



HAL
open science

Learning of pitch and time structures in an artificial grammar setting.

Jon Prince, Catherine Stevens, Mari Riess Jones, Barbara Tillmann

► **To cite this version:**

Jon Prince, Catherine Stevens, Mari Riess Jones, Barbara Tillmann. Learning of pitch and time structures in an artificial grammar setting.. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2018, 44 (8), pp.1201-1214. 10.1037/xlm0000502 . hal-02378158

HAL Id: hal-02378158

<https://hal.science/hal-02378158>

Submitted on 21 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning of pitch and time structures in an artificial grammar setting

Jon B. Prince¹, Catherine J. Stevens², Mari Riess Jones³, & Barbara Tillmann^{2,4,5}

¹ Murdoch University, Perth, Australia

² MARCS Institute for Brain, Behaviour & Development, Western Sydney University,
Sydney, Australia

³ The Ohio State University and University of California, Santa Barbara, USA

⁴ CNRS, UMR5292; INSERM, U1028; Lyon Neuroscience Research Center, Auditory
Cognition and Psychoacoustics Team, Lyon, F-69000, France

⁵ University Lyon 1, Villeurbanne, F – 69000, France

Correspondence concerning this article should be addressed to:

Jon Prince
School of Psychology and Exercise Science
Murdoch University
90 South Street
Murdoch, WA 6150
Australia
tel (+61) 8 9360 6670
j.prince@murdoch.edu.au

Abstract

Despite the empirical evidence for the power of the cognitive capacity of implicit learning of structures and regularities in several modalities and materials, it remains controversial whether implicit learning extends to the learning of temporal structures and regularities. We investigated whether (1) an artificial grammar can be learned equally well when expressed in duration sequences as when expressed in pitch sequences, (2) learning of the artificial grammar in either duration or pitch (as the primary dimension) sequences can be influenced by the properties of the secondary dimension (invariant versus randomized), and (3) learning can be boosted when the artificial grammar is expressed in both pitch and duration. After an exposure phase with grammatical sequences, learning in a subsequent test phase was assessed in a grammaticality judgement task. Participants in both the pitch and duration conditions showed incidental (not fully implicit) learning of the artificial grammar when the secondary dimension was invariant, but randomizing the pitch sequence prevented learning of the artificial grammar in duration sequences. Expressing the artificial grammar in both pitch and duration resulted in disproportionately better performance, suggesting an interaction between the learning of pitch and temporal structure. The findings are relevant to research investigating the learning of temporal structures and the learning of structures presented simultaneously in two dimensions (e.g., space and time, space and objects). By investigating learning, the findings provide further insight into the potential specificity of pitch and time processing and their integrated versus independent processing, as previously debated in music cognition research.

Introduction

The cognitive capacity of implicit learning has been shown across a wide range of domains, notably for the visual modality (e.g., objects, locations, shapes, letters, Cleeremans & McClelland, 1991; Reber, 1967) and the auditory modality (e.g., tones, environmental sounds, timbres, voices, syllables, Bigand, Perruchet, & Boyer, 1998; Howard & Ballas, 1980; Tillmann & Poulin-Charronnat, 2010). It has also been shown that movement and action can contribute to this learning (e.g., Deroost & Soetens, 2006; Remillard, 2003). However, there is less research on implicit learning of temporal structures (e.g., duration patterns, such as rhythms); indeed, data on this topic are conflicting, showing either learnability of time patterns or not.

The two main experimental paradigms for investigating implicit learning are Serial Response Time (e.g., Nissen & Bullemer, 1987) and Artificial Grammar Learning (AGL; e.g., Reber, 1967). In the serial response time paradigm, participants overtly respond to successive stimuli in a repeating sequence, and response times decrease as participants implicitly learn the sequence (or part of it). In the artificial grammar learning paradigm, participants experience passive exposure to a variety of different sequences based on a given artificial grammar, and subsequently make judgements about novel sequences (e.g., grammaticality judgements). As the participants are not informed about the structural properties of the sequences nor instructed to learn them, any learning of their underlying structural properties should represent incidental and implicit learning. However, it is possible to acquire conscious knowledge in this paradigm (Dienes, Altmann, Kwan, & Goode, 1995; Dienes & Scott, 2005), in which case although the learning is implicit and incidental, some explicit knowledge might be acquired.¹ To date, the serial response time paradigm has been the more popular paradigm for assessing learning of temporal structures. However, it has been criticized

because of the potential contribution of motor learning (e.g., Deroost & Soetens, 2006) and for inadvertent confounds involving correlations with responses, task, or with other aspects of the material (Meier & Cock, 2010). In the present study, we investigated learning of temporal structure using the artificial grammar learning paradigm and compared it to learning of the same artificial grammar structure applied to pitch.²

Findings from the serial response time paradigm reveal that participants learn a sequence based on temporal features only when the time structure is correlated with additional information, such as a structured series of tones (Buchner & Steffens, 2001) or spatial locations (Shin & Ivry, 2002). For example, Shin and Ivry's participants indicated the spatial location of a visual target, embedded in repeating sequences of spatial location and/or timing. Participants' reaction times improved over blocks, and their performance suffered when the location structured was randomized in a test block. However, randomizing the timing structure only affected performance if the location sequence in the preceding learning blocks reinforced the timing structure (i.e., had the same structure, thus correlated). That is, learning of the temporal structure was unsuccessful if unaccompanied by matching structure in spatial location. Such results imply that the learning of a temporal pattern is not independent of action (i.e., movements related to response selection) and that timing is an integrated part of action representation, allowing also for enhanced sequence learning when time structures are correlated with other (non-temporal) forms of structure (see also Gobel, Sanchez, & Reber, 2011).

However, the aforementioned studies did not ascertain how learning of a temporal sequence might be affected by an accompanying (non-temporal) event structure. For example, can temporal structures be learned when accompanied by an invariant (i.e., neutral) series of pitches as well as when accompanied by a randomly ordered series of different pitches? Some serial response time studies suggest this is the case (Brandon, Terry, Stevens, & Tillmann,

2012; Schultz, Stevens, Keller, & Tillmann, 2013). In contrast to earlier studies (Buchner & Steffens, 2001; Shin & Ivry, 2002), more recent studies reveal that learning of temporal structures is possible with an uncorrelated event structure. They showed that learning of temporal structures occurs with an invariant secondary event structure (by using the same tone repeatedly, Salidis, 2001), as well as with a random sequence of events as the secondary dimension (e.g., Brandon et al., 2012; Schultz et al., 2013). Note that in the earlier studies (Buchner & Steffens, 2001; Shin & Ivry, 2002), the temporal structures were implemented with response-to-stimulus time intervals, which is the time between the participant's response and the presentation of the next event. This contrasts with the later studies (Brandon et al., 2012; Schultz et al., 2013) that implemented temporal structures with inter-onset-intervals between events (thus purely based on the stimulus and independent from participants' response times). For example, in Brandon et al. (2012), participants listened to syllables whose onsets were presented in rhythmic patterns that had an underlying duple or triple beat structure, embedded in a repeating sequence. Participants improved their performance in a syllable discrimination task over time until the temporal pattern structure changed in a test block, indicating learning. Schultz et al. (2013) also used time structures that aligned with a consistent temporal period (i.e., a regular beat, or meter), which can facilitate attending and structure learning (e.g., Jones, 2008) compared to non-metrical (irregular) sequences. Using a different experimental paradigm, namely a serial recall task where participants explicitly reproduced a preceding sequence, Ullen and Bengtsson (2003) provide evidence for temporal sequence learning and its independence from sequential event structure learning. Participants repeatedly reproduced sequences of keypresses, which consisted of either a random sequence of keys with a regular temporal structure, or an ordered sequence with random temporal structure. Learning (decreased errors) occurred in both of these conditions, suggesting

independent learning of an event structure despite the fact that the stimulus sequences had a random temporal structure.

To our knowledge, no study has yet shown that temporal structure can be learned when it has been generated by an artificial grammar and when processed during the exposure phase only by listening and without associated movements. An artificial grammar learning experiment typically consists of two phases – an exposure and a test phase. In the exposure phase, participants are exposed to artificial sequences constructed from a grammar, typically a finite-state grammar. A finite-state grammar (Figure 1) is defined by a set of nodes (states) connected by a configuration of links, where each link reflects a state-to-state transition that determines the output of one symbol of a set (for example, a letter, a visual form or a tone). After the exposure phase, participants are first informed of the rule-governed nature of the exposure sequences; next, they receive a test phase in which they must differentiate novel sequences that follow the same rules as previously presented sequences from other novel sequences that do not follow these rules. In this test phase, listeners judge whether the new sequences obey the grammatical rules of the exposure phase sequences. Learning is reflected by test performance above chance level (Reber, 1967) or by faster grammaticity judgments of grammatical sequences than for ungrammatical sequences (Buchner, 1994; Kinder, Shanks, Cock, & Tunney, 2003).

The artificial grammar learning paradigm allows for investigating perceptual learning only, without related action sequences. It permits assessment of learning of a complex structural system (i.e., the artificial grammar), rather than merely learning a repeated individual sequence (as in the serial response time and serial recall paradigms). However, no study has directly compared the learning of artificial grammar structures instantiated in the temporal dimension with the learning of artificial grammar structures based on events drawn from a non-temporal dimension. Nor is there any research investigating the learning of an

artificial grammar in both temporal and event structures simultaneously (in correlated fashion). When focusing only on perceptual learning, one candidate of event structures becomes particularly interesting in light of music cognition research: the pitch dimension. Pitch and time are the primary form-bearing dimensions in Western tonal music, and research on their integration has raised numerous questions regarding their independence or interdependence for perception and memory (e.g., Ellis & Jones, 2009; Jones & Ralston, 1991; for a review, see Prince, Thompson, & Schmuckler, 2009). Here we investigated the learning of an artificial grammar that is applied to sequences of either durations, pitches, or both (in a correlated fashion, which is a congruent and thus redundant combination of pitch and time elements of the grammar). The present study builds on previous research having used the artificial grammar learning paradigm to investigate the learning of pitch structures (Altmann, Dienes, & Goode, 1995; Loui, Wessel, & Hudson Kam, 2010; Tillmann & Poulin-Charronnat, 2010), as well as the potential influence of temporal structure (regular, irregular, isochronous, or metric timing) on the learning of pitch structure (Selchenkova, François, et al., 2014; Selchenkova, Jones, & Tillmann, 2014) – but none of these studies has tested yet whether an artificial grammar instantiated in time can be learned and how it combines with simultaneous learning of the same artificial grammar instantiated in pitch.

In sum, this study has three main aims. The first was to investigate whether an artificial grammar can be learned as well in time as in pitch. A second aim was to ascertain if a joint artificial grammar, applied in correlated fashion to time and pitch, might affect sequence learning (i.e., does such a correlation boost learning of the artificial grammar?). A third goal was to examine the extent to which learning of an artificial grammar in either pitch or time depends on variability in the accompanying (secondary) dimension. We accomplished this by (a) neutralizing the secondary dimension using either isochronous (invariant time) or isotonous (invariant pitch) sequences (Experiments 1 and 2), and (b) randomizing the

variations of the events of the secondary dimension (Experiment 3). In all experiments, above-chance accuracy of grammaticality judgements in a test phase (following a passive exposure phase) was our indicator of successful artificial grammar learning.

Experiment 1

In Experiment 1, we tested for artificial grammar learning in three conditions. In one condition, the artificial grammar was in pitch (AGp), resulting in sequences with different pitches, played at a constant speed (isochronous). In a second condition, the artificial grammar was in duration (AGd), resulting in sequences with different tone durations, all played at the same pitch (isotonous). In a third condition, the artificial grammar was in pitch and duration (AGpd), combining pitch and duration structures in correlated fashion. All were between subject conditions. In addition, to limit effects of musical enculturation, we avoided existing pitch structures and temporal structures of Western music (with which participants are familiar through passive enculturation; e.g., Bigand, Tillmann, Poulin-Charronnat, & Manderlier, 2005). Specifically, none of the pitches corresponded to the equal-tempered system³, and none of the durations matched the simple integer ratios that characterize a stable metrical framework used in Western tonal music. To typical Western listeners, this means the pitches sounded out-of-tune in the AGp and AGpd conditions, and the temporal sequences sounded erratic (eliminating a regular beat) in the AGd and AGpd conditions. Previous research that has shown learning of this artificial grammar in pitch used tones of the equal-tempered system (Tillmann & Poulin-Charronnat, 2010), even though the sequences involved did not evoke a musical key (using only notes from a single musical scale) – a basic characteristic of Western music which could aid perception and memory (Schulze, Dowling, & Tillmann, 2012; Watkins, 1985). With regard to time, using an irregular structure

eliminates the learning benefit based on metricality and metric binding (also observed for implicit learning, see Selchenkova, Jones, et al., 2014).

Method

Participants. Sixty students (30 from Western Sydney University and 30 of Murdoch University), with an average age of 24.3 (SD = 7.8), provided informed consent and participated in Experiment 1. They completed a background questionnaire which asked them to report the years of musical training they had, in both an individual (private lessons) and group (school lessons) format. We used the years of individual training as our index of formal musical training, which was an average of 2.0 years (SD = 3.9). Participants were randomly assigned to one of three conditions: AGp (artificial grammar in pitch, with isochronous durations), AGd (artificial grammar in duration, with isotonic pitch), or AGpd (artificial grammar in both pitch and duration simultaneously, in correlated fashion). Musical expertise did not differ across conditions, $F(2, 57) < 1$.

Material. The material was based on the finite-state grammar used by Tillmann and Poulin-Charronnat (2010; see Figure 1), notably by extracting numerical patterns from the stimuli listed in their appendix and applying them to a new set of pitches and a set of durations. That is, we used the same grammar, but changed the alphabet. For example, one of their sequences consisted of the notes $a\#-c-d-a-f\#-a$, which can be represented as the numerical pattern $2-3-4-1-5-1$. We applied these patterns to create new sets of sequences based upon an alphabet of mistuned (non-equal tempered) pitches and an alphabet of irregular durations. These alphabets allow creating artificial musical system based on building blocks/events unfamiliar to listeners (as discussed in the Introduction). Applying these alphabets to the artificial grammar of Tillmann and Poulin-Charronnat yields a group of new

sequences. All grammatical exposure, grammatical test, and ungrammatical test sequences (N = 37 for each type) followed the patterns derived from Tillmann and Poulin-Charronnat. Sequences were either five or six tones in length. Figure 1 depicts the artificial grammar used in this study, and Table 1 presents the accompanying pitch and duration values associated with the numerical indices shown in Figure 1. Figure 2 visually depicts the realization of the above pattern for the three experimental conditions used in this study (AGp, AGd, and AGpd). The stimuli and a spreadsheet denoting the patterns from Tillmann and Poulin-Charronnat, and the exact Hz and ms values they yield using the new alphabet, are available as supplementary material.

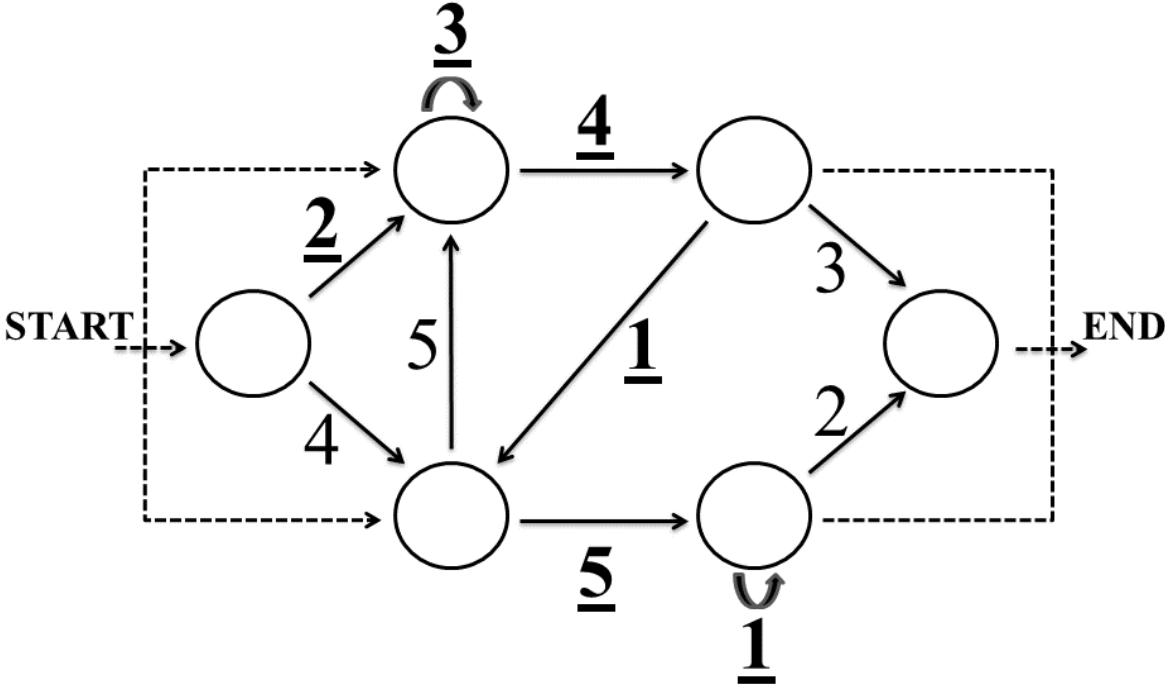


Figure 1. Artificial (finite-state) grammar from Tillmann and Poulin-Charronnat (2010), presented with numerical indices (1-5). Replacing these indices with the values in Table 1 created sequences that presented the artificial grammar in pitch, duration, or both. Please note numbers are only placeholders representing duration or pitch, not indicators of presentation order.

Table 1

Pitch and duration value indices from Figure 1

	Index values				
	1	2	3	4	5
Pitch values (Hz) Experiments 1-3	480	520	560	605	665
Duration values (ms) Experiment 1	250	330	440	580	800
Duration values (ms) Experiments 2-3	60	110	250	550	920

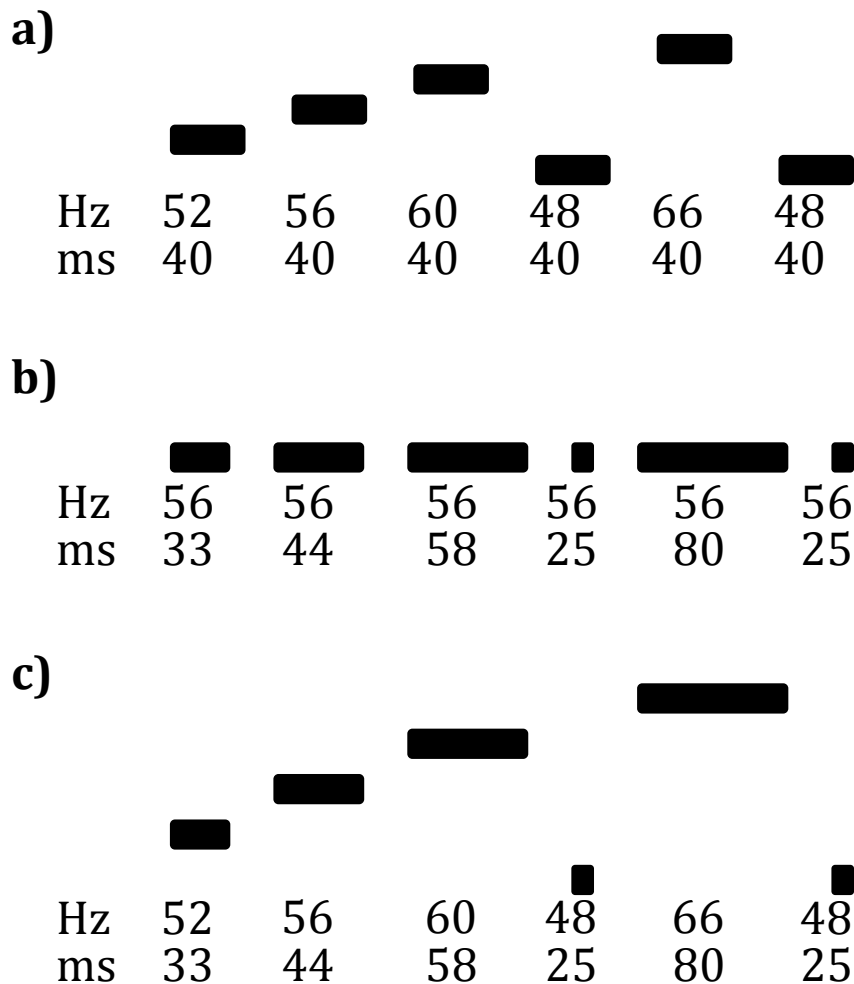


Figure 2. Example sequences based on applying the pattern 2-3-4-1-5-1 to pitch (depicted as vertical height) and/or duration values (depicted as length) to generate three conditions. Note AGp (a) is isochronous, such that all tones are of the same length; AGd (b)

is isotonous, such that all tones are the same pitch height, and AGpd (c) combines both to create a correlated instantiation of the pattern.

Pitches and durations were selected based on pilot testing for discriminability of the elements. Participants (N=9) indicated which of two tones separated by a 5-second silent interval was either longer (duration block) or higher (pitch block). For the block testing for duration discriminability, presented tones (all at 405 Hz) differed by Weber fractions ⁴ of .1, .2, .3, or .4, with base values of 350, 500, 650, and 815 ms. For the block testing for pitch discriminability, presented tones (all with a duration of 638 ms) differed by Weber fractions of .01, .02, .04, or .06, with base values of 521, 567, 617, and 671 Hz. Results are depicted in Table 2, and showed that to be clearly discriminable (here performance at least 90%), Weber fractions needed to be at least .04 for pitch, and at least .30 for duration. Based on these findings, pitch and duration sets for subsequent testing were derived such that adjacent values differed by at least these Weber fractions, and should therefore be clearly discriminable for pitch and duration.

Table 2

Weber fractions tested in pilot discriminability test and observed performance expressed in mean accuracy of correct responses (standard deviation in parentheses).

	Weber Fraction	Accuracy % (SD)
Pitches	0.01	56.9 (24.8)
	0.02	80.6 (16.5)
	0.04	95.8 (8.3)
	0.06	90.3 (11.0)

	0.10	72.2 (21.4)
	0.20	86.1 (16.5)
Durations	0.30	94.4 (8.4)
	0.40	97.2 (7.9)

The pitch set had tones at the following pitch heights: 480, 520, 560, 605, and 665 Hz (see Table 1). To ensure that their relationships did not correspond to Western equal-temperament, and the tone combinations were inconsistent with any scale of Western tonal music, the Weber fraction difference between adjacent values varied between .08 and .1. The duration set consisted of the following durations: 250, 330, 440, 580, and 800 ms. These durations were selected to avoid the simple-integer ratios characteristic of familiar Western music, leading to the absence of a metrical framework when combined into a sequence. The Weber fraction difference between adjacent values varied between .32 and .38. The pitch-duration set combined the values of the pitch and duration sets in the following way: a note at 480Hz with duration of 250 ms, 520 Hz with 330 ms, 560 Hz with 440 ms, 605 Hz with 580 ms, and 665 Hz with 800 ms (see Table 1).

All stimuli were pure tones with a 10 ms linear ramp at onset and offset. For the pitch sequences (AGp), all tones were 400 ms long, followed by a 100 ms silent gap (inter-stimulus interval, or ISI), making the inter-onset interval (IOI) 500 ms (i.e., an isochronous sequence). The pitch of the tones varied according to the patterns derived from the material of Tillmann and Poulin-Charronnat (2010), see Table 1 and supplementary material. For the duration sequences (AGd), the pitch of all tones was 566 Hz (the mean Hz of the pitch set) and the ISI remained constant at 100 ms. Consequently, the AGd condition had the artificial grammar in both the durations of the tones *and* the IOIs (except for the final tone, as there was no interval

following the final event of a sequence). The durations and IOIs varied according to the sequence (given patterns derived from Tillmann & Poulin-Charronnat, 2010). For the pitch-duration sequences (AGpd), both the pitches and the durations (and thus also the IOIs) varied according to the sequence (again using the same patterns derived from Tillmann & Poulin-Charronnat, 2010). The average length of the AGp stimuli was 3.45 s (SD = .29), and for AGd and AGpd stimuli 3.49 s (SD = .46). See Figure 2 for a visual depiction of AGp, AGd, and AGpd sequences.

Ungrammatical sequences in the test phase were created by changing a single element in the sequence such that it violated the artificial grammar rules (i.e., did not follow the arrows in Figure 1). As with the exposure and grammatical test sequences, the patterns for these sequences were adapted from the stimuli of Tillmann and Poulin-Charronnat (2010). These authors comprehensively demonstrated that the grammatical and ungrammatical test sequences were identical in terms of descriptors that can influence implicit learning, such as the frequency of elements or bigrams. Thus any learning of the artificial grammar cannot be attributed to the detection of simple feature violations in the ungrammatical sequences. The changed tone varied in position – for 12% of the five-tone sequences it was in position 1 (start of the sequence), 12% in position 2, 6% in position 3, 47% in position 4, and 24% in position 5. For the six-tone sequences, 5% had the changed tone in position 1, 25% in position 2, 10% in position 3, 10% in position 4, 25% in position 5, and 25% in position 6.

Apparatus. All stimuli were generated using custom scripts written in Matlab. The experimental interface was programmed in Matlab with the Psychophysics toolbox (Brainard, 1997). The experiment was run in a quiet room on a PC running Windows 7, and participants heard the sequences over headphones at a comfortable loudness level.

Procedure. Participants were tested individually. For the exposure phase, each exposure sequence (all grammatical) was presented twice, with order randomized (leading to 74 trials),

and participants indicated whether they heard a given sequence for the first or second time by pressing one of two vertically aligned keys on the computer keyboard. This was the cover task used in Tillmann and Poulin-Charronnat (2010), intended to ensure attentive listening to the sequences. Above-chance in this task would provide assurance that (1) the sequence elements were sufficiently distinct that participants could perceive the differences, and (2) participants were attending to the sequences. To rule out differences in the test phase based on difference in attention in the exposure phase, we expected equal performance across condition. Participants were informed that there was an equal number of trials presented for the first or second time. Key assignments were counterbalanced across participants. At the end of the exposure phase, the test phase began after a five-minute break. Participants were told that all sequences they had heard had been created by a set of rules, and that they will now listen to new sequences. In each trial, these sequences were presented in pairs, but only one pair member followed these rules. Participants had to indicate whether the first or second sequence of the pair followed the rules by pressing one of two keys on the computer keyboard. They were informed that time to give their response was limited to 5 seconds, with an alerting sound signal after three seconds, reminding them that two seconds were left to respond. We collected confidence ratings to examine the extent of participants' conscious knowledge of the sequence structure. After each response, participants pressed a key on the computer keyboard to indicate how certain they were of their response on a 5-point scale (1=not certain/sure at all, 5 = very certain). Note that in the paired sequences, the grammatical and ungrammatical member differed by a single duration and/or pitch. There were 37 trials in the test phase (one for each grammatical/ungrammatical test pair), and the order of presentation of sequences in the exposure phase and of pairs in the test phase was randomized for each participant.

This research was approved by the Murdoch University Human Research Ethics Committee (approval number 2014/023).

Results

For exposure and test phases, we performed one-way ANOVAs on the percentage of correct responses, with the factor of Condition (AGp, AGd, AGpd) as a between-participants factor. For the test phase (Figure 3), accuracy was above chance (50%) only for AGp and AGpd, $t(19) = 4.09, 7.26$ (respectively), all p values $< .001$, but not for AGd, $t(19) = 0.13, p = .902$. The ANOVA showed that the main effect of condition was significant, $F(2, 57) = 13.43, p < .001, MSE = .010, \eta_p^2 = .320$. Pairwise comparisons revealed that accuracy for AGd was lower than performance for AGp ($p = .006, power = .92$), and AGpd ($p < .001, power = 1$), and that accuracy was higher for AGpd than AGp ($p = .022, power = .8$). Bonferroni corrections confirmed these findings except for the AGp/AGpd difference, which just fell short of significance ($p = .067$). This suggests that learning was only successful when the artificial grammar was presented in pitch or both pitch and duration, but not when presented in duration only. Further, learning was nominally more successful when the artificial grammar was presented in both pitch and duration than when only in pitch.

For the cover task of the exposure phase (Table 3), accuracy was above chance (50%) for all three conditions: $t(19) = 6.51$ (AGp), 4.43 (AGd), 6.38 (AGpd); all p values $< .001$. The ANOVA showed that the main effect of condition failed to reach significance, $F(2, 57) = 2.95, p = .060, MSE = .004, \eta_p^2 = .094$, suggesting that the exposure task was equally difficult across materials. However Table 3 shows an appreciable difference between the AGd and AGpd conditions ($p = .023$, but after Bonferroni correction for multiple comparisons, $p = .07$). These marginal results suggest it is possible that the exposure task was more difficult in the AGd condition than in the AGpd condition (see next section on test phase accuracy). Nevertheless, all conditions performed above chance (50%), meaning the duration patterns were sufficiently distinct for participants to succeed in the experimental task.

Table 3

Exposure phase accuracy in Experiments 1-3. Performance expressed in mean accuracy of correct responses (standard deviation in parentheses).

Experiment	Condition	Accuracy % (SD)
1	Pitch (AGp)	58.8 (6.0)
	Duration (AGd)	55.4 (5.5)
	Both (AGpd)	60.0 (7.0)
2	Pitch (AGp)	63.0 (6.8)
	Duration (AGd)	62.4 (8.9)
	Both (AGpd)	64.9 (6.3)
3	Pitch (AGp)	60.5 (10.8)
	Duration (AGd)	61.1 (8.1)
	Both (AGpd)	62.4 (10.2)

Exposure and test performance did not correlate significantly for any condition: AGp $r(18) = .05$, $p = .82$; AGd $r(18) = -.21$, $p = .38$; AGpd $r(18) = .17$, $p = .47$. For all conditions, musical expertise (years of formal training) did not correlate with accuracy levels in the exposure phase: AGp $r(18) = .06$ ($p = .813$); AGd $r(18) = .31$ ($p = .190$); AGpd $r(18) = .40$ ($p = .078$), and did not correlate with test phase accuracy: AGp $r(18) = .34$ ($p = .148$); AGd $r(18) = .03$ ($p = .893$); AGpd $r(18) = .19$ ($p = .430$).

To evaluate the unconscious status of what is learned when using grammaticality judgments, the “zero correlation criterion” proposed by Dienes et al. (1995; Dienes & Scott, 2005) was computed. The zero correlation criterion compares the means of confidence ratings for correct and incorrect responses; a difference of 0 means that participant confidence in

incorrect decisions was just as great as their confidence in correct decisions, that is, participants were unaware of what they used to make their decision. Confidence ratings for correct trials were higher than confidence ratings for incorrect trials for AGp, $t(19) = 3.32, p = .004$ (M correct = 3.02, $SD = .71$; M incorrect = 2.75, $SD = .68$), and for AGpd, $t(19) = 2.11, p = .049$ (M correct = 3.42, $SD = .76$; M incorrect = 3.09, $SD = .78$), but only marginally higher for AGd, $t(19) = 1.88, p = .075$ (M correct = 3.15, $SD = .64$; M incorrect = 3.03, $SD = .66$).⁵ A second method for assessing the zero correlation criterion is to compare the mean accuracy for trials that received low confidence ratings (<3) to the mean accuracy for trials that received higher confidence ratings (≥ 3). One AGp participant did not provide any ratings above 2, and so was excluded from this analysis. In this measure, accuracy was higher for high-confidence trials than low-confidence trials for AGp, $t(18) = 2.57, p = .019$ (M high = .65, $SD = .15$; M low = .49, $SD = .23$), but not for AGd, $t(19) = 1.31, p = .206$ (M high = .56, $SD = .15$; M low = .49, $SD = .21$), nor for AGpd, $t(19) = 0.92, p = .367$ (M high = .68, $SD = .12$; M low = .61, $SD = .29$).

The zero correlation criterion scores (Dienes et al., 1995; Dienes & Scott, 2005) suggest that despite the incidental learning in the exposure phase participants acquired some conscious judgement knowledge, at least for the AGp condition, and somewhat for the AGpd condition. However, there was no evidence of conscious knowledge of the rule structures from the debriefing reports in which participants were asked to “Write down the rules of the system that you think that you have discovered.” There was also no articulation of rule structure from an additional prompt of “If you have used a strategy, please describe it here.”

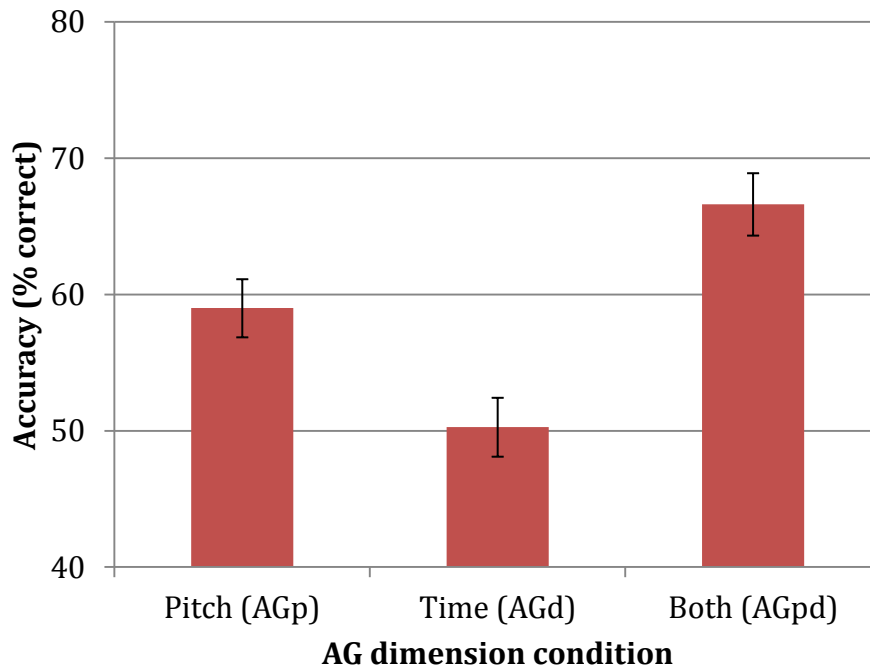


Figure 3: Accuracy for the test phase of Experiment 1 as a function of condition (the artificial grammar in pitch (AGp), duration (AGd) or both (AGpd)). Error bars are standard error of the mean.

Discussion

One goal of Experiment 1 was to assess learning of an artificial grammar when instantiated in time (played by isotonous pitches) or in pitch (played with an isochronous timing) as well as learning of a correlated joint artificial grammar applied to both duration and pitch. The results revealed that participants learned the artificial grammar when it was expressed in pitch sequences (i.e., performance significantly exceeded chance for AGp), whereas they did not when the artificial grammar was expressed in duration sequences, with accuracy levels at chance and significantly lower than those for AGp. The second goal of Experiment 1 was test learning of a joint artificial grammar (AGpd). The correlated pitch-duration artificial grammar was learned (as reflected by the significant above-chance level

performance for AGpd). Moreover, this condition benefitted from the correlated structure, with average AGpd performance being slightly higher than AGp performance (although just failing to reach significance with a Bonferroni correction).

The results of AGp extend the findings of Tillmann and Poulin-Charronnat (2010) to a pitch set that is not based on the equal-tempered scale and that used pure tones instead of spectrally rich piano tones. They also provide converging evidence with Loui et al. (2010), which used another type of grammatical system based on an unfamiliar musical scale (i.e., Bohlen-Pierce scale). Altogether, these findings confirm that incidental exposure can give rise to learning of tone structures, which reflect one example of non-verbal materials that are therefore non-referential, and difficult to verbalize because they do not allow using similarities with existing materials (e.g., words, acronyms).

The results further show that an artificial grammar in pitch can not only be learned with invariant durations (here AGp, also in Tillmann and Poulin-Charronnat, 2010), but also in the presence of non-metrical temporal variations (AGpd). That is, in the AGpd sequences, isochrony (of the AGp sequences) was replaced with irregular (i.e., non-metrical) timing, and yet listeners learned the AG (indeed, even better than in AGp). Thus, the findings further confirm and extend the findings of Selchenkova, Jones, et al. (2014) who showed that learning of an artificial pitch grammar implemented in temporally structured (in their case, metrical) sequences was better than learning of the artificial pitch grammar implemented in sequences with irregular (non-metrical) timing. In the AGpd condition of the present study, the artificial grammar in pitch was supported by an additional implementation of the artificial grammar in duration, thus providing more temporal structure than that in the isochronous (non-informative timing) AGp condition. In the case of Selchenkova et al., they attributed the benefit of temporal structure to entrainment with the regular stimulus rhythms guiding listeners' attending to the 'right' time of forthcoming aspects of sequence (e.g., Jones, 2008;

Large & Jones, 1999). Note that for Selchenkova et al., the temporal manipulation of the pitch sequences was independent from the artificial pitch grammar structures. By contrast, in the present research, pitch structure was correlated with time structure in the AGpd condition, and performance in this condition was slightly better than performance in the AGp condition (with invariant timing). Such findings suggest an advantage of correlated structures (with the same artificial grammar in both pitch and duration) for learning, although the failure of learning the artificial grammar when only in duration (AGd) may have reduced the strength of this advantage.

Overall, Experiment 1 shows that listeners learned artificial grammars in pitch when they appear in an isochronous presentation or when accompanied by a correlated artificial grammar in duration. What they did not learn was the same artificial grammar when applied to durations with a repeating pitch.

Experiment 2

Experiment 2 investigated whether (a) the AGd can be learned when the durations involved are more distinctive than in Experiment 1, and/or (b) whether with such a modification used for the correlated pitch-duration structures of AGpd would enhance the learning advantage of AGpd over AGp. Accordingly, Experiment 2 replicates Experiment 1, with a more distinctive duration set (the pitch set was as in Experiment 1). To support our aim to provide the best possible scenario for observing learning of the artificial grammar in duration sequences, we increased the length of the exposure phase by using three repetitions of the grammatical exposure materials instead of two repetitions (note that this increase of the exposure phase was applied to AGp, AGd, and AGpd). Tillmann and Poulin-Charronnat

(2010) had started with three repetitions (Experiment 1) and reduced to two repetitions (Experiment 2), still showing the same amount of learning for an artificial grammar in pitch.

Method

Participants. One-hundred seven students⁶ of Murdoch University, with an average age of 24.5 years ($SD = 7.5$) and an average of 1.9 years of formal musical training ($SD = 3.2$), participated in Experiment 2 after having given informed consent; none had participated in Experiment 1. As in Experiment 1, participants were randomly assigned to AGp ($N=34$), AGd ($N=36$), or AGpd ($N = 37$). Musical expertise did not differ across condition, $F(2, 104) = 1.69, p = .19, MSE = 10.31$.

Stimuli and apparatus. Stimuli and apparatus were as described for Experiment 1, except for the definition of the duration set to generate AGd. To make the sequences more distinctive, the duration set was changed to the following values: 60, 110, 250, 550, and 920 ms (see Table 1). The Weber fractions for adjacent values in this set ranged from .67 to 1.27, and were thus higher than for Experiment 1. The ISI remained at 100 ms.

Procedure. The procedure was as described for Experiment 1, except that exposure sequences were presented three times in random order for each participant (leading to 111 exposure phase trials). The number of test trials remained at 37.

Results

For the test phase (Figure 4), accuracy was above chance (50%) for all conditions: AGp $t(33) = 3.89, p < .001$; AGd $t(35) = 2.31, p = .027$; AGpd $t(36) = 9.97, p < .001$. The main effect of condition was significant, $F(2, 104) = 18.28, p < .001, MSE = .011, \eta_p^2 = .260$. Pairwise comparisons revealed that performance did not differ between AGp and AGd ($p =$

.128, power = .60), but performance of AGpd was higher than performance of AGp and AGd, respectively (both p values < .001, power = 1). This pattern remained when applying a Bonferroni correction for multiple comparisons. This suggests that learning was successful in all conditions, regardless of whether the artificial grammar was presented in pitches, durations, or both. Additionally, learning was more successful when the artificial grammar was presented in both pitch and duration.

For the exposure phase (Table 3), accuracy in the cover task was above chance (50%) for the three conditions: $t(19) = 11.1$ (AGp), 8.38 (AGd), 14.39 (AGpd); all p values < .001. The ANOVA (as in Experiment 1) revealed that there was no significant effect of condition, $F(2, 104) = 1.14$, $p = .323$, $MSE = .006$, $\eta_p^2 = .021$, suggesting that the exposure phase task was equally difficult across materials.

Exposure and test performance did not correlate significantly for any condition: AGp $r(32) = .27$, $p = .13$; AGd $r(33) = -.01$, $p = .94$; AGpd $r(35) = -.02$, $p = .88$. For all conditions, musical expertise did not correlate with exposure phase accuracy: AGp $r(32) = .14$ ($p = .431$); AGd $r(34) = -.16$ ($p = .928$); AGpd $r(35) = -.18$ ($p = .293$), and did not correlate with test phase accuracy: AGp $r(32) = .01$ ($p = .997$); AGd $r(34) = .09$ ($p = .616$); AGpd $r(35) = .10$ ($p = .558$).

With regards to the zero correlation criterion, confidence ratings in the AGp condition for correct trials were marginally higher than confidence ratings for incorrect trials, $t(33) = 1.80$, $p = .081$ (M correct = 3.05, $SD = .75$; M incorrect = 2.94, $SD = .86$). For AGd and for AGpd, this difference was significant, $t(35) = 2.40$ $p = .022$ (M correct = 3.25, $SD = .57$; M incorrect = 3.14, $SD = .65$) and $t(36) = 5.31$, $p < .001$ (M correct = 3.29, $SD = .69$; M incorrect = 2.94, $SD = .73$), respectively. The second zero correlation criterion measurement revealed that accuracy was no higher for high-confidence trials than low-confidence trials for AGp, $t(33) = 0.18$, $p = .862$ (M high = .60, $SD = .17$; M low = .57, $SD = .24$), also for AGd, $t(35) = 0.65$, p

= .521 (M high = .55, SD = .11; M low = .51, SD = .21). For AGpd, this difference was significant, $t(36) = 2.99$, $p = .005$ (M high = .70, SD = .17; M low = .54, SD = .26). These findings suggest that participants acquired some conscious knowledge (for the AGpd condition, and less so for the AGd condition), despite the incidental learning design of the exposure phase and the lack of evidence of conscious structural knowledge in the debriefing reports.

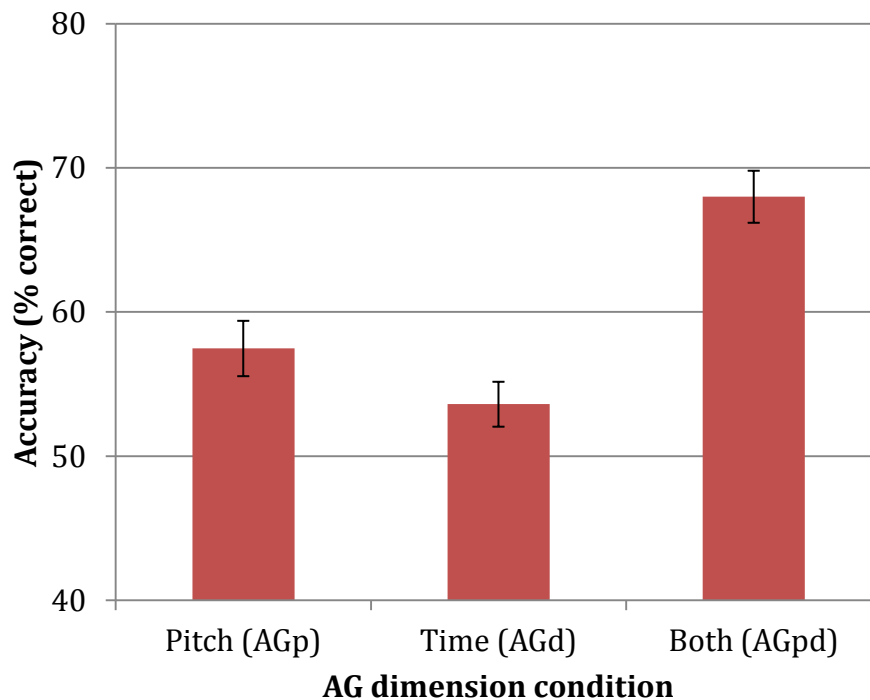


Figure 4. Accuracy for the test phase of Experiment 2 as a function of condition (the artificial grammar applied to pitch (AGp), duration (AGd) or both (AGpd)). Error bars are standard error of the mean.

Discussion

Experiment 2 used the same experimental set-up as Experiment 1, but used a more distinctive duration set and increased exposure for all artificial grammar conditions, aiming to provide most favorable conditions to show learning also for the AGd, if possible. These changes resulted in significantly above-chance learning in all three conditions. Further, performance of the combined version (AGpd) was better than performance of the single-dimension versions (AGp and AGd, which did not differ significantly, even though mean performance was slightly higher for AGp than for AGd).

With more distinctive durations and a longer exposure phase than Experiment 1, participants learned the artificial grammar in duration (AGd), and there were nominal improvements in AGpd learning. By implementing the two changes (duration distinctiveness and increased exposure) simultaneously, we cannot determine their relative contributions. However, the increased exposure of Experiment 2 had no benefit for AGp test phase accuracy, (in fact, a nominal decrease occurred between the two experiments, i.e., from 59% to 57%), thus the more distinctive durations provide a better explanation for the improvements in AGd and AGpd accuracy. The pilot study showed high discrimination accuracy (see Table 2) between single items for a Weber fraction of .30 for durations and a Weber fraction of .04 for tones. For the AGd condition, the Experiment 1 values were close to those established in pilot testing, with Weber fractions between .32 and .40, but the AGp condition used a Weber fraction of .08, larger than the pilot testing value of .04. This latter choice was due to the need to avoid Weber fractions of semitone differences of Western tonal equal-tempered scale (i.e., the adjacent keys of a piano keyboard), which were based on .06. Aiming to reach more comparable Weber fractions (i.e., approximately doubling the Weber fraction observed in the pilot study, as done for the pitch dimension), Experiment 2 thus changed the AGd set by using a Weber fraction with a minimum of .67 and showed learning of the artificial grammar when applied only to duration.

Experiment 2 showed learning in the AGd condition, in spite of the fact that the rhythms involved were not regular. The results thus highlight the power of an implicit learning paradigm even for temporal structures that do not conform to regular (metrical) structures, which would provide processing benefits (as predicted by the Dynamic Attending Theory, see Jones, 2016, and shown by Selchenkova, Jones, et al., 2014 for AGp learning). In Experiment 2, AGd sequences were isotonous (invariant pitch). This represents the mirror situation of the isochronous (invariant durations) AGp sequences, and also of Tillmann & Poulin-Charronnat's (2010) investigation of AGp learning. However, this AGd condition differs from previous research on implicit learning of temporal structure, which used serial response time paradigms (Brandon et al., 2012; Buchner & Steffens, 2001; Schultz et al., 2013; Shin & Ivry, 2002). These studies used another structure on the event dimension that was ambiguously related or shifted in structure (thus not correlated) with the temporal structure, or had random variations of the secondary (event) dimension (using tones, syllables or locations). In other words, such experiments used ambiguous or shifted structure relations. They yielded no learning of the time structure although this may have been partly due to - or reinforced by - the response-stimulus-interval methodology and the use of non-metrical timing. In contrast, learning of temporal structure was shown in serial response time experiments that used random variations of the event dimension (syllables or locations) and their implementation in a fixed IOI framework with metrical structures (Brandon et al., 2012; Schultz et al., 2013; Selchenkova, François, et al., 2014). Together, these previous findings raise the question of whether an artificial grammar in durations could be learned with random pitch information (i.e., variant secondary dimension), when presented without metrical structure. For AGp, we used isochronous sequences in Experiments 1 and 2 (as in Tillmann & Poulin-Charronnat, 2010), and an artificial grammar in pitch can be learned even with variable time sequences (Selchenkova, Jones, et al., 2014). Such temporal variations influence pitch learning: regular

rhythms with strongly metrical structures led to better pitch learning than irregular temporal structures, and also better than isochronous sequences (Selchenkova, François, et al., 2014; Selchenkova, Jones, et al., 2014). However, no previous study has addressed the mirror situation for the learning of artificial grammar structure in time.

Experiment 3

The two lines of research findings just reviewed lead to questions about the impact of the structure of the non-relevant (secondary) dimension in tasks requiring learning of an artificial grammar in pitch or duration. Accordingly, Experiment 3 investigated learning of an artificial grammar in pitch or duration (AGp and AGd, as in Experiment 2) when the non-relevant (secondary) dimension had random variations (of either durations or pitches, respectively). For comparison, we retained the AGpd condition, which was identical to Experiment 2.

Method

Participants. One-hundred five students of Murdoch University, with an average age of 25.0 years ($SD = 7.5$) and an average of 2.1 years of formal musical training ($SD = 3.3$), participated in Experiment 3 after having given written informed consent; none had participated in Experiments 1 and 2. As in Experiments 1 and 2, participants were randomly assigned to AGp ($N=34$), AGd ($N=35$), or AGpd ($N = 36$). Musical expertise did not differ across condition, $F(2, 102) = 1.01, p = .37, MSE = 11.18$.

Stimuli and apparatus. Stimuli and apparatus were as described for Experiment 2, with the following changes: For the AGp condition, each pitch sequence was as described in Experiment 2, but instead of being presented isochronously, the pattern used to determine the

pitch values was randomly reordered for the purpose of selecting the duration values. Put differently, the AGp condition used the same stimuli as in the AGpd condition, but the order of the durations was scrambled. For the AGd condition, each duration sequence was as described in Experiment 2, but instead of being presented isotonously, the pitch pattern was randomized. Thus the AGd condition used the same stimuli as in the AGpd condition, but the order of the pitches was scrambled. The ISI remained at 100 ms. The AGpd condition was exactly the same as in Experiment 2.

Procedure. The procedure was as described for Experiment 2.

Results

For the test phase (Figure 5), accuracy was above chance (50%) for AGp, $t(33) = 4.30$, $p < .001$, and for AGpd, $t(35) = 10.62$, $p < .001$, but not for AGd $t(34) = 1.35$, $p = .186$. The main effect of condition was significant, $F(2, 102) = 33.04$, $p < .001$, $MSE = .009$, $\eta_p^2 = .393$. Pairwise comparisons revealed that there was no significant difference between the AGp and AGd conditions ($p = .106$, power = .46), but AGpd outperformed both AGp and AGd (both p values $< .001$, power = 1). This pattern remained when applying a Bonferroni correction for multiple comparisons. This suggests that, like Experiment 1, learning was only successful when the artificial grammar was presented in pitch or both pitch and duration, but not when presented in duration only.

For the exposure phase (Table 3), accuracy was above chance (50%) for the three conditions: $t(33) = 5.66$ (AGp), 8.12 (AGd), 7.31 (AGpd); all p values $< .001$. The ANOVA analysis showed no significant effect of condition, $F(2, 102) < 1$, ns , suggesting that the exposure phase cover task was equally difficult across materials.

Exposure and test performance did not correlate significantly for AGp, $r(32) = .09$, $p = .62$ or for AGd $r(33) = -.01$, $p = .99$; but for AGpd there was a significant correlation $r(34) = .36$, $p = .03$. This may well be spurious given that this relationship was not significant in Experiments 1 and 2. For all conditions, musical expertise did not correlate with exposure phase accuracy: AGp $r(32) = .18$ ($p = .317$); AGd $r(33) = -.08$ ($p = .651$); AGpd $r(34) = .09$ ($p = .595$), and did not correlate with test phase accuracy: AGp $r(32) = .06$ ($p = .722$); AGd $r(33) = .18$ ($p = .289$); AGpd $r(34) = -.26$ ($p = .125$).

In the zero correlation criterion analysis, confidence ratings for correct trials were higher than confidence ratings for incorrect trials for AGp, $t(33) = 2.25$, $p = .032$ (M correct = 2.95, $SD = .78$; M incorrect = 2.84, $SD = .79$), and for AGpd, $t(35) = 5.18$, $p < .001$ (M correct = 3.32, $SD = .53$; M incorrect = 2.89, $SD = .63$), but they did not differ for AGd, $t(34) = -.42$, $p = .680$ (M correct = 3.00, $SD = .81$; M incorrect = 3.02, $SD = .79$). Similarly, the second zero correlation criterion measure showed that accuracy was higher for high-confidence trials than low-confidence trials for AGp, $t(33) = 2.12$, $p = .041$ (M high = .58, $SD = .14$; M low = .49, $SD = .24$) and AGpd, $t(35) = 4.54$, $p < .001$ (M high = .74, $SD = .13$; M low = .51, $SD = .27$), but not for AGd, $t(34) = 0.13$, $p = .897$ (M high = .49, $SD = .15$; M low = .48, $SD = .22$). These findings suggest that participants in the AGp and AGpd conditions acquired some conscious knowledge, despite the incidental learning design of the exposure phase and the lack of evidence of conscious structural knowledge in the debriefing reports.

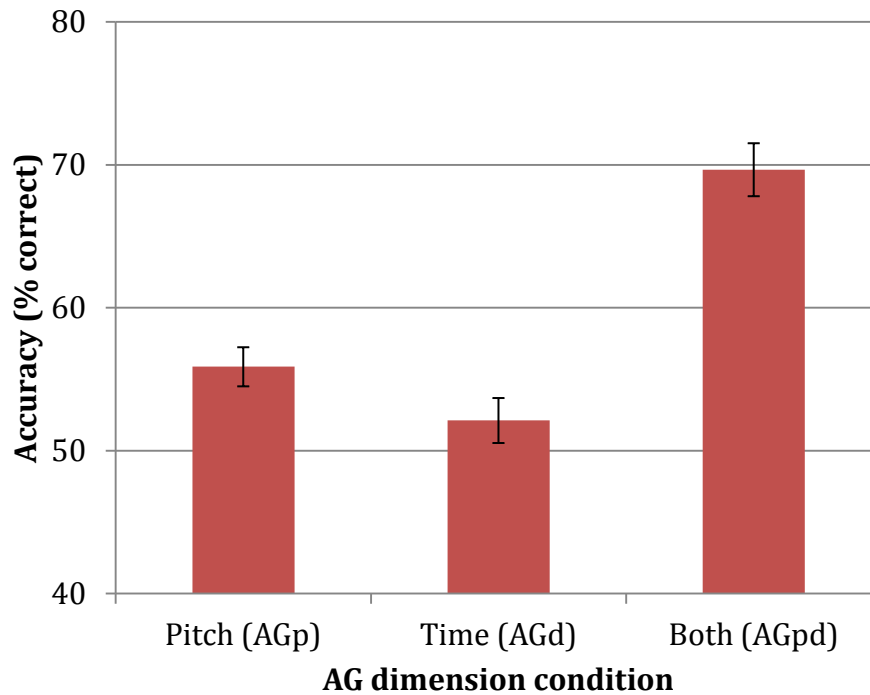


Figure 5. Accuracy for the test phase of Experiment 3 presented as a function of condition (the artificial grammar was in pitch (AGp), duration (AGd) or both (AGpd)). Error bars are standard error of the mean.

Discussion

Experiment 3 investigated the learning of AGp and AGd when accompanied by random variation in the non-relevant (secondary) dimension (either durations or pitches, respectively).

For AGp, results of the test phase revealed performance significantly above chance, which suggests that the learning of an artificial grammar in pitch resists random timing variations. With our present paradigm (measuring percentages of correct performance in a grammaticality judgment task), we did not observe a cost of the random timing variations in comparison to an isochronous timing implementation as had been reported by Selchenkova, Jones, et al. (2014) with response times in a priming task. It is possible that response times

show greater sensitivity than accuracy measures to the potential influence of timing presentations (note, however, that the response time data of Selchenkova et al. (2014) also revealed at least partial learning of the pitch artificial grammar with irregular timing). Furthermore, both Selchenkova et al. studies used tones corresponding to the equal-tempered scale, which might conflict more strongly with irregular timing than the present pitch sequences, which were built with tones that did not correspond to the Western musical system.

For AGd, however, the random variations in the non-relevant (secondary) dimension (pitch) resulted in test accuracy at chance level in Experiment 3, despite successful learning with invariant pitch in Experiment 2. Thus although AGp learning in Experiment 3 was resistant to random duration variations, the AGd learning was not resistant to random pitch variations (that is, AGd was at chance performance). Together with the successful learning of time structures in serial response time paradigms despite random syllable variations (e.g., Brandon et al., 2012), this finding suggests that (1) learning a time-based artificial grammar is more demanding than time-structure learning in an serial response time paradigm (with a single short sequence cycling repeatedly over exposure) and/or (2) that pitch variation exerts stronger interference than syllable variations on time-structure learning. For AGpd, which was the same as Experiment 2, we replicated both the above-chance test phase performance and the superior performance in AGpd than the two other conditions (AGp, AGd).

General discussion

Our present study aimed to investigate whether (1) an artificial grammar expressed in duration can be learned as well as the same artificial grammar in pitch, (2) applying the artificial grammar to both duration and pitch (in correlated fashion) boosts learning, and (3)

the presence of unstructured variability in the secondary dimension (invariant versus randomized) influences learning structures of the primary dimension (single dimension learning).

The findings reveal that both AGp and AGd can be learned when the secondary dimension is invariant, that is, with isochronous or isotonous presentations, respectively. Further, doubling the Weber fraction obtained for discriminating isolated presentations (for tones and durations) at ceiling performance was necessary for learning AGd. Indeed, AGd learning was observed in Experiment 2, but not in Experiment 1. The longer exposure phase in Experiment 2 might have contributed to the above-chance AGd learning, however note that the increased exposure did not lead to improved test performance for AGp (see Experiment 2 Discussion section).

In contrast to AGp, which was learned both with an invariant durations (isochronous) and random durations, AGd showed significant above-chance level performance only with invariant pitch (isotonous), but not with a random pitch (Experiments 2 and 3). This observation can be integrated with prior findings investigating the learning of temporal structures. Although earlier research used the serial response time paradigm and not the artificial grammar learning paradigm, these studies offer data that reveal learning of temporal structure when it was accompanied by an invariant secondary dimension (e.g., the same pitch presented repeatedly in Salidis, 2001) or a pseudo-random implementation (e.g., syllables in Brandon et al., 2012), but not with a different structure that was unrelated or ambiguously related (e.g., tones in Buchner & Steffens, 2001; spatial locations in Shin & Ivry, 2002). In our present study, we did not go so far as to use different artificial grammars in pitch and time, but given that temporal structure learning was unsuccessful with random pitches, it is unlikely to succeed when combined with a different pitch structure. .

It is interesting to point out that Shin and Ivry (2002) reported that although temporal learning was observed only in the correlated condition (but not the ambiguous, shifted condition), spatial learning was observed in both conditions (i.e., correlated and ambiguous). This asymmetry is similar to that observed in our study: although temporal structure learning was observed only when pitch was invariant (and not when random), pitch structure learning was observed both for invariant and randomized durations. In contrast to our findings (Experiment 2), Shin and Ivry (2002) did not observe learning of the temporal pattern independently from learning of the spatial pattern, which was related to a movement pattern (i.e., chaining of key presses). This observation led the authors to conclude that temporal pattern learning cannot occur independently from action learning, even though it can enhance the spatial (action) learning, also supporting a hypothesis of the integration of spatial-temporal information in learning and perception. The inherent relation between spatial and movement sequences in the serial response time paradigm precludes testing the learnability of these dimensions separately. In contrast, the artificial grammar learning paradigm does not suffer from this limitation, thus providing a favorable testing situation for the investigation of temporal structure learning, notably as it is not confounded with action sequences.

The artificial grammar learning paradigm is also useful in that, despite explicit judgements of grammaticality, it provides an opportunity to assess implicit or incidental learning in the exposure phase where participants were not informed about the AG in the sequences nor instructed to learn any rules. As implemented in our study, the findings revealed that this incidental learning during the exposure led to knowledge that was not entirely implicit. Specifically, confidence rating scores showed that participants were more confident in answers on trials they answered correctly than trials on which they were incorrect. This suggests that although the learning during the exposure phase was incidental (i.e., not actively directed towards the artificial grammar), participants nonetheless acquired

some conscious knowledge of the sequence structure. Selchenkova, François, et al. (2014) also found evidence for conscious learning of an artificial grammar in pitch in confidence ratings, marginally more so for temporally structured (metrical) sequences than for an isochronous presentation.

The zero correlation criterion tests thus showed some evidence of explicit knowledge of the grammatical structures, and are largely congruent with the test phase accuracy results. That is, when test phase accuracy was above chance (i.e., artificial grammar learning was successful), there was evidence of some explicit knowledge based on the zero correlation criterion rating measure (with AGp in Experiment 2 being the exception). The accuracy-based zero correlation criterion measure partly aligned in this regard, but was more conservative. That is, this measure only provided evidence of explicit knowledge when learning (test phase accuracy) was above chance *and* the rating-based zero correlation criterion showed evidence of explicit knowledge. Furthermore, in two cases (AGpd in Experiment 1, and AGd in Experiment 2), only one of the zero correlation criterion measures showed evidence for explicit knowledge. Combined with the fact that the post-experiment questionnaire revealed no conscious knowledge of the artificial grammar, these findings suggest that participants' acquired explicit knowledge was weak and/or vague. Nevertheless, the tendency for AGpd to show stronger evidence of explicit knowledge implies that indicating the artificial grammar simultaneously via pitch and duration sequences makes it more obvious and creates greater confidence in judgements. Future research is necessary to examine what is implicit and/or explicit in the acquired knowledge, perhaps by using additional methods such as process dissociation procedure (PDP) or generation procedures often used with the serial reaction time tasks (Destrebecqz & Cleeremans, 2001; Perruchet, 2008).

Why was the AGd condition test phase accuracy above chance for only one of the three experiments, whereas the AGp condition consistently performed above chance? One could

argue that unlike the pitch variations, the temporal variations were not sufficiently distinct for the participants to learn them. This explanation seems unlikely given the fact that the values used in both dimensions were well beyond threshold levels of discriminability as judged by the pilot study of isolated elements. Furthermore, the temporal structure should have been even easier to extract in the experiments than the pilot study would suggest. This is because when the durations were used in sequence (Experiments 1-3), the artificial grammar structure was also present in the form of inter-onset intervals, because the silent gap between notes was constant. This is because inter-onset interval, duration, and inter-stimulus interval are inherently linked. Manipulating inter-onset intervals without changing duration means the inter-stimulus interval has to vary (mirroring the inter-onset interval information). If instead duration is variable and the inter-stimulus interval is constant (as in the current study), then inter-onset interval has to vary along with duration. That is, the artificial grammar structure is available through both duration and inter-onset interval sequences. Giving two indications of the same artificial grammar means that participants could learn it through one or both of these forms. In contrast, in the pilot study the participants only judged relative duration of two items, without any additional indication from inter-onset intervals. Nevertheless, learning the artificial grammar was most difficult when communicated in durations (or, inter-onset intervals) only. It could be that the absence of a regular temporal framework made learning of AGd more difficult and thus prone to pitch interference (as per Selchenkova, Jones, et al., 2014). However, if lack of structure were to blame, then this should also decrement learning in pitch, because the pitch structure of tonality was absent (in fact, the sequences did not even conform to equal temperament). The fact that learning was successful in the AGp condition despite the presence of random durations suggests that lack of structure (be it meter or tonality) did not influence participants' ability to learn the artificial grammar.

Another explanation is that despite our attempts to avoid Western musical structure (e.g., equal temperament, regular metrical structure) in our stimuli, participants may have invoked pre-learned biases from their lifetime of exposure to Western music. Western music features more variability in pitch than in duration. For instance, in the 5,364 German folksongs from the Essen collection (Schaffrath, 1995), melodies have an average of 6.7 (SD = 1.0) unique pitch classes (collapsing across octave), and an average of 3.9 (SD = 1.2) unique durations. By comparison, our sequences had an average of 3.5 (SD = .84) unique elements (pitches for AGp, durations for AGd). The greater variability in pitch may result in encultured listeners prioritizing pitch over time, as demonstrated by experiments testing the relative influence of these dimensions in metrical grouping and goodness-of-fit judgements (Ellis & Jones, 2009; Prince, 2014b; Prince et al., 2009). In those studies, after using baseline experiments to find equally effective levels of pitch and temporal manipulations when tested separately, pitch was more influential when the dimensions were recombined (in the subsequent experiments). Other research showed that detecting pitch changes in a sequence was easier than detecting loudness changes, despite matching their discriminability when tested in isolation, supporting a prioritized status of pitch in perceptual processing. (Cousineau, Demany, & Pressnitzer, 2009). Despite efforts to equalize pitch and time (i.e., via discriminability) in the present study, for the perceivers, pitch might remain more influential than time, leading to the observed findings. Accordingly, it would be difficult for participants to ignore the random pitch information and extract the temporal structure of the AGd, whereas participants exposed to the AGp would automatically attend to the relevant (i.e., structured) dimension of pitch and thus be able to ignore the irrelevant random time information (Experiment 3).

An additional explanation for the AGd performance in Experiment 3 comes from recent research on sequence complexity (Silva, Petersson, & Castro, 2017). In an exposure phase, participants reproduced via keypress either simple sequences (high and low tones mapped to

up and down arrow keys) or complex sequences (four vowels mapped to up/down/left/right keys), with both types having an embedded temporal pattern. In a subsequent test phase, participants reproduced the sequence timing, using a single key. Only the participants who heard simple sequences in the initial phase showed learning of the temporal pattern, suggesting that the accompanying information in the complex condition presented a barrier to learning the temporal information.

A related intriguing finding from the current study was the drastically larger duration Weber fractions required for successful artificial grammar learning (Experiment 2). From the pilot study, the duration Weber fractions were roughly equivalent to ten times the size of the pitch Weber fractions. This aligns with the existing literature on frequency and duration discrimination (France et al., 2002; Grondin, 1993). However, the Weber fractions required for successful learning in the AGd condition were yet larger. There are several possible explanations for this finding. First, whereas pitch contour is a critical component of melodic perception (Dowling, 1978; Schmuckler, 2004), a temporal analogue of contour in duration patterns is not – few papers even mention the concept (Monahan, Kendall, & Carterette, 1987; Prince, 2014a; Schulkind, 1999). There is research on how tempo slows at the end of a piece (cf. Honing, 2003), however in comparison to the ubiquitous nature of melodic contour, this is a more ornamental feature of music than a fundamental structural principle. Second, Brown (1997) found that temporal tasks are more susceptible to increasing processing load than nontemporal tasks – whereas concurrent pursuit rotor tracking or visual search tasks decrease performance in a timing task (i.e., generating a 2-5 s patterned sequence of durations), the other tasks did not observe a decrease in performance (although mental arithmetic showed some disruption). Thus the processing of detailed temporal information is likely to be less resistant to irrelevant information than other forms of auditory information, such as pitch. Third, a prioritization of pitch via enculturation (as discussed with regards to the Essen

collection) may lead to the constraint that learning an artificial grammar in the temporal domain requires greater distinctiveness between items than learning an artificial grammar in the pitch domain. Our data cannot distinguish between these explanations; moreover, it is highly probably that it is based on a combination thereof.

In the AGpd condition, the artificial grammar was expressed in both pitch and duration, and learning was better than in AGp and AGd conditions, with artificial grammars presented in a single dimension (that is, in pitch or in duration only). This result showed that presenting a single artificial grammar simultaneously (and in correlated fashion) in two dimensions benefits learning, even when the structure was not learned above chance level for one of the dimensions (i.e., AGd in Experiments 1 and 3). When further analyzing the data patterns of the three experiments, the results reveal an overadditivity of performance, thus suggesting an interaction between the learning of pitch and temporal structure, and despite the fact that the structure was not learned separately in one dimension. In Experiment 1, AGp was 9 percentage-points above chance level (50%), AGd at chance level (0 difference from chance level), while AGpd was 17 points above chance level – thus outperforming a simple addition of above-chance performance for single-dimension artificial grammars (i.e., $9+0 < 17$). A similar pattern emerged in Experiment 2, with AGp 7 percentage-points above chance level, AGd 4 percentage-points, but AGpd 18 percentage points (instead of 11 as would have been predicted by additivity). Finally, this difference was the largest in Experiment 3, with AGp 6 percentage points above chance level, AGd 2 percentage points, but AGpd 20 percentage points (instead of 8 as would have been predicted by additivity). This overadditivity in the data pattern differs from serial reaction time findings observed for either learning of spatial and temporal information (Shin & Ivry, 2002) or spatial and object information (Mayr, 1996). For spatial and temporal learning (as tested by Shin & Ivry, 2002), temporal structure cannot be learned independently, but instead it boosts the learning of the spatial dimension when

correlated. For spatial and object learning (Mayr, 1996), each single dimension can be learned as efficiently as the two dimensions combined in correlated fashion, thus suggesting independent systems for the learning of spatial and non-spatial regularities.

In these experiments we inadvertently mapped higher pitches to longer durations (and lower pitches to shorter durations) for the AGpd condition. It might be argued that using a haphazard mapping of pitch height and duration could give a different result pattern across conditions. However, this is unlikely for two reasons. First, performance in the cover task of the exposure phase was equal across all conditions (AGp, AGd, and AGpd), even though only AGpd had this joint mapping. Second, this mapping did not provide an advantage in the test phase because all AGpd sequences followed this pattern, including the ungrammatical test sequences. That is, the mapping did not provide a method to differentiate between grammatical and ungrammatical test sequences, so the same pattern of results across condition should emerge regardless of the pitch-duration mapping in the AGpd condition.

The observed overadditive findings suggest that learning improved based on the particular combination of pitch and temporal information, above and beyond their independent contributions, such that learning of pitch and time structures are not independent, but interact in processing (i.e., perception, memory, and learning). This result is particularly interesting in the context of the debate in music cognition research concerning the potential integration of pitch and time processing, which has alternately suggested that these dimensions are processed independently (e.g., Palmer & Krumhansl, 1987a; Palmer & Krumhansl, 1987b; Schellenberg, Stalinski, & Marks, 2014) or interactively (e.g., Ellis & Jones, 2009; Jones, 1987). Attempts to reconcile the divergent findings on this issue include a processing-stage model (independent in early processing, and interactive later on, e.g., Thompson, Hall, & Pressing, 2001; Tillmann & Lebrun-Guillaud, 2006), and a dimensional salience model (interactive relations are more easily observed when dimensions are equally

matched, e.g., Prince, 2011). The present study is the first to apply a learning paradigm to this question, suggesting that the pitch and temporal structure were processed interactively in this context. Investigating pitch-time integration with a learning paradigm in addition to perception and production contexts (for reviews, see Jones & Pfordresher, 1997; Prince & Pfordresher, 2012), thus provides a new perspective for our understanding of underlying dimensional interactions. Based on our present findings, the processing-stage model would thus predict that the present implementation of learning allows for the emergence of interactive processing in later processing stages (testing of learning after the exposure phase with holistic grammaticality judgments). Future research should test the prediction of independence when investigating earlier processing stages of pitch and time, such as the measurement of event processing during the exposure phase as well as during the perception of single events in the test phase (in particular, with more sensitive measurements, such as response times or electrophysiological recordings).

Footnotes

¹. We investigate incidental learning of structures without explicit instruction to learn these structures. Perruchet (2008) states that “a feature which is part of virtually all definitions of implicit learning (IL) is the incidental nature of the acquisition process. IL proceeds without people’s intention to learn.” (p. 610). Berry and Dienes (1993) have a similar definition, in which learning is implicit “when a person typically learns about the structure of a fairly complex stimulus environment, without necessarily intending to do so, and in such a way that the resulting knowledge is difficult to express” (p. 2). Here we refer to implicit learning with the restricted definition of an exposure phase with incidental learning only.

². In the context of this article, we use the term structure to refer generally to an organized relation between elements. This structure could take many forms, such as an artificial grammar, a serial pattern (e.g., from a serial response time task), temporal regularity (metrical timing), musical key (tonality), or others. As we test artificial grammar learning in this article, this is the most common – but not exclusive – form of structure we refer to.

³. Equal temperament refers to the standardized set of frequencies that correspond to musical notes in Western music. A non-equal tempered set of pitches means that the notes do not align with these frequencies and their regular spacing, thus sounding “in the cracks” between adjacent keys on a piano.

⁴. Weber fractions refer to the amount of change in stimulus magnitude as a proportion of the base value. For example, changing the pitch of a 400 Hz tone to 404Hz represents a Weber fraction of .01, because $(404-400)/400 = .01$.

^{5.} To check for violations of statistical assumptions, we examined the normality of these data. The only measure to violate normality (Shapiro-Wilk) was the ratings for correct trials in the AGpd condition of Experiment 3 ($p = .045$). Nevertheless, we used non-parametric tests (Wilcoxon signed rank tests, sign tests, marginal homogeneity tests), to verify the findings, and all tests replicated the t-statistics. This was also the case for Experiments 2 and 3. Additionally, there were no overall differences in confidence ratings (neither for correct nor incorrect trials) between conditions in any experiment.

^{6.} To foreshadow the findings, the smaller sample size of Experiment 1 (compared to Experiments 2 and 3) raises the probability of a Type II error in Experiment 1. To address this issue we randomly selected 20 participants from each condition of Experiment 2 and Experiment 3, to equate sample sizes across all experiments to the smaller sample size of Experiment 1. We then re-ran the statistical analyses, and these analyses confirmed the pattern of findings, ruling out the possibility of a Type II error based on sample size differences.

References

- Altmann, G.T.M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning Memory and Cognition*, *21*, 899-912. doi: 10.1037//0278-7393.21.4.899
- Berry, D.C., & Dienes, Z. (1993). Towards a characterization of implicit learning. In D. C. Berry & Z. Dienes (Eds.), *Implicit learning: Theoretical and empirical issues* (pp. 1-18). Hillsdale, NJ: Lawrence Erlbaum.
- Bigand, E., Perruchet, P., & Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Cahiers De Psychologie Cognitive-Current Psychology of Cognition*, *17*, 577-600.
- Bigand, E., Tillmann, B., Poulin-Charronnat, B., & Manderlier, D. (2005). Repetition priming: Is music special? *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*, 1347-1375. doi: 10.1080/02724980443000601
- Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436. doi: 10.1163/156856897X00357
- Brandon, M., Terry, J., Stevens, C.J., & Tillmann, B. (2012). Incidental learning of temporal structures conforming to a metrical framework. *Frontiers in Psychology*, *3*. doi: 10.3389/fpsyg.2012.00294
- Brown, S.W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59*, 1118-1140. doi: 10.3758/bf03205526
- Buchner, A. (1994). Indirect effects of synthetic grammar learning in an identification task. *Journal of Experimental Psychology: Learning Memory and Cognition*, *20*, 550-566. doi: 10.1037//0278-7393.20.3.550
- Buchner, A., & Steffens, M.C. (2001). Simultaneous learning of different regularities in sequence learning tasks: Limits and characteristics. *Psychological Research/Psychologische Forschung*, *65*, 71-80. doi: 10.1007/s004260000052

- Cleeremans, A., & McClelland, J.L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235-253. doi: 10.1037/0096-3445.120.3.235
- Cousineau, M., Demany, L., & Pressnitzer, D. (2009). What makes a melody: The perceptual singularity of pitch sequences. *Journal of the Acoustical Society of America*, *126*, 3179-3187. doi: 10.1121/1.3257206
- Deroost, N., & Soetens, E. (2006). Spatial processing and perceptual sequence learning in srt tasks. *Experimental Psychology*, *53*, 16-30. doi: 10.1027/1618-3169.53.1.16
- Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, *8*, 343-350. doi: 10.3758/bf03196171
- Dienes, Z., Altmann, G.T.M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning Memory and Cognition*, *21*, 1322-1338. doi: 10.1037/0278-7393.21.5.1322
- Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*, 338-351. doi: 10.1007/s00426-004-0208-3
- Dowling, W.J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341-354. doi: 10.1037/0033-295X.85.4.341
- Ellis, R.J., & Jones, M.R. (2009). The role of accent salience and joint accent structure in meter perception. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 264-280. doi: 10.1037/a0013482
- France, S.J., Rosner, B.S., Hansen, P.C., Calvin, C., Talcott, J.B., Richardson, A.J., & Stein, J.F. (2002). Auditory frequency discrimination in adult developmental dyslexics. *Perception & Psychophysics*, *64*, 169-179. doi: 10.3758/bf03195783
- Gobel, E.W., Sanchez, D.J., & Reber, P.J. (2011). Integration of temporal and ordinal information during serial interception sequence learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*, 994-1000. doi: 10.1037/a0022959

- Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Attention, Perception, & Psychophysics*, *54*, 383-394. doi: 10.3758/BF03205274
- Honing, H. (2003). The final ritard: On music, motion, and kinematic models. *Computer Music Journal*, *27*, 66-72. doi: 10.1162/014892603322482538
- Howard, J.H., & Ballas, J.A. (1980). Syntactic and semantic factors in the classification of nonspeech transient patterns. *Perception & Psychophysics*, *28*, 431-439. doi: 10.3758/bf03204887
- Jones, M.R. (1987). Dynamic pattern structure in music - recent theory and research. *Perception & Psychophysics*, *41*, 621-634. doi: 10.3758/BF03210494
- Jones, M.R. (2008). A dynamic approach to attending: New issues. *The Journal of the Acoustical Society of America*, *124*, 2431-2431. doi: 10.1121/1.4782501
- Jones, M.R. (2016). Musical time. In S. Hallam, I. Cross & M. Thaut (Eds.), *2nd ed.* (2nd ed. ed., pp. 125-141, Chapter xv, 950 Pages): Oxford University Press, New York, NY.
- Jones, M.R., & Pfordresher, P.Q. (1997). Tracking musical patterns using joint accent structure. *Canadian Journal of Experimental Psychology*, *51*, 271-291. doi: 10.1037/1196-1961.51.4.271
- Jones, M.R., & Ralston, J.T. (1991). Some influences of accent structure on melody recognition. *Memory & Cognition*, *19*, 8-20. doi: 10.3758/BF03198492
- Kinder, A., Shanks, D.R., Cock, J., & Tunney, R.J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology: General*, *132*, 551-565. doi: 10.1037/0096-3445.132.4.551
- Large, E.W., & Jones, M.R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119-159. doi: 10.1037/0033-295X.106.1.119
- Loui, P., Wessel, D., & Hudson Kam, C. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, *27*, 377-388. doi: 10.1525/mp.2010.27.5.377

- Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 350-364. doi: 10.1037/0278-7393.22.2.350
- Meier, B., & Cock, J. (2010). Are correlated streams of information necessary for implicit sequence learning? *Acta Psychologica*, *133*, 17-27. doi: 10.1016/j.actpsy.2009.08.001
- Monahan, C.B., Kendall, R.A., & Carterette, E.C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics*, *41*, 576-600. doi: 10.3758/BF03210491
- Nissen, M.J., & Bullemer, P. (1987). Attentional requirements of learning - evidence from performance-measures. *Cognitive Psychology*, *19*, 1-32. doi: 10.1016/0010-0285(87)90002-8
- Palmer, C., & Krumhansl, C.L. (1987a). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 116-126. doi: 10.1037/0096-1523.13.1.116
- Palmer, C., & Krumhansl, C.L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics*, *41*, 505-518. doi: 10.3758/BF03210485
- Perruchet, P. (2008). Implicit learning. In J. Byrne (Ed.), *Cognitive psychology of memory* (Vol. 2, pp. 597-621).
- Prince, J.B. (2011). The integration of stimulus dimensions in the perception of music. *Quarterly Journal of Experimental Psychology*, *64*, 2125-2152. doi: 10.1080/17470218.2011.573080
- Prince, J.B. (2014a). Contributions of pitch contour, tonality, rhythm, and meter to melodic similarity. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 2319-2337. doi: 10.1037/a0038010

- Prince, J.B. (2014b). Pitch structure, but not selective attention, affects accent weightings in metrical grouping. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 2073-2090. doi: 10.1037/a0037730
- Prince, J.B., & Pfordresher, P.Q. (2012). The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psychologica*, 141, 184-198. doi: 10.1016/j.actpsy.2012.07.013
- Prince, J.B., Thompson, W.F., & Schmuckler, M.A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? *Journal of Experimental Psychology: Human Perception and Performance* 35, 1598-1617. doi: 10.1037/a0016456
- Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-&. doi: 10.1016/s0022-5371(67)80149-x
- Remillard, G. (2003). Pure perceptual-based sequence learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29, 581-597. doi: 10.1037/0278-7393.29.4.581
- Salidis, J. (2001). Nonconscious temporal cognition: Learning rhythms implicitly. *Memory & Cognition*, 29, 1111-1119. doi: 10.3758/BF03206380
- Schaffrath, H. (1995). The Essen folksong collection in kern format. from <http://essen.themefinder.org/>
- Schellenberg, E.G., Stalinski, S.M., & Marks, B.M. (2014). Memory for surface features of unfamiliar melodies: Independent effects of changes in pitch and tempo. *Psychological Research*, 78, 84-95. doi: 10.1007/s00426-013-0483-y
- Schmuckler, M.A. (2004). Pitch and pitch structures. In J. Neuhoff (Ed.), *Ecological psychoacoustics* (pp. 271-315). San Diego, CA: Elsevier Science.
- Schulkind, M.D. (1999). Long-term memory for temporal structure: Evidence from the identification of well-known and novel songs. *Memory & Cognition*, 27, 896-906. doi: 10.3758/BF03198542

- Schultz, B.G., Stevens, C.J., Keller, P.E., & Tillmann, B. (2013). The implicit learning of metrical and nonmetrical temporal patterns. *The Quarterly Journal of Experimental Psychology*, *66*, 360-380. doi: 10.1080/17470218.2012.712146
- Schulze, K., Dowling, W.J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception*, *29*, 255-267. doi: 10.1525/mp.2012.29.3.255
- Selchenkova, T., François, C., Schön, D., Corneyllie, A., Perrin, F., & Tillmann, B. (2014). Metrical presentation boosts implicit learning of artificial grammar. *PLoS ONE*, *9*, e112233. doi: 10.1371/journal.pone.0112233
- Selchenkova, T., Jones, M.R., & Tillmann, B. (2014). The influence of temporal regularities on the implicit learning of pitch structures. *Quarterly Journal of Experimental Psychology*, *67*, 2360-2380. doi: 10.1080/17470218.2014.929155
- Shin, J.C., & Ivry, R.B. (2002). Concurrent learning of temporal and spatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 445-457. doi: 10.1037//0278-7393.28.3.445
- Silva, S., Petersson, K.M., & Castro, S.L. (2017). The effects of ordinal load on incidental temporal learning. *Quarterly Journal of Experimental Psychology*, *70*, 664-674. doi: 10.1080/17470218.2016.1146909
- Thompson, W.F., Hall, M.D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 128-140. doi: 10.1037/0096-1523.27.1.128
- Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research*, *70*, 345-358. doi: 10.1007/s00426-005-0222-0
- Tillmann, B., & Poulin-Charronnat, B. (2010). Auditory expectations for newly acquired structures. *Quarterly Journal of Experimental Psychology*, *63*, 1646-1664. doi: 10.1080/17470210903511228

Ullén, F., & Bengtsson, S.L. (2003). Independent processing of the temporal and ordinal structure of movement sequences. *Journal of Neurophysiology*, *90*, 3725-3735. doi: 10.1152/jn.00458.2003

Watkins, A.J. (1985). Scale, key, and contour in the discrimination of tuned and mistuned approximations to melody. *Perception & Psychophysics*, *37*, 275-285. doi: 10.3758/BF03211349