Generality and specificity in the effects of musical expertise on perception and cognition

Daniel Carey [1,2]

Stuart Rosen [3]

Saloni Krishnan [1,2]

Marcus T. Pearce [4]

Alex Shepherd [2]

Jennifer Aydelott [1,2]

Frederic Dick [1,2]

1. Centre for Brain and Cognitive Development, Birkbeck College, University of London, UK
2. Department of Psychological Sciences, Birkbeck College, University of London, UK
3. Department of Speech, Hearing and Phonetic Sciences, UCL, London, UK
4. School of Electronic Engineering and Computer Science, Queen Mary, University of London, UK

Acknowledgements. We acknowledge the generous support of staff and students at the string and keyboard departments of the Royal College of Music and Guildhall School of Music and Drama, London, UK. We thank Mike Coleman and Dr. Lorna Halliday, Department of Speech, Hearing and Phonetic Sciences, UCL, for technical assistance and helpful discussion concerning results, respectively. We thank Dr. Matthew Longo, Dr. Jeremy Skipper, and Dr. Adam Tierney, for helpful comments on earlier drafts of the manuscript. The research was funded by EC grant MC-ITN-264301 (TRACKDEV).

Address for correspondence:

Daniel Carey
Centre for Brain and Cognitive Development
Birkbeck College
London, WC1E 7HX
UK
Email: d.carey@bbk.ac.uk
Tel. +44 207 079 0703

Abstract

Performing musicians invest thousands of hours becoming experts in a range of perceptual, attentional, and cognitive skills. The duration and intensity of musicians' training – far greater than that of most educational or rehabilitation programs – provides a useful model to test the extent to which skills acquired in one particular context (music) generalize to different domains. Here, we asked whether the instrument-specific and more instrument-general skills acquired during professional violinists' and pianists' training would generalize to superior performance on a wide range of analogous (largely non-musical) skills, when compared to closely matched non-musicians. Violinists and pianists outperformed non-musicians on fine-grained auditory psychophysical measures, but surprisingly did not differ from each other, despite the different demands of their instruments. Musician groups did differ on a tuning system perception task: violinists showed clearest biases towards the tuning system specific to their instrument, suggesting that long-term experience leads to selective perceptual benefits given a training-relevant context. However, we found only weak evidence of group differences in non-musical skills, with musicians differing marginally in one measure of sustained auditory attention, but not significantly on auditory scene analysis or multi-modal sequencing measures. Further, regression analyses showed that this sustained auditory attention metric predicted more variance in one auditory psychophysical measure than did musical expertise. Our findings suggest that specific musical expertise may yield distinct perceptual outcomes within contexts close to the area of training. Generalization of expertise to relevant cognitive domains may be less clear, particularly where the task context is non-musical.

Keywords: expertise; musicians; perception; cognition; generalization

Highlights

1. We studied generalization of expertise to auditory perceptual and broader cognitive skills.
2. Musician groups perceived auditory signal differences more finely than non-musicians.
3. Musician groups differed in fine perception given a training-relevant context.
4. Musical expertise did not strongly generalize to several cognitive measures (e.g., auditory scene analysis).
5. 5. Sustained attention predicted variance in fine perception of AM depth above expertise.

Generality and Specificity in the Effects of Musical Expertise on Perception and Cognition

Perceptual and cognitive skills can be shaped and enhanced through our experience with the world (e.g., Goldstone, 1999; Palmeri & Gauthier, 2004). Pursuit of expertise in a given domain is a particularly striking example: groups as diverse as chess masters, physicians, athletes and musicians spend thousands of hours training and practicing, honing perceptual, cognitive and motor skills critical to success in their field (see Ericsson, 2006; Palmeri et al., 2004; Chi, 2006, for review). Are expert-level perceptual and cognitive skills specific to the trained context? Could these skills also transfer to general or abstracted contexts, and might they also interact or influence each other?

Expert musicians are an ideal population for addressing these questions. Professional instrumentalists typically begin training very early in life and follow rigid practice regimens, often totaling 10,000+ hours of lifetime practice by early adulthood (e.g., Ericsson et al., 1993). Critically, instrumentalists are faced with clear perceptual and cognitive demands. They must finely perceive and control their instrument's acoustic signal, sustain attention to their output, reproduce complex and variable sound sequences, and carefully analyze the output of other musicians. Importantly, the perceptual and performance demands faced by particular instrumentalists differ widely – for example, violinists must attend to and adjust intonation during performance, whereas pianists have no such control over intonation. If training demands drive perceptual and broader cognitive outcomes, then differences in these outcomes between particular instrumentalist groups can provide a useful means of accounting for specificity versus transfer of skills (see Strait & Kraus, 2014). Moreover, the different demands faced by instrumentalist groups provide a testing ground to explore how finely honed auditory perception and top-down skills such as auditory attention might interact. Distinct instrumentalist groups with similar training extents also offer a way to control for differences in self-selection, motivation, or personality that can vary between musicians and non-musicians (see Herholz & Zatorre, 2012; Schellenberg, 2004; Corrigall et al., 2013).

Indeed, perceptual and cognitive outcomes associated with musical expertise have been studied extensively (see Kraus & Chandrasekaran, 2010, for review); yet many studies have examined perceptual and cognitive skills separately, with relatively small and/or heterogeneously trained samples. This is partly due to the difficulties of researching expert musician cohorts (e.g., recruitment, study time constraints, etc.) Few studies have investigated interactions between cognitive and perceptual outcomes relevant to musical training, or assessed predictive relationships between fine perceptual and higher cognitive skills such as attention (but see Strait et al., 2010; Parbery-Clark et al., 2009b). To our knowledge, no single study has examined the effects of expertise with one instrument versus another on musically-relevant perceptual and cognitive performance; this would facilitate tests of experience-specific perceptual advantages alongside tests of cognitive outcomes that may relate to musical expertise, together with tests of perception-cognition interactions. As we show in a selective review of the extensive literature concerning perceptual and cognitive benefits related to musical expertise, relatively little research has measured both fine perceptual and broader cognitive outcomes in the same expert individuals. Thus, no study yet has explored whether musicians that train on different instruments might show differences in perceptual and cognitive skills that reflect some of the specific constraints of the instrument they play, or indeed whether those perceptual and cognitive skills might interact. The present study therefore aimed to address this gap in understanding (see 1.3).

1.1 Musicianship and auditory perception

A considerable body of research suggests that musicians tend to out-perform non-musicians in perceiving fine differences in a number of basic auditory properties, including frequency and/or pitch (Spiegel & Watson, 1984; Micheyl et al., 2006; Kishon-Rabin et al., 2001; Amir, Amir & Kishon-Rabin, 2003; Nikjeh, Lister & Frisch, 2009; Koelsch, Schröger & Tervaniemi, 1999;

Parbery-Clark et al., 2009b), tone interval size (Zarate, Ritson & Poeppel, 2012, 2013; Siegel & Siegel, 1977), temporal interval size (Rammsayer & Altenmüller, 2006; Cicchini et al., 2012; Ehrle & Samson, 2005), and timbre (Pitt, 1994). Below, we review evidence for lower-level and contextually-relevant perceptual advantages in differently trained musician cohorts.

1.1.1 Instrument- and musical-genre-specific effects on auditory perception. Expert musicians' fine-grained perceptual abilities may be driven – at least in part – by the demands of the kind of music they perform or the instrument they play. For instance, classically-trained musicians can discriminate finer differences in frequency compared to rock or jazz musicians (Kishon-Rabin et al., 2001; but see Vuust et al., 2012 and footnote 1). Percussionists reproduce temporal intervals less variably than string musicians and non-musicians (Cicchini et al., 2012); string musicians match frequency differences less variably than percussionists (Hoffman et al., 1997); and trained vocalists tend to sing pitches less variably than instrumentalists (Nikjeh et al., 2009). Relatedly, electro and magnetoencephalography (EEG & MEG) data indicate enhanced cortical responses in musicians to the timbre of the specific instrument they perform (versus an instrument they do not), both in adults (Pantev et al., 2001; Shahin et al., 2003) and children (Shahin et al. 2004; 2008; Trainor et al., 2003). Moreover, string and woodwind players – who constantly monitor and adjust the pitch they are producing – can discriminate frequency differences more finely than musicians who play fixed pitch instruments like piano (Micheyl et al., 2006; Spiegel & Watson, 1984).

Bowed string instruments like violin also differ from fixed-pitch instruments like piano in that string players make extensive use of vibrato – a periodic but non-sinusoidal oscillation in the frequency and amplitude of a given note (Papich & Rainbow, 1974; see Mellody & Wakefield, 2000, for discussion of violin vibrato signal properties). Violinists manipulate vibrato (i.e., rate and depth of amplitude modulation [AM] and frequency modulation [FM]) for expressive and stylistic reasons. There is some evidence that musicians might be sensitive to signal changes associated with vibrato (e.g., AM depth; Fritz et al., 2010; see footnote 2). Yet no single study has examined whether violinists' experience in controlling these signal modulations means they can perceive such cues more finely than other musicians – such as pianists. Unlike violinists, expert pianists cannot control depth or rate of amplitude or frequency modulation. Instead, one of the primary expressive tools used by pianists is changing the velocity and acceleration of piano key strikes, which alters the attack envelope (i.e., onset rise time) of the resulting sound (see Goebl, 2001, 2005, for discussion; see also Wessel, 1979). Yet string instrument sounds also vary in attack envelope – for instance, between plucked and bowed sounds (see Gordon, 1987; Rosen & Howell, 1981). Given that pianists manipulate onset rise time to very fine extents and control only this cue (together with offset and damping), we might predict that pianists would show enhanced sensitivity to onset rise time compared to violinists (who manipulate many other cues, as outlined above). Conversely, we might expect violinists to show improved acuity for AM and FM depth compared to pianists, since violinists manipulate these cues extensively, whereas pianists cannot.

The different demands of fixed and non-fixed-pitch instruments allow us to test whether musicians' refined perceptual abilities are specific to the acoustic properties of their instrument. In the current experiments, we tested whether the differences in violinists' and pianists' control and use of AM depth, FM depth and onset rise time rates translate to differences in their ability to perceive subtle changes in these basic auditory parameters (when removed from a musical context). We also used a visual psychophysical (color hue perception) task to control for any possible musician perceptual advantage unspecific to the auditory modality (musicians and non-musicians should not differ on a visual task unrelated to musical expertise).

1.1.2 Contextual effects on musicians' perception of 'low-level' acoustic parameters. Musical notes often occur in harmonic contexts, where several notes are played at once (as in a C Major chord). The fundamental frequencies of these notes are adjusted according to a variety of tuning systems that govern the exact spacing of the frequencies relative to each other. Fixed pitch instruments like piano typically use the 'equal-tempered' tuning system, where each semitone on the keyboard is equally spaced according to a fixed complex integer ratio – one that pianists cannot alter without recourse to a professional piano tuner. In contrast, non-fixed pitch instruments like violin commonly use ‘just tempered’ tuning, where notes within a musical scale are tuned according to the resonance structure of naturally vibrating systems (see 2.3.6.1), and where semitones have different spacing based on their position within the harmonic scale. Thus, unlike the case with the piano where ‘a C# is just a C#’, a violinist playing a C# may tune it differently depending on whether it occurs in an A major, D major, or E major harmonic context.

A handful of studies have investigated how finely string player and pianists are able to perceive differences in tuning, and how closely they hew to the tuning system most relevant to their instrument. Loosen (1994) found that trained violinists and pianists adjusted the pitch of major scale notes to most closely match the frequency spacing of the tuning system specific to their instrument (Pythagorean versus equal tempered tuning, respectively; see footnote 3). Non-musicians showed no specific biases towards any tuning system, presumably due to their lack of training (see also Loosen, 1995). Using a small sample, Roberts and Mathews (1984) reported similar musician group effects for perception of chords adhering to just intonation and equal temperament; surprisingly, they also report that pianists and string players sometimes showed large deviances toward the tuning system not specific to their instrument.

Despite these results, few studies have rigorously assessed how musicians trained with fixed versus non-fixed pitch instruments perceive very subtle (e.g., less than 10% of a semitone) deviations from their relevant tuning systems in a harmonic context (that is, when notes occur simultaneously, as in a chord). This question has implications for the extent to which distinct musical expertise hones fine-grained perception in a training-specific context. Thus, in the present study we included a chord tuning perception paradigm. This provides a strong test of the compliance between perceptual sensitivity – both lower-level and contextual – and specific instrumental training experience. We also related chord tuning perception to fine-grained perceptual thresholds, allowing us to examine whether specific expertise would reflect differential reliance on acoustic cues in judging tuning.

1.2 Musicianship, attention, and cognition

Mastering a musical instrument and playing it with others requires more than sensitivity to – and control of – fine frequency, temporal and harmonic features. Musicians must learn to sustain attention to sound streams for very long periods of time, responding quickly and consistently for some sounds but not others. Similarly, musicians must rapidly and accurately recall and reproduce regular sequences of sounds, both during practice and performance. The complexity of ensemble performance may also spur changes in associated cognitive skills such as auditory scene analysis. For instance, during a symphony, a violinist might have to wait without playing for several minutes (all the time counting beats), starting to play immediately after hearing a motif played by the bassoon. The violinist must therefore perform a very sophisticated kind of auditory scene analysis: she must listen attentively for a single note or sequence of notes played by the bassoon, and will have to distinguish the bassoon from dozens of other instruments playing at the same time.

Given the complex demands associated with musicianship, we tested whether instrumentalists' expertise in sustained attention (e.g., during practice and performance) generalized from the musical realm to broader indices of sustained attention to sound. We further asked whether musicians that typically spend more time in ensemble performance – violinists relative to pianists (see ST1) – might generalize this experience to non-musical indices of complex auditory scene analysis. We also asked whether musicians' experience with reproducing sounds and sequential motifs might generalize to novel yet regular sequences. In the following sections, we review evidence of musician and instrument specific advantages across these cognitive domains.

1.2.1 Auditory attention, and influence on perception. Sustained attentional abilities in musicians are relatively understudied. Evidence suggests that musicians outperform non-musicians on auditory but not visual sustained attention measures (Strait et al., 2010); however, one recent study also showed a musician advantage on visual sustained attention metrics (Rodrigues et al., 2013). These results conform to research with other highly skilled populations such as chess players, birders, and memory experts, showing that experts differ from non-experts in both their attention to key stimulus features, and their ability to sustain such attention over extended periods (see Palmeri et al., 2004; Green & Bavelier, 2012). Such potential differences in attentional abilities are not only interesting in their own right, but are particularly important in understanding what might drive musicians' advantages in lower-level auditory perception (see Strait & Kraus, 2011b, and Zhang et al., 2012, for discussion). For instance, attention is known to modulate auditory detection (e.g., via attentional cuing to specific frequency bands; Mondor & Bregman, 1994; Justus & List, 2005; Larkin & Greenberg, 1970; Greenberg & Larkin, 1968), and attention can interact with the saliency of acoustic cues in auditory search tasks (Kayser et al., 2005). Nevertheless, recent data show that musicians can process the pitch direction of local and global auditory patterns more accurately than non-musicians, regardless of the direction of attentional focus (Ouimet et al., 2012).

While the role of attention with respect to musicians' perception remains debated (e.g., Baumann et al., 2008; Koelsch et al., 1999), research has shown that musicians differ from non-musicians in the way that attention modulates electrophysiological indices of auditory perception (e.g., Tervaniemi et al., 2005, 2009; Seppänen et al., 2012; see also Marie et al., 2011). Compared to non-musicians, musicians show increased N2b component amplitudes for attended intensity, frequency and duration deviances in speech and musical sounds (Tervaniemi et al., 2009), and significant reductions in P3b amplitudes when attending to subtle pitch deviances (Seppänen et al., 2012). Further, auditory sustained attention performance correlates with perceptual metrics like backward masking and speech-in-noise (Strait et al., 2010; see also Strait et al., 2012b). Thus, attentional differences between musicians and non-musicians may account for group differences in the detection of potentially less salient acoustic cues (Strait et al., 2010; Strait & Kraus, 2011b; Fujioka et al., 2006). Therefore, in the present study we used a novel measure of auditory sustained attention. This allowed us to investigate how musicians and non-musicians might differ in attentional abilities, and crucially, whether individual differences in auditory attention (in musicians and non-musicians) predict differences in the perception of changes in basic acoustic features. Given that both pianists and violinists typically spend considerable time sustaining attention toward instrument output (e.g., during practice), we did not hypothesize any specific musician group difference in this ability.

1.2.2 Auditory scene analysis. In order to perform successfully with a musical ensemble, musicians must analyze and then use multiple streams of information from an exceedingly complex auditory scene (see Nager et al., 2003). As noted above, musicians' experience in segregating such complex auditory streams (e.g., picking out a melody line amidst changing harmony; Bregman, 1990) may benefit their auditory scene analysis abilities in non-musical contexts.

There is some evidence to support this hypothesis. Zendel & Alain (2009, 2013) have shown that musicians segregate harmonic complexes better than non-musicians and more often report hearing a harmonic as a separate auditory object when mistuned by as little as 2%. Orchestral conductors – whose primary role is to analyze, interpret, and manipulate a colossal auditory scene – show enhanced selectivity in attending to spatially segregated auditory signals (noise bursts), when compared to both pianists and non-musicians (Nager et al., 2003). Musicians' long experience in musically-based scene analysis may also be a causal factor in their enhanced ability to comprehend speech when the speech signal is masked by noise (classic 'energetic' masking) or multi-talker babble (energetic plus so-called 'informational masking' – see footnote 4; Parbery-Clark et al., 2009a, 2009b, 2011; Strait & Kraus, 2011b; Strait et al., 2012b; but see also Patel, 2011). However, recent data suggest that musicians and non-musicians do not differ in susceptibility to informational and energetic masking during speech-in-noise perception (Ruggles et al., 2014).

Nevertheless, Oxenham et al. (2003) have shown that musicians are less susceptible to informational masking compared to non-musicians, as demonstrated using tone detection performance with masking sounds occurring at fixed frequencies (no informational masking) or variable frequencies (more informational masking). However, it is unclear if musicians can generalize such resilience to energetic or informational masking when analyzing 'everyday' auditory scenes. Therefore, in the present study we tested our musicians and non-musicians using an established naturalistic auditory scene analysis paradigm (Leech et al., 2009; Krishnan et al., 2013; Gygi & Shaffiro, 2011). Our design also allowed us to explore whether an instrumental group who play more regularly in large ensemble (violinists) might be more resistant to informational/energetic masking than a group who often perform solo or in smaller ensembles (pianists). Thus, we predicted an advantage for violinists compared to pianists on our naturalistic listening task.

1.2.3 Sequence perception and reproduction. As mentioned above, one of the fundamental challenges of musical performance is perceiving and reproducing auditory sequences that repeat over time (Koelsch et al., 2002; van Zuijen et al., 2004; see also Rohrmeier et al., 2011; Loui, Wessel & Kam, 2010; Dick et al., 2011; Patel, 2003, for discussion). These sequences can vary greatly in length, speed, and the basic unit of analysis (e.g., a single motif versus a phrase built from motifs). They can also vary in how predictably they repeat: sequences might consist of an exact repetition of a simple short motif, or variations of a sequence interspersed with non-sequential material (see Pearce et al., 2010). This experience with processing hierarchical sequences may underlie musicians' enhanced detection of deviances from regular auditory sequences. Compared to non-musicians, musicians show larger mismatch negativity (MMN) amplitudes to extra tones added to the end of regular pitch sequences (when the pitch of each sequence ascends or remains fixed; van Zuijen et al., 2004, 2005). Further, musicians show larger increases in MMN responses over time than non-musicians in response to low-probability tone sequences that violate more highly probable sequence structures (Herholz et al., 2011; see footnote 5).

There is also some evidence that musicians are better at actively reproducing sequences, and at abstracting the statistical structure of probabilistic sequences. Using an active sequence reproduction task modeled after the audiovisual 'SIMON' game, Taylor-Tierney et al. showed that musicians reproduced audio-only sequences better than non-musicians; however, groups did not differ on audiovisual sequences (Taylor-Tierney et al., 2008; see also Karpicke & Pisoni, 2004; cf. Conde et al., 2012). Further, Shook et al. (2013) found that expert musicians were better than less skilled musicians at passively learning the statistical structure of sequences of tone pips varying in duration. Similarly, relative to non-musicians, musicians have larger P2 amplitudes to novel sung melodies they have not previously heard versus familiar sung melodies heard during an exposure phase (Francois and Schön, 2011). In contrast, Rohrmeier et al. (2011) found no difference between musicians and non-musicians on a sequence familiarity judgement task, after passive exposure to tone sequences built from a finite state grammar (see also Loui et al., 2010).

These results provide some evidence of an expert advantage for encoding and reproduction of auditory sequences. Yet an open question concerns whether musicians might be better at detecting sequence regularities and whether this influences their reproduction. We thus developed a novel audiovisual sequencing paradigm (after Taylor-Tierney et al., 2008), testing whether different musician groups would show improved ability to reproduce novel sequences, compared to non-musicians. We also tested whether a short period of listening to some of the auditory regularities before the sequencing task might influence or bias participants' sequencing performance.

1.3 The Present Study

Here, we test the compliance between the demands of expert training on a musical instrument, and associated cognitive and perceptual outcomes. If instrument expertise yields improvements in perceptual and cognitive performance, such outcomes may be tied to the specific demands posed by a particular instrument. In testing this account, we recruited matched cohorts of violinists, pianists and non-musicians. We used an extensive battery of auditory psychophysical measures to probe differences in fine-grained auditory perceptual thresholds associated with long-term training on specific instruments. We also tested whether cognitive abilities potentially related to expertise (sustained attention, auditory scene analysis, sequencing) would extend to non-musical metrics, and whether performance on these tasks would relate to lower-level perceptual skills. Previous research has found largely piecemeal evidence for differences between musicians and non-musicians on several of these perceptual and cognitive tasks. Our goal was to establish whether specific instrumental expertise may yield perceptual refinements in one instrumental group but not another, along with broader improvements to cognitive skills that might reflect generalization of expertise. Moreover, we aimed to explore predictive relationships between perceptual and cognitive performance, and to relate any such relationships to the effects of long-term training on a specific instrument, or to musical expertise in general.

Method

2.1 Participants

Participants (*N* = 72) were 24 violinists, 24 pianists and 24 non-musicians (descriptive statistics displayed in Table 1), matched for gender. All were right-handed as determined by the Edinburgh Handedness Inventory (mean [SD]: violinists – 82.2 [19.3]; pianists – 84.4 [13.6]; non-musicians – 85.4 [12.5]; Kruskal–Wallis: χ^2 (2, $n = 72$) = 0.01, $p > 0.9$). None reported any history of auditory or uncorrected visual impairment, or of neurological disease or insult.

Table 1: Descriptive statistics for non-musician, violinist and pianist samples (each *n* = 24)

Group	Mean age (SD)	Age range (years)	Mean years training (SD)	Years training range	Mean lessons onset age (SD)	Total accumulated lifetime practice hours (SD)
Non-musicians	22.9 (2.8)	19–29	2.1 (1.5) *	0.25–5 *	9.5 (2.8) *	N/A
Violinists	23.1 (3.1)	19–30	16.9 (3.8) **	11–27.5	5.3 (1.9) **	10,927.6 (4520.4) **
Pianists	21.3 (2.5)	18–26	15.3 (3.8) **	8–21	5.7 (2.2) **	9,900.6 (5050.7) **

* non-musicians with training (*n* = 17)

** violinists and pianists not significantly different

2.1.1 Musicians.

Violinists (6 males, 18 females) and pianists (7 males, 17 females) were recruited from conservatories in London and through an employment website for freelance musicians. All but one violinist and one pianist were completing, or had completed, a performance degree. The violinist and pianist who had not completed a performance degree had practice histories similar to their respective samples. Violinists and pianists did not differ significantly in years of training, t (46) = 1.5, $p = 0.14$, age of onset of lessons, $z = 0.6$, $p > 0.5$, or total accumulated lifetime practice, t (44) = 0.7, $p = 0.47$ (see Table 1). Violinists and pianists had experience of playing other instruments (notably piano for violinists; see Tables 2 & 3); however, all reported these instruments as secondary, and reported not practicing those instruments at the time of the study (see footnote 6).

None of the pianists had violin training. All musicians had trained extensively with classical repertoire.

Table 2: Violinists' (*n* = 24) descriptive data and musical training histories

Participant	Gender	Age	Violin training (years)	Current daily practice (hours per day)	Other instruments	Other instruments - years played	Education (highest level)
v1	F	23	19	3	Viola	3	Performance MA
v2	F	22	14	4	Piano	6	PD - 2nd year
v3	M	19	12	4	Piano	7	PD - 1st year
v4	F	20	17	4.5	Piano; Viola; Trumpet	7; 7; 7	PD - 2nd year
v5	F	23	19	4.5	Piano	2	Performance MA
v6	F	20	12	5	Piano; Viola	12; 4	PD - 2nd year
v7	F	19	14	1	Piano	missing data	PD - 2nd year
v8	F	25	20	5	None		Performance MA
v9	M	21	17	1	Piano	9	BA & Private lessons
v10	M	24	21	4	Piano; Alto Saxophone	12; 13	Performance MA
v11	F	28	21	4	Piano; Clarinet; Viola	2; 6; 3	Performance MA
v12	M	26	20	6	Piano	5	Performance MA
v13	F	25	11	5	Viola	1	Performance MA
v14	F	21	18	4.5	Viola	6	PD - 2nd year
v15	M	28	20	2.5	None		Performance MA
v16	F	30	27.5	5	Piano	20	Performance MA
v17	F	25	18	3	Piano; Viola	missing data	Performance MA
v18	F	22	14	3	Piano; Viola	1.5; 1	PD - 2nd year
v19	F	23	16	2.5	Bass Guitar	6	PD - 2nd Year
v20	M	22	17	6	Piano; Viola; Voice	5; 2; 7	Performance MA
v21	F	19	12	3	Piano	8	PD - 2nd year
v22	F	20	13	3	Piano	2	PD - 2nd year
v23	F	26	17	1.5	Piano	5	PD - complete
v24	F	23	16	4	Piano; Trumpet	2; 2	PD - 3rd year

PD – Performance degree; note: all Performance MAs were ongoing at the time of the study.

Table 3: Pianists' (*n* = 24) descriptive data and musical training histories

Participant	Gender	Age	Piano training (years)	Current daily practice (hours per day)	Other instruments	Other instruments - years played	Education (highest level)
p1	F	23	19	5	None		Performance MA
p2	F	19	12	4.5	Guitar	0.25	PD - 1st year
p3	M	19	12	3.5	Clarinet; Voice	4; 3	PD - 2nd year
p4	F	19	16	2.5	Cello	7	PD - 1st year
p5	F	25	19	4	None		Performance MA
p6	F	24	20	6.5	Clarinet	8	Performance MA
p7	F	20	16	4	Voice; Gamelan	3; 1	PD - 1st year
p8	M	21	15	4	None		PD - 1st year
p9	M	18	9	2	Organ; Double Bass	4; 4	PD - 1st year
p10	F	22	18	5	Voice	10	Performance MA
p11	F	22	17	4	Harpsichord; Zither	2; 8	Performance MA
p12	M	26	21	5	None		Performance MA
p13	F	20	15	5	None		PD - 2nd year
p14	F	19	12	5	None		PD - 1st year
p15	M	20	8	4	Drums	0.25	PD - 2nd year
p16	F	19	15	4	None		PD - 1st year
p17	F	23	18	6	Cello	5	Performance MA
p18	F	19	10	5.5	Cello	1	PD - 2nd year
p19	M	18	10	6	Harpsichord	2	PD - 1st year
p20	F	22	14	3.5	Voice	3	PD - 3rd year
p21	F	22	18	5.5	None		Performance MA
p22	M	20	14.5	2.5	French Horn	1	PD - 1st year
p23	F	25	20	1	None		PD - complete
p24	F	25	18	1	Drums; guitar	6; 2	MSc & Private lessons

PD – Performance degree; note: all Performance MAs were ongoing at the time of the study.

2.1.2 Non-musicians. Non-musicians were recruited from a local participant pool and from courses across the University of London. All had completed or were enrolled in a university degree (see footnote 7). Non-musicians described any previous experience with musical instruments and any years of practice and/or lessons (see Table 4). Seven non-musicians (4 female, 3 male) had never played any musical instrument or taken music lessons. Seventeen participants (13 female, 4 male) had taken elementary music lessons during childhood or adolescence, but had not attended

Table 4: Non-musicians' (*n* = 24) descriptive data and musical training histories

Participant	Gender	Age	Musical training (years)	Instrument	Years since practised	Education (highest level)
nm1	F	24	4	Piano	14	MSc - complete
nm2	F	24	0			MSc - ongoing
nm3	F	22	0.5	Viola	12	MSc - ongoing
nm4	F	20	0.25	Saxophone	9	Degree - 2nd year
nm5	F	29	0			PhD - 1st year
nm6	F	28	3	Piano	18	MA - ongoing
nm7	F	21	1	Recorder	9	Degree - 2nd year
nm8	F	20	0			Degree - 2nd year
nm9	F	21	0.5	Guitar	6	Degree - 2nd year
nm10	M	27	5	Piano	16	PhD - 2nd year
nm11	F	19	1	Piano	10	Degree - 2nd year
nm12	M	26	0			Degree - complete
nm13	M	19	0			Degree - 1st year
nm14	M	21	3	Violin	9	Degree - 2nd year
nm15	M	22	3	Cornet	9	Degree - complete
nm16	F	22	3.5	Piano; Violin	12	MSc - ongoing
nm17	F	24	3	Saxophone	10	Degree - complete
nm18	F	21	0.5	Piano	8	Degree - 2nd year
nm19	F	23	1	Keyboard	19	MA - complete
nm20	F	26	2	Piano	14	MA - ongoing
nm21	F	23	0			Degree - 3rd year
nm22	F	25	4	Violin	13	Degree - complete
nm23	M	22	1	Voice	10	Degree - 2nd year
nm24	M	21	0			Degree - 2nd year

a formal music college or practiced daily over an extended period. On average, those non-musicians with musical experience had not practiced for 11.8 years (SD = 3.6; range = 6–19 years) prior to the study.

2.2 Materials

The study received ethical approval from the local ethics committee at Birkbeck College. Participants completed most of the experimental battery (auditory psychophysical thresholding, audio-visual sequencing task [SIMON], tuning system perception task, Environmental Auditory Scene Analysis [EnvASA] task, Sustained Auditory Attention to Response Task [SAART]) inside a sound attenuated booth. Two further assessments (visual psychophysical thresholding and pure tone audiometry) were conducted in a separate, quiet testing environment. All sounds were presented at a comfortable level fixed for all participants. Testing equipment, software and hardware are detailed in supplemental methods (SM1).

2.3 Procedure

Participants read an information sheet and provided voluntary informed consent before beginning the experimental battery. Rest breaks were provided between tasks as required. Tasks were always run in the order described below to avoid differential effects of fatigue. Total battery duration was approximately three hours. Test-retest reliability methods and analyses for previously unpublished measures (auditory psychophysics, tuning perception task, SIMON, SAART) are presented in supplemental methods (SM2) and supplemental results (SR2).

2.3.1 Practice history questionnaire. Musicians provided data for their current practice hours, practice history across ages (daily practice hours from 3–4 years, up to 19+ years), and hours weekly spent in ensemble. Lifetime practice history data were determined by multiplying estimates of daily practice hours at each age range (3–4 yrs, 5–6 yrs, etc., up to 19+ yrs) to produce yearly estimates. The years from 19+ to musicians' current age minus 1 year were multiplied by the year estimate for 19+ (e.g., for a 25-year-old musician, year estimate for 19+ was multiplied by 5), and added to current daily practice. These estimates were summed for each participant to produce total accumulated lifetime practice (based on Ericsson et al., 1993). One violinist failed to return a practice history questionnaire. A further violinist's estimated accumulated practice exceeded 40,000 hours; the participant was identified as an outlier and excluded from practice data analysis.

Musicians' practice hours were used as predictors for each experimental measure to determine the influence of both practice at specific early ages and total accumulated lifetime practice on musicians' psychophysical and cognitive task performance. We defined two binary variables as separate regressors: musicians who did/did not report practicing at 3–4 years of age, and musicians who did/did not report practicing 1 hour or more per day at 7–8 years (see footnote 8). These regressors were defined to account for the influence of practice at early stages in development on later perceptual and/or cognitive outcomes. We used total accumulated lifetime practice hours as a further separate continuous regressor. Musicians' total accumulated lifetime practice hours did not significantly predict performance on any task (all $p > 0.1$) either when entering or removing group (violinist/pianist) as an additional predictor; we therefore do not discuss this measure further.

2.3.2 Goldsmith's Musical Sophistication Index – Musical Training Sub-scale. All participants completed the 9-item Musical Training sub-scale from the Goldsmiths Musical Sophistication Index (Gold-MSI) (Müllensiefen, Gingras, Stewart, & Musil, 2011), an extensively normed self-rating questionnaire. Three items assessed musician status and competence as a performer according to a 7-point Likert scale. Six items assessed years engaged in training-related activities. The sub-scale yielded a single score (range: 9–63) indexing extent of musical training (see footnote 9). Supplemental table [ST] 1 displays musical training sub-scale means for each group; group comparisons are displayed in ST 2.

2.3.3 Absolute pitch assessment. In addition to self-report, musicians' absolute pitch (AP) ability was assessed by presenting them with three sinusoidal tones (495 Hz [B5]; 733.348 Hz

[F#5]; 660 Hz [E5]). After presentation of each sinusoid, musicians were asked to name the musical note they had just heard. Seven violinists reported AP, but only three named all three tones correctly. Two violinists named two tones correctly each and two violinists named a single tone correctly each. Nine pianists reported AP and seven named all the tones correctly; the other two pianists named one and two tones correctly respectively. Data were not analyzed statistically due to the small and unbalanced sample sizes.

2.3.4 Auditory psychophysical tasks. Three tasks assessed discrimination of onset envelope rise time, the detection of amplitude modulation (AM) and the detection of frequency modulation (FM). All tasks presented standard and test stimuli, where test sounds varied adaptively along logarithmically spaced continua. Decrementing through the steps in each continuum reduced the difference between the test and standard stimuli.

2.3.4.1 Stimuli. All experiments used a complex sawtooth pulse waveform (f0 = 220Hz; first 50 harmonics), sequentially run through a series of resonators of varying center frequency (cf1 = 500 Hz; cf2 = 1500 Hz; cf3 = 2500 Hz; all bandwidths = 100 Hz). For AM and FM experiments, unmodulated standard sounds were 250 ms in duration (20 ms linear rise and fall times). Rise time standard sounds had a fixed linear onset time of 15 ms. Standard and test rise time sounds had a fixed linear offset time of 350 ms (total duration = 750 ms).

For AM and FM tasks, the depth of modulation was varied over 99 test stimuli. Comparison stimuli in the AM detection task (all with a modulation rate of 8 Hz) ranged from a modulation index difference of -1.9 dB (max) to -26.0 dB (min) (i.e., 20log [*m*], where m is modulation index [range: 0.8–0.05]). Comparison stimuli in the FM detection task (all with modulation rate of 4 Hz) ranged from 16 Hz maximum peak excursion, to a potential minimum of 0.16 Hz (peak cents excursion from f0: 121.5–1.25 cents). AM depth and FM depth parameters were motivated by previous analyses of violin vibrato signals; amplitude depth variations of 15 dB, frequency modulation rates of 5–6 Hz, and frequency excursions of approximately 15 cents were found to be typical (Mellody & Wakefield, 2000). The rise time experiment varied linear onset rise of the amplitude envelope (119 test stimuli). Comparison stimuli in the rise time task ranged from 100 ms (maximum), to 15.24 ms (potential minimum).

2.3.4.2 Auditory psychophysical procedure. All tasks employed an adaptive three-alternative (3AFC) procedure tracking 79.4% response accuracy (Levitt, 1971). A one-down one-up procedure preceded the first reversal, followed by a three-down one-up procedure (Baker & Rosen, 2001; Hazan et al., 2009). Each trial presented two standard sounds and one test sound (inter stimulus interval [ISI] = 500 ms). The position of the test sound varied randomly between the three intervals across trials. Each task used a visual display with three cartoon frogs located at the left, center and right of the screen. Each frog produced a sound in turn (left to right). Participants selected the frog they perceived as being the 'odd one out' on each trial. Step size varied adaptively up to the third reversal across all three tasks. The initial three step sizes and total number of test stimuli were increased for the rise time task relative to the AM and FM tasks. These modifications (following pilot testing with an expert listener) ensured sufficient fine-grained rise time increments and prevented ceiling effects in musicians.

Participants completed the rise time task first, followed by the AM and FM tasks. Order remained fixed over all participants to minimize inter-individual differences due to differential practice or fatigue effects. Participants completed one full tracking run for each task as practice. The first three trials of every run also served as practices (i.e., their outcome did not influence the adaptive procedure or psychometric function). Within a given track, trials were presented until seven reversals were obtained, or 50 trials were completed (whichever occurred first). Threshold from each track was determined as the mean of the final four reversals, except in the following case: if a participant reached 50 trials before achieving a fourth final reversal on a track, the mean of the final three reversals was taken as threshold (Banai, Sabin & Wright, 2011), with the threshold verified by examining the psychometric function.

Participants completed a minimum of two experimental tracks during a given task. Once two tracks were completed, the experimenter inspected both track thresholds and psychometric functions. If participants' thresholds for the first two tracks were within four steps or less of each other and four final reversals were reached on both, the task was deemed complete. If the first two track thresholds exceeded four continuum steps relative to each other and/or only three final reversals were reached on either track, participants completed a third track. Thresholds were measured in this manner to maximize the efficiency of the psychophysical procedure and reduce the number of tracks run.

Psychophysical tracks and psychometric functions were re-inspected blind to subject and group once data from all participants were collected. A discrepancy of 10 continuum steps or more between a track threshold and the 79.4% point on the psychometric function (curve fitted using logistic regression) was deemed erroneous and the track was excluded. If a participant had completed two initial experimental tracks where thresholds were within four steps of each other, final threshold was taken as the mean of those two tracks. Where three experimental tracks were completed successfully, the median of those three tracks was taken as final threshold. If a participant tracked successfully on the initial practice for an experiment, but completed an experimental track erroneously, the practice track was taken as a valid data point; the median of threshold values from the valid experimental tracks and the practice track was then taken as threshold. Participants with two or more erroneous tracks for any task were not included in that analysis. On the basis of these criteria, participants were excluded from psychophysical analyses as follows: rise time – 3 violinists, 3 pianists and 8 non-musicians (final *n*'s: 21 violinists, 21 pianists, 16 non-musicians); AM depth – 1 violinist, 1 pianist, 1 non-musician (final *n*'s: 23 per group); FM depth – 3 non-musicians (final *n*'s: violinists & pianists both 24, 21 non-musicians).

We also analyzed potential changes in thresholds over four repeated runs. However, not all participants completed four runs for each experiment, so group sample sizes for these analyses were unequal (Rise time: 13 violinists, 16 pianists, 12 non-musicians; AM depth: 12 violinists, 16 pianists, 12 non-musicians; FM depth: 15 violinists, 16 pianists, 11 non-musicians). To ensure that MANOVA results were not driven by differences in group n's, MANOVA models were assured by

Box's M test of equality of covariance matrices (Stevens, 1996). Results were also verified by matching groups with larger Ns to the smallest group N for that task. This was achieved by drawing six random samples of participants from the larger group(s) for that task. We then entered each random sample into a separate MANOVA analysis with the group it was matched to, allowing for consistency of results to be checked across random samples (see 3.1).

2.3.5 Sequence reproduction task (SIMON). Participants performed an audio-visual sequence reproduction task, modeled after the SIMON interactive game. The task assessed non-instrumentally specific reproduction of multi-modal sequences, allowing for comparison across musician and non-musician groups. Additionally, we investigated the influence of passive exposure to ordered tone sequences on subsequent sequence reproduction.

2.3.5.1 Stimuli. Participants were presented with an octagonal figure containing four wedge-shaped 'buttons' (red, blue, green and yellow). Each button was paired with a fixed 300 ms sinusoidal tone (red button, 262 Hz [C4]; blue button, 327.5 Hz [E4]; green button, 393 Hz [G4]; yellow button, 524 Hz [C5]). Tones formed the notes of a C major chord. All tones had 50 ms onset and offset ramps, normalized for equal RMS amplitude (presented at a comfortable level fixed for all participants). Each button was illuminated simultaneously with the associated tone.

Test sequences were sampled from two probabilistic 'languages', referred to here as language 1 and language 2. Sequences from each language were composed of triplet units. Each SIMON sequence consisted of seven triplets from one of the languages. Language 1 triplets were: C4-E4-G4; E4-G4-C5; G4-C5-C4; C5-C4-E4. Language 2 was the reverse of language 1 (triplets: G4-E4-C4; C5-G4-E4; C4-C5-G4; E4-C4-C5). A triplet could occur more than once in the same sequence, but never consecutively. ISI between presented sequence items varied according to sequence length during the task (length < 4 items: 500 ms ISI; length < 6: 300 ms ISI; length > 6: 200 ms ISI). The interval between response completion and the next sequence iteration (ITI) was 800 ms after the first trial, and 300 ms thereafter.

2.3.5.2 Procedure. Prior to the SIMON task, participants listened to a concatenated stream of 690 SIMON tones that followed the triplet structure of either language 1 or language 2. Participants were informed they would listen to a stream of sounds, but that they did not need to focus on them. While listening, participants completed the Edinburgh Handedness Inventory and a questionnaire concerning their language background. Participants were unaware of any relationship between the passive familiarization and the SIMON task.

The SIMON task was presented following this listening period. Each SIMON trial began with a single on-screen button lighting up, paired with its matching tone (e.g., red button; C4). Participants responded by pressing the appropriate color-coded button on a Logitech Precision Gamepad; with each button press, the corresponding on-screen button illuminated and its matching tone played. If participants responded correctly, the second trial was presented. The second trial presented the same first item (e.g., red; C4) followed by the next triplet item (e.g., blue; E4). Participants had to reproduce the items in the order they were presented by the computer. Sequences incremented one item in length with each correct reproduction of the items presented. A given sequence was terminated if participants failed to reproduce items in the same order as presented by the computer. After a reproduction error, a screen was displayed showing the number of items the participant had reproduced on that sequence.

At the beginning of the experiment, participants completed two practice sequences of six items. If a participant reproduced fewer than five items on either practice, practices were re-run until a minimum of five items were achieved for both. Ten experimental SIMON sequences were then presented (five sequences each from language 1 and 2, pseudorandomly interleaved). Two fixed pseudorandom sequence orders were counterbalanced across participants. Rates of errors made on the very first sequence item (i.e., where no items were correctly reproduced for a sequence), were assessed blind to group, to ensure participants completed similar numbers of sequences for each language (i.e., both familiar and non-familiar). Criterion for exclusion was set at more than one sequence where no items were reproduced, across the 10 experimental sequences;

one non-musician failed to reproduce any items for two sequences and was excluded. Mean sequence lengths were log transformed prior to analyses to correct for positive skew.

2.3.6 Tuning system perception task. The task assessed perception of tuning of major chords. Just and equal tempered tuning systems were compared to each other, as well as to chords that deviated to some degree from either tuning system. The purpose was to assess ratings of ‘in-tuneness’ based on the relevance of tempering to one instrumental class (standard for fixed pitch instruments like piano), contrasting with relevance of just temperament to other instruments (e.g., non-fretted string instruments like violin).

2.3.6.1 Stimuli. All chords were A major triads, with a root, major third, perfect fifth, and octave. Chord stimuli were generated using complex sawtooth pulse waves (as in the auditory psychophysical tasks, but with the number of harmonics reduced to the first 10 and a duration of 1 s).

The just intonation tuning system is based on the natural harmonic resonances of vibrating systems, and relates note frequencies according to simple, small-integer ratios (e.g., 5:4; Duffin, 2007). In contrast, the system of equal temperament relates adjacent semi-tones according to a fixed constant ($^{12}\sqrt{2}$), creating irrational numeric ratios between note frequencies (e.g., 5.13:4; Loosen, 1995; Hopkin, 1996). This results in greater beating between partials, compared to just intonation (Teki et al., 2012; Duffin, 2007).

The just intonation chord was formed as root = 220 Hz (A3), major 3rd = 275 Hz (C#4), 5th = 330 Hz (E4) and octave = 440Hz (A4). This justly tuned chord was compared with chords where tempering of the major third varied: +15 cents (approximating equal temperament), -15 cents, +7.5 cents and -7.5 cents. (Although equal tempered major thirds are tempered by +13.7 cents relative to just intonation, studies have indicated +15 cents as a perceptual anchor when contrasting both tuning systems; Roberts & Mathews, 1984; Platt & Racine, 1985; Kopiez, 2003). Additionally, each tempered chord was compared to every other tempered chord. The outcome measure for each chord pair was the proportion of trials on which a given chord was chosen as most in-tune. For example, for the just vs. equal pair, proportions greater than 0.5 indicated just intonation was chosen; proportions less than 0.5 indicated equal temperament was chosen. Six of all possible chord pairs presented were selected apriori for analyses: just vs. equal (+15 cents); just vs. -15 cents; just vs. +7.5 cents; just vs. -7.5 cents; equal vs. -15 cents; equal vs. +7.5 cents. These pairs were of most theoretical interest, in comparing both tuning systems, and comparing each system to varying tempering of the major 3rd.

2.3.6.2 Procedure. Participants completed a two alternative forced choice (2AFC) task, where two chords were presented per trial. Twelve instances of each possible chord pairing were presented as trials. Participants fixated a central cross presented against a white background. Four practice trials with feedback were presented (major 3rd of a C major triad mistuned by ±30 cents, compared with major 3rd tempered by +4 cents). 120 test trials followed, with rest screens every 20 trials. On each trial, participants indicated which chord of the pair they perceived as being most in tune. The 'in tune' chord was explained to non-musicians as the chord sounding most consonant or musically acceptable (early pilot work indicated close similarity in results when non-musicians judged which chord was most in-tune, relative to making preference judgements for each pair). Participants used a Logitech Precision Gamepad to indicate which chord was most in tune. The experiment allowed 3 seconds for response from the onset of the second chord, followed by a 1.5 s ITI. Failure to respond within 3 seconds was deemed a non-response; this was followed by a further 1 second ITI before the beginning of the next trial. Two fixed pseudorandom orders of trials were counterbalanced across participants. Position of each chord (i.e., first or second) was counterbalanced across the 12 instances of each pairing in each fixed order. Participants' total non-responses across trials were assessed blind to group. Non-responses were examined to ensure sufficient numbers of observations were included for each chord pairing (minimum of nine per pair, per participant), and to provide a marker of deviation from task instructions. The inclusion criterion was set at the non-response total within two SDs of the group non-response mean. One non-musician and one violinist fell outside this criterion and were excluded.

2.3.7 Environmental auditory scene analysis (EnvASA) task. The EnvASA paradigm measured environmental sound detection within natural auditory scenes (see Leech et al., 2009). Each trial presented one to three short environmental target sounds, followed by a stereophonic auditory background scene. Participants identified each auditory target within the auditory background scene as soon as they detected it. Signal-to-noise ratio (SNR) of targets relative to backgrounds was manipulated at four levels: +3 dB, 0 dB, -3 dB, -6 dB. Congruency of targets relative to backgrounds was also manipulated (e.g., a cow 'moo' target was congruent with a farmyard auditory scene, but incongruent with an office scene). The number of auditory backgrounds also varied, with either a single stereophonic background or two different backgrounds presented dichotically. The dependent variable was percentage of sound targets correctly identified per condition. The inclusion criterion was set at 80% of trials correct or better for the single background, congruent, +3 dB trials (i.e., easiest condition); all participants met this requirement.

2.3.8 Sustained auditory attention to response task (SAART). The SAART was a speeded response switching task, indexing sustained auditory attention (similar to the sustained visual attention task of Manly et al., 1999).

2.3.8.1 Stimuli. Stimuli were nine short environmental sounds taken from Leech et al., (2009). Non-targets were: dog bark, bike bell, camera shutter, basketball bounce, ice cube 'clink', door slam, glass shatter, and frog; targets were a bird call. Durations of the individual sounds ranged from 545–678 ms.

2.3.8.2 Procedure. Participants fixated a central cross against a white background. Each sound began immediately after the response to the preceding stimulus. Two fixed orders of 162 stimuli were counterbalanced across participants. For both orders, the first 81 stimuli (nine instances of each sound) varied pseudorandomly; target sounds never occurred consecutively. The remaining 81 trials presented nine instances of all stimuli; however, targets were preceded by a regular pattern among sounds (at positions target minus 3 and target minus 2). Effects of this pattern on responses are not relevant to the present paradigm and will be discussed elsewhere; results are confined to the first 81 pseudorandom sounds. Participants completed a practice of 18 pseudorandomly arranged sounds (two targets). The 162 experimental trials followed as a single block. Participants responded as quickly as possible with the left index finger for all non-targets, and with the right index finger for targets. A response error on any trial was followed by a 500 ms on-screen error message. Non-response within 2.1 seconds of any sound also produced a 1 s on-screen error message. RTs below 60 ms were deemed early response errors and removed from analyses. RTs for correct trials only were analyzed (log transformed, to correct for positive skew). Total error rates across targets and non-targets were assessed blind to group. Error rates were examined to ensure consistency in the numbers of observations included in calculating mean target and non-target RTs. The inclusion criterion was set at the total error rate within two SDs of the group mean total error rate; two violinists and two non-musicians exceeded this criterion and were excluded.

2.3.9 Pure tone audiometry. Pure tone audiometric thresholds in dB HL were measured using an automated air-conduction thresholding procedure, based on the Hughson–Westlake ascending thresholding method (‘up 5 dB, down 10 dB’). Participants’ ears were tested in turn (left first), for frequencies of 1, 2, 3, 4, 6 and 8 kHz, followed by 500 and 250 Hz. Pure tone audiometry was not run for one violinist due to equipment failure. Pure tone thresholds for all participants were within the normal range, with no significant effects of ear, group, or interactions between these factors (all $p > 0.25$; see Supplemental Table [ST] 11).

2.3.10 Visual psychophysical thresholding. Ahead of visual psychophysical assessment, participants were screened for normal visual acuity with a scaled Lighthouse near visual acuity chart viewed at 40 cm, and for normal color vision using Ishihara plates. Participants then completed the baseline task from Tibber and Shepherd (2006). Participants discriminated increment (purple) and decrement (yellow) color hues from neutral. The task was selected owing to the low relevance of colour discrimination to the training musician groups typically receive. Two adaptive psychophysical staircases were interleaved (one for increment and one for decrement stimuli), and each terminated once 13 reversals occurred. Thresholds were determined as the mean of the final four reversals for each staircase. Staircases were inspected blind to group once all data were collected. Participants with floor level thresholds or who failed to achieve any reversals were not included in analyses. Twenty participants failed to track or displayed floor performance on decrement (yellow) staircases (7 non-musicians, 8 violinists, 5 pianists). Since the decrement staircase was not of theoretical relevance to the present study, analysis was confined to the increment (purple) staircase. Two pianists failed to track on the increment staircase and were removed from analysis. Increment thresholds were expressed as the difference between the co-ordinates of the purple, derived from each staircase, and the neutral when plotted in a log transformed Macleod-Boynton colour space; analyses were performed on these difference values (see Tibber & Shepherd, 2006).

2.4 Data analyses

Non-parametric statistics are reported where data were not normally distributed and could not be corrected for deviations from normality by transformation. Greenhouse–Geisser corrected degrees of freedom and *p* values are reported where any within-subject variables violated the assumption of sphericity. Where post-hoc multiple comparisons were performed, *p* values were corrected using the false discovery rate method (FDR-corrected $\alpha = 0.05$; Benjamini & Hochberg, 1995).

Results

3.1 Auditory Psychophysical Thresholds (Figures 1a-c; Figures 2a-c)

First, we asked whether there were group differences in each auditory psychophysical measure and whether musician groups trained with different instruments differed in their thresholds for specific acoustic features.

3.1.1 Rise time. Rise time thresholds differed significantly across groups, χ^2 (2, $n = 58$) = 15.06, $p = 0.0005$ (Kruskal–Wallis). Planned comparisons showed that non-musicians had higher thresholds than either violinists and pianists (V vs. NM, $z = 3.31$, $p = 0.0009$, Cohen's $d = 1.0$; P vs. NM, $z = 3.50$, $p = 0.0005$, Cohen's $d = 1.2$), but musician groups did not differ from each other ($p = 1.0$).

3.1.2 AM depth. AM depth thresholds differed significantly across groups, χ^2 (2, $n = 69$) = 6.63, $p = 0.036$ (Kruskal–Wallis). Planned comparisons showed non-musicians had significantly higher thresholds than pianists ($z = 2.35$, $p = 0.019$, Cohen's $d = 0.8$), and marginally higher thresholds than violinists ($z = 1.95$, $p = 0.054$, Cohen's $d = 0.6$); musician groups did not differ significantly ($p = 0.49$).

3.1.3 FM depth. FM depth thresholds were significantly different across groups, χ^2 (2, $n = 69$) = 11.03, $p = 0.004$ (Kruskal–Wallis). Again, planned comparisons showed non-musicians had higher thresholds than either musician group (V vs. NM, $z = 2.94$, $p = 0.003$, Cohen's $d = 0.9$; P vs. NM, $z = 2.83$, $p = 0.005$, Cohen's $d = 0.8$) and musician groups did not differ significantly ($p = 0.92$).

In sum, musicians were more sensitive than matched non-musicians to fine distinctions in onset envelope, amplitude modulation depth and frequency modulation depth. However, we saw no evidence of the predicted differences between musician groups. We then asked whether participants' performance changed across runs, and whether non-musicians' final runs might show thresholds similar to musicians' first runs (Micheyl et al., 2006; Kishon-Rabin et al., 2001). As noted in Methods (see 2.3.4.2), because not all participants completed four runs, group sizes were smaller and more unequal, so models were checked using Box's M test of equality of covariance matrices (Stevens, 1996), and results further verified using randomly selected samples with matching Ns.

3.1.4 Rise time (log transformed to correct for positive skew). As shown in Figure 2a, pianists' and non-musicians' sensitivity to rise time envelopes improved significantly over the four runs; violinists showed only marginal improvements. This was reflected in a group x run interaction (see Table 5), verified by post-hoc comparisons between each run (ST 6 and indicated in the figure) and by analyses of random samples (ST 3). In general, both pianists and non-musicians showed improvements from the first pair to the second pair of runs, whereas violinists showed only marginal improvements. Non-musicians' final runs did not differ significantly when compared with violinists and pianists' first runs, χ^2 (2, $n = 41$) = 3.0, $p = 0.22$ (Kruskal–Wallis). In other words, by their fourth run, non-musicians had improved to within the range of the musicians' first attempt.

3.1.5 AM depth. All groups' detection of AM depth improved across the four runs (Figure 2b), as shown by the main effects of run (interaction with group non-significant), verified by analyses of random samples (see Table 5 & ST 4). Thresholds from 1st and 2nd runs were significantly higher than those from 3rd or 4th runs; later runs did not differ significantly (see ST 6). As in the rise time analysis, non-musicians' final run did not differ significantly from the first run completed by musicians, χ^2 (2, $n = 40$) = 1.85, $p = 0.4$ (Kruskal–Wallis).

3.1.6 FM depth. There was limited improvement in FM depth detection across runs (Figure 2c), with no interaction between run and group (see Table 5 and ST 6); the effect of run was also significant in just one random sample (see ST 5). Post-hoc comparisons showed only thresholds from run 1 and run 4 differed significantly (ST 6). As in the other two experiments, non-musicians' final run was not significantly different from musicians' first run, χ^2 (2, $n = 42$) = 1.59, $p = 0.45$ (Kruskal–Wallis).

Table 5: MANOVA analyses of auditory psychophysical thresholds across run and group for each task, with effect of run split by group for rise time task

Model	Wilk's λ	df	F	p	η_p^2
Rise time					
Run	0.289	(3, 36)	29.49	< 0.0001	0.711
Group		(2, 38)	13.03	< 0.0001	0.407
Run x Group	0.491	(6, 72)	5.13	0.0002	0.299
AM depth					
Run	0.473	(3, 35)	13.02	< 0.0001	0.527
Group		(2, 37)	7.07	0.003	0.276
Run x Group	0.811	(6, 70)	1.3	0.28	0.099
FM depth					
Run	0.772	(3, 37)	3.64	0.021	0.228
Group		(2, 39)	2.76	0.076	0.124
Run x Group	0.914	(6, 74)	0.57	0.76	0.044
Rise time					
Run - Violinists	0.509	(3, 10)	3.21	0.07	0.491
Run - Pianists	0.33	(3, 13)	8.81	0.002	0.67
Run - Non-musicians	0.086	(3, 9)	31.83	< 0.0001	0.914

3.2 Visual psychophysical thresholds

It is possible that the musician advantages in the auditory psychophysical measures might be due to overall better performance on challenging psychophysical tasks, rather than reflecting a true difference in auditory perceptual abilities. To test this, participants also completed a color hue psychophysical task. In contrast to the auditory psychophysical results, a one-way ANOVA showed no effect of group on visual color hue (increment) thresholds, F (2, 67) = 1.76, $p = 0.18$, $\eta_p^2 = 0.049$ (see Figure 1d).

3.3 Tuning system perception

We next asked whether extensive training with non-fixed pitch (violin) or fixed pitch (piano) instruments would differentially affect musicians' perception of chord tuning, and whether non-musicians would show a qualitatively different profile of tuning perception. Tests of differences of group means from chance for each chord pair are shown in ST 7 (one-sample Wilcoxon Signed Rank Test (WRST)). Proportion of in-tune choices for each chord pairing were analyzed across groups (Kruskal-Wallis and post-hoc WSRT; Table 6 & Figure 3).

Table 6: Kruskal-Wallis and post-hoc group comparisons across tuning perception task pairs (all post-hoc comparisons false discovery rate-corrected [α = 0.05] for each chord pair)

Model	Just vs. Equal (+15)	Just vs. -15	Just vs. +7.5	Just vs. -7.5	Equal (+15) vs. -15	Equal vs. +7.5
Kruskal-Wallis χ^2 (2, n = 70)	24.24 ***	30.98 ***	12.12 **	24.87 ***	36.98 ***	27.89 ***
Post-hoc (WSRT)						
NM vs. V	$z = 4.44$ * Cohen's $d = 1.9$	$z = 5.18$ * Cohen's $d = 2.5$	$z = 3.04$ * Cohen's $d = 1.0$	$z = 4.65$ * Cohen's $d = 2.1$	$z = 5.45$ * Cohen's $d = 3.1$	$z = 4.85$ * Cohen's $d = 1.7$
NM vs. P	$z = 2.07$, *n.s.*	$z = 3.43$ * Cohen's $d = 1.3$	$z = 0.94$, *n.s.*	$z = 1.86$, *n.s.*	$z = 4.57$ * Cohen's $d = 1.8$	$z = 3.27$ * Cohen's $d = 1.0$
V vs. P	$z = 3.63$ * Cohen's $d = 1.3$	$z = 3.02$ * Cohen's $d = 0.8$	$z = 2.85$ * Cohen's $d = 0.9$	$z = 3.54$ * Cohen's $d = 1.2$	$z = 2.22$, *n.s.*	$z = 2.87$ * Cohen's $d = 0.7$

* $p < 0.05$ (FDR-corrected); ** $p < 0.005$; *** $p < 0.0001$; *n.s.* - non-significant

Violinists selected chords in just intonation – that most relevant to their instrument – when paired with all other chord tunings (with one exception), and did so significantly above chance levels (see Figure 3, panels 1–4; ST 7). The sole exception was just intonation paired with the moderately sharpened +7.5 cents chord (see 3.8). Violinists selected equal temperament as most in tune only when it was paired with the chord deviating the most from both tuning systems (-15 cents). Interestingly, when choosing between an equal tempered (+15 cents) chord versus the moderately sharpened one (+ 7.5 cents), violinists chose the latter – that closer to just intonation (Figure 3, panels 5 & 6).

Pianists selected equal tempered chords – adhering to their instrument-relevant tuning system – significantly above chance when paired with the -15 cents chord. However, this was not the case when equal tempered chords were compared with justly tuned chords. Indeed, pianists selected a smaller extent of tempering (+ 7.5 cents) significantly above chance when paired with either equal temperament or just intonation (Figure 3, panels 3 & 6; ST 7). Pianists only selected just intonation (i.e., their *less* relevant tuning system) significantly more often when matched with the -15 cents chord (Figure 3, panel 2; ST 7). Thus, pianists showed bias toward lesser extents of tempering than typical of their relevant tuning system (equal temperament), choosing their less familiar system only when matched with a tuning deviance.

Finally, non-musicians showed a strong and significant bias against choosing justly tuned chords, with exception of the just vs. -15 cents pair (see Figure 3, panels 1–4; ST 7). Neither did non-musicians select equal temperament significantly above chance when paired with the -15 cents or +7.5 cents chords (Figure 3, panels 5 & 6; ST 7).

Violinists' and non-musicians' choices differed significantly for every chord pair (see Table 6 & Figure 3). Violinists' choices also differed significantly from pianists' choices for every pair, except equal temperament vs. -15 cents (see Table 6; Figure 3).

Unlike violinists, pianists did not differ significantly from non-musicians when judging justly tuned chords versus all others. The only exception was for the justly tuned chord paired with the -15 cents chord; for that pair, pianists selected just intonation significantly more than non-musicians did (see Table 6). Pianists but not non-musicians also showed strong selection of the equal-tempered chord when compared with the -15 cents chord. Finally, pianists – like violinists – chose the +7.5 cents tempered chord on a significantly greater proportion of trials when paired with an equal tempered chord, and did so significantly more than non-musicians did (Table 6).

3.4 SAART

Here, we asked whether musician groups and non-musicians would differ in their ability to sustain auditory attention. We first tested potential differences in reaction time and accuracy to both rare auditory targets and more frequent non-target sounds. We found no significant group differences in overall RTs, F (2, 65) = 0.32, p = 0.73, η_p^2 = 0.01, target response accuracy, F (2, 65) = 0.47, p = 0.63, η_p^2 = 0.01 (one-way ANOVA) or non-targets response accuracy, χ^2 (2, n = 68) = 3.94, p = 0.14 (Kruskal–Wallis) (see Figure 4). RTs to targets and non-targets did differ (F (1, 65) = 9.95, p = 0.002, η_p^2 = 0.133) with mean target RTs slower than for non-targets (Figure 4). However, there was no significant interaction of target/non-target and group, F (2, 65) = 0.59, p = 0.56, η_p^2 = 0.02.

We then asked whether groups differed in a further metric of sustained attention, namely the variability of their reaction times to non-targets (i.e., standard deviation of non-target RTs). Here, groups differed marginally, F (2, 65) = 3.08, p = 0.053, η^2 = 0.086 (one-way ANOVA). Pianists were marginally less variable than non-musicians (i.e., SDs reduced; z = 2.23, p = 0.08, Cohen's d = 0.7), but did not differ from violinists (z = 0.82, p = 0.42). Violinists and non-musicians also did not differ significantly (z = 1.28, p = 0.31, all tests FDR-corrected; Figure 4, panel 5).

3.5 SIMON

We asked whether musicians would outperform non-musicians in multi-modal sequence reproduction, and whether their sequence reproduction would improve when they were passively familiarized with the sequential regularities. A 2 (familiar/non-familiar) x 3 (group) ANOVA on log-transformed mean sequence lengths showed no significant effect of group, F (2, 68) = 2.42, p = 0.096, η_p^2 = 0.07 (Figure 5). There was no main effect of familiarity, F (1, 68) = 0.08, p = 0.77, η_p^2 < 0.01, and no familiarity x group interaction, F (2, 68) = 0.82, p = 0.45, η_p^2 = 0.02. In sum, we found no significant evidence of enhanced general sequencing abilities in musicians, and none for participants being able to reproduce longer sequences when familiarized with the statistical regularities underlying those sequences.

3.6 EnvASA

Next, we investigated whether musical expertise would modulate identification accuracy of environmental sound targets within naturalistic, attentionally demanding auditory scenes, and whether musicians would be more resilient to informational or energetic masking. A 2 (congruent/incongruent) x 2 (single/dual background) x 4 (-6, -3, 0, +3 dB SNR levels) x 3 (group) ANOVA on accuracy rate showed significant main effects of background, congruency and SNR, as well as significant congruency x background and background x SNR interactions (Table 7 and Supplemental Figure [SF] 1). The pattern of effects was as expected given previous studies using this task (see Leech et al., 2009; Krishnan et al., 2013). Contrary to our predictions that musicians would show an advantage in scene analysis and in detection performance under masking conditions, there was no significant main effect of group, nor were there any significant interactions with group (all $F < 1.25$, $p > 0.25$, $\eta_p^2 < 0.04$).

Table 7: Significant ANOVA effects for percentage accuracy across EnvASA conditions

Effect	df	F	*p*	η_p^2
Background	(1, 69)	36.92	< 0.0001	0.349
Congruency	(1, 69)	22.99	< 0.0001	0.25
SNR	(2.304, 158.98)	60.93	< 0.0001	0.469
Congruency x Background	(1, 69)	13.21	0.001	0.161
Background x SNR	(1, 69)	22.99	< 0.0001	0.25

3.7 Cross task analyses

A major focus of this study was to understand whether expertise-related changes in fine-grained auditory perception might be associated with individual differences in more cognitively mediated skills, such as sustained auditory attention, audiovisual sequencing, and auditory scene analysis.

In particular, we asked how individual differences in sustained attention abilities might predict performance on auditory psychophysics tasks, and whether differences between musicians and non-musicians on these perceptual tasks might be partly driven by attentional effects (e.g., Strait et al., 2010). We thus used musician versus non-musician status and sustained auditory attentional metrics as predictors of auditory psychophysical threshold performance.

We also asked whether low-level perceptual abilities – particularly perceiving frequency and amplitude modulation depth – might relate to individual differences in perception of musical chord tempering (i.e., a perceptual task of contextual relevance). This was motivated by the importance of frequency discrimination and detection of beating to tuning perception (Spiegel & Watson, 1984; Vos, 1984; Teki et al., 2012). Thus, we examined correlations between FM depth and AM depth psychophysical thresholds and chord selection within the tuning perception task.

3.7.1 Psychophysical tasks, SAART, SIMON & EnvASA. Auditory psychophysical task thresholds were all significantly positively correlated, but did not correlate significantly with visual psychophysical thresholds (see ST 8).

Auditory psychophysical thresholds were also positively correlated with sustained attention performance (see Table 8). Standard deviations of RTs to SAART non-targets were positively correlated with all auditory psychophysical thresholds – i.e., the lower the standard deviation, the lower the psychophysical threshold – but did not correlate significantly with visual psychophysical thresholds (Table 8). RTs to SAART non-targets also correlated positively with rise time and FM depth thresholds – the lower the RT, the lower the psychophysical threshold – but did not correlate significantly with AM depth or visual psychophysical thresholds (Table 8).

Table 8: Non-parametric correlations between psychophysical tasks and SAART non-target RTs and SDs (false discovery rate-corrected; * $p < 0.05$)

Pair	Spearman's ρ	FDR-corrected p
SAART Non-target RTs & AM depth	0.2	0.16
SAART Non-target RTs & FM depth	0.392	0.02 *
SAART Non-target RTs & Onset rise time	0.381	0.02 *
SAART Non-target RTs & Visual (increments)	-0.23	0.14
SAART Non-target SDs & AM	0.312	0.04 *
SAART Non-target SDs & FM	0.44	0.01 *
SAART Non-target SDs & Onset rise time	0.355	0.03 *
SAART Non-target SDs & Visual (increments)	-0.03	0.86

Auditory psychophysical thresholds were not significantly correlated with SIMON mean sequence length or EnvASA accuracy (average, or at each level of SNR and background, all $p > 0.10$ with FDR correction).

Supporting these analyses, a principal components analysis across all measures showed that auditory psychophysical tasks and sustained attention metrics (SDs and RTs) loaded to similar extents on a single component, accounting for 28.5% of variance ($p < 0.0001$; no other components were significant with a turn in the scree plot after this component; see ST 9). Envasa, SIMON, tuning perception and visual psychophysical measures showed weaker loadings on the component.Because we found significant relationships between auditory psychophysical and sustained auditory attention measures, we assessed whether musician versus non-musician status would still predict auditory psychophysical thresholds when variance due to sustained attention performance was accounted for. Therefore we ran stepwise regressions with musician status (binary predictor; musician groups collapsed) and sustained attention (SAART non-target RTs and non-target SDs) as predictors of auditory psychophysical thresholds.

Both rise time and FM depth thresholds were best predicted by musician status with either SAART non-target RTs or SAART non-target SDs in the regression model. SAART RTs were only marginally predictive of rise time thresholds, and just reached significance as a predictor of FM depth thresholds. SAART SDs were a non-significant predictor for both psychophysical tasks (Table 9, rows 1–12). In contrast, AM depth thresholds were best predicted by SAART non-target SDs; musician status accounted for only marginal unique variance ($p = 0.06$). However, a model with musician versus non-musician status and non-target RTs showed that both were significant predictors of AM depth thresholds, but accounted for less variance than the model with musician

Table 9: Stepwise regression models with musician/non-musician status and SAART performance as predictors of auditory psychophysical thresholds

Model	Adj. R^2	β	df	F	*p*
Rise time	0.304				
Musician vs. Non-musician		8.43	(1, 52)	17.67	0.0001
SAART RT SDs		56.99	(1, 52)	1.94	0.17
Rise time	0.328				
Musician vs. Non-musician		8.97	(1, 52)	22.66	< 0.0001
SAART RTs		36.73	(1, 52)	3.88	0.054
FM depth	0.178				
Musician vs. Non-musician		2.52	(1, 62)	8.80	0.004
SAART RT SDs		27.85	(1, 62)	2.81	0.1
FM depth	0.196				
Musician vs. Non-musician		2.75	(1, 62)	11.61	0.001
SAART RTs		17.51	(1, 62)	4.27	0.043
AM depth	0.225				
Musician vs. Non-musician		0.7	(1, 62)	3.61	0.06
SAART RT SDs		27.3	(1, 62)	12.8	0.001
AM depth	0.15				
Musician vs. Non-musician		1.0	(1, 62)	7.07	0.01
SAART RTs		11.09	(1, 62)	6.18	0.016

versus non-musician status and SAART non-target SDs (see Table 9, rows 13–18).Thus, lower rise time and FM depth thresholds for musicians did not appear to be driven by individual differences in sustained attention (at least as indexed by the SAART measures); in contrast, individual differences in one metric of sustained attention (response variability) captured more variance in AM depth thresholds than did musician status.

3.7.2 Tuning system perception, FM depth and AM depth. Given their potential importance to tuning perception, we asked whether individual differences in sensitivity to envelope (AM depth) and frequency (FM depth) cues might predict how participants perceive chord tuning. Neither violinists' nor non-musicians' performance on FM or AM depth tasks correlated with chord tuning choices for any chord pairs (ST 10). However, pianists' FM depth thresholds were significantly predictive of their choice of just intonation vs. -15 cents tuning (see ST 10, row 8). Pianists with lower FM depth thresholds tended to choose just intonation (their less familiar system) as more in tune than the -15 cents chord (a large tuning deviation). Follow-up regression analyses showed pianists' FM depth thresholds significantly predicted their chord choice for just intonation vs. -15 cents [$F\ (1, 22) = 5.96$, $p = 0.02$, adjusted $R^2 = 0.177$; $\beta = 0.018$]; this relationship was not significant for the violinist group [$F\ (1, 21) = 2.64$, $p = 0.12$, adjusted $R^2 = 0.07$; $\beta = 0.007$; post-hoc test comparing violinists' and pianists' regression coefficients significant, $z = 2.93$, $p = 0.003$ (two-tailed) (Paternoster et al., 1998)]. Pianists' FM depth thresholds also correlated marginally (after FDR correction) with their choices between other chord pairs (with exception of Equal vs. +7.5 cents; ST10). Like the other groups, pianists' AM depth thresholds did not correlate with their tuning choices for any chord pair (see ST10).

General Discussion

4.1 Overview of results

Expert musicians perceive basic acoustic features more finely than non-musicians – although with some practice, non-musicians can get within striking distance of musicians' baseline perceptual performance. Violinists and pianists manipulate these acoustic features in fundamentally different ways, but did not differ in their perceptual sensitivity to these features. Instrument-specific perceptual differences only emerged when subtle frequency differences were presented in a musically relevant context – i.e., when these frequency differences mapped on to the tuning system most relevant to the performer's instrument. Thus, musical expertise – regardless of instrument – may enhance general aspects of lower-level auditory perception to a similar extent. Instrument-specific perceptual sharpening is most evident in musically-relevant harmonic contexts, and in some cases can be predicted by individual differences in frequency modulation sensitivity (in pianists).

Despite their years of experience in reproducing long sequences of notes from memory, segregating multiple complex sound streams, and attending and responding quickly to complex sounds, musicians differed little (if at all) from non-musicians on our measures of sequence reproduction, auditory scene analysis, or sustained auditory attention. However, in both musicians and non-musicians, auditory attention predicted fine perception of certain acoustic cues (AM depth), suggesting that top-down attentional mechanisms may indeed modulate fine-grained perception of some acoustic signal properties (further to Strait et al., 2010).

4.2 Basic psychoacoustic measures

As expected given past results (Kishon-Rabin et al., 2001; Micheyl et al., 2006; Strait et al., 2010; Parbery-Clark et al., 2009b; Teki et al., 2012) we found musicians to be more sensitive than non-musicians to changes in three fundamental acoustical parameters: attack envelope (onset rise time), frequency excursion (FM depth), and carrier amplitude (AM depth). Musicians' finer perceptual skills did not extend to a visual measure or reflect a general advantage on psychophysical tasks in that they did not differ from controls in discriminating gradations in color hue – a perceptual skill not associated with musical expertise.

Contrary to our expectations, the thousands of hours our violinists spent attending to and manipulating the depth of pitch and amplitude modulations (through fine tuning of intonation and vibrato) did not translate into greater sensitivity to perceiving AM or FM depth differences when compared directly to pianists, who cannot control frequency or pitch modulation. Conversely, pianists – whose primary expressive tools are attack and decay envelope – were not more sensitive than violinists to fine differences in rise times. (It is worth noting that violin pizzicato and struck piano touch have similar attack envelopes; see Barbancho et al., 2009; Goebl et al., 2005). These findings extend previous evidence of finer neural response timing to sound (speech phone) onset in musicians versus non-musicians (Parbery-Clark et al., 2012). However, our results contrast with data showing selectively improved acuity for acoustic cues specific to the instrument played (Micheyl et al., 2006; Spiegel & Watson, 1984).

One explanation for this unexpected finding is that pianists might have compensated for not being able to control AM and FM depths and rates through attentive listening to string instrumentalists and vocalists during ensemble playing or accompanying. However, violinists in the present study spent significantly greater time in ensemble performance than did pianists did (see ST 2), making this account a less than compelling one.

It is also possible that violinists listening to and adjusting vibrato quality may not attend to AM and FM as separate parameters, but instead may attend to the strength of the *covariation* between FM and AM, as in the case of deep, rapid vibrato (see Mellody & Wakefield, 2000, for discussion of covarying FM and AM parameters in vibrato signals). Further studies are required to determine if expert pianists and violinists differ in perceptual acuity when both rates and depths of AM and FM are varied concurrently (see Moore & Sek, 1994a, for discussion of concurrent AM and FM perception).

Another surprising finding was how quickly non-musicians as a group reached similar perceptual thresholds to those achieved by musicians in their first runs. While previous studies report that training non-musicians on psychoacoustic tasks can greatly improve frequency discrimination thresholds (Micheyl et al., 2006; see also Kishon-Rabin et al., 2001; Bosnyak et al., 2004), as well temporal interval discrimination (Wright et al., 1997; 2010), it was striking that non-musicians on average would approximate violinists' and pianists' initial perceptual thresholds for such musically-relevant acoustical properties. Nevertheless, it is important to highlight that in most cases, musicians' thresholds also improved significantly over the tracking runs (notably onset rise time thresholds in pianists and AM depth thresholds in both expert groups; see Figure 2). In all tasks, musicians' final thresholds were still lower than non-musicians'. This suggests that while short-term perceptual learning can influence fine acuity, it appears not to outstrip effects of musical expertise – at least over the relatively brief testing periods used here (for discussion, see Ahissar et al., 2009). Indeed, previous studies have shown that 4-8 hours of training are needed before non-musicians achieve f0 difference limens on par with musicians (Micheyl et al., 2006). It is also interesting to note that in the present study, we observed relatively reduced extents of learning across runs for FM depth thresholds (although non-musicians did still tend to reach musicians' baseline levels; further to Kishon-Rabin et al., 2001). This may indicate that perceptual acuity for temporal rather than complex spectral cues is relatively more malleable over very brief periods.

Finally, despite non-musicians' vastly different experience with producing and perceiving sound, many non-musicians' average thresholds were similar to musicians' (see Figure 1). Our musicians might have perceived differences in frequency, amplitude, and attack more finely than non-musicians had the carrier signal been a musical timbre (rather than the non-musical timbre of the sawtooth carrier used here). Musicians show finer perception of pitch and interval cues compared to non-musicians when musical timbre covaries (Pitt, 1994; Platt & Racine, 1985; but see Zarate et al., 2013), and enhanced neural responses to the timbre of the instrument played (Margulis et al., 2009; Strait et al., 2012a; Pantev et al., 2001). We are currently investigating the last possibility, as the results from the tuning sensitivity experiment (discussed below) show the importance of context on perception.

4.3 Contextual effects on experts' auditory perception

In contrast to the lack of low-level psychoacoustic differences across musician groups, and some evidence of overlap between musicians' and non-musicians' thresholds, there were qualitative differences in the way that violinists, pianists, and non-musicians perceived frequency ratios, in agreement with the demands and conventions of their instrumental expertise (or lack thereof). Indeed, previous studies have indicated that preferences for harmonic over inharmonic spectra correlate with years of musical training (McDermott et al., 2010). Our results extend these findings, showing that the instrument musicians train with has a strong influence on their ratings of harmonic tuning – particularly when considering very fine differences in interval size (see Loosen, 1994; 1995).

Violinists showed strong biases towards their instrument-relevant tuning system (i.e., just intonation); the only exception was when their relevant system was paired with a slightly sharpened major third (+7.5 cents; see Figure 3). This slight sharpening can be acceptable to string players and other non-fixed pitch instrumentalists (Roberts & Mathews, 1984; Hall & Hess, 1984; Kopiez, 2003; Platt & Racine, 1985). However, we found some (albeit weak) evidence that violinists who started to practice early in life (at 3–4 years) were more likely to choose the just tempered chord as opposed to the slightly sharp chord (see SR.1). While the power to detect this effect was suboptimal (due to the split of the violinist cohort), we tentatively suggest that early training might drive very fine sensitivity to components of harmonic complexes (further to Roberts & Mathews, 1984; Hall & Hess, 1984; Vos, 1986). Such a finding might be explored in future studies comparing the tuning sensitivities of musicians (e.g., violinists) specifically differing in the age of onset of their training (see Steele et al., 2013, for discussion).

Perhaps due to expert pianists' experience accompanying string players as well as the fixed nature of piano tuning, pianists as a group did not reliably distinguish between their relevant tuning system and their less familiar system (i.e., equal vs. just temperament – see also Spiegel & Watson, 1984; Micheyl et al., 2006). But unlike non-musicians, pianists did reliably choose more 'in-tune' chords (just or equal tempered) when paired with out-of-tune triads (with the middle note adjusted -15 cents relative to just tuning). Moreover, the degree to which pianists' chose the in-tune chord was predicted by their FM (but <u>not</u> AM) depth thresholds – a relationship that was completely absent in the data from violinists or non-musicians. This suggests that individual differences in low-level auditory acuity can have an impact on highly context-dependent perceptual judgments. But, this appears to occur only when the perceptual skill is relevant to the task and when the level of expertise in making those judgments is neither non-existent (as in non-musicians) nor over-practiced (as in violinists) (see Nikjeh et al., 2009). The lack of relationship between AM depth thresholds and tuning perception shown here suggests it may be a less robust perceptual correlate of mistuning; indeed, Teki et al. (2012) found that trained listeners (piano tuners) identify mistuning through fine perception of AM *rate* within specific frequency windows.

4.4 Sustained attention and perceptual performance

The acquisition of expertise may rely in part on developing sustained attentional abilities, particularly directed toward training-relevant stimuli or task goals (e.g., Tervaniemi et al., 2005; see Palmeri et al., 2004, for discussion). We found limited evidence that our musicians differed from non-musicians in this regard, with pianists – but not violinists – marginally less variable in their response times compared to non-musicians.

However, sustained auditory attention <u>did</u> predict significant variance in AM depth thresholds – and beyond what could be accounted for by musical expertise alone. This suggests that sustained auditory attention skills can contribute to fine acoustic perception (further to Ahissar et al., 2009) – but that these attentional skills are modality-delimited, as shown by the lack of relationship between the SAART measures and visual psychophysics performance (see Braga et al., 2013, for a recent demonstration of the modality-specific nature of attentional systems). Our PCA analyses also found that auditory psychophysical performance loaded with sustained auditory attention performance on a single component, thereby further supporting a relationship between auditory attention and some fine perceptual abilities in both musicians and non-musicians (Strait et al., 2010, 2012b; Strait & Kraus, 2011b; Tervaniemi et al., 2005; see also Zhang et al., 2012).

4.5 Auditory scene analysis

Musicians spend many hours in hugely complex auditory environments (e.g., ensembles and symphony orchestras). For instance, violinists and pianists playing with orchestras must listen for particular motifs generated by single sound sources that will be masked by dozens of other sound generators, and that may exceed the target sound in amplitude and salience. An open question is whether these advanced musical scene analysis abilities would extend to detecting and identifying familiar sounds in everyday auditory scenes, particularly under informational and energetic masking conditions. To our surprise, we found no evidence that musicians and non-musicians performed differently, under even the most demanding listening conditions. Moreover, we did not find that our violinist cohort – who spent significantly greater time in ensembles (see ST 2) – performed any better than our pianist cohort. These results contrast with previous reports of enhanced musician performance under the demands of competing speech (Parbery-Clark et al., 2009a, 2009b, 2011; Strait et al., 2012b), sources of informational masking (Oxenham et al., 2003; see footnote 10), backward masking (Strait et al., 2010), and detection of auditory objects (Zendel & Alain, 2009, 2013). Our findings also contrast with previous evidence that specific expertise with ensemble settings benefits selective attention to spatially segregated sounds (Nager et al., 2003). Recent findings suggest musician advantages for speech perception may emerge most clearly when listening demands are presented binaurally or with spatial segregation (Parbery-Clark et al., 2013; Strait et al., 2012b). However, the complex, binaural nature of the scenes presented in our task (particularly the dual backgrounds) failed to reveal any musician advantage. Moreover, a recent investigation of musician versus non-musician performance on measures of voiced and unvoiced speech perception in noise (Ruggles et al., 2014) failed to show any musician advantage – a finding partly in agreement with our non-linguistic results.

What might account for the difference between current and past results? First, it is possible that lower target/background SNRs (e.g., Gygi & Shaffiro, 2011) would have increased task difficulty and therefore have allowed group differences to emerge, particularly in dual background conditions (Leech et al., 2009). We should note that average performance in the high SNR and single background conditions was relatively high, and therefore may have caused ceiling effects. However, even at the lowest SNR (-6 dB; mean accuracies reduced to 70-80% in the dual background condition; see SF1) we did not find any hint of a musician advantage. A further possibility is that musicians' expertise in detecting, identifying, and attending to auditory targets is limited to targets that share characteristic acoustic and spatial cues of musical instruments in an ensemble – characteristics that can differ dramatically from other sound sources (for discussion, see Dick et al., 2011; Nager et al., 2003). Thus, it may be that musicians' expertise in scene analysis is context-specific, with limited benefit to non-musical auditory environments. Indeed, lack of skill transfer has also been observed in some cases of visual scene expertise (see Green & Bavelier, 2012, for discussion).

4.6 Sequence perception and reproduction

Playing a musical instrument fundamentally involves encoding and reproducing sequentially organized units of sound, as well as recognizing and using regularities in those sequences (e.g., Koelsch et al., 2002; see Bharucha et al., 2006, for review). Predicting generalization of such skills, we expected that musicians would reproduce longer multi-modal sequences than non-musicians. We also predicted that familiarity with the auditory structure of half of the sequences might allow participants – particularly musicians – to learn and use that structure to aid reproduction.

We found little evidence in favor of our hypotheses. The lack of a robust musician advantage for such a seemingly ‘musical’ task is somewhat puzzling. It may be that our participants did not rely on the tones to reproduce the sequence, and relied on the visuospatial component of the task. This would tally with the results of Taylor-Tierney et al. (2008), who found musician advantages only for audio and not audiovisual sequences (but cf. Conde et al., 2012). However, very recent unpublished data from a sequencing experiment in our laboratory – one with a smaller, less expert, and more heterogeneous musician sample – showed a musician advantage for both audiovisual and audio-only sequence reproduction. It may be that cohort effects are in part behind these inconsistent results, especially in terms of the non-musician group, which in the present study was well-matched to the musician groups in educational level and motivation. In particular, uncontrolled variation in sustained attentional abilities in non-musicians may underlie such conflicting results. Indeed, in the present study, sustained auditory attention (measured through SAART non-target RTs) was significantly related to SIMON mean sequence length ($\rho = -0.404$, $p = 0.01$), whereas musician status was not.

We found no evidence that musicians or non-musicians were able to reproduce longer sequences when they had been familiarized with the auditory structure of the sequences beforehand. Contrary to expectation, this suggests that phases of brief, passive auditory experience do not transfer to a later active, multi-modal task. The lack of group differences is in keeping with previous results showing that musicians and non-musicians do not differ in learning the underlying structure of sequences following periods of passive experience (Rohrmeier et al., 2011; cf. Shook et al., 2013; see also Reber, 1993, for discussion). As suggested by Loui et al. (2010), novel sequential regularities may present challenges for trained listeners, particularly in the face of existing knowledge of Western harmony (see also McMullen Jonaitis & Saffran, 2009). Our experts’ detailed (and likely explicit – see Hannon & Trainor, 2007) knowledge of Western tonal relations may therefore have interfered with learning or using the familiarized statistical regularities within our tone sequences (Loui et al., 2010; McMullen Jonaitis & Saffran, 2009). This suggests that learning of novel, regular auditory structures may be limited by prior expert knowledge or expectations.

4.7 Expertise and Generalization

As discussed in the above sections (4.1–4.6), we found large effects on auditory perception that were related to musical expertise (Cohen's *d* between 0.6 and 1.2 for psychophysical thresholds, and between 0.7 and 3.1 for chord tuning perception). In contrast, we found little evidence of benefit of musical expertise to auditory cognitive skills, despite the broad relevance of many such skills to both musical performance and practice. While task factors and variables such as personality likely play a role (Corrigall et al., 2013), our results nevertheless offer a point of contrast with many previous studies indicating transfer of cognitive skills arising from musical experience. Why might we have failed to find differences between groups across cognitive measures? One possible explanation is the close matching of our cohorts for levels of education. As outlined in methods, all of our controls had attained or were studying for a third level degree (several were MSc or PhD students; see Table 4). Our reasoning for this was that factors such as motivation, diligence and personality (e.g., Ericsson et al., 1993; Corrigall et al., 2013) might serve to confound comparisons of musicians and non-musicians across cognitive tasks. Due to the current battery's duration (3+ hours), we were unable to measure full-scale IQ; this is a limitation of the present design and will be remedied in future studies. Nevertheless, when we analyzed enrolment in higher education, we found no significant differences in the number of undergraduate versus post-graduate students across our cohorts; further, including this education variable as an additional regressor in our analyses revealed that it did not account for any significant variance (see footnote 7). Moreover, several studies in children and adults have shown that versions of the SIMON task are moderately to strongly correlated with measures of working memory and 'fluid intelligence'

(Baniqued et al., 2013; Cleary, Pisoni & Geers, 2001; Krishnan et al., in review). Based on these data, we could anticipate that a difference in broader intellectual level between our cohorts might manifest on the SIMON task; however, we did not find any significant group differences on this task. In the absence of standardized measures of cognitive performance (prevented due to time limitations), we suggest that the close matching for enrolment in third level education (particularly at high levels of post-graduate study) and very similar performance on the SIMON task may afford some control of broader intellectual level across our cohorts.

Although we found limited evidence that musical expertise generalized to broader cognitive metrics, such negative results are important in and of themselves. The question of benefits related to musical expertise and training has been explored for several decades, often yielding reports of positive generalization (for review, see Moreno & Bidelman, 2014; White et al., 2013). Necessary corollaries of conducting such tests include homogeneity within-samples *and* careful matching across samples, to detect an effect of musical skill if it exists and to minimize the possibility that such an effect is driven by uncontrolled nuisance factors. The present study is notable for the close matching within and across groups on a variety of nuisance variables, and the number of carefully normed experimental measures (11; see supplemental results 2). This is especially important when interpreting ‘negative’ findings. It is impossible to verify the null hypothesis (i.e., that musicians and non-musicians do not differ on cognitive measures). To further explore these null results, we ran equivalence tests across our cognitive measures; this allows us to test whether the observed group differences were significantly different from values deemed to be trivial (see Hoenig & Heisey, 2001). For each task, we first tested the groups using values of ±1 SD of the full distribution (i.e., including all groups). We found that in almost all cases the observed data were significantly different from test values of ±1 SD – i.e., the group difference was significantly smaller than the one standard deviation range (see ST 15). However, we found that virtually all of the observed differences were not significantly different from test values of ±0.5 SD (see ST 16). This suggests that the extent of group differences – if there were true differences – were lower than

±1 SD but fell within the range of ±0.5 SD. These additional equivalence tests, together with the robust differences we found on perceptual metrics suggest that if there was any potential effect of musicianship on the cognitive tasks, that it was relatively small. .

However, it is important to distinguish between experimental manipulations involving musical training assignment and correlational designs (as employed here). Indeed, assignment to musical training has been found to yield structural changes in auditory and motor cortices that correlate with performance on melody discrimination and finger tapping tasks, respectively (Hyde et al., 2009). While such results indicate near transfer, further studies demonstrate far transfer: school-aged children assigned to one year of keyboard or vocal training showed significant gains in full-scale IQ (versus peers assigned to drama lessons; Schellenberg, 2004). Further, Moreno and colleagues demonstrated far transfer in two studies: assignment of children to musical training versus a control activity (visual art classes) led to significant increases in negativity of ERP amplitudes in response to speech pitch violations (Moreno et al., 2008), and improvements in verbal IQ and executive functioning (response inhibition; Moreno et al., 2011). Nevertheless, in line with the present study, Hyde et al. (2009) found no evidence of far transfer of musical training to abilities such as verbal or non-verbal IQ. Further, recent data from adults suggest no far transfer of musicianship to speech-in-noise perception (Ruggles et al., 2014).

Although the above results suggest some transfer attributable to musical training (see also Lappe et al., 2008; Besson et al., 2011), one remaining question is whether the occurrence of transfer is selective to specific points in development. Can musical training-related cognitive differences persist beyond childhood? Or does early musical training afford children an initial advantage on some cognitive tasks, with non-musically trained children attaining similar performance at subsequent points in development (for instance, as they progress through formal education and reach adolescence/adulthood)? Our study does not allow direct investigation of these issues. Nevertheless, we could suggest that given the limited broader expertise benefits demonstrated here in adults, the latter hypothesis may be plausible. Future longitudinal studies of children assigned to music lessons and control activities may help to address these questions (see Costa-Giomi, 1999). If musicianship is to be studied as a model for plasticity – or as an intervention for hearing, attentional or language difficulties – then potential limitations of generalization must be understood.

4.8 Conclusions

Experience-dependent accounts of auditory perceptual learning and cognitive performance can be explored using expert musician groups with qualitatively different training profiles. Such differences in experience allow investigation of whether training demands lead to specific or more general perceptual and cognitive advantages, and thus offer insight into the generalization of human learning. In comparing non-musicians to two expert musician groups trained under very different acoustic and performance constraints, we found a profile of enhancements relatively specific to the area of training. Expert musician groups differed in their perception within a training-relevant context, yet showed no differences in lower-level auditory perceptual skills. These findings indicate that auditory perception may be honed most specifically within contexts close to the area of training, suggesting a role for context in delineating how expert musician groups diverge.

In exploring expertise generalization, we conclude that musical expertise may not benefit skills such as auditory scene analysis or auditory learning and sequencing when contextually removed from musical stimuli or performance situations. Our results nevertheless provide some evidence of interactions between cognitive skills and perceptual acuity: top-down attentional abilities may partly account for fine acuity for certain auditory signal features in both experts and non-experts. These findings hold implications for the extent to which musical training may be an effective intervention for learning or language-related difficulties (for discussion, see Parbery-Clark et al., 2013; Strait et al., 2012b; Kraus & Chandrasekaran, 2010). Musical training could yield benefits to difficulties related to fine-grained listening, but perhaps may provide greatest benefit when integrated with attentional skill training.

Our study provides among the first examinations of perceptual and cognitive skills in musician cohorts trained on very different instruments, whilst also allowing insight into perception-cognition interactions within the same individuals – both expert and non-expert. Our findings contribute to a growing understanding of learning as influenced by specific profiles of long-term experience, and provide further evidence of interaction between fine-grained perception and top-down attention. These results invite future efforts to explore the mechanisms through which long-term experience may guide learning outcomes and spur transfer of learning to broader perceptual and cognitive abilities.

References

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364,* 285–299.

Amir, O., Amir, N., & Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *Journal of the Acoustical Society of America*, *113*, 1102–1108.

Baker, R.J., & Rosen, S. (2001). Evaluation of maximum-likelihood threshold estimation with tone-in-noise masking. *British Journal of Audiology, 35,* 43–52.

Banai, K., Sabin, A.T., & Wright, B.A. (2011). Separable developmental trajectories for the abilities to detect auditory amplitude and frequency modulation. *Hearing Research, 280,* 219–227.

Baniqued, P.L., Lee, H., Voss, M.S., Basak, C., Cosman, J.D., … Kramer, A.F. (2013). Selling points: what cognitive abilities are tapped by casual video games? *Acta Psychologica, 142*(1), 74–86.

Barbancho, I., de la Bandera, C., Barbancho, A. M., & Tardon, L. J. (2009). Transcription and expressiveness detection system for violin music. *Proceedings of the IEEE conference on Acoustics, Speech and Signal Processing (ICASSP 2009),* 189–192.

Baumann, S., Meyer, M., & Jäncke, L. (2008). Enhancement of auditory evoked-potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience, 20,* 2238–2249.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57,* 289–300.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: common processing, attention and memory. *Frontiers in Psychology, 2,* doi: 10.3389/fpsyg.2011.00094.

Bharucha, J.J., Curtis, M., & Paroo, K. (2006). Varieties of musical experience. *Cognition, 100,* 131–172.

Bosnyak, D.J., Eaton, R.A., & Roberts, L.E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex, 14,* 1088–1099.

Braga, R. M., Wilson, L.R., Sharp, D.J., Wise, R.J.S., & Leech, R. (2013). Separable networks for top-down attention to auditory non-spatial and visuospatial modalities. *NeuroImage*, *74,* 77-86.

Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound*. London: MIT Press.

Chi, M.T.H. (2006). Laboratory methods for assessing experts' and novices' knowledge. In K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 167-184). Cambridge: Cambridge University Press.

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D.C. (2012). Optimal encoding of interval timing in expert percussionists. *The Journal of Neuroscience, 32,* 1056–1060.

Conde, V., Altenmüller, E., Villringer, A., & Ragert, P. (2012). Task irrelevant auditory feedback facilitates motor performance in musicians. *Frontiers in Psychology, 3,* doi: 10.3389/fpsyg.2012.00146

Cornsweet, T.N. (1962). The staircase-method in psychophysics. *The American Journal of Psychology, 75,* 485–491.

Corrigall, K.A., Schellenberg, E.G., & Misura, N.M. (2013). Musical training, cognition and personality. *Frontiers in Psychology, 4,* 1–12; doi: 10.3389/fpsyg.2013.00222.

Costa-Giomi, E. (1999). The effects of three years of piano instruction on children's cognitive development. *Journal of Research in Music Education, 47,* 198–212.

Dick, F., Lee, H.L., Nusbaum, H., & Price, C. (2011). Auditory-motor expertise alters 'speech selectivity' in professional musicians and actors. *Cerebral Cortex, 21,* 938–948.

Duffin, R.W. (2007). *How equal temperament ruined harmony (and why you should care).* New York: W.W. Norton & Co.

Durlach, N.I., Mason, C.R., Shinn-Cunningham, B.G., Arbogast, T.L., Colburn, H.S., & Kidd, G.R. (2003). Information masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *Journal of the Acoustical Society of America, 114,* 1-12.

Ehrlé, N., & Samson, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain and Cognition*, *58*, 133–147.

Ericsson, K.A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 685-706). Cambridge: Cambridge University Press.

Ericsson, K.A., Krampe, T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406.

Francois, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex, 21,* 2357–2365.

Fritz, C., Woodhouse, J., Cheng, F.P.H., Cross, I., Blackwell, A.F., & Moore, B.C.J. (2010). Perceptual studies of violin body damping and vibrato. *Journal of the Acoustical Society of America, 127,* 513–524.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L.J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain, 129,* 2593–2608.

Goebl, W. (2001). Melody lead in piano performance: expressive device or artifact? *Journal of the Acoustical Society of America, 110,* 563–572.

Goebl, W., Bresin, R., & Galembo, A. (2005). Touch and temporal behavior of grand piano actions. *Journal of the Acoustical Society of America, 118,* 1154–1165.

Goldstone, R. P. (1999). Perceptual learning. *Annual Review of Psychology, 49,* 585–612.

Gordon, J.W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America, 82,* 88–105.

Green, C.S., & Bavelier, D. (2012). Learning, attentional control and action video games. *Current Biology, 22,* R197–R206.

Greenberg, G.Z., & Larkin, W.D. (1968). Frequency response characteristic of auditory observers detecting signals of a single frequency in noise: the probe signal method. *Journal of the Acoustical Society of America, 44,* 1513–1523.

Gygi, B., & Shaffiro, V. (2011). The incongruency advantage for natural sounds presented in natural auditory scenes. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 551–565.

Hall, D.E., & Hess, J.T. (1984). Perception of musical interval tuning. *Music Perception, 2,* 166–195.

Hannon, E.E., & Trainor, L.J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Sciences, 11,* 466–472.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: is there any evidence of a true deficit? *Journal of Speech, Language and Hearing Research, 52,* 1510–1529.

Herholz, S.C., & Zatorre, R.J. (2012). Musical training as a framework for brain plasticity: behaviour, function and structure. *Neuron, 76,* 486–502.

Herholz, S.C., Boh, B., & Pantev, C. (2011). Musical training modulates encoding of higher-order regularities in the auditory cortex. *European Journal of Neuroscience, 34,* 524–529.

Hoenig, J.M., & Heisey, D.M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician, 55*(1), 19–24.

Hoffman, G., Mürbe, D., Kuhlisch, E., & Pabst, F. (1997). Unterschieded des auditiven frequenzdiskriminationsvermogens bei musikern verschiedener fachbereiche. *Folia Phoniatrica et Logopaedica, 49,* 21–25.

Hopkin, B. (1996). *Musical instrument design: Practical information for instrument making*. Tucson: Sharp Press.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical Training Shapes Structural Brain Development. *Journal of Neuroscience*, *29*(10), 3019–3025.

Justus, T., & List, A. (2005). Auditory attention to frequency and time: an analogy to visual local-global stimuli. *Cognition, 98,* 31–51.

Karpicke, J.D., & Pisoni, D. B. (2004). Using immediate memory span to measure implicit learning. *Memory and Cognition, 32,* 956–964.

Kayser, C., Petkov, C.I., Lippert, M., & Logothetis, N.K. (2005). Mechanisms for allocating auditory attention: an auditory saliency map. *Current Biology, 15,* 1943–1947.

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *Journal of Basic Clinical Physiology and Pharmacology, 12,* 125–143.

Kleiner, M., Brainard, D., & Pelli, D. (2007). "What's new in Psychtoolbox-3?" *Perception, 36,* ECVP Abstract Supplement.

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport, 10,* 1309–1313.

Koelsch, S., Schmidt, B.H., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: an event-related brain potential study. *Psychophysiology, 39,* 657–663.

Kopiez, R. (2003). Intonation of harmonic intervals: adaptability of expert musicians to equal temperament and just intonation. *Music Perception, 20,* 383–410.

Krishnan, S., Leech, R., Aydelott, J., & Dick, F. (2013). School-age children’s environmental object identification in natural auditory scenes: Effects of masking and contextual congruence. *Hearing Research,* http://dx.doi.org/10.1016/j.heares.2013.03.003

Krishnan, S., Alcock, K.J., Carey, D., Bergstrom, L., Karmiloff-Smith, A., & Dick, F. (in review). Fractionating nonword repetition: the contributions of short-term memory and oromotor praxis are different. *Developmental Science.*

Krampe, R.T., & Ericsson, K.A. (1996). Maintaining excellence: deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General, 125,* 331–359.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11,* 599–605.

Lappe, C., Herholz, S.C., Trainor, L.J., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *The Journal of Neuroscience, 28,* 9632–9639.

Larkin, W.D., & Greenberg, G.Z. (1970). Selective attention in uncertain frequency detection. Perception and Psychophysics, 8(3), 179–184.

Leech, R., Gygi, B., Aydelott, J., & Dick, F. (2009). Informational factors in identifying environmental sounds in natural auditory scenes. *Journal of the Acoustical Society of America, 126,* 3147–3155.Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49,* 467–477.

Loosen, F. (1994). Tuning of diatonic scales by violinists, pianists and non-musicians. *Perception & Psychophysics, 56,* 221–226.

Loosen, F. (1995). The effect of musical experience on the conception of accurate tuning. *Music Perception, 12,* 291–306.

Loui, P., Wessel. D.M., & Kam, C.L.H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception, 27,* 377–388.

Manly, T., Robertson, I.H., Galloway, M., & Hawkins, K. (1999). The absent mind: further investigations of the sustained attention to response. *Neuropsychologia, 37,* 661–670.

Marie, C., Magne, C., & Besson, M. (2011). Musicians and metric structure of words. *Journal of Cognitive Neuroscience, 23,* 294–305.

Margulis, E.H., Mlsna, L.M., Uppunda, A.K., Parrish, T.B., & Wong, P.C.M. (2009). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Human Brain Mapping, 30,* 267–275.

McDermott, J.H., Lehr, A.J., & Oxenham, A.J. (2010). Individual differences reveal the basis of consonance. Current Biology, 20, 1035–1041.

McMullen Jonaitis, E., & Saffran, J. (2009). Learning harmony: the role of serial statistics. *Cognitive Science, 33,* 951–968.

Mellody, M., & Wakefield, G.H. (2000). The time-frequency characteristics of violin vibrato: Modal distribution analysis and synthesis. *Journal of the Acoustical Society of America, 107,* 598–611.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*, 36–47.

Mondor, T.A., & Bregman, A.S. (1994). Allocating attention to frequency regions. *Perception and Psychophysics, 56,* 268–276.

Moore, B.C.J. (2012). *An Introduction to the Psychology of Hearing* (6th ed.) Brill: The Netherlands.

Moore, B.C.J., & Sek, A. (1994). Discrimination of modulation type (amplitude modulation and frequency modulation) with and without background noise. *Journal of the Acoustical Society of America, 96,* 726–732.

Moreno, S., & Bidelman, G.M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research, 308,* 84–97.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E.G., Cepeda, N.J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science, 22,* 1425–1433.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S.L., & Besson, M. (2008). Musical training influences linguistic abilities of 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex, 19,* 712–723.

Müllensiefen, D., Gingras, B., Stewart, L., & Musil, J. (2011). The Goldsmiths Musical Sophistication Index (Gold-MSI): Technical Report and Documentation v0.9. London: Goldsmiths, University of London.

Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., & Münte, T.F. (2003). The fate of sounds in conductors' brains: an ERP study. *Cognitive Brain Research, 17,* 83–93.

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *Journal of the Acoustical Society of America*, *125*, 328–338.

Ouimet, T., Foster, N.E.V., & Hyde, K.L. (2012). Auditory global-local processing: effects of attention and musical experience. *Journal of the Acoustical Society of America, 132,* 2536–2544.

Oxenham, A.J., Fligor, B., Mason, C., & Kidd, G. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America, 114,* 1543–1549.Palmeri, T.J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5,* 291–303.

Palmeri, T.J., Wong, A.C.N., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Sciences, 8,* 378–386.

Pantev, C., Roberts. L.E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport, 12,* 1–6.

Papich, G., & Rainbow, E. (1974). A pilot study of performance practices of twentieth century musicians. *Journal of Research in Music Education, 22,* 24–34.

Parbery-Clark, A., Strait, D.L., Hittner, E., & Kraus, N. (2013). Musical training enhances neural processing of binaural sounds. *The Journal of Neuroscience*, *33*, 16741–16747.

Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Frontiers in Aging Neuroscience, 4,* doi: 10.3389/fnagi.2012.00030

Parbery-Clark, A., Strait, D.L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia, 49,* 3338–3345.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009a). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience*, *29*, 14100–14107.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009b). Musician enhancement for speech-in-noise. *Ear and Hearing*, 30, 653–661.

Patel, A.D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2,* doi: 10.3389/fpsyg.2011.00142

Patel, A.D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6,* 674–681.Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology, 36,* 859–866.

Pearce, M.T., Ruiz, M.H., Kapasi, S., Wiggins, G.A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computations, behavioural, and neural manifestations of musical expectation. *NeuroImage, 50,* 302–313.

Pitt, M.A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 976–986.

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and non-musicians. *Music Perception, 24,* 37–48.

Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* Oxford: Oxford University Press.

Roberts, L., & Mathews, M.V. (1984). Intonation sensitivity for traditional and non-traditional chords. *Journal of the Acoustical Society of America, 75,* 952–959.

Rodrigues, A.C., Loureiro, M.A., & Caramelli, P. (2013). Long-term musical training may improve different forms of visual attention ability. *Brain and Cognition, 82,* 229–235.

Rohrmeier, M., Rebuschat, P., & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition, 20,* 214–222.

Rosen, S., & Howell, P. (1981). Plucks and bows are not categorically perceived. *Perception and Psychophysics, 30,* 156–168.

Ruggles, D.R., Freyman, R.L., & Oxenham, A.J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *Plos One, 9,* e86980.

Schellenberg, E.G. (2004). Music lessons enhance IQ. *Psychological Science, 15,* 511–514.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*(7), 688–694.

Seppänen, M., Pesonen, A.K., Tervaniemi, M. (2012). Music training enhances the rapid plasticity of P3a/P3b event-related brain potentials for unattended and attended target sounds. *Attention, Perception and Psychophysics, 74,* 600–612.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience*, *23*(12), 5545–5552.

Shahin, A., Roberts, L. E., & Trainor, L. J. (2004). Enhancement of auditory cortical development by musical training in children. *NeuroReport*, *15*(12), 1917–1921.

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., & Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage*, *41*(1), 113–122.

Shook, A., Marian, V., Bartolotti, J., & Schroder, S.R. (2013). Musical experience influences statistical learning of a novel language. *The American Journal of Psychology, 126,* 95–104.

Siegel, J.A., & Siegel, W. (1977). Absolute identification of notes and intervals by musicians. *Perception and Psychophysics, 21,* 143–152.

Skoe, E., & Kraus, N. (2013). Musical training heightens auditory brainstem function during sensitive periods in development. *Frontiers in Psychology, 4,* 1–15, doi: 10.3389/fpsyg.2013.00622.

Smith, D.W., Aouad, R.K., & Keil, A. (2012). Cognitive task demands modulate the sensitivity of the human cochlea. *Frontiers in Psychology, 3,* doi: 10.3389/fpsyg.2012.00030.

Spiegel, M. F., & Watson, C.S. (1984). Performance on frequency□discrimination tasks by musicians and non-musicians. *Journal of the Acoustical Society of America*, *76*, 1690–1696.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). New Jersey: Lawrence Erlbaum Associates.

Steele, C. J., Bailey, J. A., Zatorre, R J, & Penhune, V. B. (2013). Early musical training and white-matter plasticity in the corpus callosum: evidence for a sensitive period. *Journal of Neuroscience*, *33*(3), 1282–1290.

Strait, D.L., & Kraus, N. (2014). Biological impact of auditory expertise across the lifespan: Musicians as a model of auditory learning. *Hearing Research, 308,* 109–121.

Strait, D.L., Chan, K., Ashley, R. & Kraus, N. (2012a). Specialization among the specialized: Auditory brainstem function is tuned to timbre. *Cortex, 48,* 360–362.

Strait, D.L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012b). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language, 123,* 191–201.

Strait, D.L., & Kraus, N. (2011a). Playing music for a smarter ear: cognitive, perceptual and neurobiological evidence. *Music Perception, 29,* 133–146.

Strait, D.L., & Kraus, N. (2011b). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in Psychology, 2,* doi: 10.3389/fpsyg.2011.00113

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*, 22–29.

Taylor-Tierney, A., Bergeson, T.R., & Pisoni, D.B. (2008). Effects of early musical experience on auditory sequence memory. *Empirical Musicology Review, 3,* 178–186.

Teki, S., Kumar, S., von Kriegstein, K., Steweart, L., Lyness, C.R., Moore, B.C.J., Capleton, B., & Griffiths, T. D. (2012). Navigating the auditory scene: an expert role for the hippocampus. *The Journal of Neuroscience, 32,* 12251–12257.

Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., & Friederici, A.D. (2009). Top-down modulation of auditory processing: effects of sound contect, musical expertise and attentional focus. *European Journal of Neuroscience, 30,* 1636–1642.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005) Pitch-discrimination accuracy in musicians vs. non-musicians – an event- related potential and behavioral study. *Experimental Brain Research, 161,* 1–10.

Tibber, M.S., & Shepherd, A. J. (2006). Transient-tritanopia in migraine: evidence for a large field retinal abnormality in blue-yellow opponent pathways. *Investigative Ophthalmology & Visual Science, 47,* 5125–5131.

Trainor, L.J., Shahin, A., & Roberts, L.E. (2003). Effects of musical training on the auditory cortex in children. *Annals of the New York Academy of Sciences, 999,* 506–513.

Vos, J. (1984). Spectral effects in the perception of pure and tempered intervals. *Perception and Psychophysics, 35,* 173–185.

Vos, J. (1986). Purity ratings of tempered fifths and major thirds. *Music Perception, 3,* 221–257.

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., & Tervaniemi, M. (2012). The sound of music: differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia, 50,* 1432–1443.

Wessel, D.L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3,* 45–52.

White, E.J., Hutka, S.A., Williams, L.J., & Moreno, S. (2013). Learning, neural plasticity and sensitive periods: implications for language acquisition, music training and transfer across the lifespan. *Frontiers in Systems Neuroscience, 7,* 1–18. doi: 10.3389/fnsys.2013.00090

Wright, B.A., Buonomano, D.V., Mahncke, H.W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *The Journal of Neuroscience, 17,* 3956–3963.

Wright, B.A., Wilson, R.M., & Sabin, A.T. (2010). Generalization lags behind learning on an auditory perceptual task. *The Journal of Neuroscience, 30,* 11635–11639.

Zarate, J.M., Ritson, C.R., & Poeppel, D. (2013). The effect of instrumental timbre on interval discrimination. *Plos One, 8,* e75410.

Zarate, J.M., Ritson, C.R., & Poeppel, D. (2012). Pitch-interval discrimination and musical expertise: is the semi-tone a perceptual boundary? *Journal of the Acoustical Society of America, 132,* 984–993.

Zendel, B., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, *21*, 1488–1498.

Zendel, B., & Alain, C. (2013). The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *Journal of Cognitive Neuroscience, 25,* 503–516.

Zhang, Y.X., Barry, J.G., Moore, D.M., & Amitay, S. (2012). A new test of attention in listening predicts auditory performance. *Plos One, 7,* e53502.

van Zuijen, T., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2004). Grouping of sequential sounds-an event-related potential study comparing musicians and nonmusicians. *Journal of Cognitive Neuroscience, 16,* 331–338.

van Zuijen, T., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2005). Auditory organisation of sound sequences by a temporal or numerical regularity - a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research, 23,* 270–276.

Footnotes

1. The difference between fixed and non-fixed pitch instrumentalists' perception of frequency may also have accounted for the genre effects reported by Kishon-Rabin et al. (2001) in that all but one of their 'contemporary' musicians played only fixed-pitch or fretted instruments, while all the 'classical' musicians played wind, brass, or string instruments where adjusting intonation is a crucial aspect of playing (see Micheyl et al., 2006, for discussion).
2. Experiment 1 from Fritz et al. (2010), compared perception of vibrato amplitude in a small sample of string players (*n* = 4) and non-string players (referred to as 'other musicians'; *n* = 11); the groups of musicians did not differ in their perception of change in depth of vibrato signal amplitude. Further, modification of the distribution of harmonics within the auditory signal (through applying a filter to mimic violin resonance properties) did not improve perception of vibrato.
3. Pythagorean tuning (a tuning system that derives from relating notes according to a circle of perfect fifths; Loosen, 1994) is also used by string instrumentalists such as violinists. As with just intonation, it cannot be employed by fixed pitch instrumentalists (e.g., pianists). Loosen's (1994) findings suggested violinists showed greater deviance in adjusting to scales that were tuned in just intonation, compared to scales tuned to the Pythagorean system. Just intonation is explored in the current study, further to the work of Roberts and Mathews (1984).
4. Energetic masking is defined by Moore (2012) as occurring when the neural activity evoked by the signal plus the masker is the same as (or very similar to) the neural activity evoked by the masker alone. Moore (2012) defines informational masking as occurring where the signal and masker are confused by the listener, or where there is perceptual difficulty in segregating both signal and masker. Note that informational masking has also been defined by Durlach and colleagues (2003) as reflecting a difficulty in attending to a relevant signal where there is uncertainty concerning the signal's identity.

5. While the studies discussed with respect to mismatch negativity (MMN) suggest enhanced musician responses to violations of sound sequence structure, we should also highlight that a variety of studies show enhancements at relatively earlier stages of auditory processing in musicians. Schneider et al. (2002) found enhanced early MEG component responses (N19m and P30m) in professional and amateur musicians compared to non-musicians (presumably reflecting contributions from auditory cortex generators). A variety of studies by Kraus and colleagues (e.g., Parbery-Clark et al., 2011; 2009a; Strait et al. 2012a; Skoe & Kraus, 2013) also suggest musician enhancement at relatively earlier auditory processing stages, based on auditory brainstem response indices.
6. The number of violinists and pianists who reported playing other instruments did differ [χ^2 (1, n = 48) = 6.15, p = 0.013, 22/24 violinists, 15/24 pianists]. Violinists typically reported that their second instrument (primarily piano; see table 2) was a requirement of their performance degree and was studied for less than half as long as violin. Similarly, almost all pianists had much more practice with piano than their second instrument (see table 3).
7. Since the present experimental battery was 3+ hours in duration, we were not able to assess full-scale IQ for each participant. However, all of our participants had completed formal education to high-school standard (i.e., UK A-level or equivalent). Moreover, all but two musicians (one violinist and one pianist) were currently enrolled in or had completed a performance degree; further, all non-musicians were enrolled in or had completed at least one third-level degree. We therefore matched our cohorts as closely as possible for extent of enrolment in formal education. One anonymous reviewer suggested that non-musicians might have more experience with formal education compared to musicians (two of our non-musicians were PhD students, five were studying for an MSc or MA, one had completed an MSc, and one had completed an MA). However, 11 of our violinists and 8 of our pianists were completing a performance MA further to their performance degree. Such qualifications demand academic study of technical aspects of music theory (e.g., counterpoint, chorale harmony, formal analysis) as well as study of subjects

such as musicology (in addition to rigorous technical training on their chosen instrument). An analysis of level of education across groups (status as an undergraduate or post-graduate student as a nominal dependent measure with two levels) showed no evidence of any significant differences across the cohorts [χ^2 (2, n = 72) = 0.46, p = 0.8]. Moreover, when including education level (undergraduate/postgraduate) as a predictor within regression models along with group, we found that education level did not account for any significant variance (all $p > 0.64$), whereas group reached significance in the same models as described in results. As such, we think it unlikely that a difference in extent of formal education could account for the lack of cohort differences across cognitive tasks shown here.

8. The practice at 3–4 years regressor was reduced to binary form since the considerable skew in the distribution of practice at that age (approximately half of the participants in each group had not practiced at 3–4 years) meant it was not appropriate as a continuous regressor. Similarly, the practice of 1+ hour per day at 7–8 years variable was treated as binary, since the relatively low (and skewed spread of) hours of practice time at this age made it unusable as a continuous regressor.
9. The MSI training subscale data we collected were largely unsuited to regression analyses, since the highly-trained musician cohorts had near ceiling values on the musical training subscale (see ST1); our non-musician data showed considerably lower scores (as expected, since by definition our non-musicians had minimal training). This led to a distribution that was approximately bimodal, and that captured similar variance as the binary group (expert/non-expert) regressor that we did use in our analyses. Nonetheless, we did analyze the non-musician cohort alone, using MSI training scores as a regressor for performance on the perceptual and cognitive tasks. Our analyses showed only one significant result: MSI scores accounted for variance in non-musicians' FM depth thresholds [$F(1, 19) = 4.5$, $p = 0.047$, adjusted $R^2 = 0.149$]. However, with FDR correction across the full set of models, this result did not reach significance; we would therefore suggest it is best interpreted cautiously.

10. It is worth noting that musicians' resilience to informational masking in the Oxenham et al. (2003) multi-tone masker paradigm may have been facilitated by their being able to attend to the unchanging frequency of the target – a possibility we are currently exploring.

Supplemental Methods

SM. 1 Materials

Auditory psychophysical thresholding was conducted using custom software (SHaPs; Department of Speech, Hearing and Phonetic Sciences, UCL), run using a HP Pavilion dv2000 laptop computer with Windows XP. The remaining tasks were presented on a MacBook Pro laptop computer (OS 10.7.3), using the Psychophysics Toolbox (version 3; Kleiner, Brainard & Pelli, 2007) running in Matlab (2010a; 32-bit). Auditory stimuli were presented through Sennheiser HD-380 Pro headphones, via ESI UGM 96 24-bit external sound card, connected to the HP laptop and MacBook Pro by USB. All sounds were presented at a comfortable level fixed for all participants. Visual psychophysical thresholding was conducted using a custom C language program (Tibber & Shepherd, 2006), running on a Mac G3 tower with OS 9.2, and Sony Trinitron 27” monitor. Pure tone audiometry was completed using an Otovation Otopod M2 portable audiometer, with Symphony audiometric software running in Windows 7 on a Dell Precision T3500 desktop computer.

SM. 2 Test-retest Reliability Analyses

Participants for test-retest experiments were recruited from local participant pools in two phases (see below). All participants (*N* = 46; mean age ± SD: 27.304 ± 9.097; range: 19-51 years; male: 13; female: 33) were right-handed by self-report and reported no history of auditory impairment or neurological insult. All had less than 5 years’ experience with any musical instrument and none had trained formally with an instrument or voice.

In the first phase, participants (*n* = 21; mean age: 28.4 ± 8.9 [SD]; range: 19-47 yrs; male: 5; female: 16) completed each of the psychophysical experiments (ramp onset time, AM depth, FM depth), in addition to a response inhibition (*n* = 17) or response switching (*n* = 4) version of the SAART, along with the tuning system perception task. Participants completed two tracks for each psychophysical thresholding task (fixed in the order AM [x2], FM [x2], ramp onset [x2]), along with the SAART and the tuning perception task. Once participants had completed each of these experiments, the same experiments were run a second time during the same session in the same order. One participant who completed the phase one test-retest battery performed a preference judgement version of the tuning task, rather than the tuning system accuracy task. Order of experimental task completion was counterbalanced across participants.

In the second phase, participants (*n* = 25; mean age: 26.4 ± 9.4 [SD]; range: 19-51 yrs; male: 8; female: 17) completed test-retest reliability assessment for the SIMON task, interleaved with a response switching version of the SAART (*n* = 16), or a preference judgement version of the tuning system task (*n* = 9). Two participants provided test-retest data for the response switching SAART and the tuning system preference task, but did not complete the SIMON task. As in phase 1, order of task completion was counterbalanced across participants.

Supplemental Results (SR)

SR. 1 Musicians' practice hours early-in-life and task performance

We asked if instrumental practice early in life would account for musicians' performance across all tasks. We used two separate binary predictors: 1) whether the participant had started practicing by 3–4 years (y/n), and 2) whether the participant had practiced one or more hours per day at 7-8 years (see 2.3.1). Early practice significantly predicted only a single outcome variable: violinists who began formal practice at 3–4 years were more likely than later-beginning violinists to choose just intonation when paired with the (slightly sharp) +7.5 cents chord [F (1, 19) = 5.31, p = 0.033, adjusted $R^2 = 0.177$; β = -0.16; 77% (SD=28%) of early-starting violinists chose just intonation versus 44% (SD=34%) of later-starting violinists]. There was no such significant effect in early-practicing pianists (β = 0.023, $p > 0.7$; test of difference between regression coefficients marginal – $z = 1.80$, $p = 0.07$, two-tailed, post-hoc). The same relationship – albeit marginally significant – was observed for violinists practicing 1 hour or more per day at 7–8 years [F (1, 19) = 3.80, $p = 0.066$, adjusted $R^2 = 0.123$; β = -0.16], but not pianists (β = -0.036, $p > 0.5$; difference between regression coefficients non-significant – $z = 1.17$, $p = 0.12$, two-tailed, post-hoc). Although weak, these effects suggest that those violinists who began practice earlier in life may have possessed a more finely-honed ability to discriminate their instrument-specific tuning system from a very subtle deviation from that system.

SR. 2 Test-retest Reliability Analyses

SR. 2.1 Auditory Psychophysics

Stimuli were identical to those described in methods (see 2.3.4.1). As in methods (2.3.4.2), thresholds for each run were measured as the mean of the final 3 or 4 reversals. For each psychophysical experiment, Spearman's correlations over all possible pairs of runs showed moderate to high test-retest reliability for thresholds [range (ρ): 0.56 to 0.9, all $p < 0.05$; see ST 12]. However, participant n's differed across tasks, since of the full sample ($n = 21$), not all participants completed all runs adequately (precluding correlations across all possible run pairs for every participant) [onset rise time: $n = 10$ (11 excluded; 5 tracked erroneously on one run each, 5 tracked erroneously on two runs each, and 1 failed to track on all runs); AM depth $n = 15$ (6 excluded; five tracked erroneously on one run each, one tracked erroneously on two runs); FM depth: $n = 19$ (2 excluded; erroneous tracking on one run each)]. Note however that despite the small *n*'s for correlational analyses, ρ coefficients were relatively high; further, inspection of scatter plots suggested tight clustering of points with strong positive linear relationships for each experiment.

SR 2.2 Tuning Perception Task

With the exception of one participant, all participants completing the task during phase one ($n = 20$) indicated the chord of the pair they perceived as most in tune on each trial. The remaining participants ($n = 10$) were required to choose the chord of the two that they preferred on each trial. Stimuli and procedure for both tasks were identical to that described in 2.3.6.2.

Test-retest correlations (Spearman's) for proportion values from the phase 1 tuning system judgement task are presented in ST 13. Moderate test-retest correlations were observed for just intonation paired with the -15, + 7.5, and -7.5 cents tuning deviances; a modest correlation was also noted for the equal tempered (+15 cents) chord paired with the +7.5 cents chord [range (ρ): 0.45-0.51, all $p < 0.05$). Non-significant correlations were found for just intonation paired with equal temperament ($\rho = 0.3$, $p = 0.2$), and equal temperament paired with the largest tuning deviance (-15 cents) ($\rho = 0.17$, $p = 0.5$).

To rule out the possibility that participants did not understand the task instructions, we ran a test-retest condition where participants indicated their preferred chord of each pair on each trial (see ST 14). However, test-retest correlations were only improved for the just intonation vs. equal tempered pair ($\rho = 0.79$, $p = 0.006$) and the just intonation vs. -7.5 cents pair ($\rho = 0.68$, $p = 0.03$). For the remaining pairs, test-retest correlations were non-significant (see ST 14). This suggested that non-musicians were not necessarily more consistent in their responses when making preference rather than 'in-tune' judgements for the chord pairs.

SR 2.3 SAART

Two versions of the paradigm were evaluated: a response switching version (n = 20) (used in the present study) and a response inhibition version (n = 17). Results for the latter were similar to the response switching version and will be reported elsewhere. The response switching version used the identical stimuli and procedure as described in 2.3.8.

Participants' mean and SDs of RTs were calculated for correct responses to target and non-target sounds over the first 81 trials; accuracies to target sounds were also analysed. Two participants were excluded from test-retest analyses (one had mean non-target RTs across both runs > 3 SDs above cohort mean; one responded correctly to < 80% of non-targets; analysis n = 18). Test-retest correlations showed high reliability of target mean RTs ($\rho = 0.81$, $p < 0.0001$) and non-target mean RTs ($\rho = 0.69$, $p = 0.0014$); however SDs of RTs to targets ($\rho = -0.13$, $p > 0.62$) and non-targets ($\rho = 0.34$, $p > 0.17$) did not reach significance when each was correlated across runs 1 and 2. Target accuracies were significantly positively correlated across runs ($\rho = 0.572$ $p = 0.013$).

SR 2.4 SIMON

SIMON stimuli and procedure were similar to above (2.3.5), but participants only completed the game (i.e., without the pre-game listening phase). Participants (n = 23) first completed one of two pseudrorandom orders of sequences (order 1 or 2), and then completed the other order at retest. Mean sequence lengths per testing run were calculated (i.e., averaging over all 10 sequences in each run). ANOVA analysis with factors of sequence run and group (i.e., order 1 first vs. order 2 first) showed a significant main effect of run $F(1, 21) = 9.271$, $p = 0.006$, $\eta_p^2 = 0.306$, but no significant effect of group nor any significant interaction (both $F < 1.58$, $p > 0.22$). Means sequence lengths showed a small decline between runs 1 and 2 (mean difference ± SD: -0.16 ± 1.54) perhaps due to a fatigue effect over runs. A multiple regression model with test-retest run and the difference in sequence length between runs as predictors showed run 1 sequence length was a significant predictor of run 2 sequence length, F (2, 20) = 4.7, $p = 0.021$, adj. $R^2 = 0.252$ [run 1 sequence length: $t(21) = 3.02$, $p = 0.007$; run1-run2: $t(21) = 0.03$ $p > 0.97$]. Mean sequence lengths for runs 1 and 2 were also significantly positively correlated ($\rho = 0.59$, $p = 0.003$), suggesting good test-retest reliability.

Supplemental table 1: MSI musical training subscale means and SDs

	Non-musicians	Violinists	Pianists
Mean	16.58	57.42	54.54
SD	7.71	2.38	2.45

Supplemental table 2: Kruskal-Wallis and post-hoc Wilcoxon Signed Rank Test comparisons for MSI musical training subscale scores

Model	Test statistic
Kruskal–Wallis χ^2 (2, *n* = 70)	24.24 ***
Post-hoc (WSRT)	
NM vs. V	$z = 5.95$ ***
NM vs. P	$z = 5.95$ ***
V vs. P	$z = 3.87$ ***

*** $p < 0.0001$

Note: the difference between violinists and pianists was driven by violinists' increased weekly hours spent in orchestras (violinists: 6.9 ± 5.8 [SD]; pianists: 0.6 ± 1.4 [SD]; $z = 4.89$, $p < 0.0001$, Cohen's $d = 1.5$) and small ensembles (violinists: 6.6 ± 5.2 [SD]; pianists: 3.7 ± 5.4 [SD]; $z = 2.78$, $p = 0.0054$, Cohen's $d = 0.5$).

Supplemental table 3: MANOVA analyses of groups' rise time psychophysical task performance (*n* = 36), for samples drawn at random from violinist and pianist groups, matched to non-musicians' *n*

Model	Wilk's λ	df	F	*p*	η_p^2
Sample 1					
Run	0.245	(3, 31)	31.83	< 0.0001	0.755
Group		(1, 33)	12.97	< 0.0001	0.44
Run x Group	0.485	(6, 62)	4.5	0.001	0.303
Sample 2					
Run	0.293	(3, 31)	24.92	< 0.0001	0.707
Group		(1, 33)	10.64	< 0.0001	0.392
Run x Group	0.476	(6, 62)	4.651	0.001	0.344
Sample 3					
Run	0.254	(3, 31)	30.4	< 0.0001	0.746
Group		(1, 33)	12.92	< 0.0001	0.439
Run x Group	0.485	(6, 62)	4.51	0.001	0.304
Sample 4					
Run	0.279	(3, 31)	26.72	< 0.0001	0.721
Group		(1, 33)	16.99	< 0.0001	0.507
Run x Group	0.397	(6, 62)	6.07	< 0.0001	0.37
Sample 5					
Run	0.287	(3, 31)	25.61	< 0.0001	0.713
Group		(1, 33)	12.9	< 0.0001	0.439
Run x Group	0.4	(6, 62)	5.99	< 0.0001	0.367
Sample 6					
Run	0.295	(3, 31)	24.73	< 0.0001	0.705
Group		(1, 33)	11.03	< 0.0001	0.401
Run x Group	0.469	(6, 62)	4.753	< 0.0001	0.315

Supplemental table 4: MANOVA analyses of groups' AM depth psychophysical task performance (*n* = 36), for samples drawn at random from pianist group, matched to violinists and non-musicians' *n*

Model	Wilk's λ	df	F	*p*	η_p^2
Sample 1					
Run	0.47	(3, 31)	11.67	< 0.0001	0.53
Group		(1, 33)	5.58	0.008	0.253
Run x Group	0.824	(6, 62)	1.1	0.4	0.092
Sample 2					
Run	0.498	(3, 31)	10.41	< 0.0001	0.502
Group		(1, 33)	6.21	0.005	0.273
Run x Group	0.774	(6, 62)	1.4	0.2	0.12
Sample 3					
Run	0.493	(3, 31)	10.63	< 0.0001	0.507
Group		(1, 33)	7.53	0.002	0.313
Run x Group	0.658	(6, 62)	2.41	0.037	0.189
Sample 4					
Run	0.477	(3, 31)	11.32	< 0.0001	0.523
Group		(1, 33)	7.18	0.003	0.303
Run x Group	0.813	(6, 62)	1.1	0.36	0.098
Sample 5					
Run	0.503	(3, 31)	10.22	< 0.0001	0.497
Group		(1, 33)	5.46	0.009	0.249
Run x Group	0.82	(6, 62)	1.1	0.39	0.094
Sample 6					
Run	0.5	(3, 31)	10.36	< 0.0001	0.5
Group		(1, 33)	7.79	0.002	0.321
Run x Group	0.784	(6, 62)	1.3	0.26	0.114

Supplemental table 5: MANOVA analyses of groups' FM depth psychophysical task performance (*n* = 33), for samples drawn at random from violinist and pianist groups, matched to non-musicians' *n*

Model	Wilk's λ	df	F	*p*	η_p^2
Sample 1					
Run	0.729	(3, 28)	3.47	0.029	0.271
Group		(1, 33)	2.35	0.112	0.136
Run x Group	0.908	(6, 56)	0.459	0.8	0.047
Sample 2					
Run	0.765	(3, 28)	2.86	0.055	0.235
Group		(1, 33)	2.92	0.069	0.163
Run x Group	0.884	(6, 56)	0.59	0.7	0.06
Sample 3					
Run	0.807	(3, 28)	2.24	0.106	0.193
Group		(1, 33)	3.93	0.03	0.208
Run x Group	0.79	(6, 56)	1.17	0.3	0.111
Sample 4					
Run	0.789	(3, 28)	2.5	0.08	0.211
Group		(1, 33)	2.48	0.1	0.142
Run x Group	0.934	(6, 56)	0.3	0.9	0.34
Sample 5					
Run	0.852	(3, 28)	1.62	0.2	0.148
Group		(1, 33)	3.06	0.062	0.17
Run x Group	0.864	(6, 56)	0.7	0.6	0.071
Sample 6					
Run	0.811	(3, 28)	2.17	0.1	0.189
Group		(1, 33)	1.9	0.17	0.113
Run x Group	0.898	(6, 56)	0.516	0.79	0.052

Supplemental table 6: Post-hoc pairwise comparisons across rise time, AM depth and FM depth psychophysical thresholding tracks for participants completing 4 runs total (all comparisons false discovery rate-corrected; $\alpha = 0.05$)

Run	1st vs. 2nd	1st vs. 3rd	1st vs. 4th	2nd vs. 3rd	2nd vs. 4th	3rd vs. 4th
Rise time						
Pianists ($n = 16$)	$t(15) = 2.95$ * Cohen's $d = 1.5$	$t(15) = 4.30$ * Cohen's $d = 2.2$	$t(15) = 5.47$ * Cohen's $d = 2.8$	$t(15) = 1.87$ *n.s.*	$t(15) = 2.66$ * Cohen's $d = 1.4$	$t(15) = 1.61$, *n.s.*
Non-musicians ($n = 12$)	$t(11) = 1.52$, *n.s.*	$t(11) = 3.92$ * Cohen's $d = 2.0$	$t(11) = 6.08$ * Cohen's $d = 2.2$	$t(11) = 2.9$ * Cohen's $d = 3.1$	$t(11) = 4.86$ * Cohen's $d = 2.5$	$t(11) = 1.76$, *n.s.*
AM						
All subs ($n = 40$)	$t(39) = 1.87$, *n.s.*	$t(39) = 4.95$ * Cohen's $d = 2.6$	$t(39) = 5.57$ * Cohen's $d = 2.9$	$t(39) = 2.81$ * Cohen's $d = 1.5$	$t(39) = 3.69$ * Cohen's $d = 1.9$	$t(39) = 0.97$, *n.s.*
FM						
All subs ($n = 42$)	$t(41) = 0.66$, *n.s.*	$t(41) = 0.17$, *n.s.*	$t(41) = 2.59$ * Cohen's $d = 1.3$	$t(41) = 0.54$, *n.s.*	$t(41) = 2.21$, *n.s.*	$t(41) = 2.05$, *n.s.*

* $p < 0.05$ (FDR-corrected); *n.s.* - non-significant

Supplemental table 7: One-sample Wilcoxon signed rank tests of difference of group mean from chance, per tuning pair (false discovery rate-corrected [α = 0.05] per tuning pair)

	Just vs. Equal (+15)	Just vs. -15	Just vs. +7.5	Just vs. -7.5	Equal (+15) vs. -15	Equal vs. +7.5
NM	- 73.5 *	- 59.5, *n.s.*	- 80 *	- 93.5 *	26.5, *n.s.*	- 5, *n.s.*
P	- 6.5, *n.s.*	101 *	- 66.5 *	- 24.5, *n.s.*	137 *	-109 *
V	132 *	138 *	45.5, *n.s.*	118.5 *	138 *	-120.5 *

* $p < 0.05$ (FDR-corrected); *n.s.* - non-significant

Supplemental table 8: Non-parametric correlations across auditory and visual psychophysical tasks (false discovery rate corrected; * $p < 0.05$)

Pair	Spearman's ρ	FDR-corrected p
AM depth & FM depth	0.411	0.02 *
AM depth & Rise time	0.323	0.04 *
FM depth & Rise time	0.415	0.02 *
AM & Visual (increments)	-0.1	0.48
FM & Visual (increments)	-0.19	0.21
Rise time & Visual (increments)	-0.23	0.14

Supplemental table 9: Loadings from principal components analysis (PCA) across all tasks, with additional loadings from tuning system and EnvASA PCA

Full PCA		Tuning system PCA + *			EnvASA PCA + **	
Variable	PC 1	Variable	PC 1	PC 2	Variable	PC 1
AM Depth	0.5207	Just vs. Equal	0.8282	0.4103	Single-Low (-6 & -3 dB)	0.1654
FM Depth	0.5841	Just vs. -15	0.5699	0.6849	Single-High (0 & +3 dB)	0.4452
Rises time	0.6110	Just vs. +7.5	0.9094	-0.0122	Dual-Low (-6 & -3 dB)	1.0
SIMON	-0.4057	Just vs. -7.5	0.8806	0.3546	Dual-High (0 & +3 dB)	0.4640
SAART SDs	0.5844	Equal vs. -15	0.0096	0.7512		
SAART RTs	0.5910	Equal vs. +7.5	-0.5248	-0.6663		
Visual (increments)	-0.3870					
Tuning Component 1	-0.2887					
Tuning Component 2	0.2234					
EnvASA Component	-0.1385					

+ Data for the tuning perception and EnvASA tasks were first reduced with separate PCAs; only significant components were retained, verified by the turn point in the scree plot. Varimax rotation of axes was applied within the tuning perception analysis; the single EnvASA component was not rotated.

* Tuning system PCA components both significant at $p < 0.0001$, accounting for 66.91% (PC1) and 19.17%(PC2) variance, respectively.

** EnvASA PCA component significant at $p < 0.001$, accounting for 44.7% of variance; note collapsed levels of SNR for variables entered into EnvASA PCA; low: -6 dB & -3 dB SNR collapsed; high: 0 dB & +3 dB SNR collapsed.

pc: principal component.

Supplemental table 10: Non-parametric correlations between tuning task pairs, AM depth and FM depth thresholds, across groups (false discovery rate-corrected; * $p < 0.05$)

	Pianists		Violinists		Non-musicians	
Pair	ρ	FDR-corrected *p*	ρ	FDR-corrected *p*	ρ	FDR-corrected *p*
AM depth & Just vs. Equal	-0.361	0.19	-0.28	0.32	-0.04	1.0
AM depth & Just vs. -15	-0.43	0.14	-0.02	0.94	-0.03	1.0
AM depth & Just vs. +7.5	-0.12	0.61	-0.42	0.16	-0.06	1.0
AM depth & Just vs. -7.5	-0.228	0.32	-0.36	0.2	0.002	1.0
AM depth & Equal vs. -15	-0.321	0.21	0.13	0.68	-0.10	1.0
AM depth & Equal vs. +7.5	0.422	0.14	0.43	0.16	0.35	0.68
FM depth & Just vs. Equal	-0.411	0.08	-0.19	0.47	-0.10	0.8
FM depth & Just vs. -15	-0.743	0.01 *	0.06	0.8	-0.21	0.63
FM depth & Just vs. +7.5	-0.391	0.08	-0.33	0.25	-0.22	0.63
FM depth & Just vs. -7.5	-0.494	0.05	-0.19	0.47	0.02	0.93
FM depth & Equal vs. -15	-0.428	0.08	0.43	0.16	-0.19	0.63
FM depth & Equal vs. +7.5	0.30	0.16	0.41	0.16	0.21	0.63

Supplemental table 11: Means and standard deviations of pure tone audiometric thresholds (dB HL) for each group across frequencies (left and right ears collapsed)

	250 Hz	500 Hz	1 kHz	2 kHz	3 kHz	4 kHz	6 kHz	8 kHz
NM	12.92 (7.93)	10.73 (8.55)	2.74 (5.32)	2.81 (6.52)	2.29 (6.16)	-0.10 (6.01)	16.80 (8.34)	12.60 (9.68)
P	12.92 (6.82)	10.0 (7.37)	4.69 (5.38)	2.50 (6.12)	2.40 (5.24)	0.31 (7.0)	14.06 (8.43)	6.25 (7.34)
V	14.24 (6.63)	10.43 (7.82)	3.70 (6.30)	2.47 (6.84)	2.28 (6.21)	1.20 (6.30)	17.45 (7.82)	7.72 (6.90)

Supplemental table 12: Test-retest reliability correlations for all possible pairs of runs, for each psychophysical thresholding experiment (columns indicate run numbers)

	FM1 & FM2	FM1 & FM3	FM1 & FM4	FM2 & FM3	FM2 & FM4	FM3 & FM4
Spearman ($n = 19$)	0.767, $p < 0.0001$ **	0.686, $p = 0.0012$ **	0.558, $p = 0.013$ *	0.687, $p = 0.0012$ **	0.555, $p = 0.014$ *	0.745, $p = 0.0003$ **

	AM1 & AM2	AM1 & AM3	AM1 & AM4	AM2 & AM3	AM2 & AM4	AM3 & AM4
Spearman ($n = 15$)	0.853, $p < 0.0001$ **	0.630, $p = 0.012$ *	0.763, $p = 0.001$ **	0.699, $p = 0.004$ **	0.699, $p = 0.004$ **	0.740, $p = 0.002$ **

	Rise1 & Rise2	Rise1 & Rise3	Rise1 & Rise4	Rise2 & Rise3	Rise2 & Rise4	Rise3 & Rise4
Spearman ($n = 10$)	0.745, $p = 0.013$ *	0.839, $p = 0.002$ **	0.782, $p = 0.008$ **	0.681, $p = 0.03$ *	0.903, $p < 0.0001$ **	0.742, $p = 0.014$ *

* $p < 0.05$ ** $p < 0.01$

Supplemental table 13: Spearman correlation coefficients for test-retest reliability analyses of tuning perception paradigm (*n* = 20) across testing runs 1 and 2 (for each possible chord pairing; 'in-tune' judgements).

	Just vs. +15	Just vs. -15	Just vs. +7.5	Just vs. -7.5	+15 vs. -15	+15 vs. +7.5
Spearman	0.302 $p = 0.2$ n.s.	0.454 $p = 0.044$ *	0.462 $p = 0.040$ *	0.518 $p = 0.019$ *	0.167 $p = 0.48$ n.s.	0.485 $p = 0.03$ *

* $p < 0.05$

Supplemental table 14: Spearman correlation coefficients for test-retest reliability analyses of tuning perception paradigm (*n* = 10) across testing runs 1 and 2 (for each possible chord pairing; preference judgements).

	Just vs. +15	Just vs. -15	Just vs. +7.5	Just vs. -7.5	+15 vs. -15	+15 vs. +7.5
Spearman	0.793 *p* = 0.006 **	0.498 *p* = 0.14 n.s.	0.380 *p* = 0.28 n.s.	0.681 *p* = 0.03 *	0.131 *p* = 0.72 n.s.	-0.429 *p* = 0.22 n.s.

* $p < 0.05$ ** $p < 0.01$

Supplemental table 15: Results for equivalence tests (test value ± 1 SD)

Task	Test	Test value	Diff. in means	Std. err. diff	Upper threshold	Lower threshold
SIMON	V vs. NM	± 2.4	0.58	0.72	$t = -2.51$, $p = 0.008$ *	$t = 4.12$, $p < 0.0001$ *
	P vs. NM	± 2.4	1.38	0.67	$t = -1.53$, $p = 0.067$	$t = 5.67$, $p < 0.0001$ *
	V vs. P	± 2.4	-0.8	0.68	$t = -4.73$, $p < 0.0001$ *	$t = 2.37$, $p = 0.011$ *
SAART (RT SDs)	V vs. NM	± 0.146	-0.06	0.04	$t = -4.8$, $p < 0.0001$ *	$t = 1.92$, $p = 0.03$ *
	P vs. NM	± 0.146	-0.1	0.04	$t = -5.84$, $p < 0.0001$ *	$t = 1.0$, $p = 0.16$
	V vs. P	± 0.146	0.04	0.04	$t = -2.64$, $p = 0.006$ *	$t = 4.67$, $p < 0.0001$ *
EnvASA						
Single-low	V vs. NM	± 8.2	-0.52	1.93	$t = -4.52$, $p < 0.0001$ *	$t = 3.98$, $p = 0.0001$ *
	P vs. NM	± 8.2	-3.02	2.56	$t = -4.39$, $p < 0.0001$ *	$t = 2.03$, $p = 0.02$ *
	V vs. P	± 8.2	2.5	2.53	$t = -2.25$, $p = 0.02$ *	$t = 4.23$, $p < 0.0001$ *
Dual-low	V vs. NM	± 9.4	3.12	2.39	$t = -2.63$, $p = 0.006$ *	$t = 5.25$, $p < 0.0001$ *
	P vs. NM	± 9.4	-0.1	3.06	$t = -3.12$, $p = 0.002$ *	$t = 3.04$, $p = 0.002$ *
	V vs. P	± 9.4	3.23	2.67	$t = -2.31$, $p = 0.01$ *	$t = 4.73$, $p < 0.0001$ *
Single-high – Dual-high	V vs. NM	± 6.2	3.61	1.75	$t = -1.48$, $p = 0.073$	$t = 5.61$, $p < 0.0001$ *
	P vs. NM	± 6.2	1.77	1.78	$t = -2.48$, $p = 0.008$ *	$t = 4.47$, $p < 0.0001$ *
	V vs. P	± 6.2	1.84	1.79	$t = -2.44$, $p = 0.009$ *	$t = 4.5$, $p < 0.0001$ *
Single-low – Dual-low	V vs. NM	± 9.7	-3.65	2.95	$t = -4.53$, $p < 0.0001$ *	$t = 2.05$, $p = 0.02$ *
	P vs. NM	± 9.7	-2.92	2.76	$t = -4.56$, $p < 0.0001$ *	$t = 2.45$, $p = 0.009$ *
	V vs. P	± 9.7	-0.73	2.64	$t = -3.95$, $p = 0.0001$ *	$t = 3.4$, $p = 0.0007$ *

Note: EnvASA Single-low and Dual-low correspond with low SNR (-3 dB, -6 dB) conditions collapsed for the single and dual background conditions, respectively. Single-high – Dual-high and Single-low – Dual low reflect the difference between background conditions for high (0 dB, +3 dB) and low (-3 dB, -6 dB) SNR conditions.

Supplemental table 16: Results for equivalence tests (test value ± 0.5 SD)

Task	Test	Test value	Diff. in means	Std. err. diff	Upper threshold	Lower threshold
SIMON	V vs. NM	± 1.2	0.58	0.72	$t = -0.85$, $p = 0.2$	$t = 2.46$, $p = 0.009$ *
	P vs. NM	± 1.2	1.38	0.67	$t = 0.27$, $p = 0.6$	$t = 3.87$, $p = 0.002$ *
	V vs. P	± 1.2	-0.8	0.68	$t = -2.96$, $p = 0.002$ *	$t = 0.6$, $p = 0.3$
SAART (RT SDs)	V vs. NM	± 0.073	-0.06	0.04	$t = -3.12$, $p = 0.0016$ *	$t = 0.24$, $p = 0.4$
	P vs. NM	± 0.073	-0.1	0.04	$t = -4.13$, $p < 0.0001$ *	$t = -0.7$, $p = 0.8$
	V vs. P	± 0.073	0.04	0.04	$t = -0.81$, $p = 0.2$	$t = 2.84$, $p = 0.003$
EnvASA						
Single-low	V vs. NM	± 4.1	-0.52	1.93	$t = -2.39$, $p = 0.01$ *	$t = 1.85$, $p = 0.035$ *
	P vs. NM	± 4.1	-3.02	2.56	$t = -2.79$, $p = 0.004$ *	$t = 0.42$, $p = 0.3$
	V vs. P	± 4.1	2.5	2.53	$t = -0.63$, $p = 0.3$	$t = 2.61$, $p = 0.006$ *
Dual-low	V vs. NM	± 4.7	3.12	2.39	$t = -0.66$, $p = 0.3$	$t = 3.28$, $p = 0.001$ *
	P vs. NM	± 4.7	-0.1	3.06	$t = -1.57$, $p = 0.062$	$t = 1.5$, $p = 0.07$
	V vs. P	± 4.7	3.23	2.67	$t = -0.55$, $p = 0.3$	$t = 2.97$, $p = 0.002$ *
Single-high – Dual-high	V vs. NM	± 3.1	3.61	1.75	$t = 0.29$, $p = 0.6$	$t = 3.84$, $p = 0.002$ *
	P vs. NM	± 3.1	1.77	1.78	$t = -0.75$, $p = 0.2$	$t = 2.73$, $p = 0.004$ *
	V vs. P	± 3.1	1.84	1.79	$t = -0.71$, $p = 0.2$	$t = 2.77$, $p = 0.004$ *
Single-low – Dual-low	V vs. NM	± 4.85	-3.65	2.95	$t = -2.88$, $p = 0.0003$ *	$t = 0.41$, $p = 0.3$
	P vs. NM	± 4.85	-2.92	2.76	$t = -2.81$, $p = 0.004$ *	$t = 0.7$, $p = 0.2$
	V vs. P	± 4.85	-0.73	2.64	$t = -2.11$, $p = 0.02$ *	$t = 1.56$, $p = 0.063$

Note: EnvASA Single-low and Dual-low correspond with low SNR (-3 dB, -6 dB) conditions collapsed for the single and dual background conditions, respectively. Single-high – Dual-high and Single-low – Dual low reflect the difference between background conditions for high (0 dB, +3 dB) and low (-3 dB, -6 dB) SNR conditions.

Figure 1: Auditory and visual psychophysical thresholds across groups. (a) onset rise time thresholds (ms); (b) AM depth thresholds (dB); (c) FM depth thresholds (cents); (d) increment colour hue thresholds (Macleod-Boynton colour space co-ordinates); NM - non-musicians (circles); P - pianists (crosses); V - violinists (diamonds); large diamonds display mean as middle horizontal line, and upper and lower bounds of 95% CI as uppermost and lowermost diamond tips, respectively; scatter plots display thresholds for individual subjects; note logarithmic axis for onset rise time thresholds (linear axes for others); group *n*s differ across tasks - see Method for description.

Figure 2: Change in group mean auditory psychophysical thresholds across tracking runs for each task; dashed lines with circles - non-musicians; dotted lines with crosses - pianists; solid lines with diamonds - violinists; error bars denote ± 1 std. error of mean; traces in (a) highlight significant post-hoc pairwise comparisons, for non-musician and pianist groups (see respective dashed and dotted traces); traces in (b) highlight significant post-hoc pairwise comparisons collapsed across groups; trace in (c) highlights significant post-hoc pairwise comparison collapsed across groups; * $p < 0.05$ (false discovery rate-corrected), for all post-hoc tests; note logarithmic axis for onset rise time thresholds (linear axes for others); group *n*s differ across panels - see Method for description.

Figure 3: In-tune choices for just and equal tempered tuning systems when paired with tuning deviances, across groups; leftmost panels display proportion of trials where chords adhering to just intonation were chosen when paired with chords deviating from just intonation (values greater than 0.5 indicate just intonation chosen; less than 0.5 indicate deviating chord chosen); rightmost panels display proportion of trials where chords adhering to equal temperament were chosen when paired with chords deviating from equal temperament (values greater than 0.5 indicate equal temperament chosen; less than 0.5 indicate deviating chord chosen); NM - non-musicians (circles); P - pianists (crosses); V - violinists (diamonds); large diamonds display mean as middle horizontal line, and upper and lower bounds of 95% CI as uppermost and lowermost diamond tips, respectively; * markers display difference of group means from chance (one-sample Wilcoxon signed rank tests), * $p < 0.05$ (false discovery rate-corrected); n.s. - non-significant.

Figure 4: group and individual performance on sustained auditory attention task (SAART); NM - non-musicians (circles); P - pianists (crosses); V - violinists (diamonds); leftmost panel displays standard deviations of reaction times to non-target sounds (seconds); middle panels display reaction times to target and non-target sounds (seconds); rightmost panels display response accuracies to target and non-target sounds (raw counts); large diamonds display means as middle horizontal line, and upper and lower bounds of 95% CI as uppermost and lowermost diamond tips, respectively; target RTs and non-target RTs differed significantly ($p = 0.002$); targets: 449.8 ms ± 0.08 (SD); non-targets: 428.8 ms ± 0.09 (SD); interaction with group n.s.

Figure 5: group and individual mean sequence length performance on SIMON sequencing task; NM - non-musicians (circles); P - pianists (crosses); V - violinists (diamonds); large diamonds display means as middle horizontal line, and upper and lower bounds of 95% CI as uppermost and lowermost diamond tips, respectively.

Supplemental Figure 1: group and individual performance on EnvASA task; upper row displays percentage of trials correct for congruent and incongruent targets for single and dual background conditions, across groups; lower panels display percentage of trials correct across levels of signal to noise (SNR), for single and dual background conditions, across groups; large diamonds display means as middle horizontal line, and upper and lower bounds of 95% CI as uppermost and lowermost diamond tips, respectively.