axtell, & Kaminski, 2009; Laury & Ono, 2010; Levinson, 2013; Martins et al., 2015; Odifreddi, 1999; Roeper, 2011; Schiemenz, 2002; Schreuder et al., 2009; Wagner, 2010). However, it seems to be most frequently (and universally) used in the domain of language (Arsenijević & Wolfram, 2010; Goldberg, 2003), and available in this domain from an early age

(Alegre & Gordon, 1996; Roeper, 2011). The latter facts have led some theorists to postulate that recursion might have evolved in the domain of language first, only later becoming available to other domains, with the caveat that its use would be dependent on language resources (Fitch et al., 2005; Hauser et al., 2002).

In previous research, we have shown that recursion is available in the visuo-spatial domain (Martins, 2012) and that its use does not require language resources, as evidenced by interference (Martins et al., 2015), developmental (Martins, Laaha, et al., 2014) and brain imaging studies (Martins, Fischmeister, et al., 2014). But even if classic language resources are not necessary to process recursion in vision, it remained an open question whether recursion is domain-specific or domain-general. The data presented here strongly suggest the latter. In this study we show for the first time not only that recursion can be processed in the non-linguistic auditory domain, but that the capacity to form recursive representations strongly correlates with the same ability in the visual and action sequencing domains. Notably, this domain generality was *not* present for the control Auditory Iteration Task, which did not correlate with the Visual Iteration Task. Thus, despite the obvious influence of domain-specific constraints in the processing of auditory information, probably located at the interface between the sensory level and higher cognition (Cohen et al., 2011; Tervaniemi et al., 2005), the capacity to build recursive representations might be instantiated by a more general and abstract code, thus constituting a core computational capacity available to many different domains of cognition.

In the future we intend to better situate the current results in the broader debate of language evolution, by incorporating specific linguistic recursion tasks into our test battery. We will also perform neuroimaging experiments with these auditory tasks and compare the results with the networks active during visual recursion. This research will increase our understanding of this potentially human-specific capacity, and further specify its role in our unique cognitive architecture.

## Acknowledgments

This work was supported by an FCT grant SFRH/BD/64206/2009 to MM, by ERC Advanced Grant SOMACCA, project number 230604 to WTF, and by a research cluster grant “Shared Neural Resources for Music and Language” to WTF (University of Vienna). The authors declare no competing financial interests.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2017.01.001>.

## References

- Alegre, M. A., & Gordon, P. (1996). Red rats eater exposes recursion in children's word formation. *Cognition*, 60, 65–82.
- Altmann, C. F., Bühlhoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology*, 13, 342–349.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131. <http://dx.doi.org/10.1016/j.tics.2011.01.003>.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369–406.
- Arsenijević, B., & Wolfram, H. (2010). Recursion as a human universal and as a primitive. *Biolinguistics*, 4(2–3), 165–173.
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, 12(5), 193–200. <http://dx.doi.org/10.1016/j.tics.2008.02.004>.
- Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, 10(9), 659–669. <http://dx.doi.org/10.1038/nrn2667>.
- Badre, D., Hoffman, J., Cooney, J. W., & D'Esposito, M. (2009). Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nature Neuroscience*, 12(4), 515–522. <http://dx.doi.org/10.1038/nn.2277>.
- Bahlmann, J., Schubotz, R., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *NeuroImage*, 42(2), 525–534. <http://dx.doi.org/10.1016/j.neuroimage.2008.04.249>.
- Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1. <http://dx.doi.org/10.18637/jss.v067.i01>.
- Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical Classification by Rank and Kinship in Baboons. *Science*, 302(5648), 1234–1236. <http://dx.doi.org/10.1126/science.1087513>.
- Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences*. <http://dx.doi.org/10.1016/j.tics.2012.12.002>.
- Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. US: American Psychological Association. <http://dx.doi.org/10.1037/0096-1523.21.6.1278>.
- Besson, M., Faïta, F., & Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience Letters*, 168(1–2), 101–105. [http://dx.doi.org/10.1016/0304-3940\(94\)90426-X](http://dx.doi.org/10.1016/0304-3940(94)90426-X).
- Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, 100(1), 100–130. <http://dx.doi.org/10.1016/j.cognition.2005.11.007>.
- Bishop, D. V. M., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D. H. (2001). Individual differences in cognitive planning on the tower of hanoi task: neuropsychological maturity or measurement error? *Journal of Child Psychology and Psychiatry*, 42(4), 551–556.
- Bouvet, L., Rousset, S., Valdois, S., & Donnadiou, S. (2011). Global precedence effect in audition and vision: Evidence for similar cognitive styles across modalities. *Acta Psychologica (Amst)*, 138(2), 329–335. <http://dx.doi.org/10.1016/j.actpsy.2011.08.004>.
- Brown, S. (2000). The ‘musilanguage’ model of music evolution. In N. E. A. Wallin (Ed.), *The origins of music* (pp. 271–300). MIT Press.
- Busi, N., Gabrielli, M., & Zavattaro, G. (2009). On the expressive power of recursion, replication and iteration in process calculi. *Mathematical Structures in Computer Science*, 19(06), 1191. <http://dx.doi.org/10.1017/s096012950999017x>.
- Byrnes, M. M., & Spitz, H. H. (1979). Developmental progression of performance on the Tower of Hanoi problem. *Bulletin of the Psychonomic Society*, 14(5), 379–381.
- Chomsky, N. (1995). *The minimalist program*. MIT Press.
- Cohen, M., Evans, K., Horowitz, T., & Wolfe, J. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic Bulletin & Review*, 18(3), 586–591. <http://dx.doi.org/10.3758/s13423-011-0074-0>.
- Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, 5(12), 539–546.
- Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception: An Interdisciplinary Journal*, 5, 197–205.
- de Vries, M. H., Christiansen, M. H., & Petersson, K. M. (2011). Learning recursion: Multiple nested and crossed dependencies. *Biolinguistics*, 5(1–2), 10–35.
- Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science*, 21(12), 1871–1877. <http://dx.doi.org/10.1177/0956797610388810>.
- Eglash, R. (1997). Bamana sand divination: Recursion in ethnomathematics. *American Anthropologist*, 99(1), 112–122. <http://dx.doi.org/10.1525/aa.1997.99.1.112>.
- Eglash, R. (1998). Fractals in African settlement architecture. *Complexity*, 4(2), 21–29.
- Eisenberg, M. (2008). Recursion—or, better computational thinking through laughter. *International Journal of Computers for Mathematical Learning*, 13, 171–174.
- Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, 1169(1), 448–458. <http://dx.doi.org/10.1111/j.1749-6632.2009.04582.x>.
- Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1), 173–215. <http://dx.doi.org/10.1016/j.cognition.2005.11.009>.
- Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598), 1933–1955. <http://dx.doi.org/10.1098/rstb.2012.0103>.
- Fitch, W. T., Hauser, M., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, 97(2), 179–210. <http://dx.doi.org/10.1016/j.cognition.2005.02.005>.
- Fitch, W. T. (2010). Three meanings of “recursion”: Key distinctions for biolinguistics. In R. K. Larson, V. M. Déprez, & H. Yamakido (Eds.), *The evolution of human language* (pp. 73–90). New York: Cambridge University Press.
- Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, 1316, 87–104. <http://dx.doi.org/10.1111/nyas.12406>.
- Fox, J., & Weisberg, S. (2010). *An R companion to applied regression*. Sage.
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological Reviews*, 91, 1357–1392. <http://dx.doi.org/10.1152/physrev.00006.2011>-Language.
- Friederici, A. D., Bahlmann, J., Friedrich, R., & Makuuchi, M. (2011). The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics*, 5(1–2), 087–104.

- Goel, V., & Grafman, J. (1995a). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, 33(5), 623–642.
- Goel, V., & Grafman, J. (1995b). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, 33(5), 623–642. [http://dx.doi.org/10.1016/0028-3932\(95\)90866-P](http://dx.doi.org/10.1016/0028-3932(95)90866-P).
- Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5), 219–224. [http://dx.doi.org/10.1016/S1364-6613\(03\)00080-9](http://dx.doi.org/10.1016/S1364-6613(03)00080-9).
- Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, 306(5695), 496–499. <http://dx.doi.org/10.1126/science.1094492>.
- Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*, 111(3), 1340–1348.
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, 32, 803–865.
- Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, 14(11), 497–505. <http://dx.doi.org/10.1016/j.tics.2010.08.005>.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.
- Hofstadter, D. R. (2000). Analogy as the core of cognition. In J. C. a. j. Gleick (Ed.), *The Best American Science Writing* (pp. 116–144).
- Hulst, H. v. d. (2010). Re recursion. In H. v. d. Hulst (Ed.), *Recursion and human language*. Berlin/New York: de Gruyter Mouton.
- Hunyady, L. (2010). Cognitive grouping and recursion in phonology. In H. v. d. Hulst (Ed.), *Recursion and human language*. Berlin/New York: de Gruyter Mouton.
- Jackendoff, R. (2003). Précis of foundations of language: Brain, meaning, grammar, evolution. *Behavioral and Brain Sciences*, 26, 651–707.
- Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception: An Interdisciplinary Journal*, 26(3), 195–204. <http://dx.doi.org/10.1525/mp.2009.26.3.195>.
- Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, 100(1), 33–72.
- Jeon, H. A. (2014). Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. *Frontiers in Systems Neuroscience*, 8, 223. <http://dx.doi.org/10.3389/fnysys.2014.00223>.
- Karlssof, F. (2010). Syntactic recursion and iteration. In H. v. d. Hulst (Ed.), *Recursion and human language* (pp. 43–67). Berlin/New York: de Gruyter Mouton.
- Koelsch, S. (2006). Significance of Broca's area and ventral premotor cortex for music-syntactic processing. *Cortex*, 42(4), 518–520. [http://dx.doi.org/10.1016/S0010-9452\(08\)70390-3](http://dx.doi.org/10.1016/S0010-9452(08)70390-3).
- Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America*, 110(38), 15443–15448. <http://dx.doi.org/10.1073/pnas.1300272110>.
- Koike, H., & Yoshihara, H. (1993). *Fractal approaches for visualization of huge hierarchies*. Paper presented at the Proceedings of the 1993 IEEE Symposium on Visual Languages.
- Komara, J. (2011). On nested simple recursion. *Archive for Mathematical Logic*, 50(5–6), 617–624. <http://dx.doi.org/10.1007/s00153-011-0236-9>.
- Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12(4), 217–230. <http://dx.doi.org/10.1038/nrn3008>.
- Latek, M., Axtell, R., & Kaminski, B. (2009). *Bounded rationality via recursion*. Paper presented at the Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems – Volume 1, Budapest, Hungary.
- Laury, R., & Ono, T. (2010). Recursion in conversation: What speakers of Finnish and Japanese know how to do. In H. v. d. Hulst (Ed.), *Recursion and human language*. de Gruyter Mouton.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative grammar of tonal music*. Cambridge, MA.
- Lerdahl, F., & Jackendoff, R. (1996). *A generative theory of tonal music*. The MIT Press.
- Levinson, S. (2013). Recursion in pragmatics. *Language*, 89(1), 149–162. <http://dx.doi.org/10.1353/lan.2013.0005>.
- Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, 4(5), 540–545. [http://www.nature.com/neuro/journal/v4/n5/supinfo/n0501\\_540\\_S1.html](http://www.nature.com/neuro/journal/v4/n5/supinfo/n0501_540_S1.html).
- Mandelbrot, B. (1977). *The fractal geometry of nature*. Freeman.
- Martins, M. D. (2012). Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598), 2055–2064. <http://dx.doi.org/10.1098/rstb.2012.0097>.
- Martins, M. D., & Fitch, W. T. (2012). *Empirical approaches to recursion*. Paper presented at the The Evolution of Language – Proceedings of the 9th International conference, Kyoto.
- Martins, M. D., Fischmeister, F. P., Puig-Waldmuller, E., Oh, J., Geissler, A., Robinson, S., ... Beisteiner, R. (2014). Fractal image perception provides novel insights into hierarchical cognition. *NeuroImage*, 96, 300–308. <http://dx.doi.org/10.1016/j.neuroimage.2014.03.064>.
- Martins, M. D., Laaha, S., Freiberger, E. M., Choi, S., & Fitch, W. T. (2014). How children perceive fractals: Hierarchical self-similarity and cognitive development. *Cognition*, 133(1), 10–24. <http://dx.doi.org/10.1016/j.cognition.2014.05.010>.
- Martins, M. D., Mursic, Z., Oh, J., & Fitch, W. T. (2015). Representing visual recursion does not require verbal or motor resources. *Cognitive Psychology*, 77, 20–41. <http://dx.doi.org/10.1016/j.cogpsych.2015.01.004>.
- Massen, J. J. M., Pašukonis, A., Schmidt, J., & Bugnyar, T. (2014). Ravens notice dominance reversals among conspecifics within and outside their social group. *Nature Communications*, 5. <http://dx.doi.org/10.1038/ncomms4679>.
- McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *The Journal of the Acoustical Society of America*, 128(4), 1943–1951.
- Miller, S. A. (2009). Children's understanding of the second-order mental states. *Psychological Bulletin*, 135(5), 749–773.
- Mithun, M. (2010). The fluidity of recursion and its implications. In H. v. d. Hulst (Ed.), *Recursion and human language* (pp. 17–41). Berlin/New York: de Gruyter Mouton.
- Mueller, S. T. (2011). *The psychology experiment building language*.
- Mullensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9(2), e89642. <http://dx.doi.org/10.1371/journal.pone.0089642>.
- Odfreddi, P. G. (1999). *Classical recursion theory*. Elsevier.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6(7), 674–681.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10(6), 717–733.
- Peirce, J. W. (2007). PsychoPy – Psychophysics software in python. *Journal of Neuroscience Methods*, 162(1–2), 8–13.
- Picard, O., Massé, A. B., Hanard, S., Marcotte, O., Chicoinse, G., & Gargouri, Y. (2010). *Hierarchies in dictionary definition space*. arXiv:0911.5703v1 [cs.CL].
- Pring, L., Woolf, K., & Tadic, V. (2008). Melody and pitch processing in five musical savants with congenital blindness. *Perception*, 37(2), 290–307.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351–360. <http://dx.doi.org/10.1038/nrn2811>.
- Rammeyer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal*, 24(1), 37–48.
- Reboul, A. (2012). Language: Between cognition, communication and culture. *Pragmatics & Cognition*, 20(2), 295–316. <http://dx.doi.org/10.1075/pc.20.2.06reb>.
- Roeper, T. (2009). *The minimalist microscope: How and where interface principles guide acquisition*. Paper presented at the Boston University Conference on Child Language.
- Roeper, T. (2011). The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics*, 5(1–2), 57–86.
- Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1), 35–53. <http://dx.doi.org/10.1080/17459737.2011.573676>.
- Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., ... Schulze-Bonhage, A. (2009). Overlap of musical and linguistic syntax processing: Intracranial ERP evidence. *Annals of the New York Academy of Sciences*, 1169(1), 494–498. <http://dx.doi.org/10.1111/j.1749-6632.2009.04792.x>.
- Schiemenz, B. (2002). *Managing complexity by recursion*. Paper presented at the Cybernetics And Systems, Wien.
- Schreuder, M., Gilbers, D., & Quené, H. (2009). Recursion in phonology. *Lingua*, 119(9), 1243–1252. <http://dx.doi.org/10.1016/j.lingua.2009.02.007>.
- Semal, C., & Demany, L. (1990). The upper limit of "musical" pitch. *Music Perception: An Interdisciplinary Journal*, 8(2), 165–175.
- Seyfarth, R. M., & Cheney, D. L. (2008). Primate social knowledge and the origins of language. *Mind & Society*, 7(1), 129–142. <http://dx.doi.org/10.1007/s11299-007-0038-2>.
- Seyfarth, R. M., Silk, J. B., & Cheney, D. L. (2014). Social bonds in female baboons: The interaction between personality, kinship and rank. *Animal Behaviour*, 87, 23–29. <http://dx.doi.org/10.1016/j.anbehav.2013.10.008>.
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 298(1089), 199–209.
- Steinke, W. R., Cuddy, L. L., & Holden, R. R. (1997). Dissociation of musical tonality and pitch memory from nonmusical cognitive abilities. *Canadian Journal of Experimental Psychology*, 51(4), 316–334.
- ten Cate, C., & Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: Natural vocalizations and artificial grammar learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598), 1984–1994. <http://dx.doi.org/10.1098/rstb.2012.0055>.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, 161(1), 1–10. <http://dx.doi.org/10.1007/s00221-004-2044-5>.
- Tillmann, B. (2012). Music and language perception: expectations, structural integration, and cognitive sequencing. *Topics in Cognitive Science*, 4(4), 568–584. <http://dx.doi.org/10.1111/j.1756-8765.2012.01209.x>.
- Tillmann, B., & Bigand, E. (2004). The relative importance of local and global structures in music perception. *The Journal of Aesthetics and Art Criticism*, 62(2), 211–222.
- Tillmann, B., Bigand, E., & Madurell, F. (1988). Local versus global processing of harmonic cadences in the solution of musical puzzles. *Psychologische Forschung*, 61, 157–174.

- Tomasello, M. (2008). *Origins of human communication*. Cambridge: MIT Press.
- Tremblay, A., and Ransijn, J. (2015). LMERConvenienceFunctions: Model selection and post-hoc analysis for (G)LMER models. R Package Version 2.10.
- Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes*, 78(3), 313–334. <http://dx.doi.org/10.1016/j.beproc.2008.02.017>.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). 0-387-95457-0. New York: Springer.
- Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *The Journal of the Acoustical Society of America*, 105(1), 339–346.
- Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*, 28(1), 183–237. <http://dx.doi.org/10.1007/s11049-009-9086-0>.
- Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know your place: Neural processing of social hierarchy in humans. *Neuron*, 58(2), 273–283. <http://dx.doi.org/10.1016/j.neuron.2008.01.025>.
- Zook, N. A., Davalos, D. B., DeLosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition*, 56(3), 286–292. <http://dx.doi.org/10.1016/j.bandc.2004.07.003>.