British Journal of Developmental Psychology (2004), 22, 569–583

www.bps.org.uk

The development of mental models for auditory events: Relational complexity and discrimination of pitch and duration

Catherine Stevens* and Melinda Gallagher

School of Psychology & MARCS Auditory Laboratories, University of Western Sydney, Australia

This experiment investigated relational complexity and relational shift in judgments of auditory patterns. Pitch and duration values were used to construct two-note perceptually similar sequences (unary relations) and four-note relationally similar sequences (binary relations). It was hypothesized that 5-, 8- and 11-year-old children would perform unary level pitch and duration discrimination tasks accurately. Relational shift predicted a poorer performance of the younger age groups on binary relation tasks; relational primacy predicted no effect of age. Accuracy was operationalized as a discrimination index (DI: hit rate minus false alarm rate). Results supported relational shift: DI for all age groups exceeded chance on unary and binary relation tasks, with significantly poorer performance by all age groups on binary relation tasks. The 5-years age group showed evidence of perceptual similarity. Relational complexity of auditory dimensions and tasks, and manipulation of domain specific musical knowledge in evaluating theories of relational processing, are discussed.

Children's ability to perceive, produce and recognize patterns in pitch and time is central to their enjoyment, understanding and production of music. Developmental studies of music perception and cognition indicate that young infants demonstrate sensitivity to melodic contour (the pattern of ups and downs in pitch) and simple frequency ratios or musical intervals (Morrongiello, Trehub, Thorpe, & Capodilupo, 1985; Schellenberg & Trehub, 1996; Trehub, 1993; Trehub, Schellenberg, & Hill, 1997), can discriminate between frequencies (Olsho, Schoon, Sakai, Turpin, & Sperduto, 1982) and can detect temporal or rhythmic changes (Dowling, 1999; Morrongiello, 1984; Trehub & Thorpe, 1989). By 4–5 years, children can discriminate the register of pitches, although discriminating between the same and different easy tonal or rhythm

*Correspondence should be addressed to Catherine Stevens, School of Psychology & MARCS Auditory Laboratories, University of Western Sydney – Bankstown, Locked Bag 1797, South Penrith DC NSW 1797, Australia (e-mail: kj.stevens@uws.edu.au) http://marcs.uws.edu.au.

patterns is not achieved until some 12 months later (Drake, 1993; Gérard & Drake, 1990; Shuter-Dyson & Gabriel, 1981). Results such as these suggest that there is gradual and systematic development in the perception and cognition of organized auditory patterns and events. Our aim is to apply a theory of cognitive development (Halford, 1993) to the auditory domain and investigate empirically assumptions of perceptual versus relational processing. We examine the efficacy of Halford's view of relational complexity in predicting and explaining children's discrimination of increasingly complex non-verbal auditory patterns.

Halford's (1993) theory of cognitive development is characteristic of a class of neo-Piagetian theories (e.g. Case, 1985; Pascual-Leone, 1970; see also Case, 1992). The essence of Halford's theory is the notion that children's thought can be considered in terms of mental models and the operations on them. The tenet that children's cognitive development involves the ability to cognize relations, and higher-order relations of relations, is also common to accounts proposed by Case (1992), Gentner (e.g. Rattermann & Gentner, 1998a, 1998b) and Goswami (1989, 1992b, 1998; Goswami *et al.*, 1998). Halford's theory is taken as a particular framework of general assumptions about relational processing and, for the first time, is tested in an auditory and temporal context. It is essential that a theory of cognitive development be evaluated using temporal patterns. Elman (1990) noted that 'problems may change their nature when cast in temporal form' (p. 186), and Goswami (1998) has emphasized the importance of temporal order and causal relations as organizing principles in children's memory. An experiment is reported that investigates the effect of relational complexity (Andrews & Halford, 2002) and age on auditory discrimination tasks wherein relational complexity is realized using patterns of varying values of pitch or duration. Although the controlled nature of the pitch and duration features limit the musicality of the stimuli and tasks, the results have implications for music training and education, and highlight the need to consider temporal and auditory processes in developmental theories.

Halford's theory of cognitive development

Halford's (1993) theory of development provides a testable framework that may be applied to analysis of children's understanding of musical patterns. Understanding, Halford argues, entails having a mental model that represents the structure of a concept or phenomenon – a representation is a mapping from a symbol system to an environment system (Halford & Wilson, 1980). The symbol system expresses the structure of a representation, whereas the environment system expresses the structure of the aspect of the world that is represented. Representation elements are mapped into elements in the world such that functions (relations, transformations) between representation elements correspond to functions (relations, transformations) between environment elements. Halford conceives of children's increasing understanding as the development of representations that are mental models.

Representations differ in their dimensionality (Andrews & Halford, 2002; Halford, Wilson, & Phillips, 1998); dimensionality is defined as the number of independent units of information required to represent a concept. One-dimensional concepts are defined as predicates with one argument, or as unary relations of the form $r(x)$. Category membership, such as CAT(Max) asserting that Max is a cat, or attribute-object bindings such as RED(car) are examples of unary relations. Two-dimensional tasks are defined as predicates with two arguments or as binary relations, $r(x,y)$. An example is the binary relation LARGER THAN(elephant, dog). Three-dimensional concepts consist of predicates with three arguments or as ternary relations, $r(x,y,z)$. Transitivity or a set of ordered triples, such as Jim is happier than Paul, Paul is happier than Dave, so Jim is happier than Dave, is an example of a three-dimensional concept expressible as a ternary relation. Four-dimensional concepts are defined as predicates with four arguments or as quaternary relations, $r(w,x,y,z)$. Examples are the composition of binary operations such as $a(b + c) = d$, or the proportion $a / b = c / d$ entailing relations between the four terms a, b, c, d (Halford *et al.*, 1998).

Cognitive tasks that involve the processing of relations and cognitive capacity limitations in children (and in adults) can be defined by the complexity of relations that can be processed in parallel (Halford, 1993; Halford *et al.*, 1998). A problem becomes more complex as the number of interacting factors increases. For example, problems that entail a binary relation are simpler than those that entail a ternary relation. Based on children's ability to complete written language and spatial tasks involving analogical reasoning and transitive inference, Halford argued that age-related increases in processing capacity enable children to process relations of increasing complexity with children acquiring the capacity necessary for unary, binary, ternary and quaternary relations at median ages of 1, 2, 5 and 11 years, respectively (Andrews & Halford, 1998, 2002; Halford, 1989, 1993; Halford & Macdonald, 1977; Maybery, Bain & Halford, 1986). Processing capacity refers to performing at a level greater than that predicted by chance. Capacity to perform a task is not all or none but has gradations. For example, a binary relational task may be performed with 98% accuracy, ternary with 70% accuracy, and quaternary at no better than chance (Andrews & Halford, 2002). Development, in terms of mental models and structure mappings, involves increases in capacity that allow children to 'become able to construct progressively more complex structure mappings as they grow older' (Halford, 1993, p. 129).

Many agree that cognitive development may be examined by scrutinizing the age at which children can process tasks that involve relations of increasing complexity (Case, 1992; Mondloch, Geldart, Maurer & Le Grand, 2003; Wiese, 2003). However, there are contentious issues associated with some of the assumptions made in Halford's (1993) theory. Goswami, Leevers, Pressley and Wheelwright (1998) distinguished between children's ability to reason analogically when presented with single (binary) versus double (ternary) relations and concluded that the number of relations in an analogy does not overload capacity (see also Goswami, 1992a). Goswami *et al.* emphasized the importance of relational familiarity providing evidence of 'learning-to-learn' effects with children's analogical reasoning performance improving significantly during the time course of an experiment. Gentner, Rattermann, Markman and Kotovsky (1995) have argued that the transition from judging similarity of entities based on objects involved, to judgment based on relations between those objects, relies on domain-specific knowledge and not development of global competence or processing capacity. Rattermann and Gentner (1998a) propose that a relational shift occurs in children's thinking, and they identify a kind of bootstrapping that occurs from judgments based on object or perceptual similarity leading to more relational judgments. Contrary to this latter view, Goswami (1992b) and Thompson (1994) find no evidence of a relational shift or transition from holistic to analytic processing. Goswami notes that young children's focus on overall object similarity, even in the face of relational differences, is a performance factor that interferes with accurate completion of relational tasks. She has noted that the types of errors made by children can differ but does not theorize about changes at particular ages or stages.

The concept of relational complexity motivates the design of the present experiment. It is a concept that may vary in nomenclature but is common across this class of theories. More specifically, the present experiment uses stimuli that are defined as unary versus binary relations in Halford's terms or object versus relational similarity in the framework of Gentner (Gentner *et al.*, 1995; Rattermann & Gentner, 1998a) and Goswami *et al.* (1998). The performance of three different age groups was recorded to investigate the separate and interacting effects of age and relational complexity in an auditory and temporal context. The comparison of unary and binary relations will shed light on sensitivity to perceptual similarity versus relational structure at 5, 8 and 11 years of age and invoke discussion of relational shift. However, specific analysis of changes in cognitive capacity, relational familiarity via domain-specific knowledge and within-experiment learning effects are not the focus of the present experiment.[1]

Constructing auditory tasks that differ in relational complexity

A unary relation consists of a mapping of one dimension of the form A:B. An auditory unary level task will involve two pitched notes (or two note durations) and a judgment whether the pitch (or duration) values are the same or different. A binary relation such as 'larger than' has two arguments; a binary analogy is of the form A:B::C:D. Binary relations underpin the structure of analogical reasoning (Gentner, 1983; Gick & Holyoak, 1980; Goswami, 1989, 1992b; Goswami *et al.*, 1998; Nelson, Barresi, & Barrett, 1992; Sternberg, 1977). The relation must be taken between A and B and recognized in C and D; two features need to be held in memory simultaneously and processed in parallel. Presently, an auditory binary task will involve the participant listening to a pattern of four sounds split into patterns of two and deciding whether the second pattern is the same as the first.

Temporal patterns and relational complexity

Constructing auditory analogues of visual tasks that differ in relational complexity is not difficult. However, the way in which the temporal dimension should be conceptualized is unclear. Is it the case that time adds another relation to a task (Elman, 1990; Mix, 1999), or the ephemeral quality of sound increases the storage required to complete a task (Szelag, Kowalska, Rymarczyk, & Pöppel, 2002; Zenatti, 1985)? Does temporal extent provide an opportunity for a child to rehearse and thereby facilitate performance (e.g. Diamond, Kirkham, & Amso, 2002; Flavell, 1992; Wilson, Wales, & Pattison, 1997), or do slow rates of presentation impede performance (e.g. Morrongiello *et al.*, 1985)? There is currently no simple solution to conceptualize and quantify fleeting patterns in relational complexity terms – research using musical sequences provides some clues.

Auditory tasks can involve perceptually bound or relative judgments. Auditory unary tasks are achieved by recognizing that the two pitch values or note lengths are the same or different in absolute terms, whereas auditory binary tasks require relative judgments. In the latter, the child judges whether the directions of pitch change or contour in each pair (low–high, high–low or no change) are the same or different and, in the case of

[1] Once the principle of relational complexity has been implemented in the auditory domain and investigated experimentally, it will be possible to scale tasks to ternary and quaternary levels. These relational levels have not been included in the present experiment for three reasons. First, ternary relation tasks often involve transitive inference (TI), and there is some unresolved controversy concerning cognitive development and TI (Andrews & Halford, 2002). Second, ternary relation tasks may be decomposed into two binary level tasks and solved at that level. Finally, when working with auditory stimuli, increasing relational complexity can be confounded with stimulus and trial length. These problems are not insurmountable but deserve thorough conceptual analysis and the development of appropriate tasks. Such progress will follow logically from investigation of the basic principles in an auditory context.

duration, whether the duration relations in each pair (long–short, short–long or no change) are the same or different. The duration task may also impose an additional load – the need to extract and compare four durations, as opposed to two (pitch) intervals.

White, Dale, and Carlsen (1990) noted that preschool children were perceptually bound, focusing on musical patterns as undifferentiated wholes. Nelson *et al.*, (1992) used spatial and musical analogies with four levels of complexity to test musical aptitude in children from kindergarten to 6th grade. Children performed more accurately in response to spatial than musical analogies: specific strategies in the spatial domain were said to precede their use in the musical domain by at least one year. The authors concluded that 'the effects of memory and the inherent temporal nature of the musical analogy tasks cause this increased difficulty and apparent delayed developmental pattern' (p. 78). Viewed from the perspective of perceptual versus relative judgments and assuming that a relational shift occurs, we expect that older children are more adept at ignoring irrelevant absolute features of pitch and duration and responding to the relative qualities of pitch direction and tone length (Bartlett & Dowling, 1980; Dowling, 1999; Hargreaves, 1986; Schwarzer, 1997; Sergeant & Boyle, 1980).

Aim, design and hypotheses

The aim was to investigate perceptual similarity versus relational complexity in the auditory domain using systematic manipulation of the musical features tone pitch and tone duration. The $3 \times 2 \times 2$ experimental design comprised the independent variables age (5 years, 8 years, 11 years), relational complexity (unary, binary) and feature (pitch, duration) with repeated measures on the latter two factors. The dependent variable was accuracy–calculated as a discrimination index (hit rate minus false alarm rate).

Three different hypotheses were derived from the literature. If musical perceptual matching and analogical reasoning translate directly into unary and binary tasks respectively, then the performance accuracy of 5-, 8- and 11-year-old children is greater than chance, with the 11-year old children performing most accurately, followed by the 8-year-old, and then the 5-year-old groups (Halford, 1993). A main effect of complexity for these age groups would emerge if ternary and quaternary tasks were used, as capacity for ternary relations is thought to develop between 3 and 8 years (Andrews & Halford, 2002). Nonetheless, there may be gradations of improvement at each age on unary and binary tasks. Hypothetical accuracy data generated according to Halford's model are shown in Fig. 1A. By contrast, if the youngest age group is perceptually bound (White *et al.*, 1990) or lacks domain-specific knowledge in response to musical and temporal patterns (Gentner *et al.*, 1995; Goswami *et al.*, 1998; Nelson *et al.*, 1992), the performance of that group on binary tasks is no greater than chance, whereas 8- and 11-year-old children perform the tasks accurately (see Fig. 1B). Finally, if relational primacy rather than a relational shift occurs (Goswami, 1989; Goswami & Brown, 1990), there is no difference between the three age groups as they perform the unary and binary tasks (Fig. 1C).

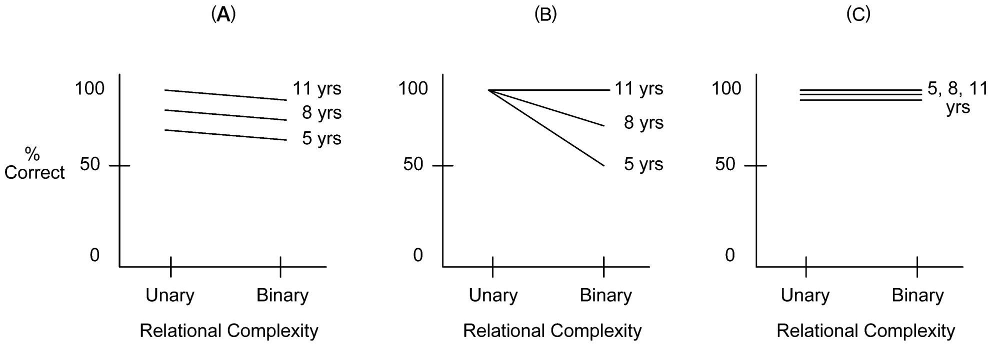

Figure 1. Hypothesized data according to: (A) Halford's (1993) relational complexity metric; (B) perceptually bound versus relative judgments; and (c) relational primacy. (A) Relational complexity predicts that all three age groups perform both unary and binary tasks better than predicted by chance (50% correct). (B) If binary tasks involve relative judgments, then 11-year-old children perform most accurately, followed by 8-year-old and then 5-year-old children. (c) If there is no relational shift with age, then the three age groups perform equally well on both levels of complexity.

Method

Participants

Participants were 86 students from local state schools in south-west Sydney. Participation was voluntary, and requirements were parentally-reported normal hearing. For control purposes, participants had no formal musical training and were from an English-speaking background. There were 27 participants (10 females, 17 males) aged 5.0–5.11 years (M = 5 years 7 months, SD = 3 months), 28 participants (13 females, 15 males) aged 8.0–8.11 years (M = 8 years 6 months, SD = 4 months), and 31 participants (14 females, 17 males) aged 11.0–11.11 years (M = 11 years 6 months, SD = 3 months).

Materials

Auditory stimuli

Auditory materials were created using SoundEdit. The pitch and duration trials of different levels of complexity consisted of consonant patterns of pure tones. The tones were selected from an octave of the equally tempered Western tonal scale of C ranging from 277.2 to 493.9 Hz, as this falls in the central region of human auditory sensitivity and comfortably within the child's singing range (White *et al.*, 1990). In the pitch condition, the note duration was held constant at 300 ms. The majority of pitch intervals were major or minor 3rds which are common in children's songs, with other intervals included for tonal variety and to maintain attention. The duration condition consisted of tones with lengths of 300 and 600 ms, with the frequency held constant at 440 Hz. The relative tone lengths were representative of crotchet and minim note durations, and exceeded the just noticeable difference for duration. The amplitude was constant, and the trials were played at a comfortable listening level.

Unary pitch trials consisted of two notes with an inter-onset interval (IOI) of 250 ms. The total trial length was 850 ms. The IOI was adjusted in trials of increasing relational complexity to minimize the inevitable complexity–trial length confound. Binary pitch trials consisted of two groups of two notes with an IOI of 125 ms within groups and 500 ms between groups. The total trial length was 1950 ms. Duration-varying trials contained the same number of notes and IOIs as those used for pitch trials at each level of relational complexity. Trial length differed according to whether trial pairs were same or different and the pattern of note durations used. Unary duration trials ranged from 850 to 1450 ms. Binary duration trials ranged from 2250 to 2850 ms.

The pitch and duration trials consisted of the two levels of relational complexity (unary, binary), and two to four practice trials preceded each level. For the unary level, two sounds were related according to a matching analogy (A:B). Same trials consisted of two notes that were either the same pitch or the same duration, e.g. 440 Hz–440 Hz or 300 ms–300 ms. Different trials contained either two different pitches (392 Hz–329.6 Hz, where the first note was higher than the second or vice versa) or two different durations (600 ms–300 ms, where the first note was longer than the second or vice versa). Different unary trials involved pitches that varied by a major or minor 3rd and durations that differed by 300 ms.

At the binary level, four note patterns (A:B::C:D) were used wherein the first and second pairs of notes were either the same or different according to a rule. For example, a same trial in the pitch condition might reflect the rule: first note lower than second note: 329.6–370 Hz—349.2–440 Hz. An example from the duration condition would be for the first note to be longer than the second note in each pair. Binary same trials contained an equal mix of easy and hard combinations. Easy same-pitch trials were those that contained only two different frequencies across the four-note pattern (e.g. 493.9–493.9 Hz—440–440 Hz). More difficult same trials contained a pitch or duration relation that was the same but expressed using intervals of different sizes (e.g. 370–329.6 Hz—493.9–370 Hz). Similarly, in same-duration trials, half of the trials contained variation within the two-note pairs that needed to be ignored, while relative lengths across the four notes were abstracted and compared. Binary different trials consisted of pitches that were higher than, or lower than, using intervals of a 2nd, 3rd, 4th or 5th, and durations that differed by 300 ms. There were 8 test trials at the unary level (4 pitch, 4 duration) and 16 test trials at the binary level (8 pitch, 8 duration) with an equal number of same and different trials presented at each level of relational complexity.

Visual stimuli

Visual stimuli consisted of on-screen instructions with a cartoon face and text. Cuisenaire 'numbers in colour' rods were used to assist with verbal instructions and to practise the tasks using concrete visual materials.

Equipment

The experiment used the experiment editor SuperLab v1.74 running on an Apple iMac 266-MHz computer with a single powered external speaker for output. A plastic box for children to indicate their responses (25.5 cm × 18 cm × 6.5 cm) was constructed, with electrician buttons coloured green for 'same', red for 'different', and purple for 'I don't know' responses (Nelson, 1984).

Procedure

Familiarization phase using visual materials

After engaging the child in conversation to help put them at ease, the experimenter said:

> We're going to play a game using sounds. I'll play some sounds and ask you whether you think the sounds are the same or different from one another. For example, they might be different because they are higher than, or longer than, the other sounds. First of all, I'd like you to show me what 'higher than' means.

The experimenter asked the child to take two of the coloured rods and arrange them so that one was higher than the other. The experimenter then selected two new rods and placed them such that one rod was higher than the other. The experimenter asked the child to point to the higher rod. The experimenter repeated this task using two new rods. In the duration condition, the relation 'longer than' replaced the higher-than relation. The experimenter then checked the child's understanding of the terms 'same' and 'different' first using two, and later four, rods. For example, the experimenter asked the child to select four rods and to make two patterns with them. The experimenter asked the child whether the patterns were the same or different. The experimenter then made two patterns, analogous to auditory binary tasks (e.g. long rod, short rod–long rod, short rod), asked the child to compare the patterns and say whether they were the same or different. Asked why the patterns were the same or different, the child responded generally on the basis of colour, height, or length, and the experimenter would explain that she was most interested in whether they were the same in height or in length. The child was asked to use the rods to form two patterns that were the same and another two patterns that were different. Each trial type (pitch, duration) and relational level (unary, binary) was preceded by the visual familiarization phase. Before the computer task began, at least two recognition and two production tasks had been completed accurately for both higher than/longer than and same/different judgments.

Practice and test phases using auditory materials

Once it had been established that the child was familiar with the task, they were introduced to the computer version. Two to four practice trials preceded each condition. Feedback was given, and progress to the next practice trial was contingent on the correct response being given. If there was any confusion regarding same and different, the experimenter asked the child why they thought that a pattern was different if it was the same, or vice versa. This dialogue continued until there was consensus. The computer task was introduced as a game that would become harder the more the child played. On the first screen, a cartoon face appeared that was named 'PK'. The child was told that PK would make some sounds in a pattern. Once the sounds were completed, the child would need to let PK know if his sounds made a same or different pattern. To do this, the child would need to press the buttons on the box – green to respond 'same', red to respond 'different' and purple to start a trial and to respond 'I don't know'.

Throughout the experiment, verbal communication was kept open, and verbal praise offered regularly and indiscriminately. The order of pitch and duration trials was counterbalanced across participants, and unary trials always preceded binary trials. Trials within conditions and individual items were randomized to distribute serial order effects. The experiment stopped if the participant responded incorrectly to two consecutive trials or if they did not want to participate any longer. Data were not included in the analysis if a participant did not finish the experiment. The experiment lasted for 30 min.

Results

Accuracy was calculated using a discrimination index (DI) consisting of Hit Rate (HR) minus False Alarm Rate (FAR), and this index was used as the dependent variable in the analyses. A hit referred to responding different when in fact the items were different. The number of trials precluded the calculation of parametric d′. DI is an effective measure of accuracy in that the number of correct 'different' responses a child makes (HR) is adjusted for the number of incorrect 'different' responses, i.e. false positives – responding different when in fact the items are the same. The maximum accuracy score attainable was +1 and the minimum -1. A score of zero reflects chance (equal hit and false alarm rates). As the analyses conducted to test specific hypotheses involved comparing DI against chance, no adjustment to α, which was set at .05, was required.

The mean HR, FAR and DI for the three age groups performing pitch unary and binary trials are listed in Table 1, and these accuracy scores in response to duration unary and binary trials are shown in Table 2. The accuracy scores in Tables 1 and 2 reflect an effect of age, with lowest HR and highest FAR always recorded by the youngest age group. The 8-year-old children performed at an intermediate level; the highest DIs, consisting of high HR and relatively low FAR, were recorded by the 11-year-old group. A general effect of age on performance was found. As age increased, accuracy on all tasks improved, as evidenced by a significant linear trend of DI for age, $F(1, 83)=40.98$, $p < .05$, Cohen's $f = .70$. This result refutes the relational primacy hypothetical data depicted in Fig. 1C. Overall, 5-year-old children were performing just above chance (mean DI = .29, *SD* = .46). Eight-year-old children recorded a mean DI of .55 (*SD* = .43) and 11-year-old children a mean DI of .71 (*SD* = .29).

Table 1. Mean hit rate (HR), false alarm rate (FAR), discrimination index (HR − FAR) and standard deviation (*SD*) as a function of age and relational complexity on pitch trials

	HR	FAR	DI (*SD*)
Unary			
5 years	.78	.19	.59 (.46)
8 years	.89	.07	.82 (.31)
11 years	.97	.03	.94 (.21)
Binary			
5 years	.65	.38	.27 (.59)
8 years	.76	.29	.47 (.52)
11 years	.96	.34	.62 (.31)

According to both the perceptual similarity hypothesis and relational complexity metric, it was hypothesized that all age groups respond significantly better than chance on unary trials. This hypothesis was supported in the context of unary pitch trials: 5 years $t(26) = 6.85$, $p < .05$, Cohen's $d = .02$; 8 years $t(27) = 13.99$, $p < .05$, Cohen's

Table 2. Mean hit rate (HR), false alarm rate (FAR), discrimination index (HR − FAR) and standard deviation (*SD*) as a function of age and relational complexity on duration trials

	HR	FAR	DI (*SD*)
Unary			
5 years	.57	.33	.24 (.41)
8 years	.77	.13	.64 (.41)
11 years	.90	.13	.77 (.29)
Binary			
5 years	.47	.40	.07 (.38)
8 years	.59	.33	.26 (.45)
11 years	.76	.24	.52 (.33)

d = .03; 11 years $t(30) = 24.37$, $p < .05$, Cohen's d = .03; and in the context of unary duration trials: 5 years $t(26) = 2.11$, $p < .05$, Cohen's d = .01; 8 years $t(27) = 6.49$, $p < .05$, Cohen's d = .02; 11 years $t(30) = 13.82$, $p < .05$, Cohen's d = .03.

It was hypothesized that, as binary pitch and duration trials involved judgments of relational similarity, 8- and 11-year-olds would perform accurately, but the 5-year-old children may be perceptually bound. In response to binary pitch trials, all age groups performed at a level that was significantly greater than that expected by chance: 5 years $t(26) = 3.41$, $p < .05$, Cohen's d = .01; 8 years $t(27) = 6.11$, $p < .05$, Cohen's d = .02; 11 years $t(30) = 12.02$, $p < .05$, Cohen's d = .02. Performance in response to binary pitch trials was significantly poorer overall relative to performance on unary pitch trials, $F(1, 83) = 46.73$, $p < .05$, Cohen's f = .75. Hit and false alarm rates (Table 1) indicate that particularly for the 8- and 11-year-old groups, binary pitch trials attracted both lower hit rates and higher false alarm rates than unary pitch trials. The data for pitch trials approximate Fig. 1A. Scrutiny of the types of false alarms made by the younger age groups did not reveal any particular pattern of errors in the pitch condition. For example, it was possible that children would tend to respond different to items that contained the same pitch contour in the presence of varying notes and intervals, that is, the more difficult same trials. However, the percentage of false alarms to these kinds of trials compared with trials that contained same contours and no interval change did not differ greatly in either the 5-year- old group (39% versus 37%) or the 8-year-old group (23% versus 32%).

Evidence of perceptually bound responding was apparent in the DI of 5-year-old children in response to binary duration trials, wherein DI was not significantly greater than chance, $t(26) = 1.00$, $p > .05$, Cohen's d = .00. The two older age groups completed binary duration trials at a level significantly greater than that expected by chance: 8 years, $t(30) = 3.02$, $p < .05$, Cohen's d = .01; 11 years, $t(30) = 8.74$, $p < .05$, Cohen's d = .02. A significant difference was also found, overall, between performance of all age groups on unary duration versus binary duration trials, F (1, 83) = 19.15, $p < .05$, Cohen's f = .48. Table 2 shows that binary duration trials attracted both lower hit rates and slightly higher false alarm rates across all age groups. The types of false alarms made by all three age groups reflected the difficulty of the same trials. Children were much more likely to respond different to same trials where there was duration variation within the two-note items in a trial but not in the relation between the two items; this type of error was made 50% of the time by the 5- and 8-year age groups and 37.5% of the time by the 11-year age group. By contrast, the 'easier' same trials, where there was no variation within two-note items, attracted false alarms 27.7%, 16.1% and 9.4% of the time by the 5-year, 8-year and 11-year age groups, respectively.

Discussion

This experiment has investigated the effect of age on response to perceptually similar versus relationally similar auditory patterns. An overall effect of age was evident, with the 11-year age group responding more accurately than the 8-year age group who, in turn, responded more accurately than the 5-year age group. The simplest tasks required that children match particular values of either the auditory dimension (note pitch) or duration (note length). All age groups performed these pitch and duration tasks significantly better than that expected by chance. Perceptual matching of this type corresponds with the unary level of Halford's (1993) relational complexity hypothesis, and the present results accord with those obtained by Halford (1993), Sergeant and Boyle (1980), and White *et al.* (1990).

Binary tasks involved relations between either pitch or duration values and could not be performed accurately if a child simply responded to absolute features of the patterns. Relative to unary tasks, binary relation tasks were completed less accurately by all age groups. This finding supports the view that a relational shift occurs (Case, 1992; Gentner *et al.*, 1995; Halford, 1993; Nelson *et al.*, 1992). The 5-, 8- and 11-year age groups were able to recognize pitch contour relations at greater-than-chance level, although Gentner's observation that young children make errors based on perceptual similarity was evident in the greater number of false alarms and misses made by 5- and 8-year-old children as they completed binary relation tasks. The youngest age group, however, was not able to recognize binary relations expressed through the duration dimension; judgments appeared to be based on perceptual similarity. Given the accurate performance of this youngest group in response to unary duration trials, we can be sure that it was not a problem of just noticeable difference. Rather, these children were unable to abstract the relation longer than or shorter than from the first two notes and the last two notes of the four-note pattern, hold the two relations in memory and compare them. Interestingly, the youngest age group was able to abstract and compare two relations in the pitch context. The duration dimension – its temporal and ephemeral quality – presents a challenge, and there are two explanations that warrant investigation.

The inability of the 5-year age group to perform binary duration tasks relates to the broader issue of conceptualizing the temporal dimension within relational complexity. One possibility is that temporal extent adds a relation to a task (e.g. Elman, 1990; Mix, 1999; Nelson *et al.*, 1992). If so, the tasks constructed in the present experiment would constitute binary and ternary relations rather than unary and binary relations, respectively. If this assumption is valid, the performance of the 5-year age group is in keeping with that predicted by Halford (1993); at 5 years, unary and binary relations can be processed, but the necessary capacity is not available to process ternary relations. A moot point here is that such an explanation is not required in the pitch context. Analysis of the relational complexity of auditory dimensions themselves is needed to assess whether the dimensions are comparable in complexity – the structure of the dimension requires quantification as well as the relational complexity of a task.

Alternatively, auditory patterns may demand greater memory. The youngest children may not have abstracted the relations between two pairs of notes in a binary duration pattern but attempted to perceptually match four notes. In the pitch condition, the interval between two notes (i.e. fall, rise or same contour) needed to be abstracted and compared. However, in the duration condition, note onset to offset needed to be derived from each of four notes, held in memory and compared. The load of four items of differing lengths is likely to have exceeded the capacity of the 5-year-old children (see Jensen & Neff, 1993; Keller & Cowan, 1994)

The general results reflect an ability of children to perceptually match auditory patterns when presented with simple unary tasks and an ability to make relational judgments that appears to increase with age. To Halford (1993), the mechanism is one of an increase in cognitive capacity that allows construction of progressively more complex structure mappings. Gentner and colleagues (Gentner & Medina, 1998; Rattermann & Gentner, 1998a) specify two mechanisms that mediate the shift of attention from common object properties to common relational structure. The first is the acquisition of relational language, and the second is progressive abstraction brought about by structural alignment. We propose that the use of auditory and increasingly musical stimuli, and the systematic manipulation of formal musical training versus musical aptitude, provide the means to analyse changes in children's increasing knowledge and language of a particular domain, independent of changes in cognitive capacity. Western tonal music has acquired its own specialist vocabulary. Children's knowledge and use of terms such as interval, contour and rhythm may be manipulated to assess whether performance on relationally complex auditory tasks, which involve one or more of these features, is facilitated by possession of domain-specific relational language. Furthermore, children who have musical aptitude but no formal training or knowledge of musical language will provide an informative comparison group.

Musical materials lend themselves to investigation of perceptual and relational similarity. Patterns in the present experiment were kept deliberately simple and constructed from dimensions that could be quantified and varied independently. However, there is scope for the creation of complex and musical patterns using the dimensions pitch, duration, timbre and loudness, and contriving diagnostic relational structures (e.g. Gérard & Drake, 1990; Schwarzer, 1997; Webster & Pflederer Zimmerman, 1983). Higher-order ternary and quaternary relational tasks can be constructed using multidimensional musical materials. Moreover, stimuli that consist of interacting musical dimensions can eliminate relational complexity and trial-length confounds. For example, a transitive inference task might consist of three sounds, A, B and C, with the relation between sounds A and B, and B and C implying a relation between A and C. The relations could include pitch (e.g. rising contour) that is independent of the timbral (instrument) relation between the notes (e.g. trumpet, piano, flute). Musical production tasks may also be explored as the means to keep task structure, procedures and response type consistent across levels of relational complexity.

Acknowledgements

This research was supported by a grant from the Australian Research Council to the first author. The authors wish to thank the students, parents and teachers who took part in the study, and John Towse, Anik de Ribaupierre and three anonymous reviewers for helpful comments and advice. SuperLab is a trademark registered to Cedrus Corporation.

References

Andrews, G., & Halford, G. S. (1998). Children's ability to make transitive inferences: The importance of premise integration and structural complexity. *Cognitive Development, 13*, 479–513.

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*, 153–219.

Bartlett, J., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 501–515.

Case, R. (1985). *Intellectual development: birth to adulthood*. Orlando, FL: Academic Press.

Case, R. (1992). *NeoPiagetian theories of child development*. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 161–196). Cambridge: Cambridge University Press.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology, 38*, 352–362.

Dowling, W. J. (1999). The development of music perception and cognition. In D. Deutsch (Ed), *The psychology of music* (2nd ed., pp. 603–625). San Diego: Academic Press.

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics, 53*, 25–33.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Flavell, J. H. (1992). Cognitive development: Past, present, and future. *Developmental Psychology, 28*, 998–1005.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65*, 263–297.

Gentner, D., Rattermann, M., Markman, A., & Kotovsky, L. (1995). Two forces in the development of relational similarity. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modelling* (pp. 263–313). Hillsdale, NJ: Erlbaum.

Gérard, C., & Drake, C. (1990). The inability of young children to reproduce intensity differences in musical rhythms. *Perception & Psychophysics, 48*, 91–101.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355.

Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cognitive Development, 4*, 251–268.

Goswami, U., & Brown, A. L. (1990). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition, 35*, 69–95.

Goswami, U. (1992a). Analogical reasoning and conceptual complexity in cognitive development: Commentary. *Human Development, 35*, 222–225.

Goswami, U. (1992b). *Analogical reasoning in children*. Hove, UK: Erlbaum.

Goswami, U. (1998). *Cognition in children*. Hove, UK: Psychology Press.

Goswami, U., Leevers, H., Pressley, S., & Wheelwright, S. (1998). Causal reasoning about pairs of relations and analogical reasoning in young children. *British Journal of Developmental Psychology, 16*, 553–569.

Halford, G. S. (1989). Cognitive processing capacity and learning ability: An integration of two areas. *Learning and Individual Differences, 1*, 125–153.

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Erlbaum.

Halford, G. S., & Macdonald, C. (1977). Children's pattern construction as a function of age and complexity. *Child Development, 48*, 1096–1100.

Halford, G. S., & Wilson, W. H. (1980). A category theory approach to cognitive development. *Cognitive Psychology, 12*, 356–411.

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, 21*, 803–864.

Hargreaves, D. J. (1986). *The developmental psychology of music*. Cambridge: Cambridge University Press.

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

Keller, T. A., & Cowan, N. (1994). Developmental increase in the duration of memory for tone pitch. *Developmental Psychology, 30*, 855–863.

Maybery, M. T., Bain, J. D., & Halford, G. S. (1986). Information-processing demands of transitive inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 600–613.

Mix, K. S. (1999). Preschoolers' recognition of numerical equivalence: sequential sets. *Journal of Experimental Child Psychology, 74*, 309–332.

Mondloch, C. J., Geldart, S., Maurer, D., & Le Grand, R. (2003). Developmental changes in face processing. *Journal of Experimental Child Psychology, 86*, 67–84.

Morrongiello, B. A. (1984). Auditory temporal pattern perception in 6- and 12-month old infants. *Developmental Psychology, 20*, 441–448.

Morrongiello, B. A., Trehub, S. E., Thorpe, L. A., & Capodilupo, S. (1985). Children's perception of melodies: The role of contour, frequency and rate of presentation. *Journal of Experimental Child Psychology, 40*, 279–292.

Nelson, D. J. (1984). The conservation of rhythm in Suzuki violin students: A task validation study. *Journal of Research in Music Education, 32*, 25–34.

Nelson, D. J., Barresi, A. L., & Barrett, J. R. (1992). Musical cognition within an analogical setting: Toward a cognitive component of musical aptitude in children. *Psychology of Music, 20*, 70–79.

Olsho, L. W., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982). Auditory frequency discrimination in infancy. *Developmental Psychology, 18*, 721–726.

Pascual-Leone, J. A. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 32*, 301–345.

Rattermann, M., & Gentner, D. (1998a). More evidence for a relational shift in the development of analogy: Children's performance on a causal mapping task. *Cognitive Development, 13*, 453–478.

Rattermann, M., & Gentner, D. (1998b). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In K. Holyoak, D. Gentner, & B. Kobinov (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences* (pp. 274–282). Sophia: New Bulgarian University.

Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: evidence from infant listeners. *Psychological Science, 7*, 272–277.

Schwarzer, G. (1997). Analytic and holistic modes in the development of melody perception. *Psychology of Music, 25*, 35–56.

Sergeant, D. C., & Boyle, J. D. (1980). Contextual influences on pitch judgment. *Psychology of Music, 8*, 3–15.

Shuter-Dyson, R., & Gabriel, C. (1981). *The psychology of musical ability* (2nd ed.). London: Methuen.

Sternberg, R. J. (1977). *Intelligence, information processing and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.

Szelag, E., Kowalska, J., Rymarczyk, K., & Pöppel, E. (2002). Duration processing in children as determined by time reproduction: implications for a few seconds temporal window. *Acta Psycologica, 110*, 1–19.

Thompson, L. A. (1994). Dimensional strategies dominate perceptual classification. *Child Development, 65*, 1627.

Trehub, S. E. (1993). Temporal auditory processing in infancy. *Annals of the New York Academy of Sciences, 682*, 137–149.

Trehub, S. E., Schellenberg, E. G., & Hill, D. S. (1997). The origins of music perception and cognition: A developmental perspective. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 103–128). Hove, UK: Psychology Press.

Trehub, S. E., & Thorpe, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences of temporal structure. *Canadian Journal of Psychology, 43*, 217–229.

Webster, P. R., & Pflederer Zimmerman, M. (1983). Conservation of rhythmic and tonal patterns of second through sixth grade children. *Bulletin of the Council for Research in Music Education, 73*, 28–49.

White, D. J., Dale, P. S., & Carlsen, J. C. (1990). Discrimination and categorization of pitch direction by young children. *Psychomusicology, 9*, 39–58.

Wiese, H. (2003). Iconic and non-iconic stages in number development: the role of language. *Trends in Cognitive Sciences, 7*, 385–390.

Wilson, S. J., Wales, R. J., & Pattison, P. (1997). The representation of tonality and meter in children aged 7 and 9. *Journal of Experimental Child Psychology, 64*, 42–66.

Zenatti, A. (1985). The role of perceptual-discrimination ability in tests of memory for melody, harmony, and rhythm. *Music Perception, 2*, 397–404.

Received 18 June 2001; revised version received 17 November 2003