



## ORBIT - Online Repository of Birkbeck Institutional Theses

---

Enabling Open Access to Birkbeck's Research Degree output

### Audition, learning and experience: expertise through development

<https://eprints.bbk.ac.uk/id/eprint/40107/>

Version: Full Version

**Citation: Carey, Daniel (2015) Audition, learning and experience: expertise through development. [Thesis] (Unpublished)**

© 2020 The Author(s)

---

All material available through ORBIT is protected by intellectual property law, including copyright law.

Any use made of the contents should comply with the relevant law.

---

[Deposit Guide](#)  
Contact: [email](#)

# Audition, Learning and Experience: Expertise Through Development

Dr. Daniel Carey

Thesis submitted for the degree of Doctor of Philosophy (Ph.D.)  
in Psychology

Birkbeck College, University of London

September 2014

Centre for Brain and Cognitive Development,  
32 Torrington Sq.,  
Birkbeck College,  
London,  
WC1E 7HX



**Signed Declaration**

I, Daniel Carey, confirm that except where due acknowledgement has been made to other sources, the work presented in this thesis is my own.

---

Daniel Carey

## **Abstract**

Our experience with the auditory world can shape and modify perceptual, cognitive and neural processes with respect to audition. Such experience can occur over multiple timescales, and can vary in its specificity and intensity. In order to understand how auditory perceptual, cognitive and neural processes develop, it is important to explore the different means through which experience can influence audition. This thesis aims to address these issues. Using an expertise framework, we explore how the auditory environment and ontogenetic factors can shape and guide perceptual, cognitive and neural processes through long- and short-term profiles of experience. In early chapters, we use expertly-trained musicians as a model for long-term experience accrued under specific auditory constraints. We find that expertise on a particular instrument (violin versus piano) yields training-specific auditory perceptual advantages in a musical context, as well as improvements to 'low-level' auditory acuity (versus non-musicians); yet we find limited generalisation of expertise to cognitive tasks that require some of the skills that musicians hone. In a subsequent chapter, we find that expert violinists (versus non-musicians) show subtle increases in quantitative MR proxies for cortical myelin at left auditory core. In latter chapters, we explore short-term sound learning. We ask whether listeners can learn combinations of auditory cues within an active visuo-spatial task, and whether development can mediate learning of auditory cue combinations or costs due to cue contingency violations. We show that auditory cue combinations can be learned within periods of minutes. However, we find wide variation in cue learning success across all experiments, with no differences in overall cue combination learning between children and adults. These experiments help to further understanding of auditory expertise, learning, development and plasticity, within an experience-based framework.

## Table of Contents

Signed Declaration.....	2
Abstract.....	3
Table of Contents.....	4
List of Tables.....	12
List of Figures.....	15
Acknowledgements.....	17
<u>Chapter 1</u> : General Introduction.....	18
<b>1.1</b> Introduction .....	19
<b>1.2</b> Perceptual and neural plasticity.....	20
1.2.1 Auditory perception and plasticity: ontogeny.....	21
1.2.2 Auditory perception and plasticity: environment.....	26
1.2.3 Plasticity, ontogeny and environment: a framework.....	28
<b>1.3</b> Long-term expertise and plasticity.....	30
1.3.1 Long-term expertise.....	30
1.3.2 Long-term expertise: structural plasticity.....	31
1.3.3 Long-term expertise: functional plasticity.....	32
1.3.4 Long-term expertise: summary.....	34
<b>1.4</b> Short-term expertise and plasticity.....	35
<b>1.5</b> Learning, generalisation and plasticity.....	38
<b>1.6</b> Cortical plasticity and myeloarchitecture.....	42
1.6.1 Myelination and development.....	42
1.6.2 Auditory cortex: processing hierarchy.....	45
1.6.3 Location and definition of auditory core.....	46
1.6.4 Implications for studies of expertise.....	49
<b>1.7</b> Thesis overview.....	50
Chapter 1 footnotes.....	53
<u>Chapter 2</u> : Generality and Specificity in the Effects of Musical Expertise on Perception and Cognition.....	55
<b>2.1</b> Introduction .....	56
2.1.1 Musicianship and auditory perception.....	58

2.1.1.1 Instrument- and musical-genre-specific effects on auditory perception.....	58
2.1.1.2 Contextual effects on musicians' perception of 'low-level' acoustic parameters.....	60
2.1.2 Musicianship, attention, and cognition.....	62
2.1.2.1 Auditory attention, and influence on perception.....	63
2.1.2.2 Auditory scene analysis.....	64
2.1.2.3 Sequence perception and reproduction.....	66
2.1.3 The present study.....	67
<b>2.2 Method.....</b>	<b>69</b>
2.2.1 Participants.....	69
2.2.1.1 Musicians.....	69
2.2.1.2 Non-musicians.....	71
2.2.2 Materials.....	73
2.2.3 Test-retest reliability.....	73
2.2.4 Procedure.....	75
2.2.4.1 Practice history questionnaire.....	75
2.2.4.2 Goldsmith's Musical Sophistication Index – Musical Training Sub-scale.....	76
2.2.4.3 Absolute pitch assessment.....	76
2.2.4.4 Auditory psychophysical tasks.....	77
2.2.4.4.1 Stimuli.....	77
2.2.4.4.2 Auditory psychophysics procedure.....	78
2.2.4.4.3 Test-retest reliability.....	81
2.2.4.5 Sequence reproduction task (SIMON).....	81
2.2.4.5.1 Stimuli.....	81
2.2.4.5.2 SIMON procedure.....	82
2.2.4.5.3. Test-retest reliability.....	83
2.2.4.6 Tuning system perception task.....	84
2.2.4.6.1 Stimuli.....	84
2.2.4.6.2 Tuning task procedure.....	86
2.2.4.6.3. Test-retest reliability.....	87
2.2.4.7 Environmental auditory scene analysis (EnvASA) task.....	88

2.2.4.8 Sustained auditory attention to response task (SAART).....	88
2.2.4.8.1 Stimuli.....	88
2.2.4.8.2 SAART procedure.....	88
2.2.4.8.3 Test-retest reliability.....	89
2.2.4.9 Pure tone audiometry.....	91
2.2.4.10 Visual psychophysical thresholding.....	91
2.2.5 Data analyses.....	92
<b>2.3 Results.....</b>	<b>93</b>
2.3.1 Auditory Psychophysical Thresholds.....	93
2.3.1.1 Rise time.....	93
2.3.1.2 AM depth.....	93
2.3.1.3 FM depth.....	93
2.3.1.4 Rise time.....	95
2.3.1.5 AM depth.....	95
2.3.1.6 FM depth.....	96
2.3.2 Visual psychophysical thresholds.....	98
2.3.3 Tuning system perception.....	98
2.3.4 SAART.....	101
2.3.5 SIMON.....	103
2.3.6 EnvASA.....	104
2.3.7 Cross task analyses.....	105
2.3.7.1 Psychophysical tasks, SAART, SIMON & EnvASA.....	106
2.3.7.2 Tuning system perception, FM depth and AM depth.....	108
2.3.8 Musicians' practice hours early-in-life and task performance.....	110
<b>2.4 General discussion.....</b>	<b>112</b>
2.4.1 Overview of results.....	112
2.4.2 Basic psychoacoustic measures.....	113
2.4.3 Contextual effects on experts' auditory perception.....	115
2.4.4 Sustained attention and perceptual performance.....	117
2.4.5 Auditory scene analysis.....	118

2.4.6 Sequence perception and reproduction.....	120
2.4.7 Expertise and generalisation.....	122
2.4.8 Conclusions.....	124
Chapter 2 footnotes.....	127
Chapter 2 Supplemental methods.....	131
Chapter 2 Supplemental tables.....	132
Chapter 2 Supplemental figures.....	146

<b>Chapter 3: Musical Training Effects and Cortical Plasticity: Relationships with Training Extent and Behavioural Performance.....</b>	<b>148</b>
<b>3.1 Introduction.....</b>	<b>149</b>
3.1.1 Musical training and cortical plasticity.....	151
3.1.2. Musical training, behavioural performance and plasticity.....	152
3.1.3. The present study.....	154
<b>3.2 Method.....</b>	<b>156</b>
3.2.1 Participants.....	156
3.2.2 Data acquisition.....	156
3.2.3 Procedure.....	160
3.2.4 Data pre-processing.....	160
3.2.5 Cortical surface reconstruction.....	163
3.2.5.1 Image synthesis.....	163
3.2.5.2 Manual image adjustment.....	164
3.2.5.3 Skull strip.....	164
3.2.5.4 Surface reconstruction.....	165
3.2.5.5 Surface adjustment.....	166
3.2.6 Data analyses.....	167
3.2.6.1 R <sub>1</sub> data extraction and mapping.....	167
3.2.6.2 Regions of interest (ROI) analyses.....	168
3.2.6.3 Whole-brain analyses.....	171
<b>3.3 Results.....</b>	<b>172</b>
3.3.1 Effects of expertise: violinists vs. non-musicians.....	172
3.3.1.1 ROI analyses: auditory core and Heschl's gyrus.....	172
3.3.1.2 ROI analyses: hand area.....	176

3.3.2. Effects of expertise: violinists' training.....	178
3.3.2.1 Violinists' training: whole-brain analyses.....	178
3.3.2.2 Violinists' training: ROI analyses.....	180
3.3.3 Performance and cortical R <sub>1</sub> : auditory psychophysical thresholds and SAART.....	181
3.3.3.1 AM depth thresholds and R <sub>1</sub> .....	182
3.3.3.2 FM depth thresholds and R <sub>1</sub> .....	183
3.3.3.3 Onset rise time thresholds and R <sub>1</sub> .....	183
3.3.3.4 SAART RT SD and R <sub>1</sub> .....	183
3.3.4 Further analyses: cortical thickness.....	184
3.3.4.1 Cortical thickness: whole-brain analyses.....	185
3.3.4.2 Cortical thickness: ROI analyses.....	185
<b>3.4 Discussion.....</b>	<b>187</b>
3.4.1. Musicianship and cortical plasticity.....	188
3.4.2. Cortical myelination, expertise and development.....	191
3.4.3. Cortical myelination: behavioural implications.....	194
3.4.4. Conclusions.....	196
Chapter 3 supplemental figures.....	198

#### Chapter 4: Short-term Auditory Learning within a Multi-modal Environment:

Theory and Normative Data.....	200
<b>4.1 Introduction.....</b>	<b>201</b>
4.1.1. Cue combination.....	202
4.1.2. Auditory cues: saliency, context and expectation.....	202
4.1.3. Developmental learning and cue competition.....	205
4.1.4. Learning mechanisms and goal-directed behaviour.....	208
4.1.5. Attention, cross-modal cuing and learning.....	210
4.1.6. The current studies.....	212
<b>4.2 Method.....</b>	<b>215</b>
4.2.1 Participants.....	215
4.2.1.1 Experiment 4a.....	215
4.2.1.2 Experiment 4b.....	215
4.2.2 Materials.....	215
4.2.3 Stimuli.....	216

4.2.3.1 Experiment 4a.....	216
4.2.3.2 Experiment 4b.....	219
4.2.4 Procedure.....	220
4.2.4.1 Experiment 4a.....	220
4.2.4.2 Experiment 4b.....	224
4.2.4.3 SAART.....	224
<b>4.3 Results.....</b>	<b>225</b>
4.3.1 Experiment 4a.....	225
4.3.1.1 2AFC results.....	225
4.3.1.2 In-task results.....	228
4.3.1.2.1 Task accuracy.....	228
4.3.1.2.2 Task RTs.....	229
4.3.2 Experiment 4b.....	229
4.3.2.1 2AFC results.....	229
4.3.2.2 In-task results.....	231
4.4.2.2.1 Task accuracy.....	231
4.4.2.2.2 Task RTs.....	232
<b>4.4 Chapter summary and discussion.....</b>	<b>233</b>
Chapter 4 footnotes.....	235

<u>Chapter 5: Short-term Auditory Learning within a Multi-modal Environment: Cue Combination and Saliency.....</u>	<u>237</u>
<b>5.1 Introduction.....</b>	<b>238</b>
<b>5.2 Method.....</b>	<b>238</b>
5.2.1 Participants.....	238
5.2.1.1 Experiment 5a.....	238
5.2.1.2 Experiment 5b.....	238
5.2.2 Materials.....	239
5.2.3 Stimuli.....	239
5.2.4 Procedure.....	240
5.2.4.1 Experiment 5a.....	240
5.2.4.2 Experiment 5b.....	245
<b>5.3 Results - experiment 5a.....</b>	<b>249</b>
5.3.1 4AFC results.....	249

5.3.2 In-task results.....	255
5.3.2.1 RTs to alien onsets.....	255
5.3.2.2 Response accuracies.....	258
5.3.2.3 Response spatial distance and variability.....	258
5.3.3 Cross task analyses.....	259
5.3.4 Interim summary.....	260
<b>5.4 Results - experiment 5b.....</b>	<b>262</b>
5.4.1 4AFC results.....	262
5.4.2 In-Game results.....	267
5.4.2.1 RTs to alien onsets.....	267
5.4.2.2 Response accuracies.....	269
5.4.2.3 Response spatial distance and variability.....	270
5.4.3 Cross task analyses.....	271
5.4.4 Interim summary.....	272
<b>5.5 Chapter summary.....</b>	<b>273</b>
Chapter 5 Footnotes.....	275
Chapter 5 Supplemental Analyses.....	277
Supplemental analyses 5.1: experiment 5a.....	277
Supplemental analyses 5.2: experiment 5b.....	278

Chapter 6: Short-term Auditory Learning within a Multi-modal Environment:

Development.....	282
<b>6.1 Introduction.....</b>	<b>283</b>
<b>6.2 Method.....</b>	<b>283</b>
6.2.1 Participants.....	283
6.2.2 Stimuli and materials.....	283
6.2.3 Procedure.....	284
6.2.3.1 Test of Word Reading Efficiency (TOWRE).....	284
6.2.3.2 Learning task (cue combination).....	285
6.2.3.3 TEA-Ch: Sky Search subtest.....	285
6.2.3.4 SAART.....	286
<b>6.3 Results.....</b>	<b>287</b>
6.3.1 4AFC results.....	287
6.3.2 In-task results.....	292
6.3.2.1 RT to alien onsets.....	293

6.3.2.2 Response accuracies.....	296
6.3.2.3 Response spatial distance and variability.....	296
6.3.3 Cross-task analyses.....	297
6.3.4. Interim summary.....	298
<b>6.4 Adult-child comparisons: experiments 5a &amp; 6.....</b>	<b>299</b>
6.4.1 4AFC results.....	299
6.4.2 In-task RTs.....	301
6.4.3 Interim summary.....	303
<b>6.5 General Discussion.....</b>	<b>304</b>
6.5.1 Cue combination and learning.....	304
6.5.2 Cue combination, competition and development.....	309
6.5.3 Auditory cues and learning.....	312
6.5.4 Attention, learning and generalisation.....	315
6.5.5 Conclusions.....	317
Chapter 6 Supplemental Analyses.....	319
Supplemental analyses 6.1: experiment 6.....	319
Supplemental analyses 6.2: adult-child comparisons.....	320
<u>Chapter 7: General Discussion.....</u>	<u>322</u>
<b>7.1 General Discussion.....</b>	<b>323</b>
<b>7.2 Long-term expertise: development, environment and plasticity.....</b>	<b>325</b>
<b>7.3 Long-term expertise: specificity versus generality of transfer.....</b>	<b>327</b>
7.3.1 Perceptual transfer.....	327
7.3.2 Broader transfer.....	329
7.3.3 Implications for transfer and plasticity.....	330
<b>7.4 Expertise and cortical plasticity.....</b>	<b>332</b>
7.4.1. Cortical R <sub>1</sub> in experts and non-experts.....	332
7.4.2. Experience-dependent plasticity and expertise.....	334
7.4.3. Expertise and cortical adaptations: auditory and motor cortex.....	335
<b>7.5 Short-term learning, plasticity and expertise.....</b>	<b>336</b>
7.5.1. Auditory cue learning and ‘expertise’.....	337
7.5.2. Auditory cue learning and development.....	340
<b>7.6 Concluding remarks.....</b>	<b>342</b>
References.....	345

## List of Tables

Table 2.1: Descriptive statistics for non-musician, violinist and pianist samples.....	69
Table 2.2: Violinists' ( $n = 24$ ) descriptive data and musical training histories.....	70
Table 2.3: Pianists' ( $n = 24$ ) descriptive data and musical training histories.....	71
Table 2.4: Non-musicians' ( $n = 24$ ) descriptive data and musical training histories.....	72
Table 2.5: MANOVA analyses of auditory psychophysical thresholds across run and group for each task, with effect of run split by group for rise time task.....	96
Table 2.6: Kruskal-Wallis and post-hoc group comparisons across tuning perception task pairs.....	100
Table 2.7: Significant ANOVA effects for percentage accuracy across EnvASA conditions.....	105
Table 2.8: Non-parametric correlations between psychophysical tasks and SAART non-target RTs and SDs.....	107
Table 2.9: Stepwise regression models with musician/non-musician status and SAART performance as predictors of auditory psychophysical thresholds.....	109
Table 3.1: Violinists' ( $n = 21$ ) descriptive data and musical training histories.....	157
Table 3.2: Non-musicians' ( $n = 19$ ) descriptive data and musical training histories.....	158
Table 4.1: Filter CFs and ERBs (both in Hz) for noise bands.....	217
Table 4.2: CFs (in Hz) of filtered noise bands that comprised each CS.....	219
Table 4.3: Frequency sweep (CF kHz) and AM rate (Hz) parameters for each 300 ms PS.....	220

Table 4.4: Rates of hits/misses and false alarms/correct rejections, with Kappa ( $\kappa$ ) data for each CS pair presented at 2AFC.....	227
Table 4.5: Rates of hits/misses and false alarms/correct rejections, with Kappa ( $\kappa$ ) data for each PS pair presented at 2AFC.....	230
Table 5.1: Confusion matrices for adults above and below 4AFC median.....	252
Table 5.2: Kappa coefficients and proportion data for accuracy, true positive, false positive, true negative, false negative and precision data for adults above and below the cohort 4AFC median.....	253
Table 5.3: Confusion matrices for adults above and below 4AFC median.....	265
Table 5.4: Kappa coefficients and proportion data for accuracy, true positive, false positive, true negative, false negative and precision data for adults above and below the cohort 4AFC median.....	266
Table 6.1: Confusion matrices for children above and below 4AFC median....	290
Table 6.2: Kappa coefficients and proportion data for accuracy, true positive, false positive, true negative, false negative and precision data for children above and below the cohort 4AFC median.....	291
Table 6.3: Arithmetic means ( $\pm$ SD) of harmonic mean RTs (s) (and pairwise test significances) across block groupings for the cohort 4AFC median split.....	295
Table 6.4: Kappa coefficients for adults and children above and below their respective cohort 4AFC medians.....	301
Table 6.5: arithmetic means ( $\pm$ SD) of harmonic mean RTs (s) across block groupings for adults and children above cohort 4AFC median split.....	302
Supplement Tables (ST):	
ST 2.1: MSI musical training subscale means and SDs.....	132
ST 2.2: Kruskal-Wallis and post-hoc Wilcoxon Signed Rank Test comparisons for MSI musical training subscale scores.....	133
ST 2.3: 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' $n$ .....	134

ST 2.4: MANOVA analyses of groups' AM depth psychophysical task performance ( $n = 36$ ), for samples drawn at random from violinist and pianist groups, matched to non-musicians' $n$ .....	135
ST 2.5: 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' $n$ .....	136
ST 2.6: Post-hoc pairwise comparisons across rise time, AM depth and FM depth psychophysical thresholding tracks for participants completing 4 runs total.....	137
ST 2.7: One-sample Wilcoxon signed rank tests of difference of group mean from chance, per tuning pair.....	138
ST 2.8: Non-parametric correlations across auditory and visual psychophysical tasks.....	139
ST 2.9: Loadings from principal components analysis (PCA) across all tasks, with additional loadings from tuning system and EnvASA PCA.....	140
ST 2.10: Non-parametric correlations between tuning task pairs, AM depth and FM depth thresholds, across groups.....	141
ST 2.11: Means and standard deviations of pure tone audiometric thresholds (dB HL) for each group across frequencies.....	142
ST 2.12: Test-retest reliability correlations for all possible pairs of runs, for each psychophysical thresholding experiment.....	143
ST 2.13: 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).....	144
ST 2.14: 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).....	145

## List of Figures

Figure 1.1: Action potential propagation along a myelinated axon.....	43
Figure 1.2: Organisation of primary and secondary subfields of primate and human auditory cortex.....	47
Figure 2.1: Auditory and visual psychophysical thresholds across groups.....	94
Figure 2.2: Change in group mean auditory psychophysical thresholds across tracking runs.....	97
Figure 2.3: In-tune choices for just and equal tempered tuning systems when paired with tuning deviances.....	99
Figure 2.4: Group and individual performance on sustained auditory attention task (SAART).....	102
Figure 2.5: Group and individual mean sequence length performance on SIMON sequencing task.....	104
Figure 3.1: A priori regions of interest (ROIs) specified for analyses of cortical $R_1$ and cortical thickness.....	170
Figure 3.2: Group average cortical $R_1$ maps for non-musician (top row) and violinist (bottom row) cohorts.....	173
Figure 3.3: Cortical depth fraction analyses of scaled $R_1$ data across violinist and non-musician cohorts, for auditory core and Heschl's gyrus ROIs.....	175
Figure 3.4: Scaled $R_1$ ROI values for hand area across cortical depths.....	177
Figure 3.5: Whole-brain analyses of effects of years of violin training in predicting cortical $R_1$ .....	179
Figure 3.6: Linear regression fits of scaled mean ROI $R_1$ values, by years of training.....	180
Figure 3.7: Linear regression fits of scaled mean ROI $R_1$ values by AM depth thresholds.....	184
Figure 3.8: Group differences in mean cortical thickness measured at auditory ROIs.....	186
Figure 4.1: Spectrograms of CS stimuli used in experiment 4a.....	218
Figure 4.2: Spectrograms of PS stimuli used in experiment 4b.....	221
Figure 4.3: Alien characters and on-image positions for experiments 4a and 4b.....	223

Figure 4.4: Proportion correct at 2AFC for each CS.....	226
Figure 4.5: Data from Figure 4.4 replotted to highlight individual differences in 2AFC performance over each CS.....	226
Figure 4.6: Mean RTs (ms) ( $\pm$ 1 std. error) for each CS across blocks.....	228
Figure 4.7: Proportion correct at 2AFC for each PS.....	230
Figure 4.8: Mean RTs (ms) ( $\pm$ 1 std. error) for each PS across blocks.....	232
Figure 5.1: Spectrograms of CS and PS stimulus combinations used in experiments 5a & 5b.....	241
Figure 5.2: Design for learning task within experiment 5a.....	243
Figure 5.3: Design for learning task within experiment 5b.....	247
Figure 5.4: 4AFC proportion correct data for experiment 5a.....	251
Figure 5.5a: arithmetic (i.e., non-harmonic) mean RTs across blocks 1-11, for those adults below (left side) and above (right side) the cohort 4AFC median.....	257
Figure 5.5b: arithmetic mean RTs across blocks, split by CS-PS pair.....	257
Figure 5.6: 4AFC proportion correct data for experiment 5b.....	264
Figure 5.7: arithmetic (i.e., non-harmonic) mean RTs across blocks 1-11, for adults below (left side) and above (right side) the cohort 4AFC median.....	269
Figure 6.1: 4AFC proportion correct data for experiment 6.....	288
Figure 6.2a: arithmetic (i.e., non-harmonic) mean RTs across blocks 1-11, for those children below (left side) and above (right side) the cohort 4AFC median.....	294
Figure 6.2b: arithmetic mean RTs across blocks, split by CS-PS pair.....	294
Supplemental Figures (SF):	
SF 2.1: Graphical summary of the procedure for each experimental task.....	146
SF 2.2: Group and individual performance on EnvASA task.....	147
SF 3.1: $R_1$ volumes for two participants.....	198
SF 3.2: Effects of Corpus Callosum scaling of $R_1$ values on ROI data.....	199
SF 5.1: 4AFC proportion correct by median split for control experiment cohort.....	280
SF 5.2: Arithmetic mean RTs across blocks for control experiment.....	281

## Acknowledgements

The completion of this thesis reflects the culmination of a very rewarding three years spent at the Centre for Brain and Cognitive Development. I am deeply grateful to those who have helped me at various stages along the way.

First and foremost, huge thanks are due to my supervisors, Dr. Fred Dick and Dr. Jennifer Aydelott. Their tireless guidance, training, criticism and encouragement helped ensure that these projects reached fruition; I feel exceptionally privileged to graduate as an Alphalab alumnus. Many thanks are also due to the CBCD faculty; their support and generous provision of the Marie Curie TRACKDEV grant funding made this work possible.

Thanks are also due to the fantastic collaborators I have had the great fortune to work with: Prof. Stuart Rosen; Dr. Marcus Pearce; Dr. Alex Shepherd; Dr. Mike Coleman; Prof. Marty Sereno; Dr. Joern Diedrichsen; Dr. Nikolaus Weiskopf; Dr. Antoine Lutti; Dr. Martina Callaghan; Prof. Lori Holt; Dr. Jason Zevin; Mr. John Rogers; Mr. James Ireland. I am grateful also for the assistance of the Royal College of Music, Guildhall School of Music and Drama, and Alleyn's Primary School (particularly Ms. Alison Wright, Ms. Jane Mines, and Ms. Carol Wells). Thanks in particular are due to the participants from these institutions, who gave so generously of their time and energy.

Many thanks also to my examiners, Dr. Lorna Halliday and Dr. Matthew Davis, for their time, care, and attention to detail in examining this thesis.

Thanks to the colleagues and friends who have made CBCD an excellent place to work; in particular, Saloni Krishnan has been a highly valued colleague and dear friend. Thanks also to Kostas, George, Kate, Rhiannon, Carina, Maria, Varun, Rebecca, Michelle, Ellie and Sarah, and to the many friends made at conferences and summer schools over the past three years.

I could not have reached this stage without the support of my family. Love and thanks to my mother, Jill, whose constant support and encouragement has been unsurpassed; and to my father, Philip, whose interest and insights I have greatly appreciated. Huge thanks are due also to Ian and to Seán – your support has been outstanding.

Finally, this thesis is dedicated to my late grandmothers, Katherine Elizabeth Carey and May Smullen, both of whom passed away whilst I was completing this PhD. I miss you both, now and always.

## Chapter 1: General Introduction

## 1.1 Introduction

We are immersed in sound virtually every day of our lives. From early in development, we are capable of processing sounds – pressure and displacement waves arriving at the eardrum due to environmental events perturbing molecules in the air. Yet our perception and cognition with regard to sound can take a lifetime to develop (Werner, 2007; Boothroyd, 1996). Throughout our lives we become remarkably adept at perceiving, experiencing and learning from sounds. We develop complex abilities to finely perceive subtle changes in incoming acoustic signals transduced by the inner and outer hair cells along the basilar membrane. We also develop complex schemas to represent the sounds we encounter. In many ways, we become experts.

How does this expertise arise? Models of experience-dependent plasticity provide a means of addressing this question. Across development, the experience we have with sounds in a variety of forms may be fundamental to guiding our auditory perception and cognition. Furthermore, this experience may critically alter the underlying cortical architecture involved in how we process sounds. The interaction we have with sounds and their relevance to our behaviour may be essential factors that influence how we develop useful sound representations, leading us toward expertise.

This thesis explores the nature of experience, development, learning and expertise within audition. Using long-term and short-term models of experience, we develop a series of experiments to address how differences in experience across the lifespan or in the laboratory can lead to expert or expert-like outcomes in auditory perception and cognition. We also explore how differences in long-term experience can relate to plastic adaptations to auditory (and other) cortical regions involved in expert performance.

In this chapter, we begin with a broad review of the literature regarding plasticity, with emphasis on auditory perceptual development and neural plasticity. We explore the nature of experience and consider what it means to be an expert. We discuss how experience accrued over both long- and short-term periods may lead to expert or 'expert-like' outcomes. We then extend our view of experience toward broader perceptual and cognitive skill; in particular, we raise the question of generalisation of auditory experience across perception and higher cognition. We review cortical substrates critically involved in auditory processing and explore the role of myeloarchitecture in the auditory system, with emphasis on plastic adaptations to myelin in experts. Finally, we provide a brief overview of the experiments and methods used to address these issues in this thesis.

## **1.2 Perceptual and neural plasticity**

The mammalian brain has been regarded as a highly malleable and plastic organ. Moreover, cognitive and perceptual processes that arise from brain systems have also been characterised as adaptable and plastic. Definitions of plasticity applied to perceptual learning emphasise enduring changes to a perceptual system that improve an organism's ability to respond to the environment (Goldstone, 1998). Similarly, definitions of plasticity with respect to cognition and behaviour highlight that plasticity reflects the capacity of a system to perform flexibly (e.g., Lövdén et al., 2010; see 1.2.3). Neural plasticity may be defined broadly as any functional or structural brain adaptation arising as a result of environmental change or some modification of the organism's internal state (May, 2011). In recent years, electrophysiological and in-vivo imaging methods have provided evidence of the brain's capability to

adapt to experience (e.g., Chang et al., 2005; Nakahara, Zhang & Merzenich, 2004; Kempermann, Kuhn & Gage, 1997; Kempermann et al., 1998; Wilbrecht et al., 2010; Trachtenberg et al., 2002; Gibson et al., 2014; Draganski et al., 2004; see Holtmaat & Svoboda, 2009, for review). Indeed, neuroplastic adaptations may take many forms; in grey matter, these include formation of new neurons and synapses, increases in dendritic branching, increases in glial cell numbers, and angiogenesis; in white matter, adaptations can include the myelination of axons, maintenance of myelin sheaths, fibre bundle organisation and genesis of oligodendrocytes (Zatorre, Fields & Johansen-Berg, 2012; Paus, 2005; Emery, 2010; Gibson et al., 2014; Baumann & Pham-Dinh, 2001).

Central to issues of mammalian perceptual and neural plasticity are considerations of ontogeny and environmental demands. We begin by discussing each in turn with particular emphasis on audition.

**1.2.1. Auditory perception and plasticity: ontogeny.** The development of auditory perception extends from infancy across much of childhood and adolescence, and occurs in parallel to changes in neural systems involved in audition (see 1.6.1).

The perception of complex sounds such as speech appears to begin prenatally. For instance, newborns adapt their rate of preferential sucking when hearing a familiar story that is read aloud by their mother during pregnancy, rather than a novel story that is not read aloud (and do so postnatally regardless of the particular voice that reads the story; DeCasper & Spence, 1986). Further evidence suggests that newborns will increase their rate of sucking when hearing speech in their native language rather than a non-native language, and do so even for low-pass filtered versions of that speech (Mehler et al., 1988). These results suggest that complex features of auditory input

(such as linguistic prosody, stress and temporal structure) may be processed by infants following experience that occurs before or very shortly after birth (see Werker & Yeung, 2005).

The ability to discriminate relatively elementary features of auditory signals also appears to begin developing within the first months of life. By 4 months of age, some infants can discriminate differences in frequency and may begin to approach frequency difference limens (i.e., the smallest perceptible frequency difference between two tones) similar to those of adults for a 1 kHz pure tone standard (Werner-Olsho et al., 1982). Nevertheless, the infant auditory system remains relatively immature in its processing of high frequency sounds (e.g., > 3 kHz), such that infants have elevated frequency difference limens compared to adults up to 6 months of age (Werner-Olsho et al., 1982). Indeed, this holds implications for the extent to which fine spectral details (e.g., within speech) can be processed during early development (Werner, 2007). At 4 months of age, infants show sensitivity to acoustic features such as frequency contour, and will look longer toward novel acoustic signal sources that have  $f_0$  contours derived from infant-directed speech prosody versus adult-directed speech prosody (Fernald & Kuhl, 1987). Infants are also capable of processing differences in temporal properties of sounds. Six month olds can discriminate temporal onsets between sound sources at differences of 25 ms (Morrongiello, Kulig & Clifton, 1984), differences in sound durations at thresholds as short as 20 ms (Morrongiello & Trehub, 1987), and differences in syllable voice onset time (i.e., the time between plosive release and voicing onset) of approximately 40 ms (Aslin et al., 1981). Infants' thresholds tend to be elevated relative to those of typical adults however (Morrongiello et al., 1984; Aslin et al., 1981).

Latter stages of infant auditory development typically reflect the emergence of more specialised perceptual processing. For instance, while infants show a preference for native speech (as discussed above), they are also capable of discriminating between non-native speech phones. However, by 12 months of age this perceptual ability diminishes, as infants show a relatively greater bias toward the speech phones of their native language (Werker & Tees, 1984; Kuhl et al., 2006; but see Kuhl, Tsao & Liu, 2003). Infants' abilities to track and extract information from the on-going auditory signal also develop toward the end of the first year. By 8 months, infants can segment synthetic word streams based on the transitional probabilities between syllables, where low probability transitions reflect word boundaries; this enables infants to discriminate between trisyllabic 'words' (e.g., 'bi-da-ku') and 'part-words' (e.g., 'da-ku-pa') that are heard in the absence of stress, prosody or differences in between-syllable temporal intervals (Saffran et al., 1996). This may be a possible mechanism infants use in segmenting continuous speech, helping them to isolate and perhaps learn words (see Werker & Yeung, 2005).

In childhood, auditory perception continues to develop, alongside more complex auditory cognitive skills. For instance, the fine-grained discrimination of differences in sound frequency, intensity and duration improves from 4 years of age through to later childhood and adulthood (Jensen & Neff, 1993; see also Moore et al., 2008; Halliday et al., 2008; Moore et al., 2010; Morrongiello & Trehub, 1987). Children show continuing development of auditory temporal perception, for instance with respect to the temporal envelope properties of syllable onsets (e.g., Nittrouer, Lowenstein & Tarr, 2013); this in turn may hold implications for broader development of phonological abilities (see Goswami et al., 2002). The perception of some auditory signal modulations also shows

protracted development across childhood. The ability to detect periodic variation in the temporal envelope of a sound (i.e., amplitude modulation) remains immature up to 12 years of age; in contrast, detection of periodic variation in frequency (i.e., frequency modulation) is mature by approximately 8 years of age (Banai, Sabin & Wright, 2011) (see footnote 1).

Importantly, children's auditory development involves improvements in the perception of a single sound where other sounds are present. Where a single target tone is presented simultaneously with other non-overlapping tones that vary randomly in frequency, children can detect the target tone at lower signal intensities at later (8-10 years) versus earlier (5-7 years) ages (Leibold & Neff, 2007; see also Leibold & Bonino, 2009). These findings suggest improvements in the isolation and perception of particular sounds during childhood, reflecting reduced susceptibility to 'informational masking' (i.e., where the presence of multiple sounds leads to confusion or uncertainty regarding a target sound; Moore, 2012; see Hall, Buss, & Grose, 2005; Krishnan et al., 2013). Moreover, broader auditory cognitive skills such as auditory working memory (i.e., the ability to manipulate auditory information held in mind) and selective auditory attention continue to improve up to late childhood (Siegel & Ryan, 1989; Banai & Ahissar, 2013; Coch, Sanders & Neville, 2005).

The preceding review has briefly charted some of the features of early auditory development that occur in the path to adulthood. One additional consideration of importance to auditory development is the age at which particular auditory experience occurs. Indeed, evidence suggests that during relatively early periods of life – so-called critical periods (CPs) – the influence of experience on the brain can be especially pronounced and may be essential to

development of typical neural circuits and functioning (e.g., Wiesel & Hubel, 1963; Huang et al., 1999; see Knudsen, 2004; White et al., 2013).

In auditory domains, a CP has been proposed for language development, during which native language input and cortical maturation contribute to language learning and progressive left-lateralisation of language function (Lenneberg, 1967; see Werker & Tees, 2005; Bates et al., 1995). More moderate forms of the hypothesis – the sensitive period hypothesis – have proposed that language acquisition is possible beyond childhood (e.g., Hurford, 1991); however, experience with a native language will tend to delimit or entrench a learner's set of representations, particularly with respect to phonology (Birdsong, 2009). As a result, native-like features of a second language (e.g., accent) may be less easily learned with increasing age (Flege, Munro & MacKay, 1995; see Zevin, 2012).

With respect to mechanisms that may guide critical or sensitive periods, the expression of neuroplastic potential within the brain appears to be greatest during early development. For instance, synaptogenesis within primary auditory cortex peaks within the first three years of life in humans (Huttenlocher & Dabholkar, 1997). Animal models have provided strong evidence that experience at specific points early in development influences neural adaptations to auditory cortical regions. For instance, rats reared with early exposure to noise stimuli during a putative CP (post-natal days 7-30) had substantially degraded frequency selectivity of primary auditory cortex tonotopic maps; however, behavioural training (post-natal days 36-90) could re-establish typical response profiles of neurons and tonotopic gradients (Zhou & Merzenich, 2007; Chang et al., 2005). Similarly, relative to controls, rats exposed to tone pips during this CP had increased tonotopic areal representations at frequencies

close to those of the tones; moreover, neural responses also showed selectivity to the temporal order in which the tone pips were presented (Nakahara et al., 2004). These results suggest relative malleability of auditory cortex within early periods of development, such that particular stimulus environments may shape cortical organisation and responses.

This brief review of development of auditory perception across infancy and childhood has shown that development yields refined auditory perceptual abilities and improved auditory cognitive skills. We have also seen that experience during early development can influence domains such as language, and may shape development of auditory cortex. In the next section, we discuss experience-dependent effects on auditory perception, and plasticity within auditory cortex arising from specific environmental experience.

**1.2.2. Auditory perception and plasticity: environment.** Experience with particular auditory environments during development may influence both perceptual and neural plasticity. One widely studied form of such experience is musical training. The intensive acoustic demands encountered while playing music (e.g., perceiving pitch, melody and rhythm) and the related training of motor skills (e.g., fine sequencing of manual movements) provide rich avenues through which experience and plasticity may be explored (Herholz & Zatorre, 2012; Zatorre et al., 2012a; Münte, Altenmüller & Jäncke, 2002).

Children with musical experience show enhanced auditory perceptual skills compared to their non-musician peers. For example, 10-13 year old children with musical training perform better in detection of subtle frequency deviances (~25 cents) for single notes within major or minor scales than children who have no musical training (Lynch & Eilers, 1991). Such perceptual abilities also extend to low-level non-musical tasks; musically-trained children

show improved frequency difference limens for pure tones compared to children without musical training (Banai & Ahissar, 2013). Children assigned to one year of music lessons also show improved rhythmic abilities and can tap to a cued beat with greater temporal precision after the cue stops, compared to children without musical training (Slater, Tierney & Kraus, 2013).

These findings have been widely replicated in adults; extensive evidence suggests that adult musicians can perceive differences in frequency and/or pitch (e.g., Kishon-Rabin et al., 2001; Micheyl et al., 2006; Weiss et al., 2014), differences in temporal intervals (Ehrle & Samson, 2005), and differences in timbre (Pitt, 1994) more precisely than non-musicians (see chapter 2; 2.1.1).

However, a major question concerns the extent to which musical experience is the dominant causal factor driving these perceptual adaptations, and whether such perceptual adaptations show related neural underpinnings (see Herholz & Zatorre, 2012). Random training assignment studies offer one means of addressing these questions.

Hyde et al. (2009) followed 6 year old children who had been assigned to musical training or no training. Pre-training data from the cohort showed no evidence of perceptual or brain structure differences between the trained and untrained children (Norton et al., 2005). After 15 months, the musically trained children showed improved abilities to discriminate differences in either the pitch or melody of brief 5 note musical phrases, compared to children not assigned to training. Moreover, children's improved discrimination performance correlated positively with increases in brain tissue deformation metrics at right Heschl's gyrus, indicating relative increases in tissue volume (Hyde et al., 2009). These data therefore suggest that musical training can causally drive both perceptual and neuroplastic adaptations.

Furthermore, Moreno et al. (2008) and Chobert et al. (2012) assigned children to either musical training or painting training for one year. Electrophysiological recordings after one year showed that children assigned to musical training had increased amplitudes of event-related potential (ERP) components in response to subtle violations in the pitch of sentence final words (Moreno et al., 2008), and to subtle changes in the f<sub>0</sub>, duration and voice onset time of a /ba/ speech phone (Chobert et al., 2012); however, control children assigned to painting training did not show these adaptations. Relatedly, children that perform specific instruments demonstrate selectively enhanced ERP amplitudes when listening to the timbre of the instrument they have learned to play (e.g., violin) compared to an instrument they cannot play (e.g., piano) (Shahin, Roberts & Trainor, 2004; see also Shahin et al., 2008). This finding has similarly been replicated with adult musicians (Pantev et al., 2001).

Taken together, these results provide evidence of a causal role for auditory experience in driving perceptual and neuroplastic enhancement. Musical training is one such domain in which auditory experience can shape both perceptual and neural adaptations, and can do so from early childhood through to adulthood.

**1.2.3. Plasticity, ontogeny and environment: a framework.** The preceding sections suggest complex roles for both ontogeny and the environment with respect to plasticity. How then can we begin to account for the complexity of mechanisms that mediate plasticity?

One recent model relevant to mechanisms of plasticity describes the relationship between environmental demands and functional capacity (Lövdén et al., 2010). If a system (behavioural and/or neural) is faced with a set of demands that exceed its current capability (i.e., functional capacity), then this

mismatch can serve to drive plasticity. Critically, the determinant of plasticity in the model is supply-demand mismatch: where the environment poses an increased demand to the organism that exceeds its functional capacity, there is impetus for plasticity; if demand remains relatively stable (within the existing functional capacity), there is no such impetus. This ‘supply and demand’ framework posits that systems strive toward maintaining a state of dynamic equilibrium that allows for flexibility across time (Lövdén et al., 2010).

Yet following the discussion above (see 1.2.1 and 1.2.2), ontogeny and environment are heavily interrelated; the expression of plasticity can therefore only occur via mechanisms that arise in the context of both factors (Karmiloff-Smith, 2012; Westermann et al., 2007). Neuroconstructivism offers a counterpoint to the ‘supply and demand’ mismatch framework outlined above; it proposes that cognitive development can be accounted for by a trajectory that arises based on many constraints (e.g., genetic, cellular, and experiential) acting on the neural development of systems that underpin cognition (Westermann et al., 2007). A key distinction between neuroconstructivism and the ‘supply and demand’ mismatch framework is their account of development: while neuroconstructivism views developmental change along a trajectory that forms a central tenet of the model, the ‘supply and demand’ mismatch framework largely accounts for plasticity within the mature adult brain (Westermann et al., 2007; Lövdén et al., 2010). It is therefore useful to consider both frameworks when exploring profiles of plasticity across the lifespan.

Studying expertise provides one means of uniting these views, and offers a model for investigating how ontogenetic and environmental factors can interact to spur plasticity over long-term (i.e., years) and short-term (i.e., days or

even hours) time frames. In the next section, we explore expertise as it pertains to this thesis, with respect to both long-term and short-term experience.

### **1.3 Long-term expertise and plasticity**

The previous sections discussed auditory plasticity with respect to ontogenetic and environmental factors. In the following, we explore plasticity in the context of an expertise framework. We consider expertise in many domains (auditory and visual) via long-term experience acquired over the lifespan, along with expertise related to short-term experience and learning (see 1.4).

**1.3.1. Long-term expertise.** As a first step, it is important to consider the nature of expertise. Perhaps unsurprisingly, the breadth and depth of literature on experts in many domains has precluded a concise definition of ‘an expert’. Basic descriptive accounts of experts emphasise: the accumulation of an extensive body of specialised knowledge (Chi, 2006); the possession of rare and exceptional skills (Krampe & Ericsson, 1996); and the intensive, lifelong training (10,000+ hours) needed to acquire such skills (Ericsson et al., 1993; Hoffman, 1996).

Moreover, experts differ fundamentally from non-experts in the depth and application of their perceptual and cognitive skills (Hoffman, 1996; Chi, 2006; Palmeri, Wong & Gauthier, 2004). Experts can abstract knowledge beyond simple heuristics, and possess rapid and accurate access to fine-grained detail and subordinate levels of description (compared to novices’ focus at the basic or superordinate level) (Tanaka & Taylor, 1991; Sowden et al., 2000; Palmeri et al., 2004; see also Ahissar et al., 2009). For instance, experts’ reaction times in deciding whether images from their expert category (e.g., birds) match a subordinate level name (e.g., ‘sparrow’) are not significantly different from their

reaction times when deciding whether those images match the basic level name (e.g., ‘bird’) (Tanaka & Taylor, 1991). Moreover, experts can apply finely honed discrimination abilities to relatively low-level perceptual conditions. Expert radiographers are significantly more accurate than novice X-ray film readers when identifying low luminance contrast dots printed on film, even where expert visual search and pattern recognition skills are task-irrelevant (Sowden et al., 2000). These data indicate specialisation of very fine perceptual abilities in experts along with cognitive skills characterised by rapid access to highly refined representations of the expert category. We next explore structural and functional plasticity related to expertise, with a focus on audition.

**1.3.2. Long-term expertise: structural plasticity.** In line with these expert-novice distinctions, neuroimaging studies of expert groups have shown structural adaptations to brain regions that appear to be involved in expert skill. London taxi drivers show greater grey matter volumes at posterior hippocampi compared to control subjects, presumably related to their expertise in spatial navigation (Maguire et al., 2000). Relatedly, diffusion tensor MR data from karate experts has shown reduced fractional anisotropy (FA) in the left superior cerebellar peduncles of experts relative to non-experts; these FA metrics further related positively to the measured latencies of hand strikes (Roberts et al., 2013).

With respect to auditory expertise, we have seen above (see 1.2.2) that musicianship is associated with auditory cortical adaptations. Indeed, a variety of neuroimaging studies have shown structural adaptations to brain regions thought to be involved in musical performance, including Heschl’s gyri (Schneider et al., 2002; 2005; Seither-Preisler et al., 2014; Gaser & Schlaug, 2003), planum temporale (Schlaug et al., 1995; Bermudez et al., 2009), primary

motor cortex (Amunts et al., 1997; Gaser & Schlaug, 2003), cerebellum (Hutchinson et al., 2002) and inferior frontal gyrus (Sluming et al., 2002) (see chapter 3). Crucially, many of these studies have also shown relationships between experience-based metrics and structural adaptations; thus, years of training or measures of musicians' proficiency can predict variance in brain structure (e.g., Schneider et al., 2002; Gaser & Schlaug, 2003; Sluming et al., 2002). Such data provide support for experience-dependent processes in guiding expert plastic adaptations following long-term training.

Structural imaging data also suggest experience-dependent plastic adaptations in non-musical auditory experts. In expert piano tuners, years of training show a positive correlation with right hippocampal volumes (perhaps due to tuners' expertise in navigating a spatially complex auditory scene whilst tuning; Teki et al., 2012). Furthermore, expert phoneticians have increased white matter density at Heschl's gyri bilaterally compared to non-experts (Golestani et al., 2011). Moreover, phoneticians' years of training can account for increases in surface area and cortical volume at left pars opercularis (a region thought to be involved in phonological processing; Golestani et al., 2011). These data further support the role of experience-dependent factors in shaping expert plastic outcomes over long-term periods.

**1.3.3. Long-term expertise: functional plasticity.** Functional neuroimaging data have also suggested differences in neural processes and systems when comparing experts and non-experts. In vision, ERP data have shown selectively increased N170 component amplitudes when dog or bird experts view images from their expert category versus another category (Tanaka & Curran, 2001) (note that in the typical population, N170 amplitude is usually selectively increased for familiar visual stimuli such as faces; Bentin &

Deouell, 2000). In addition, functional magnetic resonance imaging (fMRI) data have shown that cortical regions such as superior temporal sulcus (STS) are recruited maximally by speech stimuli in expert actors (versus a resting baseline or listening to violin music) (Dick et al., 2011). However, STS in expert violinists shows an increase in activation for violin music (compared to resting baseline), such that similar extents of activation occur at violinists' right STS for violin music and for speech (Dick et al., 2011) (notably, STS tends to show greater activation for speech stimuli in the typical population; Scott et al., 2000; Agnew et al., 2011). These studies show adaptations to neural processes in experts that reveal plasticity for stimuli from the expert domain. Moreover, this plasticity appears to manifest within functional processes and cortical regions that often specialise toward highly familiar stimulus types (e.g., faces, speech sounds) in the typical population.

Moreover, functional imaging data suggest that experts' memory for stimuli from their field reflects less effortful encoding and retrieval compared to novices. For instance, in vision, data show that behavioural recollection scores that relate to status as a car expert correlate negatively with ERP amplitude differences between recollected car images and car images judged as familiar (Herzmann & Curran, 2011; see also Chi, 2006). These findings suggest that in those subjects who are most expert, functional activity during recognition and recollection of car stimuli is more similar; in contrast, those who are less expert show greater differences in activity for stimulus recognition versus recollection (Herzmann & Curran, 2011).

In addition, functional data also suggest differences in the encoding of stimulus sequences in experts compared to novices. When presented with tones that violate the regular temporal rhythm of an isochronous sequence

during fMRI, musicians activate anterior hippocampus to a relatively greater extent than non-musicians; moreover, longitudinal data collected from a cohort of music students at the beginning and end of the academic year showed that the musicians' extent of anterior hippocampus activation to temporal deviants was greater following a year of music tuition (Herdener et al., 2010). These data provide strong evidence for experience-driven functional plasticity with respect to encoding of sequential auditory information in experts relative to non-experts.

**1.3.4. Long-term expertise: summary.** In sum, our expertise framework emphasises: the exceptional nature of expert knowledge, perception and skill; experts' rapid and deep perceptual abilities alongside cognitive advantages in the realms of their field; and plastic adaptations to neural structures and functional processes involved in expert performance.

In further exploration of expertise and plasticity, a core theme of this thesis is the investigation of perceptual, cognitive and cortical adaptations in expertly trained cohorts. Experts trained over long-term periods (from childhood to adulthood) offer a window into plasticity, through interaction between specific experience and ontogeny. A retrospective approach to studying this development can shed further light on specific (and perhaps more general) perceptual, cognitive and brain structure outcomes related to expertise (see 1.4 and 1.5). In chapter 2, we investigate perceptual and cognitive performance in two cohorts of expertly trained musicians (violinists and pianists) compared to a closely matched non-musician cohort. In chapter 3, we expand this investigation to consider cortical adaptations in the same cohorts of violinists and non-musicians.

#### 1.4 Short-term expertise and plasticity

Can short-term environmental experience result in plastic adaptations akin to expertise? Evidence from behavioural training interventions and neuroimaging studies suggests that this may be the case.

Classic studies of visual object learning have shown that even periods of brief training lead to outcomes with many of the features of expertise that we have discussed (see 1.3.1) (e.g., Tanaka, Curran & Sheinberg, 2005). For example, ten hours' training with novel visual objects ('greebles') leads to increased speed of greeble differentiation at subordinate levels (i.e., individuation of greeble category exemplars using local details such as shape and orientation of appendages). But, this increase in processing efficiency can be disrupted by changes to the configural relations amongst features (similar to effects observed in perceptual studies using highly familiar stimuli, such as faces) (Gauthier & Tarr, 1997; 2002; Bukach et al., 2012).

In addition, perceptual learning studies have shown that training environments can lead to substantial improvements in the perception of relatively low-level acoustic features, even in normal hearing children and adults. Fine-grained auditory perceptual learning (e.g., frequency discrimination) can occur within a single laboratory session lasting just a few hours, leading to perceptual thresholds on par with those of expert musicians (Micheyl et al., 2006). Indeed, following laboratory training subjects can learn to discriminate fine differences in a wide variety of acoustic features, including frequency (Halliday et al., 2008; Moore et al., 2008), intensity (Whitton et al., 2014; see also Halliday et al., 2011), inter-onset interval (Wright et al., 1997; 2010; van Wassenhove & Nagarajan, 2007), and inter-aural time and level (Wright & Fitzgerald, 2001). With training, listeners can adapt to and learn even highly

complex acoustic stimuli, such as non-native speech phones (Lim & Holt, 2011; Kuhl et al., 2003), noise-vocoded speech (Davis et al., 2005; see 1.5) and even repetitions of noise segments (Agus et al., 2010).

Recent behavioural and neuroimaging data support the complexity of auditory learning that is possible via training and reveal adaptations that may be deemed 'expert-like'. Exemplars of 'auditory greebles' (complex, spectrally variable auditory categories analogous to speech phone categories) that cued visual character onsets within a video game could be grouped and learned as auditory categories in an unsupervised fashion, following just 30 minutes of game play (Wade & Holt, 2005). Moreover, a subsequent fMRI study (Leech et al., 2009b) in which subjects trained on this game showed that auditory greeble categorisation success was positively related to pre-post training increases in activation at left STS (a cortical area that tends to respond selectively to highly learned auditory stimuli, particularly intelligible speech) (see Agnew et al., 2011; Scott et al., 2000; Scott & McGettigan, 2013).

These results suggest that complex auditory categories can be learned to 'expert-like' levels within an active training task, and moreover, can lead to adaptations to cortical responses for those stimuli such that learning success mediates the extent of cortical activity. Indeed, converging fMRI evidence (Wong et al., 2009) has shown that learning to individuate novel visual objects increases activation in cortical areas typically involved in human face processing (fusiform face area). These data further suggest that short-term perceptual training can modify neural representations of learned stimuli (Wong et al., 2009; see also James & James, 2013; Gauthier et al., 2000; Xu, 2005; but see footnote 2).

Further to these findings, studies of short-term auditory learning and 'expertise' have explored facets of brain structure and function that may account for learning success. For instance, rate of success in learning to discriminate along a native to non-native speech phone continuum (dental-retroflex) correlates with increased density of white matter underlying left (but not right) Heschl's gyrus (Golestani et al., 2007; see also Golestani et al., 2002). Similarly, increases in grey and white matter volumes at left Heschl's gyrus account for improved learning success during short-term training in linguistic pitch discrimination (Wong et al., 2008). Functional MRI data have also shown that learning rates for fine discrimination of microtonal melodies relate to the slope of the blood-oxygen level dependent (BOLD) response at early stages of training; for larger pitch differences, rapid learners tend to show steeper BOLD signal slopes at anterior superior temporal gyri earlier in training (Zatorre, Delhommeau & Zarate, 2012).

Structural imaging evidence also suggests that short-term visuo-motor training environments can spur neuroplastic adaptations. Over a period of months, training in ball juggling yields increased grey matter volumes in early visual areas (Scholz et al., 2009) and in cortical areas sensitive to visual motion (MT/V5) (Draganski et al., 2004; Boyke et al., 2008), along with increases in FA in the underlying white matter at intraparietal sulci (Scholz et al., 2009). However, these plastic changes do not persist once training ceases: adaptations diminish to pre-training baseline levels after three months (Draganski et al., 2004; Boyke et al., 2008) and one month (Scholz et al., 2009). This suggests that continued engagement in the behaviour is necessary in order to maintain the adaptations over time.

Taken together, these findings suggest that brief periods of training within laboratory environments may lead to plastic adaptations for trained stimuli at perceptual, cognitive and neural levels. The fine detail of the representations that develop, the similarity to some facets of performance in ‘real-world’ experts, the neural adaptations that result and the specificity of neural regions that modulate successful performance all suggest characteristics that resemble expertise (see Bukach, Gauthier & Tarr, 2006).

A further core theme within this thesis is the emergence of ‘expertise’ over short timeframes. In chapter 2, we further investigate patterns of short-term perceptual learning in expert musicians and in non-experts. In chapter 4, we explore the development of novel auditory representations using an interactive task, to model some of the complex experience-dependent features of real-world sound learning. In chapter 5, we extend this paradigm to learning of auditory cue combinations, allowing a model of learning in complex, multifaceted auditory environments. We further investigate questions of environment and ontogeny, by asking whether ‘expert-like’ outcomes can arise in children and adults within this paradigm (chapter 6).

### **1.5 Learning, generalisation and plasticity**

So far we have considered the roles of development, environment and expertise with respect to plasticity. We have seen that experience at certain phases of development and in particular environments can influence perceptual, cognitive and neural adaptations. We have discussed how expertise arises in the context of long-term development, yet have also seen that short-term training can yield outcomes that parallel expert adaptations.

Perhaps one of the most controversial questions in psychology and neuroscience is the specificity versus generality of learning and plasticity. That is, if learning and plasticity occur within a certain domain, are these outcomes specific to that domain? Could these adaptations extend more generally to other domains and reveal transfer?

Data relating to these issues present a complex picture. Concerning long-term auditory expertise, evidence has suggested that musicians show enhanced abilities to perceive and process speech stimuli under conditions of multi-talker babble and speech-shaped noise (Parbery-Clark et al., 2009a; 2009b; 2011; 2012; Strait et al., 2011b; see Kraus & Chandrasekaran, 2010). While these findings suggest generalisation of musicians' listening skills beyond musical contexts, more recent research has failed to replicate improved speech-in-noise perception abilities in musicians (Ruggles et al., 2014). Further studies suggest that extended periods of musical instruction may result in generalisation of that training to less directly related cognitive abilities (i.e., far transfer), including full-scale IQ (Schellenberg, 2004) and executive function (Moreno et al., 2011) (we return to these studies in chapter 2).

In contrast, some data from the visual modality show compelling evidence of expertise and training transfer. In a series of studies, Green and Bavelier found that compared to non-players, action video game players had improved visuo-spatial resolution (i.e., reduced crowding effects; Green & Bavelier, 2007), greater speed (but not accuracy) in visual perception (Dye, Green & Bavelier, 2009), and finer perception of visual motion coherence (Green, Pouget & Bavelier, 2010). Further, video game players' skills transferred to benefits on higher-level visual selective attention and attentional blink tasks (Green & Bavelier, 2003; 2012), and to an auditory spatial task

(judging the location of a tone in noise; Green et al., 2010). Moreover, in each study, results were replicated following random assignment of non-players to action video game training (alongside active control cohorts trained on non-action games). The precise mechanisms accounting for such general perceptual and cognitive transfer remain unclear, although the rapid, dynamic and immersive visual demands of action games may play a role (note that controls trained on non-action games show some transfer, but less than those who train on action games; Green & Bavelier, 2003; 2007).

Auditory learning and training studies have suggested less consistent findings of transfer and generalisation. Training in perception of a temporal interval between two tones (i.e., gap duration) yields improved temporal discrimination at a trained frequency (1 kHz) and at an untrained frequency (4 kHz); however, such training does not transfer to untrained standard temporal intervals (Wright et al., 1997; 2010). Training on backward masking (i.e., identifying a pure tone presented immediately before a noise burst) transfers to a condition where the tone occurs 10 ms before the noise, but not to forward masking (i.e., noise immediately before tone) or simultaneous masking conditions (Huyck & Wright, 2013). Similarly, auditory frequency discrimination training can yield significant improvements in thresholds relative to the trained standard frequency, both in adults and children (Moore et al., 2008; Halliday et al., 2008; Halliday et al., 2012). However, such learning does not transfer to higher-level phonological tasks (i.e., word and non-word reading, non-word repetition and rhyme judgement; Halliday et al., 2012) or discrimination with an untrained standard frequency (Halliday et al., 2008).

In auditory domains such as speech perception, listeners can generalise learning of vocoded speech across low-pass and high-pass filtered versions of

noise-vocoded sentences (Hervais-Adelman et al., 2011) (in noise vocoding, the temporal envelope of speech is extracted and smoothed over multiple non-overlapping frequency bands, used to modulate a non-speech carrier at each frequency band, and then recombined across those bands; Davis et al., 2005). However, generalisation of vocoded speech perception across carrier signals (sine wave, pulse train or noise) is less consistent (Hervais-Adelman et al., 2011). Further, training on noise-vocoded sentences containing non-words (or function words and non-words) does not generalise to lexically correct noise-vocoded sentences; this appears to suggest a role for top-down, lexical or semantic processes in mediating learning (Davis et al., 2005).

The above findings offer a complex profile of results. Across various domains, long-term expertise (e.g., in music, video games) has shown evidence of near transfer to related areas (and some evidence of far transfer – for instance, across modalities in video game players). Short-term training studies similarly have replicated some of these findings: near transfer may occur after brief periods of laboratory training, although evidence of far transfer following short-term training varies across domains (contrast the auditory and speech training results above, with video game training results).

A major goal of this thesis is to explore both specificity and generality of learning and expertise. In chapter 2, we explore whether musicians' expert listening skills also extend to non-musical domains that reflect some of the same perceptual and cognitive demands of musicianship.

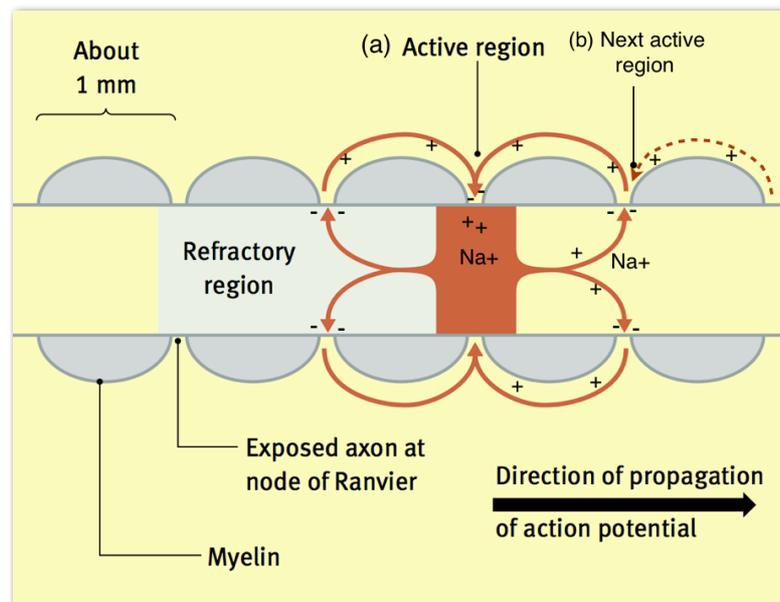
## 1.6 Cortical plasticity and myeloarchitecture

In the preceding sections we have seen that mechanisms related to expertise and plasticity are complex, and notably so in audition. In order to understand these mechanisms, it is important to explore the underlying architecture of the cortex. In so doing, we may begin to characterise profiles of cortical structure – in particular, cortical myeloarchitecture – that relate to expertise. Moreover, we may also investigate whether such structural properties of cortex relate to perceptual and cognitive performance, across experts and non-experts.

In the following sections, we review the nature of myeloarchitecture and development, the motivation for studying auditory cortex in particular, definitions of primary auditory core regions, and implications for studies of expertise.

**1.6.1. Myelination and development.** Myelin is a spiral-shaped tissue composed of many lipid layers that ensheaths axons within the central and peripheral nervous system (Emery, 2010; Baumann & Pham-Dinh, 2001). In the central nervous system (CNS), the myelin sheath is formed by glial cells called oligodendrocytes (Schwann cells in the peripheral nervous system). The sheath forms as an extension of the cytoplasmic membrane of the oligodendrocyte, which surrounds and covers the axon in sections known as internodes (Baumann & Pham-Dinh, 2001).

The role of myelin is well characterised: myelin facilitates action potential conduction velocities along axons via the process of saltatory conduction, ensuring efficient action potential propagation (Emery, 2010; Baumann & Pham-Dinh, 2001; see Figure 1.1 & footnote 3). Myelin accounts for a considerable extent of human brain tissue (~40-50% of total dry weight; Baumann & Pham-Dinh, 2001), with the majority of CNS myelin comprising subcortical white



**Figure 1.1:** Action potential propagation along a myelinated axon (cross section) by saltatory conduction. (a) Depolarisation of resting membrane potential occurs at the active region (node of Ranvier) due to opening of sodium channels and influx of  $\text{Na}^+$  ions to the intracellular region;  $\text{Na}^+$  ions then conduct along the intracellular region as electrotonic potentials that reach the next node of Ranvier (i.e., to the right of the active region). The inactivation of  $\text{Na}^+$  channels due to the refractory period from a prior depolarisation (i.e., to the left of the active region) ensures that the action potential propagates in one direction only (i.e., to the right). (b) Depolarising  $\text{Na}^+$  current reaching the next node of Ranvier leads to opening of further sodium channels and influx of  $\text{Na}^+$  ions (see dashed arrow); subsequently,  $\text{Na}^+$  efflux occurs, with  $\text{Na}^+$  flowing back across the exterior of the myelin sheath to (a). The myelin sheath aids conduction velocities by reducing capacitance and leakage of current across the membrane; opening of channels at the nodes of Ranvier only greatly increases the rate of action potential propagation. Adapted from Wareham (2005).

matter. Indeed, initial development of myelination in humans begins in sub-cortex. The cerebellar peduncles begin myelinating before birth, followed by deep cerebellar white matter at 1 month of age; the posterior limb of the internal capsule by 2-3 months; the splenium and genu of corpus callosum by 4 and 6 months, respectively; and the subcortical white matter of the frontal, parietal and

occipital lobes by 12 months (Paus et al., 2001; Partridge et al., 2004; Barkovich et al., 1988; Baumann & Pham-Dinh, 2001; Nakagawa et al., 1998).

Nevertheless, mature myelin development in humans follows a protracted course (e.g., Barnea-Goraly et al., 2005). Total brain white matter volume increases up to middle age before declining (Sowell et al., 2003); fibre myelination across pre-frontal, motor, somatosensory and occipital cortices does not peak until approximately 30 years of age (Miller et al., 2012); and total cortical white matter volume similarly peaks at approximately 30 years of age before decreasing (see Paus, 2001). With respect to audition, the acoustic radiations (afferent fibres projecting from medial geniculate mainly to layers IIIb and IV of primary auditory cortex; Hackett, 2011) typically begin myelinating by the 26th week of gestation and finish by three years of age (Nakagawa et al., 1998; Baumann & Pham-Dinh, 2001; Schnupp, Nelken & King, 2011).

Importantly, primary sensory and motor cortical regions show relatively higher extents of cortical myelin than association cortex, and tend to myelinate earlier in development (see 1.6.3). Myeloarchitectonic maps of cortex by Flechsig (see Sereno et al., 2012; Barbey & Patterson, 2011) highlighted regions that myelinate by early childhood, including primary motor, somatosensory, visual and auditory cortices. The advent of in-vivo MR methods that can detect and quantitatively measure the lipid-dense (and low tissue water) relaxation properties of myelin has greatly enhanced our ability to probe the myeloarchitecture of cortical regions, particularly primary auditory cortex (Dick et al., 2012; Lutti et al., 2014; see 1.6.3). Recent imaging studies measuring proxies for cortical myelin in-vivo have shown relatively high myelination in primary motor and somatosensory cortex (M1 & S1; Glasser &

Van Essen, 2011; Sereno et al., 2012), primary and higher visual areas (V1, V2, V3a, V6, MT; Sereno et al., 2012) as well as primary auditory cortex (A1 & R; Dick et al., 2012; Sigalovsky et al., 2006; Glasser & Van Essen, 2011), when compared to association cortex (e.g., pre-frontal regions).

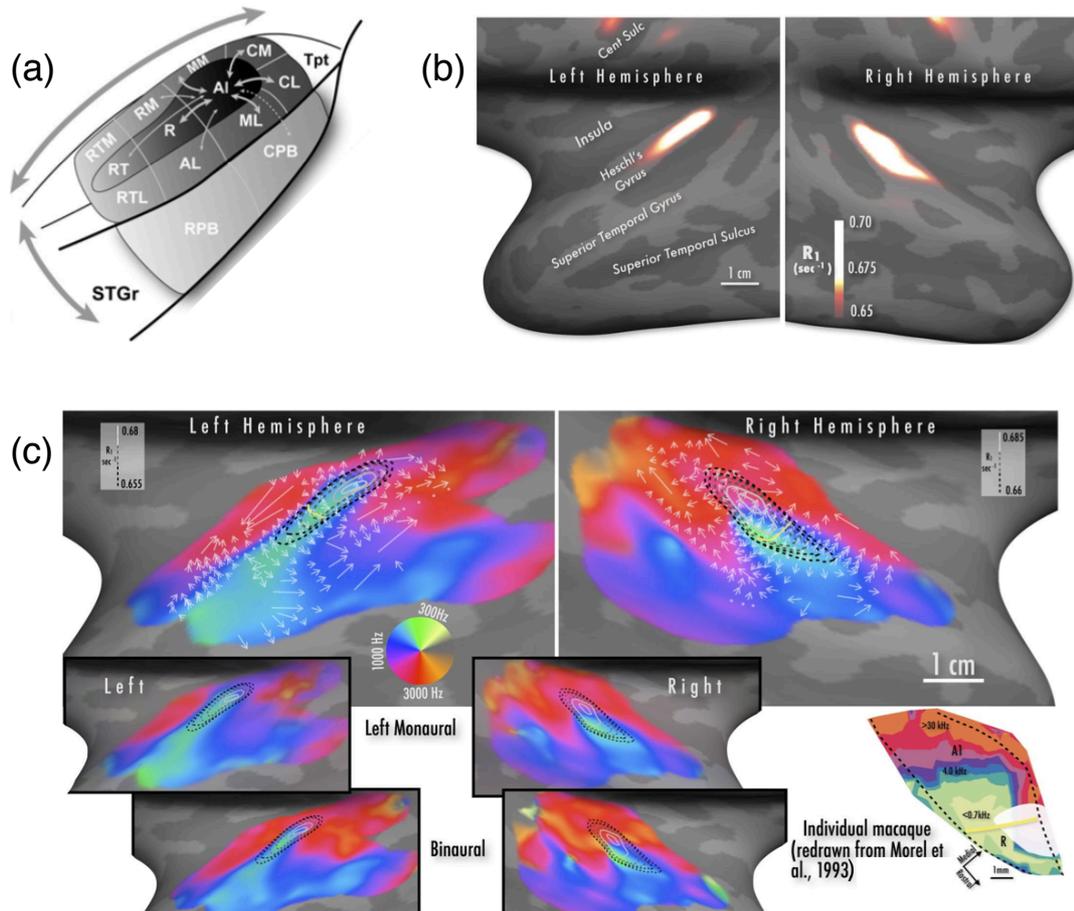
**1.6.2. Auditory cortex: processing hierarchy.** At this stage, let us consider why primary auditory cortex is important to study. In primates and humans, primary auditory cortex represents the first cortical processing stage in the ascending hierarchy of the auditory system (Kaas, Hackett & Tramo, 1999; Rauschecker & Tian, 2000; Hackett, 2011). The auditory system begins extracting sound features including spectral content and temporal onsets at the early processing stages of the cochlea and cochlear nuclei (Schnupp et al., 2011; Shamma, 2001). At further early stages, features of inputs critical to sound localisation are coded, such as inter-aural time differences at medial superior olive (Grothe & Sanes, 1993; Ashida & Carr, 2011), and inter-aural level differences at lateral superior olive (Irvine, Park & McCormick, 2001; Schnupp et al., 2011). Through a series of ascending afferents via nuclei of the lateral lemniscus, the inferior colliculi and the medial geniculate nucleus of the thalamus, these earlier auditory processing stages project to primary auditory cortex (see Hackett, 2011).

Primary auditory cortex has consistently been shown to manifest plastic adaptations when comparing experts and non-experts (e.g., Schneider et al., 2002; 2005; Golestani et al., 2011; Hyde et al., 2009; Seither-Preisler et al., 2014). Based on the position of auditory cortex as the first major cortical processing stage in the ascending hierarchy of the auditory system, manifestation of plasticity within this cortical region may be critical to the fine-grained listening skills that typify the performance of auditory experts

(Schneider et al., 2002; 2005). Thus, the study of auditory cortex may provide valuable insight into the underlying neural architecture that enables auditory expertise. Moreover, metrics of auditory cortical structure that can be related to specific markers of experience provide a means of exploring the role of experience-dependent factors in influencing plastic adaptations. In particular, relating measures of experience to tissue-specific structural markers (e.g., indices of cortical myelin) can provide insight into the mechanisms that may account for plastic adaptations to auditory cortex – and the specific neural processes involved.

Indeed, both tonotopic organisation (Dick et al., 2012; Talavage et al., 2004) and representation of temporal envelope (Herdener et al., 2013) at primary (and adjacent secondary) auditory cortex are key features of the position within the cortical processing hierarchy (see Figure 1.2c). Nevertheless, the isolation and definition of primary auditory cortex in humans has presented considerable challenges (e.g., Penhune et al., 1996).

**1.6.3. Location and definition of auditory core.** The parcellation of auditory cortex at human Heschl's gyrus has been largely based on studies of myeloarchitecture, histochemistry, and post-mortem probabilistic atlases, in order to distinguish primary (i.e., core) and non-primary subfields (Hackett, Preuss & Kaas, 2001; Hackett, 2011; Morosan et al., 2001; Rademacher et al., 2001). Much of what is known about primary subfields of auditory cortex derives from primate histological studies, which have delineated three major caudal-rostral divisions within auditory core: area A, area R, and area RT; these primary fields are also densely interconnected with adjacent non-primary subfields (Rauschecker & Tian, 2000; Hackett, 2011; 2007; Kaas & Hackett, 2000; see Figure 1.2a). Indeed, identifying homologues for these boundaries in humans



**Figure 1.2:** Organisation of primary and secondary subfields of primate (a) and human (b & c) auditory cortex. (a) major subdivisions of macaque auditory cortex (LH shown) are highlighted; note the darker A1 and R, denoting heavier myelination compared to adjacent regions; arrows indicate projections between core, belt and parabelt subregions (arrow thickness denotes density of projections; dashed arrow highlights reciprocal projection; large grey arrows show major gradients of information flow). (b) group average ( $N = 6$ ) map of cortical  $R_1$  in humans showing heavily myelinated auditory core (likely reflecting primate A1 & R). (c) group average phase-encoded tonotopic maps from the same subjects as (b), with overlaid bounds of thresholded  $R_1$  maps at auditory core (black and grey isocontours); colour scale indicates ‘best frequency’ responses of cortical neurons to auditory stimuli; arrows show direction of phase spread as CF varied. Lower insets show relative consistency of tonotopic maps across monaural and binaural stimulation conditions; right inset shows replotted tonotopic gradients across macaque auditory cortex, measured electrophysiologically. (a) adapted from Hackett (2011); (b) & (c) adapted from Dick et al. (2012).

has proved problematic. Post-mortem atlases have used cytoarchitectonic methods to define the major subfields of auditory cortex in humans (caudal-rostral: Te 1.1; Te 1.0; Te 1.2) (Morosan et al., 2001). Indices of laminar thickness indicative of a well-defined layer IV (i.e., koniocortex), together with metrics of high cell volume densities at layer IV have helped to delineate Te 1.0 as a highly probable location for human primary auditory cortex (Morosan et al., 2001; see also Rademacher et al., 2001).

As discussed above (see 1.6.1), primary auditory cortex has also been defined based on the relatively greater extent of myelination that typifies primary sensory areas (Dick et al., 2012; Sereno et al., 2012). The high lipid and low water content of cortical myelin greatly facilitates MR  $T_1$  times (Lutti et al., 2014).  $T_1$  is a time parameter that represents the exponential recovery of total longitudinal magnetisation within a given tissue or medium after an RF excitation pulse is applied; that is, it reflects the regrowth of magnetisation along the longitudinal ( $z$ ) axis as excited protons return to equilibrium after absorbing RF energy (Lutti et al., 2014; Paus et al., 2001; Gore & Kennan, 1999). Human in-vivo MR mapping methods that quantitatively measure the rate at which  $T_1$  regrowth occurs ( $R_1 = 1/T_1$ ) offer a means of indexing cortical regions that are highly myelinated (higher  $R_1$  reflects greater extents of myelin, since myelin facilitates  $T_1$  times) (Dick et al., 2012; Lutti et al., 2014). Thus, cortical mapping of  $R_1$  values at Heschl's gyrus across individuals has demonstrated that the heavily myelinated core of primary auditory cortex occupies the most medial two-thirds of Heschl's gyrus (Sigalovsky et al., 2006; Dick et al., 2012; see also Glasser & Van Essen, 2011; further to Hackett et al., 2001) (see Figure 1.2b).

**1.6.4. Implications for studies of expertise.** We have seen that myelin mapping techniques greatly assist in delineation of the boundaries of auditory core (and further myelin-rich cortical areas, including primary motor cortex). In spite of this, no single study yet has explored whether expertise is associated with specific adaptations to cortical myelin in regions critical to success in the expert domain. Indeed, given previous evidence of the importance of cortical myelin to basic perceptual processes (e.g., vision; Fornari et al., 2007), and the debilitation associated with myelin loss in disorders such as multiple sclerosis (Summers et al., 2008; Faiss et al., 2014; Baumann & Pham-Dinh, 2001), study of the optimisation of myelin in experts (relative to non-experts) may offer critical insight into the experiential mechanisms of white matter plasticity, and the relative malleability of cortical myelin itself. Further to adaptations to subcortical white matter tracts in experts (Imfeld et al., 2009; Roberts et al., 2013; see chapter 3), quantitative evidence of cortical myelin adaptations would greatly inform our understanding of cortical plasticity in experts. Moreover, combining such quantitative methods with behavioural indices of perceptual and cognitive performance may help to shed light on the behavioural significance of any measured change in cortical myelin proxies (see Zatorre et al., 2012a).

Thus, a major goal of this thesis is to explore adaptations to cortical myelin proxies in expert musicians as compared to non-experts. In chapter 3, we examine cortical myelin proxies ( $R_1 = 1/T_1$ ) across cortex, with particular focus on auditory cortical regions (auditory core as defined by Dick et al., 2012; Heschl's gyrus) in both expert violinists and non-musicians. Moreover, behavioural measures collected from the same participants (see chapter 2) allow us to probe perceptual and cognitive abilities as they relate to such

cortical myelin proxies – an important step in determining relationships between behaviour and cortical structure in experts and non-experts.

We investigated these myelin-related effects using very high-resolution MR techniques that allow for quantitative measures of the longitudinal relaxation rate,  $R_1$  ( $= 1/T_1$ ; see 1.6.3). Using the multiple flip angle technique formalised by Helms et al. (2008) and others (Weiskopf et al., 2011; Lutti et al., 2010; Dick et al., 2012; Sereno et al., 2012), we could resolve at each voxel for the rate of longitudinal relaxation ( $R_1 = 1/T_1$ ). This property offers a quantitative proxy for cortical myelin, since regions showing relatively high myelin content will have correspondingly high  $R_1$  (i.e., short  $T_1$ ; Lutti et al., 2014; Paus et al., 2001; see footnote 4).

## **1.7 Thesis overview**

The central aim of this thesis is to investigate experience-dependent learning, plasticity and expertise within audition. To these ends, we examine effects of experience across different timescales (long-term and short-term) and within a variety of cohorts (expert and non-expert; developing and adult). We address these questions at perceptual, cognitive and cortical levels, as described below.

Firstly, we investigate auditory perceptual and cognitive skills associated with long-term experience in the domain of music (chapter 2). Using two cohorts of expert musicians (violinists and pianists), we explore whether differences in the auditory demands of their respective long-term training environments relate to differences in expert-level perceptual and cognitive outcomes. We further compare the performance of our experts to a closely matched sample of non-musicians, and ask whether expert skills (perceptual and cognitive) transfer

beyond the immediate area of training to benefit non-musical auditory (and multi-modal) abilities. Our study emphasises a retrospective approach to studying long-term experience and plasticity. By investigating measures of training (e.g., duration, intensity of practice, age of training onset) in two expert groups with experience accrued under very different demands, we can begin to explore if the nature and intensity of training at particular points in development accounts for expert-level differences in perceptual and cognitive outcomes. We probed fine perceptual skills in our cohorts using adaptive psychophysical procedures (Levitt, 1971; Cornsweet, 1962) that allow for measurement of thresholds for instrument-relevant acoustic cues. We further used two-interval two-alternative forced choice discrimination procedures to examine perception of subtle fixed differences between musical chord stimuli. We explored cognitive performance and expert skill transfer across a range of tasks, examining sustained auditory attention, naturalistic auditory scene analysis and multi-modal sequence reproduction. Test-retest reliability data for novel measures developed for this thesis are described in chapter 2.

Secondly, we extend our study of expertise and plasticity and ask whether long-term training to expert level is related to differences in cortical myelin in experts relative to non-experts (chapter 3). Using subsets of the same violinist and non-musician cohorts from chapter 2, we investigated whether violinists' expertise might be associated with increases in quantitative proxies for cortical myelin ( $R_1$ ). In particular, we sought to characterise myelin profiles within auditory core regions, via analysis of effects related to expertise across cortical depths. Further, we explored cortical myelin proxies within motor hand area regions, that we predicted would differ between violinists and non-musicians due to violinists' extensive fine manual training. Furthermore, the

perceptual and cognitive metrics collected for these same violinists and non-musicians (presented in chapter 2) allowed us to test whether behavioural performance across our experts and non-experts might relate to cortical myelin proxies ( $R_1$ ) measured in those participants.

Finally, we investigate profiles of perceptual learning over short periods within the laboratory and ask whether active task performance can yield adaptations that may be thought of as 'expert-like'. We first explore profiles of auditory learning across relatively low-level perceptual tasks as completed by experts and non-experts (chapter 2). We further examine how more complex representations are learned, using a novel, multi-modal paradigm ('Space Holiday'). In particular, we explore whether listeners can learn complex, novel auditory cues to visual events over short periods of time (where cues can be considered as broadband and contextual, or punctate and object-like; see chapter 4). Moreover, we explore whether novel auditory cues may be learned as a combination that yields the most optimally informative cue (i.e., a 'scene', comprising a contextual and punctate cue; chapter 5). We explore these questions with respect to ontogeny, asking whether the ability to learn complex auditory cue combinations differs between adults and children (chapter 6). We further investigate the role of broader attentional mechanisms and ask whether individual differences in attentional ability might relate to differences in learning outcomes (chapters 4 & 5).

In chapter 7, we return to these themes, address the results from each chapter, and integrate the findings within the broader themes of experience-dependent learning and plasticity discussed in this chapter.

## Chapter 1 Footnotes

1. It is important to acknowledge that apparent differences in children's thresholds for detection of particular acoustic cues (e.g., AM depth versus FM depth) may partly reflect developmental factors that are non-sensory in nature, such as in-task attention and motivation (e.g., Moore et al., 2008).
2. The model of expertise as applied to perception of visual object categories is not without controversy. While authors such as Gauthier et al. (2000), Xu (2005) and Wong et al. (2009) have argued that FFA shows enhanced responses to visual object categories that become well-learned following periods of individuation training, others have proposed that FFA selectively processes face stimuli (e.g., Grill-Spector, Knouf & Kanwisher, 2004; Kanwisher & Yovel, 2006). Recent data also suggest normal learning of visual greebles in two clinical cases of prosopagnosia, particularly where one patient showed damage and considerably reduced fMRI activation for faces (versus other object categories) at right FFA (Rezlescu et al., 2014).
3. Myelin sheath internodes are spaced at distances of approximately 150-200  $\mu\text{m}$ , with intervening myelin-free nodes – the nodes of Ranvier (Huxley & Stämpfli, 1949; Baumann & Pham-Dinh, 2001). Since nodes of Ranvier contain high concentrations of sodium channels, the propagation of an action potential is associated with sodium channel opening (and intracellular sodium influx) at the nodes of Ranvier only (see Figure 1.1). As a result, depolarising positive ion ( $\text{Na}^+$ ) current flows along the intracellular side of the axon membrane to the next node of Ranvier, yielding further depolarisation at that node of Ranvier as the action potential propagates (Huxley & Stämpfli, 1949; Baumann & Pham-Dinh, 2001). The presence of channels at nodes of Ranvier means that current propagation occurs more rapidly (relative to

piecemeal opening of adjacent channels, as in unmyelinated axons). Moreover, since myelin sheaths have high impedance and help to reduce membrane capacitance, current largely flows intracellularly along the axon (rather than across the sheath; Huxley & Stämpfli, 1949; Baumann & Pham-Dinh, 2001). Myelin therefore facilitates the rapid conduction of action potentials across axons.

4. We operationalise  $R_1$  ( $=1/T_1$ ) as a quantitative proxy for cortical myelin throughout this thesis. It is important to note that while  $R_1$  does show higher values in both subcortical and cortical white matter compared to grey matter, several tissue properties can contribute to measured  $R_1$ . Specifically, the  $T_1$ -dependence of  $R_1$  means that  $T_2^*$  (i.e., the exponential decay of transverse magnetisation due to local magnetic field inhomogeneities; Paus et al., 2001) also has a small effect on  $R_1$  estimation. In particular, tissue properties such as iron can influence  $T_2^*$  (and its quantitative metric;  $R_2^*$ ), which in turn influences  $R_1$  (Callaghan et al., 2014). Nevertheless, in healthy adults, cortical regions high in iron content tend to reflect areas of high myelination, since oligodendrocyte cell bodies typically are associated with very high concentrations of iron (Bartzokis, 2004; Todorich et al., 2009). Thus, we acknowledge that  $R_1$  metrics may show some small contributions from tissue properties not directly associated with the  $T_1$  properties of myelin lipids or tissue water; however, such contributions likely reflect other related (i.e., oligodendrocyte) processes that are critical to maintaining myelination.

## Chapter 2: Generality and Specificity in the Effects of Musical Expertise on Perception and Cognition \*

\* Note: A peer-reviewed and edited version of this chapter is published in the journal *Cognition*. Carey, D., Rosen, S., Krishnan, S., Pearce, M.T., Shepherd, A., Aydelott, J. & Dick, F. (2014). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition*. doi: 10.1016/j.cognition.2014.12.005.

## 2.1 Introduction

Perceptual and cognitive skills can be shaped and enhanced through our experience with the world (e.g., Goldstone, 1999; Palmeri & Gauthier, 2004). As discussed in chapter 1 (see 1.3.1), 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 analyse 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 instrument 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 generality 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. 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. Moreover, 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. The present study aimed to address this gap in understanding (see 2.1.3).

### 2.1.1 Musicianship and auditory perception

A considerable body of research suggests that musicians tend to outperform 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.

**2.1.1.1 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 for piano timbre relative to pure tones (Pantev et al., 1998), and enhanced responses 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 Goebel, 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 (colour 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).

**2.1.1.2 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 players 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 reported 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 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.

### **2.1.2 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) generalised 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 generalise this experience to non-musical indices of complex auditory scene analysis. We also asked whether musicians' experience with reproducing sounds and sequential motifs might generalise to novel yet regular

sequences. In the following sections, we review evidence of musician and instrument specific advantages across these cognitive domains.

**2.1.2.1 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.

**2.1.2.2 Auditory scene analysis.** In order to perform successfully with a musical ensemble, musicians must analyse 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 analyse, 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 generalise 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 & Shafiro, 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.

**2.1.2.3 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 (François & Schön, 2011; see François et al., 2014, for similar results with N400 amplitudes; note that in both studies, behavioural indices showed no significant learning of melodies in musicians or non-musicians). Further, 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.

### **2.1.3 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 generalisation 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.

## 2.2 Method

### 2.2.1 Participants

Participants ( $N = 72$ ) were 24 violinists, 24 pianists and 24 non-musicians (descriptive statistics displayed in Table 2.1), 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:  $\chi^2(2, N = 72) = 0.01, p > 0.9$ ). None reported any history of auditory or uncorrected visual impairment, or of neurological disease or insult.

**Table 2.1:** 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

**2.2.1.1 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 2.1). Violinists and pianists had experience of playing other instruments

(notably piano for violinists; see Tables 2.2 & 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 2.2:** 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)
v1	F	23	19	3	Viola	3
v2	F	22	14	4	Piano	6
v3	M	19	12	4	Piano	7
v4	F	20	17	4.5	Piano; Viola; Trumpet	7; 7; 7
v5	F	23	19	4.5	Piano	2
v6	F	20	12	5	Piano; Viola	12; 4
v7	F	19	14	1	Piano	missing data
v8	F	25	20	5	None	
v9	M	21	17	1	Piano	9
v10	M	24	21	4	Piano; Alto Saxophone	12; 13
v11	F	28	21	4	Piano; Clarinet; Viola	2; 6; 3
v12	M	26	20	6	Piano	5
v13	F	25	11	5	Viola	1
v14	F	21	18	4.5	Viola	6
v15	M	28	20	2.5	None	
v16	F	30	27.5	5	Piano	20
v17	F	25	18	3	Piano; Viola	missing data
v18	F	22	14	3	Piano; Viola	1.5; 1
v19	F	23	16	2.5	Bass Guitar	6
v20	M	22	17	6	Piano; Viola; Voice	5; 2; 7
v21	F	19	12	3	Piano	8
v22	F	20	13	3	Piano	2
v23	F	26	17	1.5	Piano	5
v24	F	23	16	4	Piano; Trumpet	2; 2

**Table 2.3:** 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)
p1	F	23	19	5	None	
p2	F	19	12	4.5	Guitar	0.25
p3	M	19	12	3.5	Clarinet; Voice	4; 3
p4	F	19	16	2.5	Cello	7
p5	F	25	19	4	None	
p6	F	24	20	6.5	Clarinet	8
p7	F	20	16	4	Voice; Gamelan	3; 1
p8	M	21	15	4	None	
p9	M	18	9	2	Organ; Double Bass	4; 4
p10	F	22	18	5	Voice	10
p11	F	22	17	4	Harpsichord; Zither	2; 8
p12	M	26	21	5	None	
p13	F	20	15	5	None	
p14	F	19	12	5	None	
p15	M	20	8	4	Drums	0.25
p16	F	19	15	4	None	
p17	F	23	18	6	Cello	5
p18	F	19	10	5.5	Cello	1
p19	M	18	10	6	Harpsichord	2
p20	F	22	14	3.5	Voice	3
p21	F	22	18	5.5	None	
p22	M	20	14.5	2.5	French Horn	1
p23	F	25	20	1	None	
p24	F	25	18	1	Drums; guitar	6; 2

**2.2.1.2 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 and were well-matched to musicians in terms of educational background (see footnote 7). Non-musicians

described any previous experience with musical instruments and any years of practice and/or lessons (see Table 2.4). 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

**Table 2.4:** Non-musicians' ( $n = 24$ ) descriptive data and musical training histories

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

lessons during childhood or adolescence, but had not attended 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.

### **2.2.2 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 (SM.1).

### **2.2.3 Test-retest reliability**

The six newly designed experiments within the battery were assessed for test-retest reliability following initial development and pilot testing. These experiments were: psychophysical thresholding for onset rise time, amplitude modulation depth and frequency modulation depth; SAART; tuning system perception task; SIMON sequencing task. For each task, we present the results for test-retest analyses in the following sections (see 2.2.4).

Participants for test-retest experiments were recruited from local participant pools in two phases (see below). All participants ( $N = 46$ ; mean age

$\pm$  SD:  $27.304 \pm 9.097$ ; 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:  $28.4 \pm 8.9$  [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 ( $n = 17$ ) (see chapter 4) or response switching ( $n = 4$ ) version of the SAART, and 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:  $26.4 \pm 9.4$  [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.

## 2.2.4 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. A summary figure of the procedure for each task is presented in Supplemental Figure (SF) 2.1.

**2.2.4.1 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.

#### **2.2.4.2 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. Supplemental table [ST] 2.1 displays musical training sub-scale means for each group; group comparisons are displayed in ST 2.2.

**2.2.4.3 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 analysed statistically due to the small and unbalanced sample sizes.

**2.2.4.4 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.

**2.2.4.4.1 Stimuli.** All experiments used a complex sawtooth pulse waveform ( $f_0 = 220\text{Hz}$ ; first 50 harmonics), sequentially run through a series of resonators of varying center frequency ( $CF_1 = 500\text{ Hz}$ ;  $CF_2 = 1500\text{ Hz}$ ;  $CF_3 = 2500\text{ Hz}$ ; all bandwidths =  $100\text{ Hz}$ ). 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.,  $20\log [m]$ , 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  $f_0$ : 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).

**2.2.4.4.2 Auditory psychophysics 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). Numbers of subjects who completed 2 versus 3 experimental tracks for tasks was as follows: rise time – 2 tracks: 4 violinists, 4 pianists, 0 non-musicians; 3 tracks: 17 violinists, 17 pianists, 16 non-musicians; AM depth – 2 tracks: 11 violinists, 7 pianists, 9 non-musicians; 3 tracks: 12 violinists, 16 pianists, 14 non-musicians; FM depth – 2 tracks: 9 violinists, 8 pianists, 8 non-musicians; 3 tracks: 15 violinists, 16 pianists, 13 non-musicians.

We also analysed 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  $n$ s 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 2.3.1).

**2.2.4.4.3 Test-retest reliability.** Stimuli were identical to those described above (see 2.2.4.4.1). As above (2.2.4.4.2), thresholds for each run were measured as the mean of the final 3 or 4 reversals. For each psychophysical experiment, Spearman's  $\rho$  correlations over all possible pairs of runs showed moderate to high test-retest reliability for thresholds [range ( $\rho$ ): 0.56 to 0.9, all  $\rho < 0.05$ ; see ST 2.12]. 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 poorly on one run each, 5 tracked poorly on two runs each, and 1 failed to track on all runs); AM depth  $n = 15$  (6 excluded; five tracked poorly on one run each, one tracked poorly on two runs); FM depth:  $n = 19$  (2 excluded; poor tracking on one run each)]. Note however that despite the small  $n$ 's for correlational analyses,  $\rho$  coefficients were relatively high; further, inspection of scatter plots suggested tight clustering of points with strong positive linear relationships for each experiment.

**2.2.4.5 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.

**2.2.4.5.1 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.

**2.2.4.5.2 SIMON 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 colour-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.

**2.2.4.5.3. Test-retest reliability.** SIMON stimuli and procedure were identical to above (2.2.4.5.2), but participants only completed the game (i.e.,

without the pre-game listening phase). Participants' ( $n = 23$ ) 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  $F(1, 21) = 9.271$ ,  $p = 0.006$ ,  $\eta^2 = 0.306$ , but no significant effect of group nor any significant interaction (both  $F < 1.58$ ,  $p > 0.22$ ). Inspection of means indicated a small decline in sequence lengths between runs 1 and 2 (mean difference  $\pm$  SD:  $-0.16 \pm 1.54$ ) perhaps suggesting a fatigue effect over runs. As expected, 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:  $t(21) = 0.03$ ,  $p > 0.97$ ]. Similarly, mean sequence lengths for runs 1 and 2 (collapsed across groups) were significantly positively correlated ( $\rho = 0.59$ ,  $p = 0.003$ ), suggesting good test-retest reliability.

**2.2.4.6 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).

**2.2.4.6.1 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. A roving detection paradigm was used in order to keep the interval differences fixed at proportions that corresponded with the differences between the just and equal tempered tuning systems. The outcome measure for each chord pair was the proportion of trials on which a given chord was chosen as most in-tune (e.g., for the just vs. equal pair, proportions greater than 0.5 indicated just intonation tended to be chosen; proportions less than 0.5 indicated equal temperament tended to be chosen). Six of all possible chord pairs presented were selected a priori 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.

**2.2.4.6.2 Tuning task 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  $\pm 30$  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 (in test-retest reliability experiments we also explored a condition where participants were instructed to simply choose the chord of each pair that they preferred; see 2.2.4.6.3). 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.

**2.2.4.6.3. Test-retest reliability.** 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 (see 2.2.4.6.2). 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 2.2.4.6.2.

Test-retest correlations (Spearman's) for proportion values from the phase 1 tuning system judgement task are presented in ST 2.13. 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 ( $\rho$ ): 0.45-0.51, all  $p < 0.05$ ]. However, 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 (correlations displayed in ST 2.14). 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 2.14). This suggested that non-musicians were not any more consistent in their responses when making preference rather than 'in-tune' judgements for the chord pairs.

**2.2.4.7 Environmental auditory scene analysis (EnvASA) task.** The EnvASA paradigm measured environmental sound detection within natural auditory scenes (see Leech et al., 2009a). 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.

**2.2.4.8 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).

**2.2.4.8.1 Stimuli.** Stimuli were nine short environmental sounds taken from Leech et al., (2009a). 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.

**2.2.4.8.2 SAART 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 analysed (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.

**2.2.4.8.3 Test-retest reliability.** Two versions of the paradigm were evaluated: a response inhibition ( $n = 17$ ) and a response switching version ( $n = 20$ ). The response inhibition version used the identical stimuli and broadly the same procedure as described in 2.2.4.8.2, requiring participants to make speeded keyboard presses (using the space bar) to each non-target stimulus,

but to inhibit this response for the targets. Stimuli were presented randomly during the response inhibition paradigm. Twelve instances of each of the 9 stimuli were presented yielding 108 trials in total (targets occurred on 8.3% of trials). The response switching version utilised the identical stimuli and procedure as described in 2.2.4.8.2.

For the response inhibition version, participants' ( $n = 17$ ) mean reaction times and standard deviations for correct trials only were calculated across the entirety of both testing runs (i.e., across all non-target stimuli). Total accuracy (i.e., number of correct target inhibitions) was also analysed for each run. Test-retest analyses over runs 1 and 2 suggested moderately strong correlations for participants' response accuracies ( $\rho = 0.59$ ,  $p = 0.012$ ) and mean reaction times ( $\rho = 0.61$ ,  $p = 0.009$ ). Standard deviations of participants' reaction times across the entirety of both runs also displayed a moderately strong correlation ( $\rho = 0.51$ ,  $p = 0.037$ ).

For the response switching version, participants' mean and SDs of RTs were calculated for correct responses to target sounds over the first 81 trials (i.e., the pseudorandom portion of each run); mean and SDs of RTs for correct non-targets were calculated in the same manner. Accuracies to target sounds were also analysed. Two participants were excluded from test-retest analyses (one showed an overall mean RT across both runs greater than 900 ms, i.e., > 3 SDs above cohort mean; another responded correctly to less than 80% of non-targets; analysis  $n = 18$ ). Test-retest analyses showed high reliability of target mean RTs ( $\rho = 0.81$ ,  $p < 0.0001$ ) and non-target mean RTs ( $\rho = 0.69$ ,  $p = 0.0014$ ) during the pseudorandom half of the experiment over both runs; however SDs of RTs to targets ( $\rho = -0.13$ ,  $p > 0.62$ ) and non-targets ( $\rho = 0.34$ ,  $p > 0.17$ ) during the pseudorandom half of the experiment were not significantly

correlated across runs. Target accuracies during the pseudorandom half of the experiment were significantly positively correlated across runs ( $\rho = 0.572$   $p = 0.013$ ).

**2.2.4.9 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] 2.11).

**2.2.4.10 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 colour vision using Ishihara plates. Participants then completed the baseline task from Tibber and Shepherd (2006). Participants discriminated increment (purple) and decrement (yellow) colour 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).

### **2.2.5 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).

## 2.3 Results

### 2.3.1 Auditory Psychophysical Thresholds (Figures 2.1a-c; Figures 2.2a-c)

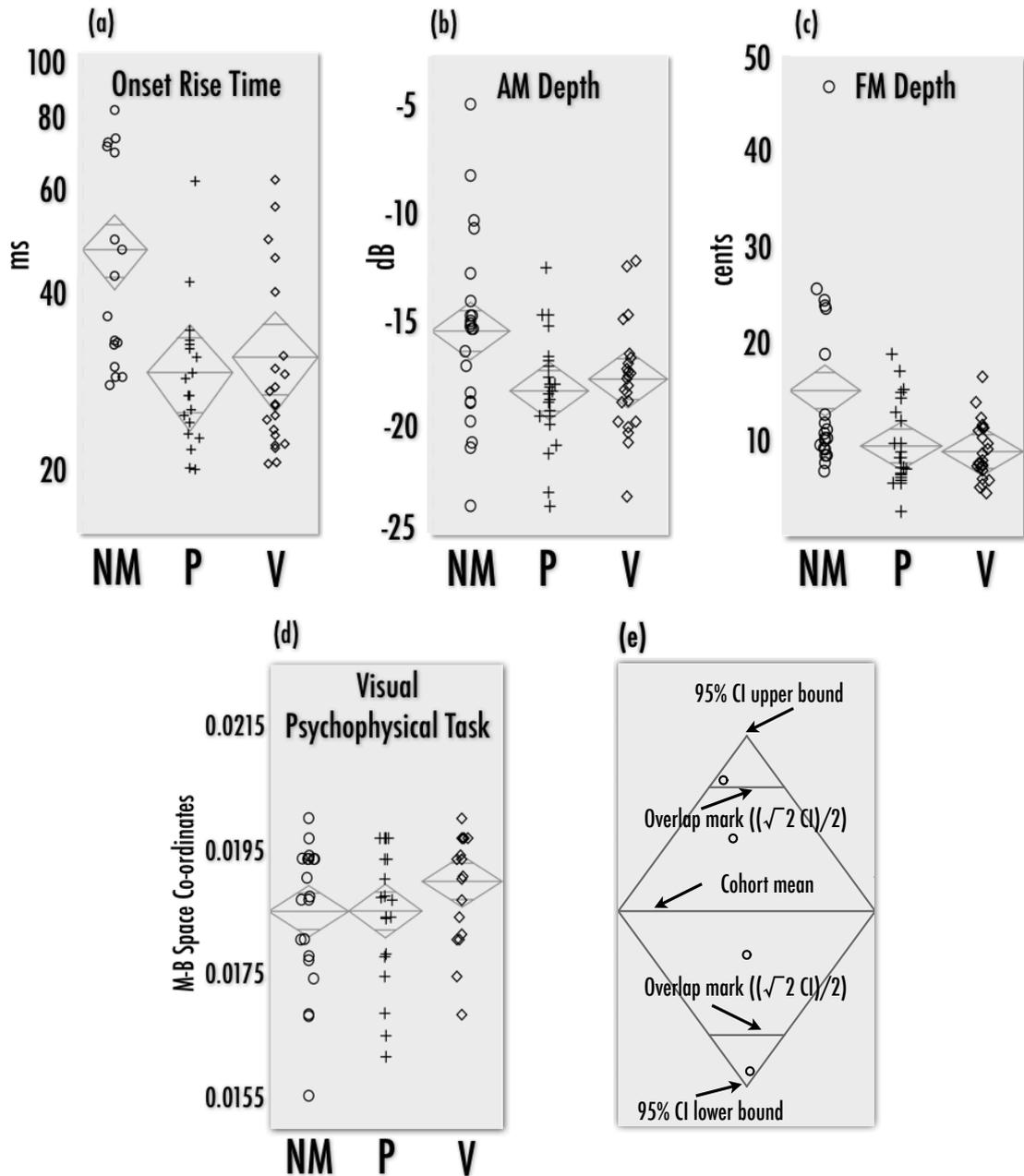
First, we asked whether there were group differences in each auditory psychophysical measure and whether musician groups trained with different instruments differed in their thresholds for specific acoustic features.

**2.3.1.1 Rise time.** Rise time thresholds differed significantly across groups,  $\chi^2(2, n = 58) = 15.06, p = 0.0005$  (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 ( $p = 1.0$ ).

**2.3.1.2 AM depth.** AM depth thresholds differed significantly across groups,  $\chi^2(2, n = 69) = 6.63, p = 0.036$  (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 ( $p = 0.49$ ).

**2.3.1.3 FM depth.** FM depth thresholds were significantly different across groups,  $\chi^2(2, n = 69) = 11.03, p = 0.004$  (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



**Figure 2.1:** 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); (e) display convention for kite plots used here and in figures elsewhere in this thesis; small points display individual subject data; note that where sample sizes are equal, overlapping marks indicate that group means are not significantly different at the 95% CI. Note logarithmic axis for onset rise time thresholds (linear axes for others); group *ns* differ across tasks - see 2.2.4.4.2 for description.

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 2.2.4.4.2), 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.

**2.3.1.4 Rise time (log transformed to correct for positive skew).** As shown in Figure 2.2a, 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 2.6 and indicated in the figure) and by analyses of random samples (ST 2.3). 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,  $\chi^2(2, n = 41) = 3.0, p = 0.22$  (Kruskal–Wallis). In other words, by their fourth run, non-musicians had improved to within the range of the musicians' first attempt.

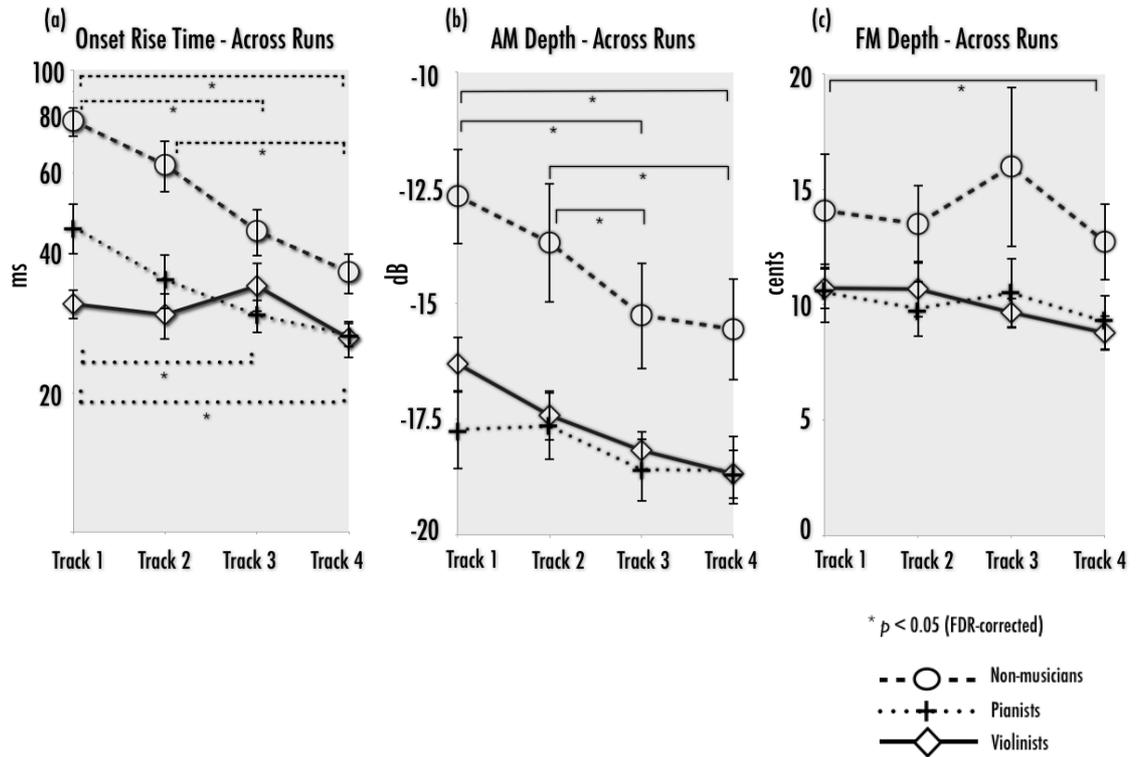
**2.3.1.5 AM depth.** All groups' detection of AM depth improved across the four runs (Figure 2.2b), as shown by the main effects of run (interaction with group non-significant), verified by analyses of random samples (see Table 2.5 & ST 2.4). Thresholds from 1st and 2nd runs were significantly higher than those from 3rd or 4th runs; later runs did not differ significantly (see ST 2.6). As in the

rise time analysis, non-musicians' final run did not differ significantly from the first run completed by musicians,  $\chi^2(2, n = 40) = 1.85, p = 0.4$  (Kruskal–Wallis).

**2.3.1.6 FM depth.** There was limited improvement in FM depth detection across runs (Figure 2.2c), with no interaction between run and group (see Table 2.5 and ST 2.6); the effect of run was also significant in just one random sample (see ST 2.5). Post-hoc comparisons showed only thresholds from run 1 and run 4 differed significantly (ST 2.6). As in the other two experiments, non-musicians' final run was not significantly different from musicians' first run,  $\chi^2(2, n = 42) = 1.59, p = 0.45$  (Kruskal–Wallis).

Table 2.5: 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 $\lambda$	df	$F$	$p$	$\eta_p^2$
<i>Rise time</i>					
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
<i>AM depth</i>					
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
<i>FM depth</i>					
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
<i>Rise time</i>					
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



**Figure 2.2:** 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  $\pm 1$  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 *ns* differ across panels - see 2.2.4.4.2 for description.

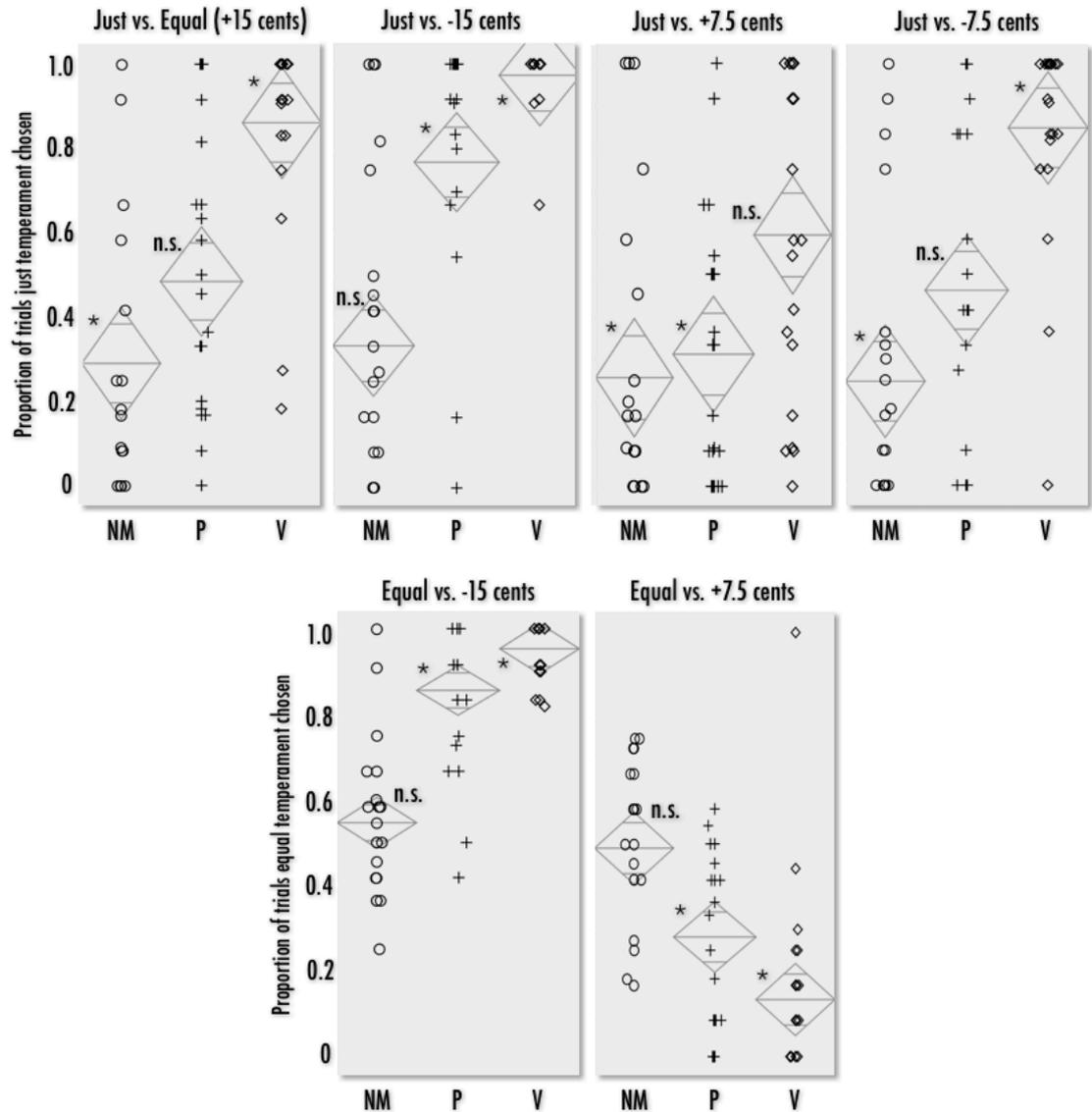
### 2.3.2 Visual psychophysical thresholds

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 colour hue psychophysical task. In contrast to the auditory psychophysical results, a one-way ANOVA showed no effect of group on visual colour hue (increment) thresholds,  $F(2, 67) = 1.76$ ,  $p = 0.18$ ,  $\eta_p^2 = 0.049$  (see Figure 2.1d).

### 2.3.3 Tuning system perception

We next asked whether expertise 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 2.7 (one-sample Wilcoxon Signed Rank Test (WRST)). Proportion of in-tune choices for each chord pairing were analysed across groups (Kruskal-Wallis and post-hoc WSRT; Table 2.6 & Figure 2.3).

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 2.3, panels 1–4; ST 2.7). The sole exception was just intonation paired with the moderately sharpened +7.5 cents chord (see 2.3.8). 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 2.3, panels 2.5 & 2.6).



**Figure 2.3:** In-tune choices for just and equal tempered tuning systems when paired with tuning deviances, across groups; upper 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); lower 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); \* 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.

**Table 2.6:** Kruskal-Wallis and post-hoc group comparisons across tuning perception task pairs (all post-hoc comparisons false discovery rate-corrected [ $\alpha = 0.05$ ] 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
<i>Kruskal-Wallis</i> $\chi^2 (2, n = 70)$	24.24 ***	30.98 ***	12.12 **	24.87 ***	36.98 ***	27.89 ***
<i>Post-hoc (WSRT)</i>						
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	$z = 2.07,$ <i>n.s.</i>	$z = 3.43^*$ Cohen's $d = 1.3$	$z = 0.94,$ <i>n.s.</i>	$z = 1.86,$ <i>n.s.</i>	$z = 4.57^*$ 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 $d = 1.2$	$z = 2.22,$ <i>n.s.</i>	$z = 2.87^*$ Cohen's $d = 0.7$

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

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 2.3, panels 2.3 & 2.6; 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 2.3, panel 2; ST 2.7). 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

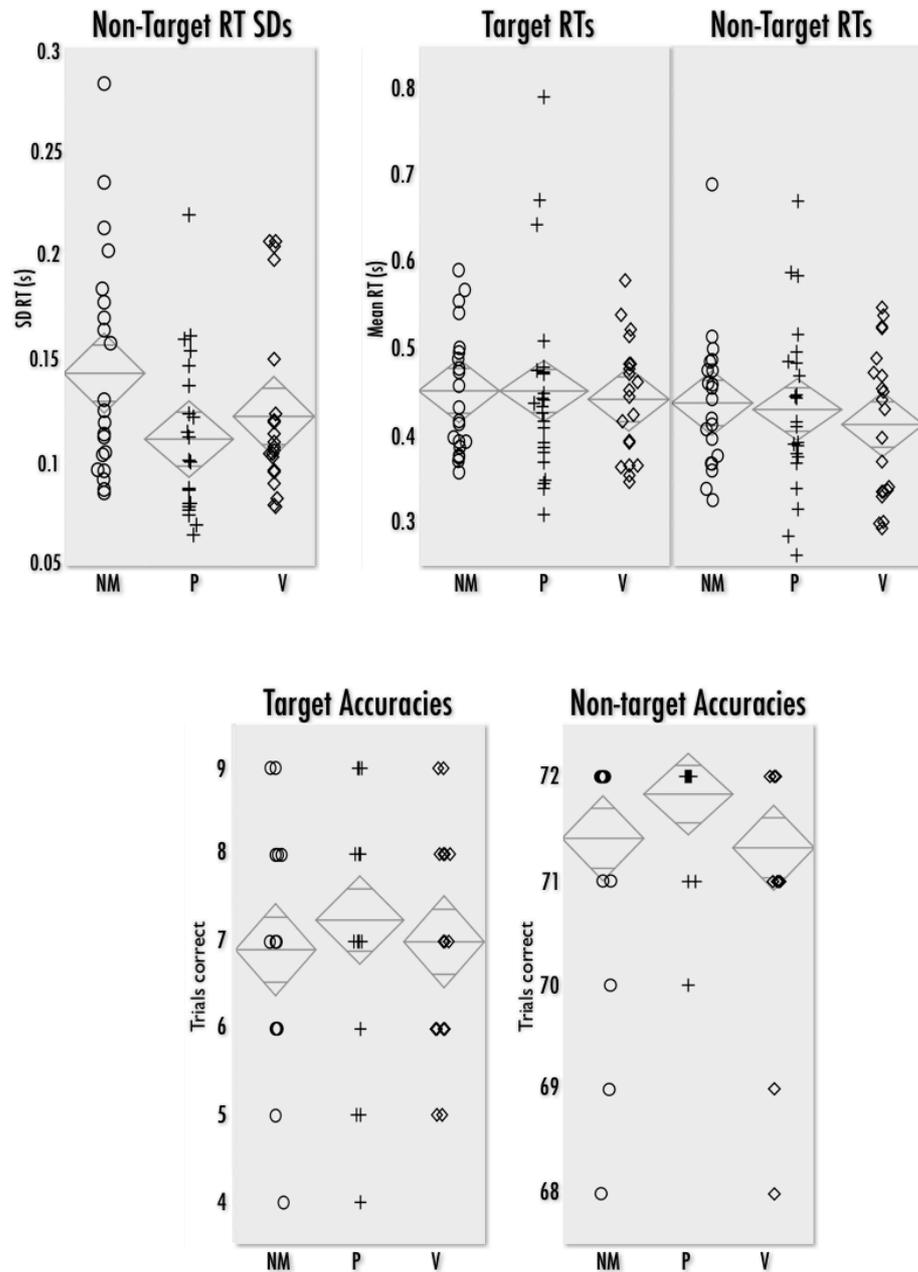
Figure 2.3, panels 1–4; ST 2.7). Neither did non-musicians select equal temperament significantly above chance when paired with the -15 cents or +7.5 cents chords (Figure 2.3, panels 5 & 6; ST 2.7).

Violinists' and non-musicians' choices differed significantly for every chord pair (see Table 2.6 & Figure 2.3). Violinists' choices also differed significantly from pianists' choices for every pair, except equal temperament vs. -15 cents (see Table 2.6; Figure 2.3).

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 2.6). 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 2.6).

#### **2.3.4 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,  $F(2, 65) = 0.32$ ,  $p = 0.73$ ,  $\eta_p^2 = 0.01$ , target response accuracy,  $F(2, 65) = 0.47$ ,  $p = 0.63$ ,  $\eta_p^2 = 0.01$  (one-way ANOVA) or non-targets response accuracy,  $\chi^2(2, n = 68) = 3.94$ ,  $p = 0.14$  (Kruskal–Wallis) (see Figure 2.4). RTs to targets and non-targets did differ,  $F(1, 65) = 9.95$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.133$ , with mean target RTs



**Figure 2.4:** 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.

slower than for non-targets (Figure 2.4). However, there was no significant interaction of target/non-target and group,  $F(2, 65) = 0.59$ ,  $p = 0.56$ ,  $\eta_p^2 = 0.02$ .

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,  $F(2, 65) = 3.08$ ,  $p = 0.053$ ,  $\eta^2 = 0.086$  (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 2.4, upper left panel).

### 2.3.5 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 familiarised 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,  $F(2, 68) = 2.42$ ,  $p = 0.096$ ,  $\eta_p^2 = 0.07$  (Figure 2.5). There was no main effect of familiarity,  $F(1, 68) = 0.08$ ,  $p = 0.77$ ,  $\eta_p^2 < 0.01$ , and no familiarity x group interaction,  $F(2, 68) = 0.82$ ,  $p = 0.45$ ,  $\eta_p^2 = 0.02$ . In sum, we found no significant evidence of enhanced general sequencing abilities in musicians, nor for participants being able to reproduce longer sequences when familiarised with the statistical regularities underlying those sequences.

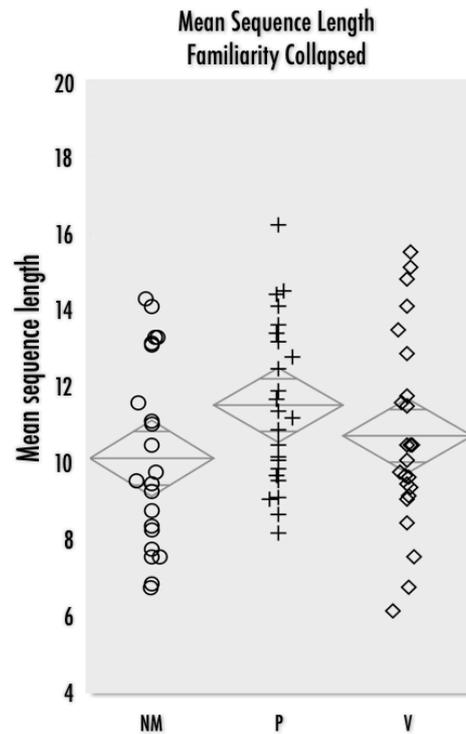


Figure 2.5: Group and individual mean sequence length performance on SIMON sequencing task; NM - non-musicians (circles); P - pianists (crosses); V - violinists (diamonds); large diamonds display means as middle horizontal line, and upper and lower bounds of 95% CI as uppermost and lowermost diamond tips, respectively.

### 2.3.6 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 2.7 and Supplemental Figure [SF] 2.2). 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 2.7: Significant ANOVA effects for percentage accuracy across EnvASA conditions

Effect	df	$F$	$p$	$\eta_p^2$
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

### 2.3.7 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.

**2.3.7.1 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 2.8).

Auditory psychophysical thresholds were also positively correlated with sustained attention performance (see Table 2.8). 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 2.8). 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 2.8).

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 2.9). Envasa, SIMON, tuning perception and visual psychophysical measures showed weaker loadings on the component.

**Table 2.8:** Non-parametric correlations between psychophysical tasks and SAART non-target RTs and SDs (false discovery rate-corrected; \*  $p < 0.05$ )

Pair	Spearman's $\rho$	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

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 2.9, 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 versus non-musician status and SAART non-target SDs (see Table 2.9, 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.

**2.3.7.2 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 2.10). However, pianists' FM depth thresholds were significantly predictive of their choice of just intonation vs. -15 cents tuning

**Table 2.9:** Stepwise regression models with musician/non-musician status and SAART performance as predictors of auditory psychophysical thresholds

Model	Adj. R <sup>2</sup>	$\beta$	df	F	p
<i>Rise time</i>	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
<i>Rise time</i>	0.328				
Musician vs. Non-musician		8.97	(1, 52)	22.66	< 0.0001
SAART RTs		36.73	(1, 52)	3.88	0.054
<i>FM depth</i>	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
<i>FM depth</i>	0.196				
Musician vs. Non-musician		2.75	(1, 62)	11.61	0.001
SAART RTs		17.51	(1, 62)	4.27	0.043
<i>AM depth</i>	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
<i>AM depth</i>	0.15				
Musician vs. Non-musician		1.0	(1, 62)	7.07	0.01
SAART RTs		11.09	(1, 62)	6.18	0.016

(see ST 2.10, 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 [ $F(1, 22) = 5.96, p = 0.02, \text{adjusted } R^2 = 0.177; \beta = 0.018$ ]; this relationship was not significant for the violinist group [ $F(1, 21) = 2.64, p = 0.12, \text{adjusted } R^2 = 0.07; \beta = 0.007$ ]; 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 ST 2.10).

### **2.3.8 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 2.2.4.1). Early practice significantly predicted only a single outcome variable: violinists who began formal practice at 3–4 years were more likely than later-starting violinists to choose just intonation when paired with the (slightly sharp) +7.5 cents chord [ $F(1, 19) = 5.31, p = 0.033, \text{adj. } R^2 = 0.177; \beta = -0.16$ ; 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 ( $\beta = 0.023, p > 0.7$ ; 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 [ $F(1, 19) = 3.80$ ,  $p = 0.066$ , adj.  $R^2 = 0.123$ ;  $\beta = -0.16$ ], but not pianists ( $\beta = -0.036$ ,  $p > 0.5$ ; 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.

## **2.4 General Discussion**

### **2.4.1 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).

## 2.4.2 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 colour 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; Goebel 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 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 2.2). 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 2.1). 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.

### **2.4.3 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 2.3). 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. 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 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 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 not 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.

#### **2.4.4 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 did 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).

#### **2.4.5 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 2.2) – 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 9), 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 & Shafiro, 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 SF 2.2) 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).

#### **2.4.6 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 generalisation 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 familiarised 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; François & Schön, 2011; François et al., 2014; 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 familiarised 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.

### 2.4.7 Expertise and generalisation

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 interval 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). 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. While we did not measure full-scale IQ (a limitation imposed by the current battery's 3+ hours duration), we believe that we matched our cohorts as closely as possible for intellectual level and general motivation (see footnote 7).

Contrary to our expectations, we did not find that musical expertise generalised to the cognitive metrics we tested here. This result is important as it suggests that even long-term intensive musical training may not strongly influence auditory and audiomotor skills that would seem to rely on similar processing mechanisms. 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). A lack of evidence for such generalisation (as in the present case) can be difficult to interpret, as the failure to find an effect of expertise may be driven by a lack of statistical power with the measure in question. While we cannot exclude this possibility in the present study, this explanation appears unlikely given the complete overlap in the distributions of scores across groups, the large N (48 professionally-trained, active young musicians and 24 non-musicians), the reliability-normed measures, and the close matching within and across groups on a variety of nuisance variables. If musicianship is to be studied as a model for plasticity – or as an intervention for hearing, attentional or language difficulties – then it is important that the limitations on generalisation be understood (Ruggles et al., 2014).

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.

Although these 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 lack of 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).

#### **2.4.8 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 generalisation 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 generalisation, 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.

This 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.

In the next chapter, we extend our investigation to consider cortical structure differences between the same cohort of violinists and cohort of non-musicians as explored in this chapter. This allows us to ask whether expert status relates to plastic enhancements to cortical regions involved in expert performance, and moreover, whether the metrics of experience and behavioural performance indexed in this chapter might account for variance in cortical structure adaptations. We probe these questions using high-resolution MR methods that provide quantitative proxies for cortical myelin ( $R_1 = 1/T_1$ ).

## Chapter 2 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 [ $\chi^2 (1, n = 48) = 6.15, p = 0.013, 22/24$  violinists,  $15/24$  pianists]. Violinists typically reported that their second instrument (primarily piano; see table 2.2) 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 2.3).
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). 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. 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.

## Chapter 2 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.

Supplemental table 2.1: 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 2.2: Kruskal-Wallis and post-hoc Wilcoxon Signed Rank Test comparisons for MSI musical training subscale scores

Model	Test statistic
Kruskal–Wallis $\chi^2 (2, n = 70)$	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:  $6.9 \pm 5.8$  [SD]; pianists:  $0.6 \pm 1.4$  [SD];  $z = 4.89$ ,  $p < 0.0001$ , Cohen's  $d = 1.5$ ) and small ensembles (violinists:  $6.6 \pm 5.2$  [SD]; pianists:  $3.7 \pm 5.4$  [SD];  $z = 2.78$ ,  $p = 0.0054$ , Cohen's  $d = 0.5$ ).

Supplemental table 2.3: 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 $\lambda$	df	$F$	$p$	$\eta_p^2$
<i>Sample 1</i>					
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
<i>Sample 2</i>					
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
<i>Sample 3</i>					
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
<i>Sample 4</i>					
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
<i>Sample 5</i>					
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
<i>Sample 6</i>					
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 2.4: 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 $\lambda$	df	$F$	$p$	$\eta_p^2$
<i>Sample 1</i>					
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
<i>Sample 2</i>					
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
<i>Sample 3</i>					
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
<i>Sample 4</i>					
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
<i>Sample 5</i>					
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
<i>Sample 6</i>					
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 2.5: 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 $\lambda$	df	$F$	$p$	$\eta_p^2$
<i>Sample 1</i>					
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
<i>Sample 2</i>					
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
<i>Sample 3</i>					
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
<i>Sample 4</i>					
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
<i>Sample 5</i>					
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
<i>Sample 6</i>					
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 2.6: 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
<b>Rise time</b>						
Pianists ( <i>n</i> = 16)	<i>t</i> (15) = 2.95 *	<i>t</i> (15) = 4.30 *	<i>t</i> (15) = 5.47 *	<i>t</i> (15) = 1.87	<i>t</i> (15) = 2.66 *	<i>t</i> (15) = 1.61,
	Cohen's <i>d</i> = 1.5	Cohen's <i>d</i> = 2.2	Cohen's <i>d</i> = 2.8	<i>n.s.</i>	Cohen's <i>d</i> = 1.4	<i>n.s.</i>
Non-musicians ( <i>n</i> = 12)	<i>t</i> (11) = 1.52,	<i>t</i> (11) = 3.92 *	<i>t</i> (11) = 6.08 *	<i>t</i> (11) = 2.9 *	<i>t</i> (11) = 4.86 *	<i>t</i> (11) = 1.76,
	<i>n.s.</i>	Cohen's <i>d</i> = 2.0	Cohen's <i>d</i> = 2.2	Cohen's <i>d</i> = 3.1	Cohen's <i>d</i> = 2.5	<i>n.s.</i>
<b>AM</b>						
All subs ( <i>n</i> = 40)	<i>t</i> (39) = 1.87,	<i>t</i> (39) = 4.95 *	<i>t</i> (39) = 5.57 *	<i>t</i> (39) = 2.81 *	<i>t</i> (39) = 3.69 *	<i>t</i> (39) = 0.97,
	<i>n.s.</i>	Cohen's <i>d</i> = 2.6	Cohen's <i>d</i> = 2.9	Cohen's <i>d</i> = 1.5	Cohen's <i>d</i> = 1.9	<i>n.s.</i>
<b>FM</b>						
All subs ( <i>n</i> = 42)	<i>t</i> (41) = 0.66,	<i>t</i> (41) = 0.17,	<i>t</i> (41) = 2.59 *	<i>t</i> (41) = 0.54,	<i>t</i> (41) = 2.21,	<i>t</i> (41) = 2.05,
	<i>n.s.</i>	<i>n.s.</i>	Cohen's <i>d</i> = 1.3	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>

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

Supplemental table 2.7: One-sample Wilcoxon signed rank tests of difference of group mean from chance, per tuning pair (false discovery rate-corrected [ $\alpha = 0.05$ ] 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, <i>n.s.</i>	- 80 *	- 93.5 *	26.5, <i>n.s.</i>	- 5, <i>n.s.</i>
P	- 6.5, <i>n.s.</i>	101 *	- 66.5 *	- 24.5, <i>n.s.</i>	137 *	-109 *
V	132 *	138 *	45.5, <i>n.s.</i>	118.5 *	138 *	-120.5 *

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

Supplemental table 2.8: Non-parametric correlations across auditory and visual psychophysical tasks (false discovery rate corrected; \*  $p < 0.05$ )

Pair	Spearman's $\rho$	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 2.9: 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 2.10: Non-parametric correlations between tuning task pairs, AM depth and FM depth thresholds, across groups (false discovery rate-corrected; \*  $p < 0.05$ )

Pair	Pianists		Violinists		Non-musicians	
	$\rho$	FDR-corrected $p$	$\rho$	FDR-corrected $p$	$\rho$	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 2.11: 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 2.12: 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 2.13: 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).

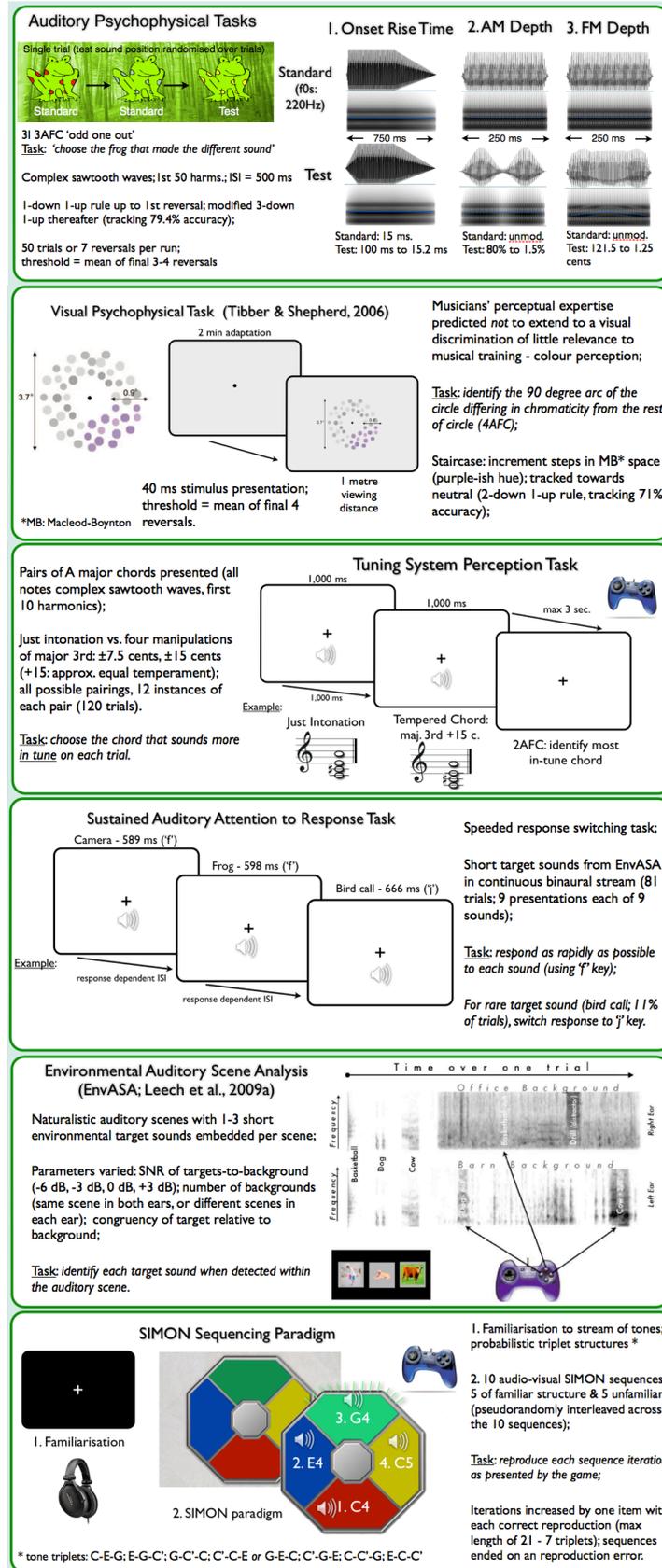
	Just vs. +15	Just vs. -15	Just vs. +7.5	Just vs. -7.5	+15 vs. -15	+15 vs. +7.5
Spearman	0.302 $p = 0.2$ n.s.	0.454 $p = 0.044$ *	0.462 $p = 0.040$ *	0.518 $p = 0.019$ *	0.167 $p = 0.48$ n.s.	0.485 $p = 0.03$ *

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

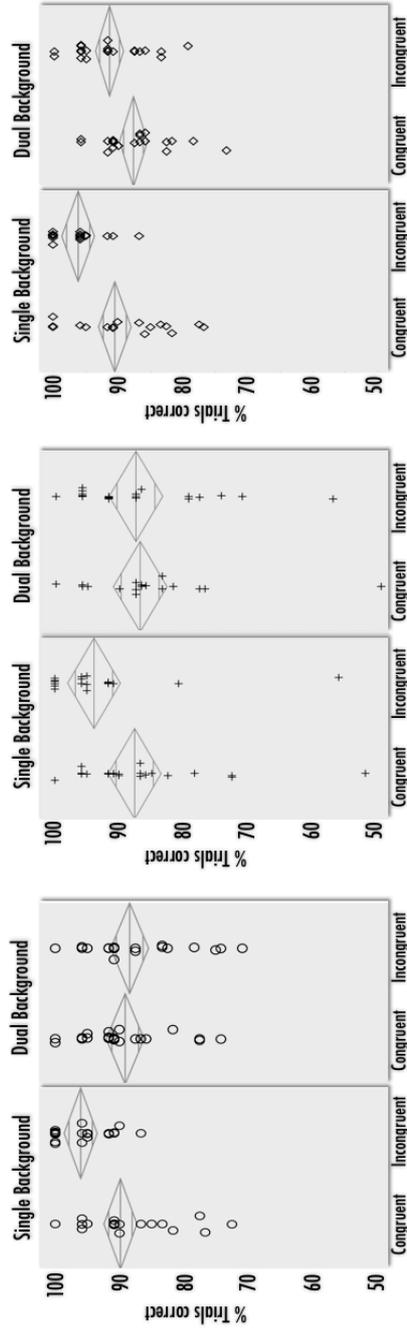
Supplemental Table 2.14: Spearman correlation coefficients for test-retest reliability analyses of tuning perception paradigm ( $n = 10$ ) across testing runs 1 and 2 (for each possible chord pairing; preference judgements).

	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.

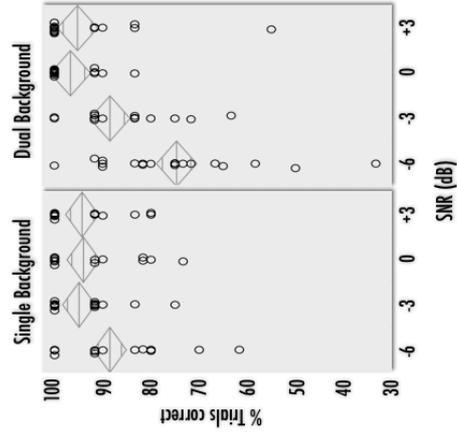
\*  $p < 0.05$  \*\*  $p < 0.01$



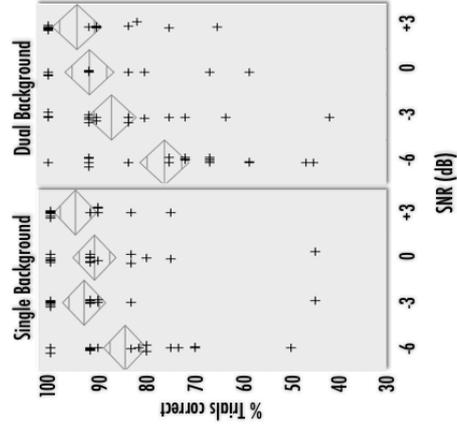
Supplemental Figure 2.1: Graphical summary of the procedure for each experimental task used within the battery.



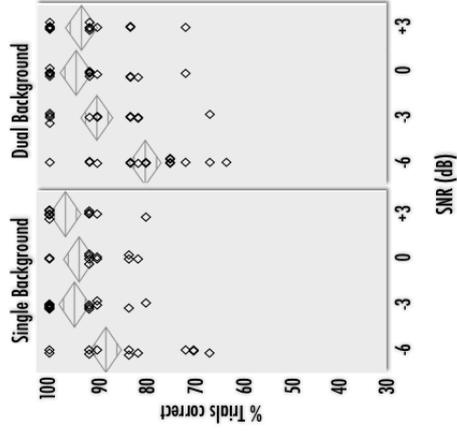
### Non-musicians



### Pianists



### Violinists



Supplemental Figure 2.2: 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.

Chapter 3: Musical Training Effects and  
Cortical Plasticity: Relationships with  
Training Extent and Behavioural  
Performance

### 3.1 Introduction

The human brain displays a remarkable capacity to adapt to the pressures and demands posed by the environment. Particular environmental experience has the potential to spur an array of changes to brain structure (e.g., Buonamano & Merzenich, 1998; Draganski & May, 2008). Yet relating complex changes in neural structure to behavioural and cognitive abilities can be problematic – particularly where behaviours are multi-faceted (see May, 2011; Ramsden et al., 2011; Zatorre et al., 2012a). Mechanistic accounts of relationships between *in vivo* metrics of brain structure and complex, real-world behavioural experience can thus be difficult to establish (see Hyde et al., 2009, for discussion).

As discussed in chapter 1, studies of expert populations offer one means of addressing these issues. Where a group spends substantial time focused on a well-defined task from an early age, some of the variability of environmental factors may be accounted for through common profiles of behaviour (e.g., Ericsson et al., 1993). Moreover, the common demands of shared experience allow clear, testable predictions as to where in the brain structural differences may emerge and what their behavioural significance might be (e.g., Teki et al., 2012; Draganski et al., 2004; Maguire et al., 2000).

Expert musicians provide an excellent model to test these assumptions. The well-defined nature of musical practice (based on established pedagogy, formal technique, ear training, repertoire, etc.) and the audio-motor demands musicians train under allow specific predictions concerning structural adaptations – and perceptual and cognitive outcomes – tied to musical training (Han et al., 2009; Hyde et al., 2009; see Zatorre et al., 2012a).

Indeed, a number of brain structure changes have been attributed to effects of musical training (see Herholz & Zatorre, 2012). However, measurement of musical-training related plasticity within the human brain has largely been based on metrics of gross or voxel-based morphometry [VBM] (Bangert & Schlaug, 2006; Amunts et al., 1997; Gaser & Schlaug, 2003; Sluming et al., 2002; Han et al., 2009; Groussard et al., 2010), cortical thickness (Bermudez et al., 2009) or diffusion tensor imaging metrics (Halwani et al., 2011; Oechslin et al., 2010; Bengtsson et al., 2005; Imfeld et al., 2009; Han et al., 2009). Measures such as VBM and cortical thickness provide indirect proxies for plasticity within tissue subtypes, but are limited as they do not allow for a single tissue-specific parameter to be measured (i.e., multiple properties of tissue classes such as vasculature, cell bodies, myelin and glial cells can contribute to the measured volumetric or thickness change; see May, 2011; Draganski & May, 2008). Similarly, changes in white matter structure as quantified with DTI metrics (FA, diffusivity, etc.) can be influenced by differences in myelination, fibre orientation, and tract volume; thus, quantitative indices derived from DTI are non-specific to a single tissue property (Zatorre et al., 2012a). Moreover, the correspondence between changes in structural brain metrics (both VBM and DTI) and related behavioural outcomes has not been investigated in some studies (e.g., Schmithorst & Wilke, 2002; Han et al., 2009).

The main goal of the present study was to provide a quantitative assessment of structural brain change indexing expression of cortical myelin, as it relates to musical expertise and long-term training. We sought to compare quantitative myelin metrics across experts and non-experts to explore whether musical training might be associated with myelin adaptations within specific auditory and motor cortical regions. Moreover, we examined whether training

metrics and behavioural indices of fine auditory perception would relate to myelin proxies within the same auditory cortical regions. Below, we review literature demonstrating both training-dependent brain structure adaptations (and related behavioural outcomes) in musicians, using non-quantitative imaging metrics.

**3.1.1. Musical training and cortical plasticity.** Studies of musical expertise and training intensity provide evidence to support experience-dependent plasticity. Investigations with adults have found that musical proficiency shows robust associations with neural structure. For instance, Gaser and Schlaug (2003) found that professional musicians had significantly greater grey matter volumes in left pre-central gyrus, left Heschl's gyrus and right superior parietal cortex compared to amateurs; amateurs similarly had enhanced grey matter volumes in these regions compared to non-musicians. Further, Schneider et al. (2002) found that performance on a standardised metric of musical aptitude related linearly to increases in grey matter volume averaged over bilateral antero-medial Heschl's gyri; professionals showed highest aptitude and greatest grey matter volumes, followed by amateurs and non-musicians (see also Schneider et al., 2005). In addition, years of formal training can account for grey matter volume increases in inferior frontal gyrus of professional musicians when controlling for age (Sluming et al., 2002).

Furthermore, learning to play one instrument versus another may reveal effects of training-dependent plasticity. Such effects have been shown via differential volumetric changes in primary motor, somatosensory and cortico-spinal tract regions in pianists compared to other musicians (Gartner et al., 2013), as well as greater tract volume in left arcuate fasciculi of singers

compared to instrumentalists (Halwani et al., 2011; but see Bengtsson et al., 2005).

**3.1.2. Musical training, behavioural performance and plasticity.** In tandem, a range of studies have shown changes in cortical and subcortical regions tied to behavioural measures that relate to musical training. Longitudinal data indicate that compared to an untrained control group, 6 year old children followed-up after 15 months of instrumental training had relative increases in voxel deformation-based metrics at right pre-central gyrus and posterior corpus callosum (suggesting relative expansion of tissue from pre- to post-training); further, voxel deformation in these regions correlated positively with improved left hand finger tapping performance (Hyde et al., 2009; see also Norton et al., 2005). Children engaged in training also showed increases in voxel deformation-based metrics at right Heschl's gyrus that correlated positively with improved melodic and rhythmic discrimination (Hyde et al., 2009). A recent longitudinal study suggested similar findings; children engaged in musical training showed increases in right Heschl's gyrus grey matter volumes that were predicted by intensity of practice between the initial scans and those taken at follow-up (13 months later) (Seither-Preisler et al., 2014). Moreover, the ratio of Heschl's gyrus to planum temporale volume at right hemisphere accounted for significant variance in children's musical aptitude at follow-up; frequency discrimination thresholds were also significantly negatively correlated with volumes of left and right Heschl's gyri at follow-up (Seither-Preisler et al., 2014).

Studies of adult musicians have also shown structural adaptations that relate to behavioural performance and indices of training. For instance, FA in adults' left temporal lobes positively correlated with synchrony of tapping with a visually-cued stimulus (in musicians with late training onset and in non-

musicians); moreover, age of training onset showed a negative relationship with FA in the same temporal regions (Steele et al., 2013). Furthermore, musicians who began training before age 7 also had increased fractional anisotropy (FA) and decreased radial diffusivity at the posterior mid-body and isthmus of the corpus callosum (compared to non-musicians and adult musicians who began after age 7) (Steele et al., 2013; see also Han et al., 2009). Musicians with greater practice intensity earlier in development also show increased FA in the posterior limbs of the internal capsule, as well as isthmus and splenium of the corpus callosum (Bengtsson et al., 2005).

Further studies have considered whether highly advanced performance ability such as absolute pitch (AP) in musicians is related to differences in neural structure, with varied findings. Several studies have shown increased left-ward asymmetry of planum temporale (PT) area and volume in musicians with AP, relative to non-musicians and non-AP musicians ( Schlaug et al., 1995; Luders et al., 2004), as well as an overall increase in left (but not right) PT volume in AP musicians versus non-musicians (Zatorre et al., 1998). However, one study showed relatively thinned cortex across a range of regions (bilateral superior frontal gyri, right pre-central and inferior frontal gyri, and left post-central gyrus) in musicians who had high AP ability versus musicians who did not (Bermudez et al., 2009; cf. Dohn et al., 2013). Further, DTI findings have revealed individual differences in musicians with absolute pitch (AP): AP musicians who had increased FA in three clusters within the superior longitudinal fasciculus tended to make more errors on a pitch identification task (no such relationship was seen in non-AP musicians; Oechslin et al., 2010). Notably, musicians with and without AP have been found to show no difference in diffusion parameters within the cortico-spinal tract (the fibre bundles linking

cortical hand areas to the spinal cord via the cerebral peduncles) (Imfeld et al., 2009).

**3.1.3. The present study.** The results described in the above sections suggest that musical training plays a key role in mediating structural change in a range of cortical regions and subcortical tracts. Such changes reflect adaptations to regions critical to fine auditory processing (Heschl's gyrus: Hyde et al., 2009; Schneider et al., 2002; Gaser & Schlaug, 2003), motor performance (pre-central gyrus and cerebellum: Hyde et al., 2009; Gaser & Schlaug, 2003; Gartner et al., 2013; Amunts et al., 1997; Hutchinson et al., 2003), and sequential processing and cognitive control (inferior frontal gyrus: Sluming et al., 2002; Bermudez et al., 2009). Similarly, enhancements to white matter tracts may be essential to musicians' fine motor performance (cortico-spinal tract: Imfeld et al., 2009), and integration of information across auditory and motor regions (arcuate fasciculus: Halwani et al., 2011; Bengtsson et al., 2005).

However, in spite of evidence of correlations between training, behavioural performance and structural brain change, no single study yet has compared detailed assessments of musical expertise and training – as well as training-relevant perceptual and cognitive skills – to quantitative metrics of brain structure. Quantitative imaging metrics (where measured signal across voxels relates to a specific tissue property), combined with detailed behavioural assessments can greatly inform understanding of how experience (e.g., musical training) relates to tissue-specific structural plasticity, and can help to explain the behavioural significance of any observed structural change.

This study addresses these issues by providing the first investigation of brain structural differences in musicians and non-musicians, using a quantitative

imaging method ( $R_1$  [ $1/T_1$ ] mapping) offering a proxy for myelination within human cortex (see Dick et al., 2012; Sereno et al., 2012; Sigalovsky et al., 2006). We investigated whether profiles of long-term violin training would be associated with differences in myelination within auditory and motor regions. Further, we conducted analyses of brain structure-behaviour relationships, based on assays of fine-grained auditory perception (psychophysical thresholds for instrument-relevant acoustic parameters), and cognitive performance (auditory attention), previously collected in the same participants. By comparing expert violinists and non-musicians, we could assess whether profiles of long-term training with very particular auditory (e.g., fine training of intonation and temporal sensitivity) and motor (e.g., fine training of left hand digits) demands would be associated with changes in myelination within auditory and motor cortex. Moreover, we could examine whether any such changes in myelination could be accounted for by perceptual or cognitive performance on a range of training-relevant measures, or by onset, duration and intensity of training.

## 3.2 Method

### 3.2.1 Participants

Participants were 21 violinists (mean age  $\pm$  SD: 23.1  $\pm$  2.9; 6 male) and 19 non-musicians (mean age  $\pm$  SD: 23.3  $\pm$  3.0; 6 male), drawn from the same cohort as recruited for experiments in chapter 2. All were right-handed (Edinburgh Handedness Inventory; mean  $\pm$  SD: violinists: 84.6  $\pm$  19; non-musicians 83.1  $\pm$  12.8;  $z = 0.99$ ,  $p = 0.32$ ). Violinists had trained extensively with violin (mean years training  $\pm$  SD: 16.7  $\pm$  3.9), and had some experience with secondary instruments (see table 3.1). None reported actively practicing their second instruments at the time of the study. Five non-musicians (2 female, 3 male) had never played any musical instrument or taken music lessons. Fourteen non-musicians (11 female, 3 male) had taken elementary music lessons during childhood or adolescence, but had not attended a formal music college or practiced daily over an extended period (see table 3.2). On average, those non-musicians with musical experience had not practiced for 12.1 years (SD = 3.8; range = 6–19 years) prior to the study.

### 3.2.2 Data acquisition

Structural images were acquired at the Wellcome Trust Centre for Neuroimaging using a 3T whole-body Tim Trio system (Siemens Healthcare) with radiofrequency body transmit and 32-channel receive head coil. Scans used for the quantitative  $R_1$  mapping protocol comprised proton density-weighted (PDw) and  $T_1$ -weighted ( $T_{1w}$ ) images. Images were acquired using an in-house multi-echo 3D FLASH pulse sequence (see Weiskopf et al., 2011; Dick et al., 2012).

**Table 3.1:** Violinists' ( $n = 21$ ) descriptive data and musical training histories

Participant	Gender	Age	Violin training (years)	Other instruments	Other instruments - years played
v1	F	23	19	Viola	3
v2	F	22	14	Piano	6
v3	M	19	12	Piano	7
v4	F	20	17	Piano; Viola; Trumpet	7; 7; 7
v5	F	23	19	Piano	2
v6	F	20	12	Piano; Viola	12; 4
v7	M	21	17	Piano	9
v8	M	24	21	Piano; Alto Saxophone	12; 13
v9	M	26	20	Piano	5
v10	F	25	11	Viola	1
v11	F	21	18	Viola	6
v12	M	28	20	None	
v13	F	30	27.5	Piano	20
v14	F	25	18	Piano; Viola	missing data
v15	F	22	14	Piano; Viola	1.5; 1
v16	F	23	16	Bass Guitar	6
v17	M	22	17	Piano; Viola; Voice	5; 2; 7
v18	F	19	12	Piano	8
v19	F	20	13	Piano	2
v20	F	26	17	Piano	5
v21	F	23	16	Piano; Trumpet	2; 2

Image parameters were as follows: voxel size: 0.8 x 0.8 x 0.81 mm<sup>3</sup>; image matrix: 280 x 320 x 208 mm; bandwidth: 460 Hz/pixel; echo times (TE) (echos 1 - 8): 1) 2.39 ms; 2) 4.75 ms; 3) 7.11 ms; 4) 9.47 ms; 5) 11.83 ms; 6) 14.19 ms; 7) 16.55 ms; 8) 18.91 ms; TR: 25.25 ms; excitation flip angles: 5° (PDw); 29° (T<sub>1w</sub>); 8 gradient echoes acquired and averaged to increase signal-to-noise ratio (see Helms et al., 2009).

**Table 3.2:** Non-musicians' ( $n = 19$ ) descriptive data and musical training histories

Participant	Gender	Age	Musical training (years)	Instrument	Years since practised
nm1	F	24	4	Piano	14
nm2	F	20	0.25	Saxophone	9
nm3	F	29	0		
nm4	F	28	3	Piano	18
nm5	F	20	0		0
nm6	F	21	0.5	Guitar	6
nm7	M	27	5	Piano	16
nm8	F	19	1	Piano	10
nm9	M	26	0		
nm10	M	19	0		
nm11	M	22	3	Cornet	9
nm12	F	22	3.5	Piano; Violin	12
nm13	F	24	3	Saxophone	10
nm14	F	21	0.5	Piano	8
nm15	F	23	1	Keyboard	19
nm16	F	26	2	Piano	14
nm17	F	25	4	Violin	13
nm18	M	22	1	Voice	10
nm19	M	21	0		

Image acquisition was sped up via 2 x GRAPPA parallel imaging in the phase encoding direction and by 6/8 Partial Fourier in the partition direction. In addition to the PDw and T<sub>1</sub>w images, a further magnetisation transfer-weighted (MTw) scan was acquired (parameters identical to the T<sub>1</sub>w and PDw scans, with exception of flip angle [9°] and TR [29.25 ms]). MTw images were included as part of the multi-parameter mapping (MPM) protocol (see Weiskopf et al., 2011; 2013).

For full sets of MPM images, a different slab orientation was used during acquisition for some subjects compared to others. Initial inspection of data acquired with the slab aligned to each cardinal axis showed susceptibility artifact that affected cortex in a subset of participants. Although eye movements were monitored during scanning runs (see below), slight movement (e.g., due to blinking) led to artifact within orbitofrontal and medial temporal lobes in some datasets. To counter this issue, the acquisition protocol was modified, by rotating each MPM image slab at 30° about the x-axis (such that the eyes lay outside the slab; see Supplemental Figure 3.1). Participants with data acquired without slab rotation were inspected blind to subject and group for evidence of susceptibility artifact; those participants that showed evidence of artifact within cortical areas were re-scanned using the rotated acquisition protocol. In total, 6 participants (3 violinists, 3 controls) showed susceptibility artifact with the unrotated acquisition and were re-scanned with the rotated protocol; 15 participants (9 violinists, 6 non-musicians) showed no evidence of susceptibility artifact with the original unrotated acquisition and were not re-scanned; 13 participants were scanned using the rotated protocol as default (6 violinists, 7 controls). A whole-brain analysis using slab rotation as a regressor of interest showed no significant differences across any vertices over either hemisphere for participants with rotated versus unrotated acquisition (uncorrected threshold;  $p < 0.001$ ).

Two further scans were collected to estimate inhomogeneities in the B1 and B0 fields. Maps of the RF transmit field ( $B1^+$ ) were used to correct the images for effects of RF transmit inhomogeneities, using the 3D echoplanar imaging spin-echo (SE)/stimulated echo (STE) method described in Lutti et al. (2010) (slice thickness: 4 mm; matrix size: 64 x 48 x 48; field-of-view: 256 x 192

x 192 mm<sup>3</sup>; bandwidth: 2298 Hz/pixel; TE<sub>SE</sub>/TE<sub>STE</sub>: 33.2 ms/67.73 ms; TR: 500 ms; flip angle SE/STE: 160-200°/80-100° by steps of 10°/5°, respectively). In addition, a map of the B<sub>0</sub> field was acquired and used to correct the B<sub>1+</sub> map for off-resonance effects (Lutti et al., 2010; see also Weiskopf et al., 2006) (voxel size: 3 x 3 x 2 mm<sup>3</sup>; slice thickness: 4mm; field-of-view: 192 x 192 mm<sup>2</sup>; 64 slices, 1mm gap; bandwidth: 260 Hz/pixel; TE 10 ms; TR: 1020 ms; flip angle: 90°). Image slabs for field maps were all non-rotated.

### **3.2.3 Procedure**

All participants provided signed voluntary informed consent prior to commencement of scanning, in line with the local ethics committee protocol. Participants were briefed on the scanning procedure and were informed of the issues posed by head movements during scanning. To reduce possible head and saccade related artifacts, participants fixated a cursor presented centrally on-screen, whilst watching a subtitled film of their choice. The cursor was located immediately above the subtitles, positioned such that subtitles could be read without breaking fixation. Participants practised reading whilst maintaining fixation for several minutes prior to scanning. In addition, participants' eye and head movements were monitored using an eye tracker (Eyelink 1000 Core System) during scanning runs. Rest breaks of several minutes were provided between scans as required. Total acquisition time was 1 hour.

### **3.2.4 Data pre-processing**

Images were pre-processed using an in-house toolbox (Voxel Based Quantification; VBQ) running in SPM 8 via Matlab (2012a; 7.14.0). Images from the multi-parameter protocol were reconstructed using the B<sub>1+</sub> maps to correct

for inhomogeneities in local flip angles (and thus non-uniformity of  $T_1$  values) within each tissue type (as described in Helms et al., 2008, 2009; see also Lutti et al., 2010).

The procedure involved first using the  $B_0$  field map to correct for susceptibility-induced geometric distortions in the  $B_1^+$  maps (see Lutti et al., 2010). The  $B_1^+$  mapping method allowed determination of correct local flip angle values by repetition of the  $B_1$  image acquisition using a range of values for the nominal flip angles (see Data Acquisition, 3.2.2); this was followed by a linear regression of nominal versus local flip angle values (Lutti et al., 2010). The square root of the residual mean square (RMS) of the linear regression model fit was calculated at each voxel and then divided by the number of nominal flip angle values, thereby producing a map of error values (RMS map). RMS maps were corrected for  $B_0$  distortions, and were then used to identify voxels that showed a poor fit with the linear regression (Lutti et al., 2010). Poorly fitting voxels were masked out of the  $B_1^+$  maps; flip angles that were omitted as a result of this masking were estimated by averaging flip angles from neighbouring voxels (RMS padding; see Lutti et al., 2010). Images from the multi-parameter protocol (PDw,  $T_{1w}$  & MTw) were reconstructed by separately averaging the gradient echoes acquired for each scan type, with subsequent correction for local flip angle inhomogeneities using the  $B_1^+$  map calculated for each subject (see Helms et al., 2008, 2009).

Quantitative  $R_1$  (i.e.,  $1/T_1$ ) maps were reconstructed according to the variable flip angle procedure described in Weiskopf et al. (2011). The procedure estimates the local  $R_1$  value by employing two different FLASH images (PDw and  $T_{1w}$ ) with different nominal flip angles [ $\alpha_1$  &  $\alpha_2$ ; in the present experiments:  $\alpha_1 = 5^\circ$  (PDw) and  $\alpha_2 = 29^\circ$  ( $T_{1w}$ )]. Following equation 1 from Weiskopf et al.

(2011), quantitative  $R_1$  values were estimated at each voxel based on the rational approximation of the Ernst equation,

$$(1) R_{1app} = \frac{1}{2} \frac{S_2\alpha_2/TR_2 - S_1\alpha_1/TR_1}{S_1/\alpha_1 - S_2/\alpha_2}$$

where  $S_1$  and  $S_2$  indicate the signal amplitude at each voxel for PDw and  $T_{1w}$  images respectively, and  $TR_1$  and  $TR_2$  the TRs for the PDw and  $T_{1w}$  images, respectively (see Helms et al., 2008). Note that the  $\alpha$  parameters within the  $R_1$  calculations were based on the corrected estimates of local flip angles as derived from the  $B1^+$  mapping procedure described above (rather than the nominal flip angle as specified on the scanner console; see Weiskopf et al., 2011).

Following reconstruction of multi-parameter images, all images were manually inspected blind to subject and group identity for any evidence of alignment difficulties, head movement or other image artifacts (e.g., image aliasing).

Six participants showed evidence of head movement artifacts and were excluded from analyses. Thus, 18 violinists (mean age  $\pm$  SD:  $22.83 \pm 2.8$ ; mean training  $\pm$  SD:  $16.7 \pm 4.1$ ; 5 male) and 16 non-musicians (mean age  $\pm$  SD:  $23.25 \pm 3.1$ ; mean training  $\pm$  SD [those with training;  $n = 13$ ]:  $2.1 \pm 1.5$ ; 4 male) were retained for cortical  $R_1$  analyses (there were no significant differences in age [ $z = 0.3$ ,  $p > 0.7$ ] or gender [ $\chi^2(1, n = 34) = 0.03$ ,  $p > 0.85$ ] between violinists and non-musicians retained for analyses).

### 3.2.5 Cortical surface reconstruction

Participants' cortical surfaces were reconstructed using FreeSurfer (5.3 for Mac OS 10.7) (see Dale et al., 1999). Use of  $R_1$  images as input to FreeSurfer can lead to localised tissue segmentation failures due to boundaries between the pial surface, dura matter and CSF showing different contrast values compared to those assumed within FreeSurfer algorithms (for discussion, see Dick et al., 2012; Sereno et al., 2012). Therefore, an in-house FreeSurfer surface reconstruction procedure was developed to overcome some of these issues. Description follows below.

**3.2.5.1. Image synthesis.** First, two synthetic FLASH volumes were created using the FreeSurfer `mri_synthesize` routine. Inputs to the routine were a scaled version of the quantitative PD scan (produced by VBQ toolbox, with negative image values removed), and a scaled and truncated  $T_1$  image (i.e., the reciprocal of the  $R_1$  values  $[1/R_1]$  as produced by the VBQ toolbox, also with negative values removed). For the first synthetic image, default FreeSurfer contrast parameters were specified. The second synthetic image was produced using the same PD and  $T_1$  input volumes, but with the following synthetic contrast parameters specified:  $TR = 20$  ms;  $\alpha = 30^\circ$ ;  $TE = 2.5$  ms. During synthesis, images were re-sampled to  $1\text{ mm}^3$  isotropic resolution in FreeSurfer. Image scaling and truncation, removal of negative values and  $T_1$  (i.e.,  $1/R_1$ ) calculation were performed on volumes using the AFNI `3dcalc` routine (Cox, 1996). Both synthetic images were then further scaled with AFNI `3dcalc`; this additional scaling yielded image intensity properties closer to the optimal intensity values required to segment tissue boundaries in FreeSurfer. The image synthesised with default contrast parameters was used as the main input to the FreeSurfer automated processing stream (following further pre-

processing steps; see below). The image synthesised with specified contrast parameters was used at a later stage as input to the FreeSurfer Talairach transformation. Finally, a version of the PD volume was produced with AFNI 3dcalc; negative numbers were removed and the range of values was truncated to eliminate high values corresponding with noise from non-neural tissue. This adjusted PD volume was used as input to the skull strip procedure (see below).

**3.2.5.2. Manual image adjustment.** Magnetic susceptibility issues (for instance, in anterior and inferior temporal regions) can lead to low image intensity values, often causing segmentation errors at grey-white and grey-CSF boundaries in FreeSurfer. To counter these difficulties, the FLASH image synthesised with default parameters was further adjusted using an in-house version of FreeSurfer (Csurf). Each subject's synthetic image was hand-adjusted using a piecewise linear normalisation procedure to linearly ramp brightness values of grey and white matter within isolated regions. Brightness values of voxels within target regions were iteratively multiplied by factors of either 1.1 or 1.2. Regions adjusted for all subjects included inferior and medial temporal lobes, temporal pole, long and short insular gyri, and ventro-medial pre-frontal cortex. Manual blink comparison between the synthetic volume and the labelled white matter surface was used to compare adjustments as each brightening iteration was applied. Care was taken to ensure that manual brightening did not cause grey and white matter to exceed the intensity value bounds specified for those tissue classes in FreeSurfer (grey matter: 50-70; white matter: 100-140). Manually brightened synthetic images were saved and used within the skull strip procedure.

**3.2.5.3. Skull strip.** Next, the subject's adjusted quantitative PD volume (see 3.2.5.1) was used as input to a customised skull strip procedure run in

Csurf. Briefly, the skull strip procedure removed the skull and regions exterior to it from the image volume, rendering an image of remaining brain tissue (including cerebellum and brainstem). First, an elliptical surface (4th or 5th geodesic subtessellation of an icosahedron) was expanded from inside the PD volume, with expansion of the surface constrained by arrival at low intensity voxels (i.e., those containing CSF and/or the inner surface of the skull). The set of voxels intersecting the faces of the resulting surface was then flood-filled from the outside, thereby constraining the brain volume to the brighter voxels inside the surface region. Using this PD volume as a mask, flood-filled voxels in the volume were used to set the corresponding voxels in the subject's synthetic image (i.e., that synthesised with default parameters) to an intensity of zero. The boundaries of the flood-filled voxels within the skull-stripped PD image were then manually adjusted to correct for any local deviations into neural tissue (particularly in regions proximal to paranasal sinuses, often prone to susceptibility artifacts). Manual adjustment involved reducing the intensity value regarded by Csurf as the threshold for cortical grey matter (typically, to a value of 40); the flood-filled boundary was then forced toward voxels with an intensity less than this value. The manual adjustment was applied to the synthetic volume; the skull-stripped synthetic volume was used as input to a custom version of the surface reconstruction pipeline (FreeSurfer recon-all).

**3.2.5.4. Surface reconstruction.** First, each subject's skull-stripped synthetic volume was intensity normalised in FreeSurfer (using the `mri_normalize` routine). Normalised images were briefly inspected to ensure grey and white matter intensity values were within the appropriate ranges (white matter: 110; grey matter: 50-70). Next, the skull-stripped default parameter synthetic volume (see skull strip) was used to mask the contrast-specified

synthetic volume (see image synthesis). This masked (i.e., skull-stripped), contrast-specified synthetic volume was then used as the input volume to an initial Talairach transformation process (run using the FreeSurfer `mri_em_register` routine). Next, a further normalisation step was performed (using the `-canorm` parameter in FreeSurfer `recon-all`); the initial Talairach transform, the skull-stripped, default parameter synthetic volume and the intensity-normalised version of that volume, were used as inputs. Following this, a multi-dimensional Talairach transformation was applied (using the `-careg` and `-careginv` parameters in `recon-all`); the normalised volume (produced by `recon-all -canorm`), the skull-stripped default parameter synthetic volume, and the initial Talairach transform were used as inputs. Finally, the full FreeSurfer `recon-all` pipeline was run for each subject (parameters specified can be found at: <https://surfer.nmr.mgh.harvard.edu/fswiki/ReconAllDevTable>; parameters used were those of the `autorecon-2` stage, and the first 6 parameters from `autorecon-3`).

**3.2.5.5. Surface adjustment.** Following reconstruction, pial and white matter surfaces were inspected blind to group (overlaid onto the normalised `recon-all` input volume), to identify local surface deviances. Seven participants' (4 violinists' and 3 non-musicians') surfaces showed minor deviances such that the pial surface underestimated the true pial boundary; those participants' synthetic volumes were re-brightened and re-run through the processes above. Re-inspection of these subjects' surfaces indicated improved segmentations relative to the first attempt that were in line with the rest of the cohort.

### 3.2.6 Data analyses

Following cortical surface reconstruction,  $R_1$  data were mapped onto participants' cortical surfaces in FreeSurfer. A priori regions of interest were specified and used to extract mean  $R_1$  values for each participant. Additionally, whole-brain vertex-wise analyses were performed. Description follows below.

**3.2.6.1.  $R_1$  data extraction and mapping.** First, all input  $R_1$  volumes were re-sampled to a finer image resolution (0.6 mm isotropic) and all subjects were rotated to the same (canonical) orientation, using the AFNI 3dwarp routine (-deoblique flag). Next, we scaled these higher resolution  $R_1$  volumes, in order to reduce any possible effects of subtle measurement biases that may have influenced group differences. The  $B1^+$  mapping and  $B0$  field mapping procedures greatly reduce effects of transmit field biases and static field inhomogeneities on  $R_1$  measurements. However, it is possible that other nuisance factors over subjects could have a very small effect on the signal measured at particular voxels. For each subject, a solution to this was to scale  $R_1$  values at every voxel in the subject's  $R_1$  volume by the ratio of the the corpus callosum  $R_1$  mean averaged over all subjects relative to the given subject's corpus callosum  $R_1$  mean (see below). Scaling each subject's data in this way provided a quantitative  $R_1$  metric at each voxel that accounted for possible additional signal bias in a given subject relative to the cohort, reducing further noise variance that was not controlled by the  $B1^+$  and  $B0$  mapping procedures. (We chose corpus callosum due to its consistently high  $R_1$  values and its position within the centre of the image volume; see Supplemental Figure 3.2).

Using the FreeSurfer subcortical parcellation for each subject, we therefore extracted mean  $R_1$  values for the entirety of each subject's corpus callosum (CC). We then calculated a grand average mean of CC  $R_1$  values

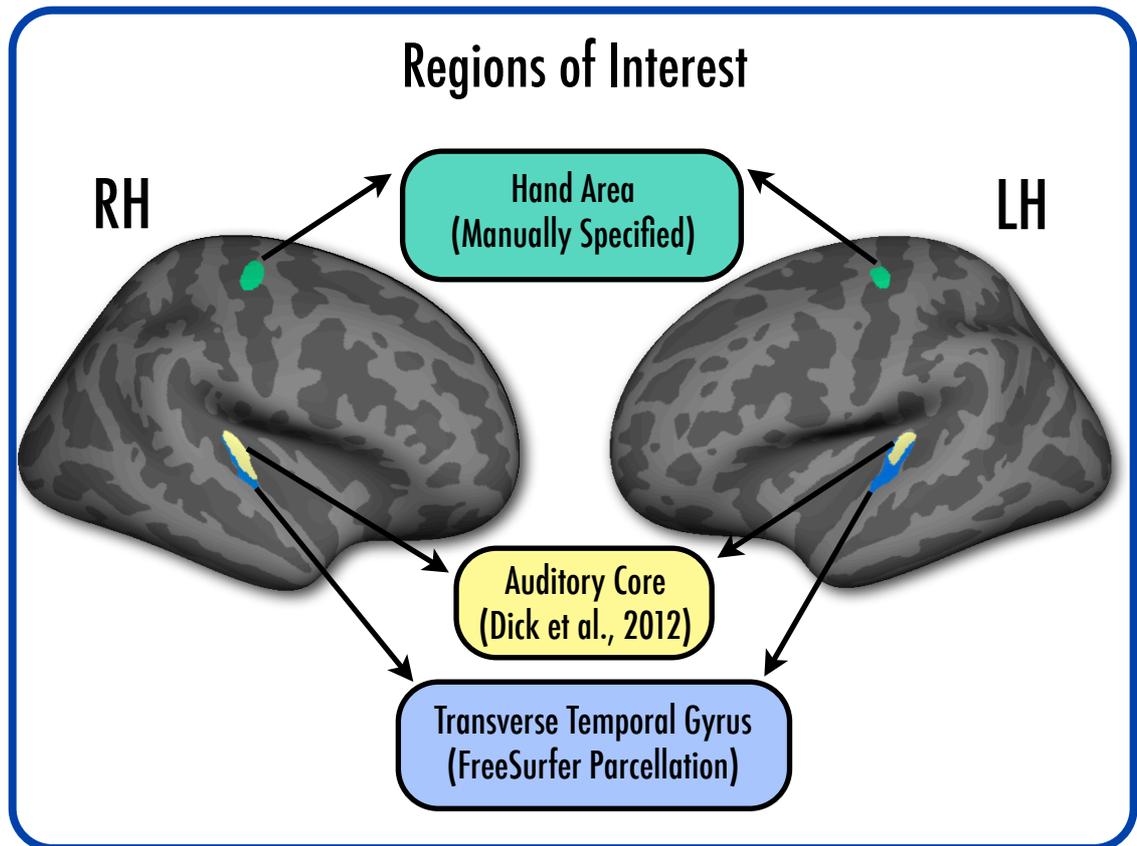
across all subjects (note that there was no significant group difference in CC mean  $R_1$  values,  $z = 0.9$ ,  $p = 0.38$ ). Finally, for each subject, we scaled  $R_1$  values at each voxel by the ratio of the cohort CC  $R_1$  grand mean relative to the subject's CC  $R_1$  mean [i.e., for a given subject: voxel  $R_1$  \* (cohort CC  $R_1$  grand mean/subject CC  $R_1$  mean)]; scaling was performed using the AFNI 3dcalc routine. Each scaled, high resolution volume was aligned to the recon-all input volume (using FreeSurfer tkregister2). Scaled high resolution  $R_1$  data were then mapped for each subject (using the FreeSurfer mri\_vol2surf routine). For each reconstructed hemisphere, data were sampled from the scaled high resolution  $R_1$  volume along the normal to each surface vertex, for cortical depth fractions from 0 (i.e., white matter surface boundary) to 1.0 (i.e., pial surface boundary) in increments of 0.1 (see Dick et al., 2012). Surface smoothing approximating a 4mm FWHM Gaussian kernel was applied to data in sampling onto each vertex (Hagler, Saygin & Sereno, 2006). The vertex-wise scaled  $R_1$  data for each hemisphere at each cortical depth fraction were saved as separate files, and used for data analyses.

**3.2.6.2. Regions of interest (ROI) analyses.** The primary goal of the present study was to explore profiles of structural change associated with long-term musical training and performance. ROIs for analyses were selected a priori. Two considerations were used to motivate ROI selection. Firstly, previous studies examining profiles of structural brain change related to musical training have robustly identified primary auditory cortex (i.e., Heschl's gyrus; Schneider et al., 2002, 2005; Gaser & Schlaug, 2003) and primary motor cortex (Gaser & Schlaug, 2003; Amunts et al., 1997; Bangert & Schlaug, 2006; Bermudez et al., 2009) as regions that manifest plastic adaptations, even with relatively short-term training (Hyde et al., 2009). Secondly, our quantitative  $R_1$  maps show high

sensitivity to profiles of cortical myelination, particularly across primary sensory and motor areas (Dick et al., 2012; Sereno et al., 2012). Given the critical importance of fine auditory processing and motor control to expert violin training, we hypothesised that adaptations reflecting increased cortical myelination would be present within both primary auditory and primary motor areas of our expert participants. Further, we hypothesised a reduction in profiles of asymmetry of hand area cortical myelination in our experts (relative to our non-musicians), owing to their extensive fine manual training of the digits of the non-dominant hand.

We therefore defined a series of ROIs across cortex. ROIs for Heschl's gyrus comprised labels from the FreeSurfer annotation cortical parcellation (covering the medial to antero-lateral extent over both hemispheres). In addition, custom ROIs were defined, using the group average map of auditory core from Dick et al. (2012); those labels covered approximately the medial two-thirds of Heschl's gyrus over each hemisphere (and thus fell within the FreeSurfer Heschl's gyrus parcellations). ROIs were also defined for hand area, by manual delineation of the hand omega on the FreeSurfer fsaverage cortical surface by a trained operator (see Figure 3.1). ROIs were saved as labels onto the FreeSurfer fsaverage brain, and were mapped onto each participant by morphing between the fsaverage spherical cortical surface and each participant's spherical surface (using the `mri_surf2surf` routine). ROIs were inspected on each participant's inflated surface, and were manually fixed where labels showed incomplete filling (using the `dilate` and `erode` tools in `tksurfer`).

ROI  $R_1$  data were extracted as the mean of scaled  $R_1$  values across all vertices within each ROI. For all ROIs, mean data were extracted at depth



**Figure 3.1:** A priori regions of interest (ROIs) specified for analyses of cortical  $R_1$  and cortical thickness.

fractions from 0.1 to 0.9 (see Dick et al., 2012; Sereno et al., 2012). We analysed data across cortical depths in an effort to capture differences that may have partly reflected the laminar organisation of each region. In particular, we anticipated that thalamocortical afferents from medial geniculate nucleus (that largely project to lower pyramidal [IIIb] and internal granular layers [IV] in mid-A1; Hackett, 2011) might account for a considerable proportion of fibres showing potential for expression of myelin plasticity; hence, we expected that we might find such effects at mid-cortical depths. We did not expect upper cortical layers of A1 – e.g., layers I & II that receive relatively fewer thalamocortical inputs – to show plastic effects. Given the agranular nature of M1 (Nolte, 2009) and its relatively large layer V (with heavy projections to

corticospinal tract; see Sherwood et al., 2004; Kaneko, 2013), we similarly anticipated that mid-cortical depths would show the highest potential for expression of myelin plasticity within hand area. As a caveat, we should note that it is not yet possible to determine the exact correspondence between cortical layers, and cortical depth fractions as measured in FreeSurfer. We therefore used the present scheme as a coarse approximation for laminar profiles within cortex.

**3.2.6.3. Whole-brain analyses.** In addition to ROI analyses, we also analysed  $R_1$  data at whole-brain level. The purpose was two-fold: firstly, to determine whether hypothesised group differences within ROIs were robust at whole-brain level; secondly, to explore potential  $R_1$  differences in cortical areas outside of our a priori ROIs. We hypothesised that cortical regions such as superior temporal gyrus and premotor cortex might show relatively increased profiles of myelination in our experts (based on previous evidence of cortical thickness and functional differences in those regions, between musicians and non-musicians; Bermudez et al., 2009; Dick et al., 2011). Scaled  $R_1$  data measured at 0.5 cortical depth were used for whole-brain analyses; all analyses were run using Qdec (MGH FreeSurfer, v. 5.3).

### 3.3 Results

As outlined in methods, data analysed were extracted from resampled high-resolution (0.6 mm<sup>3</sup> resampled) R<sub>1</sub> volumes; voxel-wise R<sub>1</sub> values were scaled for every subject by the ratio of the cohort grand mean corpus callosum (CC) R<sub>1</sub> relative to a given subject's mean CC R<sub>1</sub>. Data analysed from ROIs were the means of vertex-wise scaled R<sub>1</sub> values within the ROI at a given cortical depth. For ANOVA analyses of ROI data, Greenhouse-Geisser corrected degrees of freedom (calculated in SPSS) are reported where the sphericity assumption was violated. R<sub>1</sub> is reported here as ms<sup>-1</sup> [=1000/T<sub>1</sub>(ms)].

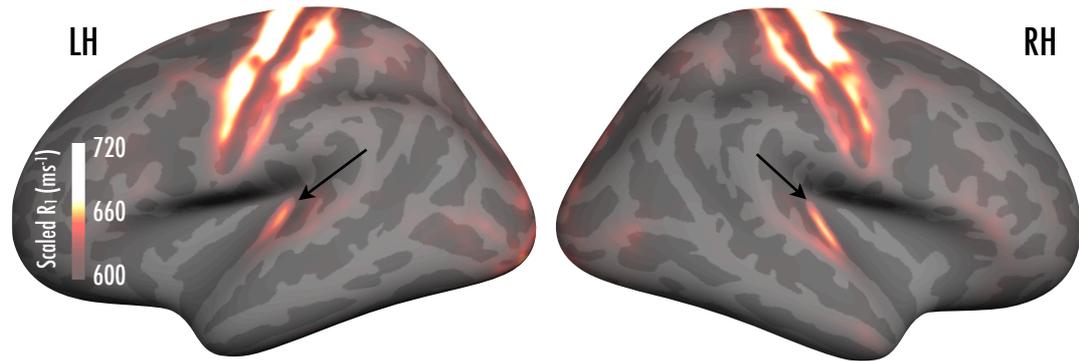
#### 3.3.1 Effects of expertise: violinists vs. non-musicians

Our primary aim was to explore differences in cortical R<sub>1</sub> values between violinists and non-musicians. We therefore analysed R<sub>1</sub> data both using *a priori* regions of interest (reflecting auditory and hand omega cortical areas) and using whole-brain analyses.

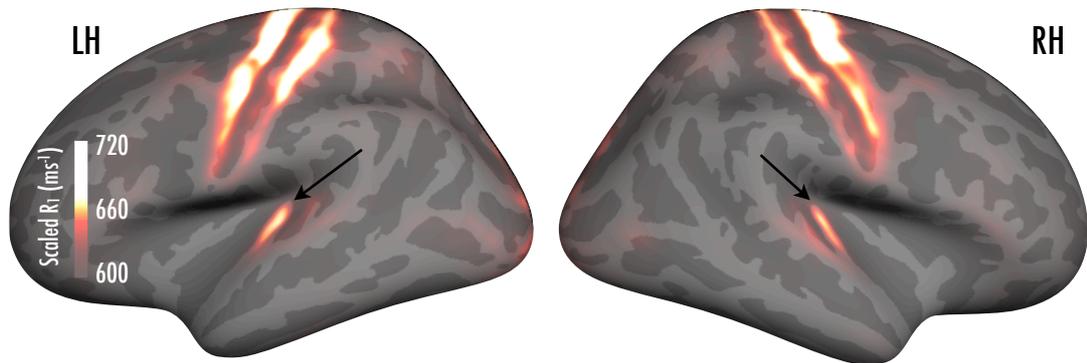
We first ran whole-brain analyses in Qdec to test for group differences in scaled R<sub>1</sub> (measured at 0.5 cortical depth). We did not find evidence of significant vertex-wise differences between violinists and non-musicians at whole-brain level, when thresholds were uncorrected ( $p < 0.001$ ) or corrected for multiple comparisons (FDR:  $p = 0.05$ ). Figure 3.2 displays average cortical R<sub>1</sub> maps for each of the groups.

**3.3.1.1. ROI analyses: auditory core and Heschl's gyrus.** While whole-brain analyses did not reveal significant group differences in R<sub>1</sub>, a central aim of our analysis was to explore group differences at auditory core and Heschl's gyrus ROIs over cortical depth fractions (see 3.2.6.2; further to Dick et al., 2012). We therefore conducted planned analyses of ROI mean scaled R<sub>1</sub>

### Non-musicians (n = 16)



### Violinists (n = 18)



**Figure 3.2:** Group average cortical  $R_1$  maps for non-musician (top row) and violinist (bottom row) cohorts. Data displayed are sampled halfway through cortex (0.5 cortical depth), and represent means of scaled  $R_1$  values at each vertex. Scale at left denotes range of  $R_1$  ( $\text{ms}^{-1}$ ) values (data displayed with a statistical midpoint of 660 ms for both groups). Arrow heads indicate medial edge of Heschl's gyrus; auditory core occupies the postero-medial two thirds of the gyrus approximately (see Dick et al., 2012). Data are presented on an average cortical surface (fsaverage).

values across cortical depths and hemispheres (0.1 to 0.9), modelling core and Heschl's gyrus separately. ANOVA models were followed-up with planned comparisons of group (uncorrected) at each depth fraction, for core and Heschl's gyrus.

A 2 (hemisphere) x 9 (cortical depth: 0.1-0.9) x 2 (violinist vs. non-musician) model of auditory core revealed a significant three-way interaction of these factors [ $F(2.39, 76.45) = 3.31, p = 0.034, \eta_p^2 = 0.094$ ], together with a significant main effect of cortical depth [ $F(1.99, 63.7) = 3985.8, p < 0.0001, \eta_p^2$

= 0.992]. Post-hoc pairwise comparisons (FDR-corrected) of the depth fraction main effect showed that as expected,  $R_1$  values differed significantly across all cortical depths (all comparisons  $p < 0.0001$ ). Splitting the three-way interaction by hemisphere revealed a close to significant depth fraction x group interaction at left hemisphere auditory core [ $F(2.34, 75.02) = 2.87, p = 0.054, \eta_p^2 = 0.082$ ], with a quadratic fit to the data showing a significant depth fraction x group interaction [ $F(1, 32) = 6.44, p = 0.016, \eta_p^2 = 0.167$ ]. There was no significant depth fraction x group interaction at right hemisphere core [ $F(1.86, 59.55) = 0.46, p = 0.62$ ].

Critically, planned group comparisons across cortical depths at left auditory core revealed close to significant differences in cortical  $R_1$  values between violinists and non-musicians at mid cortical depth fractions (0.4 and 0.5; both  $z = 1.95, p = 0.051$ ; see Figure 3.3). A marginal trend toward a group difference also emerged at left core at a cortical depth fraction of 0.3 ( $z = 1.81, p = 0.07$ ) (all tests two-tailed). Comparisons at the same cortical depths at right hemisphere core revealed no evidence of any significant group differences (all  $z < 0.33, p > 0.7$ ). Age did not account for any significant variance as a predictor of  $R_1$  at LH auditory core over depths of 0.4 and 0.5 (both  $F < 1.4, p > 0.25$ ).

A 2 (hemisphere) x 9 (cortical depth: 0.1-0.9) x 2 (violinist vs. non-musician) model of Heschl's gyrus also revealed a significant three-way interaction of these factors [ $F(2.46, 78.68) = 3.32, p = 0.032, \eta_p^2 = 0.094$ ], together with a significant main effect of cortical depth [ $F(1.84, 58.84) = 4681.24, p < 0.0001, \eta_p^2 = 0.993$ ]. Post-hoc pairwise comparisons (FDR-corrected) of the depth fraction main effect again showed that  $R_1$  values differed significantly across all cortical depths (all comparisons  $p < 0.0001$ ). However, splitting the three way interaction by hemisphere revealed no significant depth x

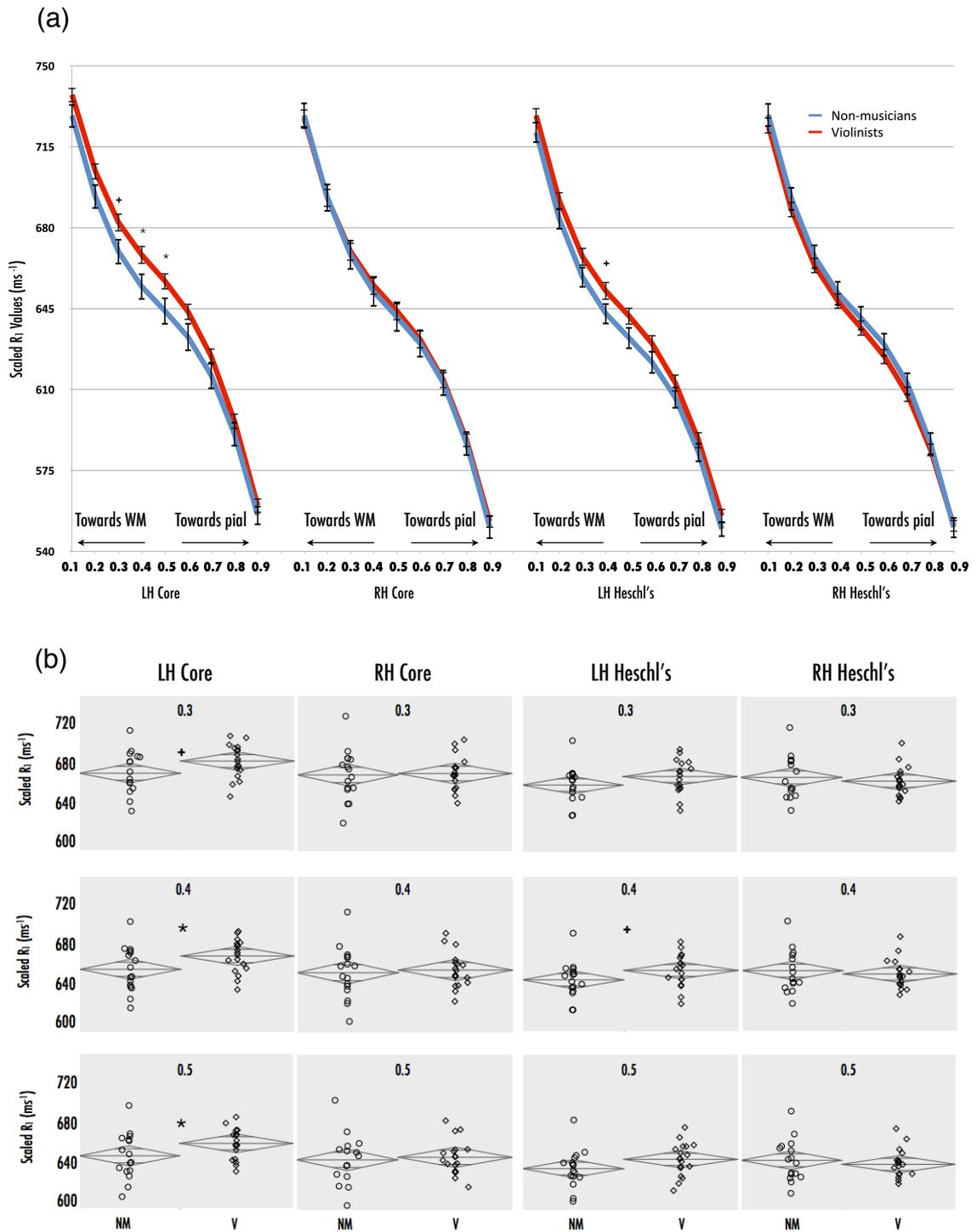


Figure 3.3: (a) Cortical depth fraction analyses of scaled  $R_1$  data across violinist and non-musician cohorts, for auditory core and Heschl's gyrus ROIs. Cortical depth fractions are indicated along x-axis, grouped by ROI. Error bars:  $\pm 1$  std. error of mean. (b)  $R_1$  data across ROIs at cortical depth fractions of 0.3 (top row), 0.4 (middle row) and 0.5 (bottom row). V - violinists; NM - non-musicians. \*  $p = 0.051$ ; +  $p \leq 0.07$  (two-tailed)

group interaction at left [ $F(1.92, 61.39) = 0.86, p = 0.42, \eta_p^2 = 0.026$ ] or right [ $F(2.04, 65.22) = 1.07, p = 0.35, \eta_p^2 = 0.032$ ] hemisphere. Given the significant group effects at cortical depth fractions of 0.4 and 0.5 over left hemisphere core, we considered whether left Heschl's gyrus would show similar effects at these depths; however, we found only a marginal group difference at LH Heschl's at 0.4 cortical depth ( $z = 1.85, p = 0.065$ ), with only a very weak trend toward a significant difference at 0.5 cortical depth ( $z = 1.6, p = 0.11$ ). Hence, group differences in cortical  $R_1$  were weaker when considering the full extent of Heschl's gyrus, than when considering the more medial auditory core (see Figure 3.3).

**3.3.1.2. ROI analyses: hand area.** In addition to expected group differences within auditory ROIs, we also hypothesised that violin expertise would be associated with significant increases in cortical  $R_1$  within motor cortex. Specifically, we anticipated that the extensive motor training violinists engage in with the digits of the non-dominant hand, would lead to reduced cortical asymmetry of  $R_1$  values; thus, right-handed non-musicians might show a LH > RH effect, whereas a reduction of such an effect would be expected for violinists.

To test this hypothesis, we performed a 2 (hemisphere) x 9 (cortical depth) x 2 (group) ANOVA on mean hand area ROI scaled  $R_1$  values. A main effect of hemisphere reached significance [ $F(1, 32) = 4.89, p = 0.034, \eta_p^2 = 0.133$ ]; however, to our surprise we found no evidence of a significant hemisphere x group interaction [ $F(1, 32) = 0.001, p = 0.97, \eta_p^2 < 0.001$ ]. Neither was there any significant main effect of group [ $F(1, 32) = 0.9, p = 0.35, \eta_p^2 = 0.03$ ]. A significant depth x group interaction did emerge [ $F(1.72, 55.16) = 4.46, p = 0.02, \eta_p^2 = 0.122$ ]; however, post-hoc comparisons across depths

(hemisphere collapsed) showed no evidence of any significant group differences (all  $z < 1.55$ ,  $p > 0.12$ ). As expected, a main effect of cortical depth was significant [ $F(1.72, 55.16) = 2471.38$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.987$ ], with pairwise tests showing the difference between every depth to be significant (all  $p < 0.0001$ , FDR-corrected). As displayed in Figure 3.4, mean  $R_1$  values were elevated for both groups at the LH hand area ROI, relative to RH, and declined across cortical depths as expected. Hence, unexpectedly, the predicted reduction in asymmetry of hand area cortical  $R_1$  for violinists versus non-musicians was not supported.

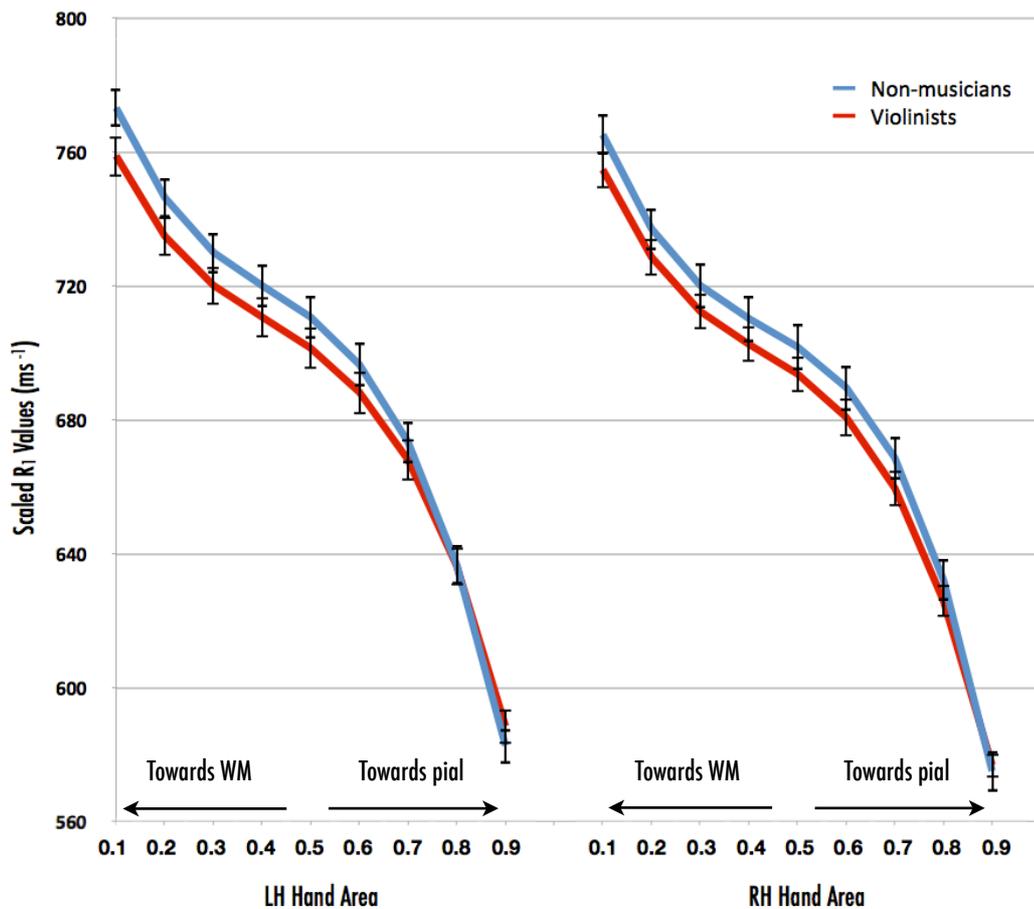


Figure 3.4: Scaled  $R_1$  ROI values for hand area across cortical depths. Note the increase in  $R_1$  values at LH relative to RH for both cohorts. Error bars  $\pm 1$  std. error of mean.

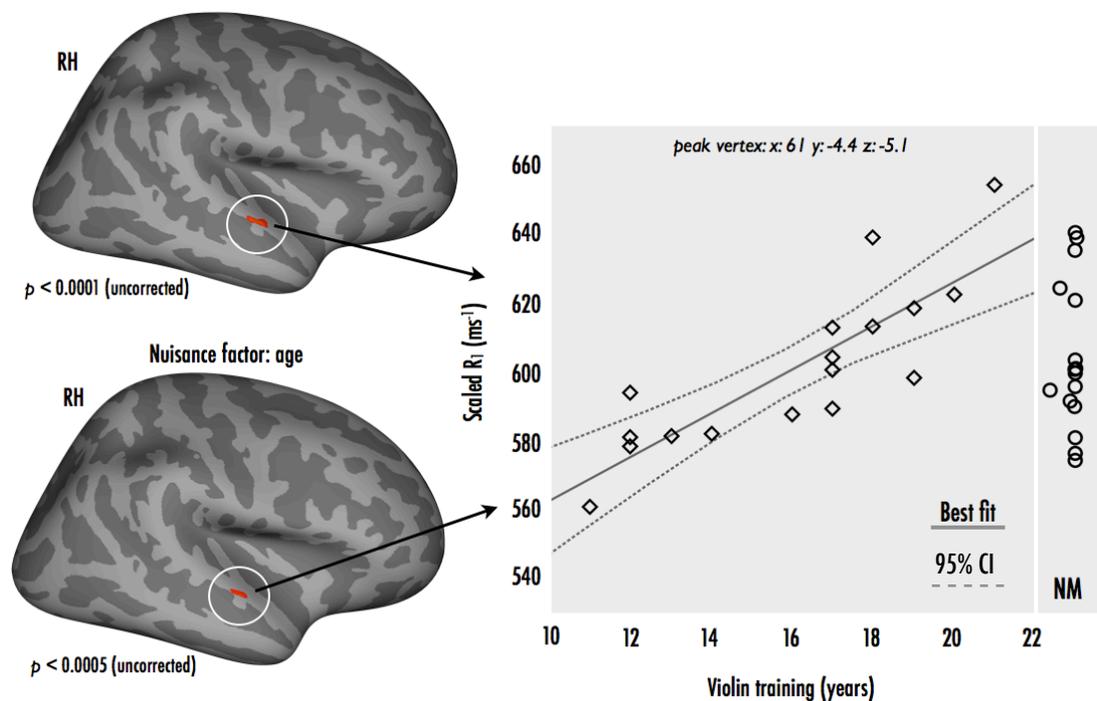
### 3.3.2. Effects of expertise: violinists' training

As a core aim of the study was to explore profiles of plasticity related to expert training, we also sought to account for the influence of violin training metrics on cortical  $R_1$ . We therefore defined a series of training variables of interest: a) years of violin training (defined from the age at which formal lessons began); b) training onset at 3-4 years (yes/no); c) current daily practice hours. We then used these variables as regressors in analyses of scaled cortical  $R_1$  (0.5 cortical depth), both at whole-brain level and within ROIs. Note that one violinist was excluded from all whole-brain and ROI regression analyses using years of training, since their training (27.5 years) placed them more than 3 SDs above group mean years of training ( $n = 17$ , for all years of training analyses).

**3.3.2.1. Violinists' training: whole-brain analyses.** We first explored whether violin training metrics would predict scaled cortical  $R_1$  within whole-brain analyses. Neither onset of training at 3-4 years nor current daily practice accounted for significant variance in cortical  $R_1$  across either hemisphere (all thresholds  $p < 0.001$ , uncorrected).

We found a cluster of  $R_1$  values at right lateral superior temporal gyrus (rSTG) that showed a significant relationship with violinists' years of training (see Figure 3.5) (uncorrected whole-brain threshold of  $p < 0.0001$ ). Since chronological age might have co-varied with years of training (and indeed, age may influence cortical  $R_1$ ), we ran the same model entering age as a regressor of no interest; the rSTG cluster again reached significance at a slightly lower whole-brain threshold ( $p < 0.0005$ , uncorrected). When applying a correction for multiple comparisons over all vertices (FDR,  $p = 0.05$ ), we found that the cluster did not survive; however, due to the cluster's discrete, circumscribed nature, FDR-correction was likely a conservative adjustment (Benjamini & Hochberg,

1995). As shown in Figure 3.5, years of training demonstrated a tight linear relationship with  $R_1$  increases at this cluster. We did not find evidence of significant  $R_1$  clusters (at  $p < 0.0001$ , uncorrected) over left hemisphere that were predicted by years of training (with or without age entered as a nuisance factor). Thus, both analyses over RH suggested increased cortical  $R_1$  values at rSTG in those violinists who had trained the longest.



**Figure 3.5:** Whole-brain analyses of effects of years of violin training (cohort mean years  $\pm$  SD:  $16.1 \pm 3.1$ ) in predicting cortical  $R_1$ . Upper left: cluster at right STG (size:  $57.4 \text{ mm}^2$ ; 115 vertices) reaches significance at  $p < 0.0001$  (uncorrected). Lower left: cluster (size:  $37.7 \text{ mm}^2$ ; 67 vertices) reaches significance at  $p < 0.0005$  (uncorrected) when entering age as a nuisance factor. Peak vertex is identical in both analyses (Talairach co-ordinates:  $x: 61.0 \ y: -4.4 \ z: -5.1$ ). Right: regression fit of scaled  $R_1$  values by years of training at peak vertex; panel at right shows non-musician  $R_1$  values at the same peak vertex (note that there was no significant group difference at the cluster).

**3.3.2.2. Violinists' training: ROI analyses.** Further to whole-brain analyses of effects of violin training in predicting cortical  $R_1$ , we examined whether ROI mean  $R_1$  would be predicted by training metrics. We found a significant relationship between violinists' years of training and ROI mean scaled  $R_1$  values at LH Heschl's gyrus [ $F(1, 16) = 7.92, p = 0.013, \text{adj. } R^2 = 0.302$ ]. However, when age was included along with years of training, the model yielded a significant fit overall [ $F(2, 14) = 5.74, p = 0.015, \text{adj. } R^2 = 0.372$ ], but only a weak trend toward an effect of years of training [ $\beta = 2.1, t(14) = 1.75, p = 0.1$ ; age:  $\beta = 2.8, t(14) = 1.63, p = 0.13$ ]. Stepwise regressions showed that age

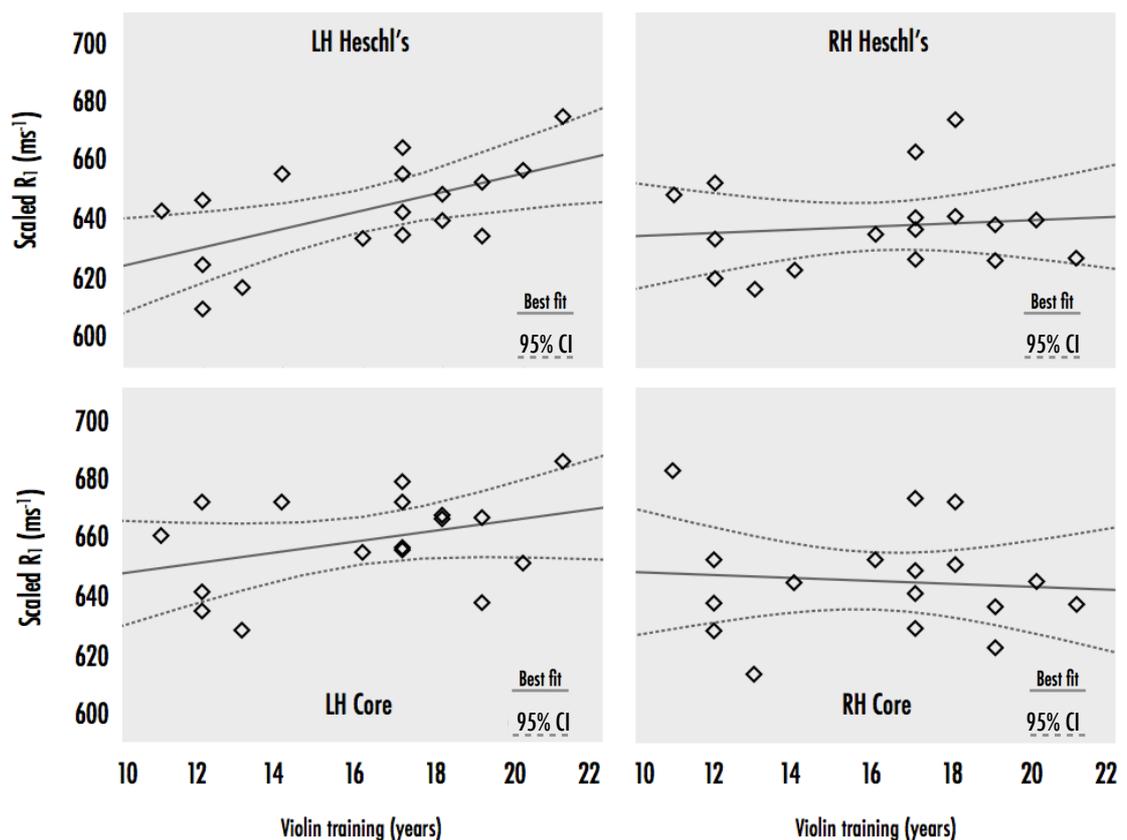


Figure 3.6: Linear regression fits of scaled mean ROI  $R_1$  values, by years of training (above: Heschl's gyrus; below: auditory core). Years of training predicted scaled mean  $R_1$  values at LH Heschl's gyrus, but not at RH Heschl's gyrus. However, inclusion of age as a further predictor showed the training effect at LH Heschl's to be less robust (see text). Solid lines: best fit; dotted lines: regression 95% CI.

accounted for only slightly less variance (adj.  $R^2 = 0.286$ ) than years of training (adj.  $R^2 = 0.302$ ).

Although we found significant violinist vs. non-musician  $R_1$  differences at LH auditory core (and only weak group differences at LH Heschl's gyrus; see 3.3.1.1), we did not find any significant effect of violinists' years of training in accounting for  $R_1$  values at LH core [ $F(1, 16) = 2.26, p = 0.15, \text{adj. } R^2 = 0.07$ ] (see Figure 3.6).

Hence, years of training accounted for some variance in  $R_1$  values at LH Heschl's gyrus, although this effect was markedly less robust when age was accounted for. Despite the group differences in  $R_1$  at LH auditory core (see 3.3.1.1), we did not find evidence that violinists' years of training accounted for their increased mean scaled cortical  $R_1$  within that ROI.

Violinists' years of training did not significantly predict ROI mean scaled  $R_1$  at any other ROIs (RH core, RH Heschl's gyrus, LH hand area, RH hand area; all  $F < 0.47, p > 0.5$ ). Violinists' current daily practice hours did not significantly predict mean scaled  $R_1$  at any ROI (all  $F < 1.2, p > 0.3$ ). Violinists' who reported practice at 3-4 years showed only very weak evidence of a trend toward greater  $R_1$  values at LH auditory core ( $z = 1.55, p = 0.12$ , two-tailed), compared to violinists who had not practised at that age; there were no other differences in mean scaled  $R_1$  between those who practised at this age versus those who did not at any other ROI (all  $z < 1.19, p > 0.23$ ).

### **3.3.3. Performance and cortical $R_1$ : auditory psychophysical thresholds and SAART**

A final goal of the present study was to relate fine auditory perception and auditory attentional skill across both experts and non-experts to cortical  $R_1$

measures. In this way, we could determine whether individual differences (indexed with continuous measures that showed sensitivity to group differences) in fine perceptual and broader attentional metrics might relate to differences in proxies for underlying cortical myelination (i.e.,  $R_1$ ). We therefore used thresholds for AM depth, FM depth and onset rise time, along with variability of sustained attention, as predictors of cortical  $R_1$  (0.5 cortical depth) across the violinist and non-musician cohorts.

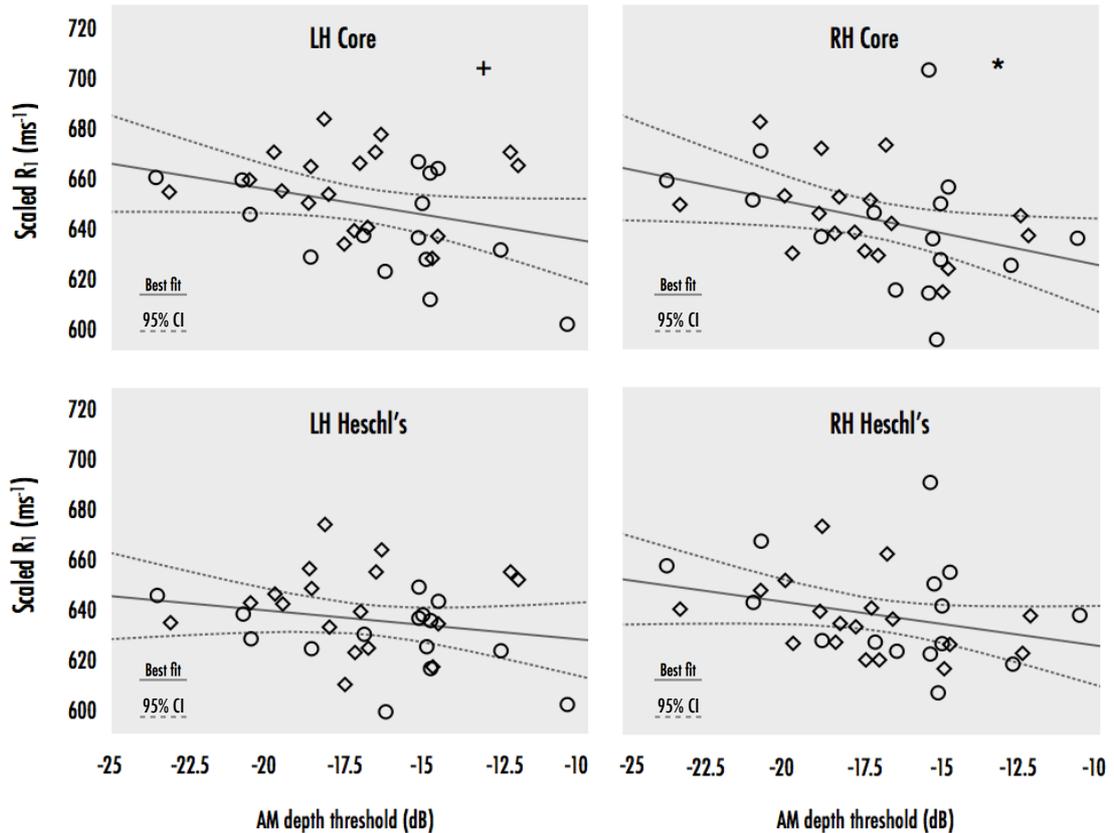
Using these metrics as regressors at whole-brain level, we found no evidence of significant relationships with cortical  $R_1$  (all analyses  $p < 0.0001$ , uncorrected). We therefore explored relationships between behavioural metrics and  $R_1$  measured within auditory ROIs (note that motor ROIs were not considered since there was no theoretical motivation for examining relationships between measures of fine auditory perception and cortical regions associated with motor skill; see Schneider et al., 2002).

**3.3.3.1. AM depth thresholds and  $R_1$ .** We first examined whether AM depth thresholds would account for variance in auditory ROI mean scaled  $R_1$  values (AM depth thresholds did not differ significantly between both groups,  $z = 1.12$ ,  $p = 0.26$ ). We found that AM depth thresholds significantly predicted ROI mean scaled  $R_1$  measured at RH auditory core [ $F(1, 31) = 4.66$ ,  $p = 0.039$ , adj.  $R^2 = 0.103$ ], with a marginal trend noted at LH auditory core [ $F(1, 31) = 3.53$ ,  $p = 0.07$ , adj.  $R^2 = 0.073$ ]. As expected, the models indicated that those achieving lower AM depth thresholds (i.e., improved performance) tended to have increased cortical  $R_1$  within these ROIs. When considering the same effects at Heschl's gyrus, we found only a very weak trend at RH [ $F(1, 31) = 2.94$ ,  $p = 0.096$ , adj.  $R^2 = 0.057$ ], and no significant effect at LH [ $F(1, 31) = 1.42$ ,  $p = 0.24$ , adj.  $R^2 = 0.01$ ]. Figure 3.7 displays the regression model fits.

**3.3.3.2. FM depth thresholds and  $R_1$ .** Next, we considered whether FM depth thresholds would predict mean  $R_1$  measured at auditory ROIs (FM depth thresholds differed significantly between both groups,  $z = 2.03$ ,  $p = 0.043$ ). However, we found no evidence that FM depth thresholds predicted ROI mean scaled  $R_1$  over either hemisphere, either at auditory core [both  $F(1, 31) < 0.39$ ,  $p > 0.54$ ] or Heschl's gyrus [both  $F(1, 31) < 0.14$ ,  $p > 0.7$ ].

**3.3.3.3. Onset rise time thresholds and  $R_1$ .** We used thresholds for onset rise time to predict mean  $R_1$  measured within auditory ROIs (onset rise time thresholds differed significantly between both groups,  $z = 2.86$ ,  $p = 0.004$ ). As with FM depth thresholds, we found no evidence that onset rise time thresholds significantly predicted ROI mean scaled  $R_1$  at auditory core [both  $F(1, 26) < 1.0$ ,  $p > 0.33$ ] or Heschl's gyrus [both  $F(1, 26) < 0.9$ ,  $p > 0.35$ ].

**3.3.3.4. SAART RT SD and  $R_1$ .** Finally, we explored whether variability of performance on a sustained auditory attentional task would predict differences in cortical  $R_1$  (SAART RT SDs did not differ significantly between groups,  $z = 0.73$ ,  $p = 0.46$ ). However, we found no evidence that SD of RTs to non-target sounds during the SAART significantly predicted mean ROI scaled  $R_1$  at LH or RH auditory core [both  $F(1, 29) < 1.6$ ,  $p > 0.23$ ], and only a very weak trend toward an effect at RH Heschl's gyrus [ $F(1, 29) = 2.83$ ,  $p = 0.1$ , adj.  $R^2 = 0.057$ ], with no significant effect at LH Heschl's gyrus [ $F(1, 29) < 0.08$ ,  $p > 0.79$ ].



**Figure 3.7:** Linear regression fits of scaled mean ROI  $R_1$  values by AM depth thresholds (above: auditory core; below: Heschl's gyrus). AM depth thresholds predicted scaled mean  $R_1$  values at RH auditory core, and marginally at LH auditory core. AM depth thresholds did not significantly predict  $R_1$  values at Heschl's gyrus over either hemisphere. Circles: non-musicians; diamonds: violinists. Solid lines: best fit; dotted lines: regression 95% CI. \*  $p < 0.05$ ; +  $p = 0.07$

### 3.3.4. Further analyses: cortical thickness

The present study aimed primarily to index expertise-related changes in cortical structure using a quantitative proxy for cortical myelination. Nevertheless, we also explored possible group differences in cortical thickness, both at whole-brain and ROI level. Our reasons for this were two-fold. Firstly, given that measured  $R_1$  varies partly as a function of cortical thickness (see Dick et al., 2012; Sereno et al., 2012), we sought to explore whether differences in cortical thickness might partly underlie the group  $R_1$  differences we found at

LH auditory core. Secondly, previous investigations of cortical thickness in musicians and non-musicians demonstrated thickness increases in musicians in regions including right pre-central gyrus, planum temporale and superior temporal gyri bilaterally (Bermudez et al., 2009; Dohn et al., 2013). Thus, we sought to further these findings with the present sample, and explore potential thickness differences across auditory core and Heschl's gyri. Thickness data were analysed at whole-brain level using Qdec (thickness values surface-smoothed, approximating 20mm FWHM); thickness data were extracted from ROIs using FreeSurfer routines (mris\_anatomical\_stats). Thickness was measured as per standard FreeSurfer procedure (i.e., distance in mm along each vertex normal between the white matter and pial surfaces; vertex-wise means were produced for ROIs).

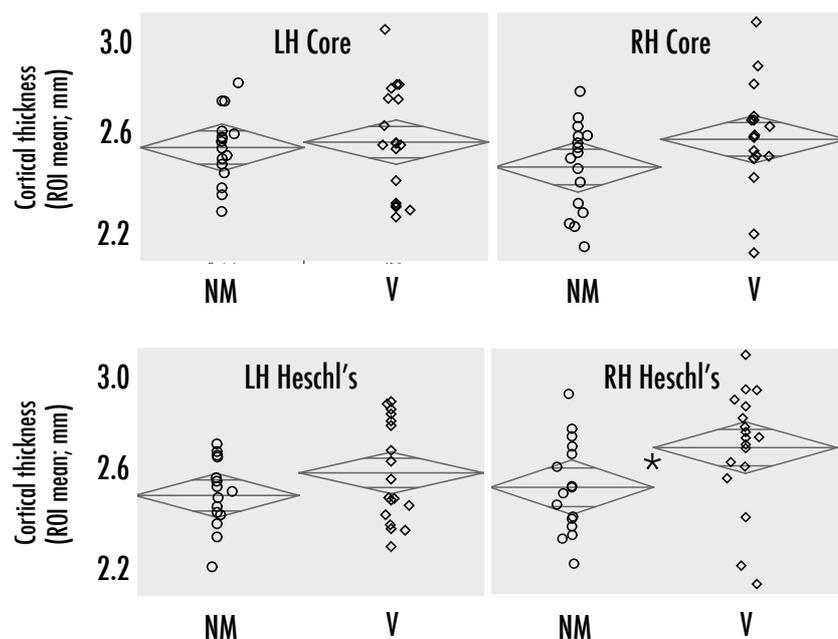
**3.3.4.1. Cortical thickness: whole-brain analyses.** We first explored possible group differences in thickness at whole-brain level. However, we found no evidence of any cortical regions that showed significant differences in thickness between violinists and non-musicians at uncorrected thresholds ( $p < 0.0005$ ) across either hemisphere; we therefore confined our analyses to cortical ROIs.

**3.3.4.2. Cortical thickness: ROI analyses.** We next considered whether thickness would differ between groups within auditory ROIs. Given that group differences in  $R_1$  were isolated to specific ROIs (see 3.3.1.1), we ran planned tests of group at each ROI. There was a significant group difference in mean cortical thickness at RH Heschl's gyrus ROI ( $z = 2.26$ ,  $p = 0.024$ ); violinists showed significantly increased mean thickness relative to non-musicians (see Figure 3.8). However, we found no significant differences in cortical thickness at LH core, RH core, or LH Heschl's gyrus (all  $z < 1.46$ ,  $p > 0.14$ ). Moreover, we

also found that neither violinists' years of training nor current practice accounted for significant variance in thickness at any auditory ROI [all  $F(1,15) < 2.23$ ,  $p > 0.15$ , adj.  $R^2 < 0.08$ ]; mean thickness did not differ significantly at any auditory ROI between those violinists who reported practising at 3-4 years and those who did not (all  $z < 0.8$ ,  $p > 0.4$ ).

In sum, these results suggest that the increase in cortical thickness at RH Heschl's gyrus in musicians was not accounted for by violin training metrics. Moreover, these findings indicate that the greater mean scaled  $R_1$  at LH auditory core in violinists than non-musicians did not reflect a concomitant group difference in cortical thickness within the same ROI.

In exploring hand area ROIs, a 2 (hemisphere) x 2 (group) ANOVA showed no significant main effect of group, no significant main effect of hemisphere, and no significant interaction of these factors [all  $F(1, 32) < 1.62$ ,  $p > 0.21$ ]. Thus, we found no significant difference between violinists and non-musicians in cortical thickness as measured at hand area ROIs.



**Figure 3.8:** Group differences in mean cortical thickness measured at auditory ROIs. Note significant group difference at RH Heschl's gyrus ROI (\*  $p < 0.05$ ).

### 3.4 Discussion

This study explored profiles of experience-related structural brain change via high-resolution quantitative MR metrics known to index cortical myelination. Comparing a subset of the same expert violinists and matched non-musicians as in chapter two, we investigated whether violinists' status as experts would relate to increases in cortical  $R_1$ . Previous studies have demonstrated volumetric (and gross morphological) changes in auditory and motor cortical areas in musicians; we therefore sought to explore differences in quantitative cortical myelin proxies in the same regions. Indeed, given the high extents of myelination at primary auditory and primary motor cortices (Sereno et al., 2012; Glasser & Van Essen, 2011) and the sensitivity of  $R_1$  to cortical myelin (Lutti et al., 2014), our high-resolution  $R_1$  metrics offered a means of detecting potential myelin differences in these regions that related to expertise.

We hypothesised that violinists would show increased cortical  $R_1$  compared to non-musicians within auditory core ROIs; we further predicted that these differences might extend laterally to cover the extent of Heschl's gyrus. We also hypothesised that the fine motor training of the left hand digits that violinists engage in might relate to increased cortical  $R_1$  at right hand area ROI (and thus a reduction in left-right asymmetry), compared to non-musicians. We expected that training metrics collected for violinists (years of training, training onset, and current practice) would account for some  $R_1$  variance, particularly for ROIs where group differences emerged. Finally, we explored whether perceptual and cognitive metrics that revealed some evidence of group differences in chapter two (thresholds for AM depth, FM depth and onset rise time; SAART RT SDs) might relate to individual differences in cortical  $R_1$ .

Expert violinists had increased cortical  $R_1$  at LH but not RH auditory core compared to non-musicians (with the differences emerging at mid cortical depths). We found limited evidence of relationships with violinists' training however: years of training predicted  $R_1$  values at LH Heschl's gyrus, but the effect was not robust when including age as a further regressor. Nevertheless, at whole-brain level, violinists with greater years of training showed significant increases in  $R_1$  at rSTG (also significant when controlling for age). Contrary to our hypothesis, we did not find a significant group x hemisphere interaction at hand area ROIs, and instead found a main effect of hemisphere. Finally, AM depth thresholds accounted for significant variance in  $R_1$  at RH auditory core, and marginally at LH core; variability of sustained attention performance, and thresholds for FM depth and onset rise time were non-significant predictors.

We first consider our findings with respect to use of musicians as a model for cortical plasticity. We then explore our results from the perspective of cortical myelination, expertise and developmental change. Finally, we consider behavioural implications of cortical myelination in experts and non-experts.

**3.4.1. Musicianship and cortical plasticity.** Of central importance to this study was the exploration of experience-related structural changes in cortical areas critically involved in musicianship.

Our results revealed a profile of moderately increased  $R_1$  in violinists versus non-musicians that were isolated to a small subset of cortical regions. Further to findings of greater grey (Schneider et al., 2002; 2005; Hyde et al., 2009) and subcortical white matter (Schneider et al., 2002) volumes at Heschl's gyri, we found increased  $R_1$  at mid cortical depths over left hemisphere auditory core for violinists versus non-musicians. While broadly agreeing with the white matter volume effect found at antero-medial Heschl's gyri by Schneider et al.

(2002) (who averaged over both hemispheres), our effect differs in that it resided within cortex and was left-lateralised (we found no group  $R_1$  differences at RH core or Heschl's gyrus). While our group  $R_1$  differences were left-lateralised, we nevertheless observed a group difference in cortical thickness at right Heschl's gyrus. Although not central to the aims of the present study, this effect nevertheless agrees with previous findings of increased grey matter density at right primary auditory cortex and increased cortical thickness at right planum temporale, in musicians versus non-musicians (Bermudez et al., 2009; see also Dohn et al., 2013).

Further to investigating effects of expertise at auditory cortical regions, we also explored the influence of expert violin training on cortical myelin proxies at primary motor cortex. The relative asymmetry involved in violinists' motor training (requiring fine co-ordination of the non-dominant hand digits) provides a strong test of experience-dependent training effects (e.g., Elbert et al. 1995). To our surprise, we found no evidence that violinists had increased cortical  $R_1$  at right hand area (nor any reduction in left-right asymmetry) compared to non-musicians, either for whole-brain or ROI analyses (in fact, ROI means suggested a slight increase in hand area scaled  $R_1$  values for non-musicians compared to violinists, although this difference was not significant; see Figure 3.4). Moreover, we also found no evidence of group differences in cortical thickness (a broader, non-quantitative metric) at whole-brain level or at hand area ROIs (cf. Bermudez et al., 2009). This result was unexpected, given previous findings of grey matter density increases (Gaser & Schlaug, 2003; Hyde et al., 2009) and gross morphological differences (Amunts et al., 1997; Bangert & Schlaug, 2006) at hand area in musicians versus non-musicians. Indeed, one study of string musicians and non-musicians also showed right-

lateralised increases in MEG dipole moments at hand area in string players, reflecting the fourth finger of the left hand (Elbert et al., 1995).

The lack of reduction in hand area  $R_1$  asymmetry in the violinists studied here might have arisen due to bilateral transfer. Indeed, motor learning is known to yield transfer effects to the untrained hand: learning may transfer to the dominant hand when the non-dominant hand is trained, and to a great extent compared to the reverse (i.e., dominant hand training; Phillips et al., 2013; but see Teixeira, 2000). Such an account, formalised as the callosal access model (Taylor & Heilman, 1980), predicts that non-dominant hand access to motor programs must occur via the corpus callosum, whereas the dominant hand holds direct cortico-spinal tract (i.e., non-callosal) access. If non-dominant hand learning must occur by transfer via corpus callosum, then it is perhaps possible that long-term training of the non-dominant hand (as in our violinists) could spur structural brain differences within the dominant hemisphere (possibly reflected in the LH main effect shown here). Indeed, electrical stimulation of non-dominant hand area can yield facilitatory potentials in dominant hand area, likely arising as a result of callosal transfer (Hanajima et al., 2001). However, a recent study examining motor synchronisation learning showed that post-training tapping synchrony performance correlated negatively with fractional anisotropy in bilateral cortico-spinal tract and bilateral superior longitudinal fasciculi (Steele et al., 2012). Such data appear to indicate a strong role for bilateral fibre integrity in manual motor learning, although callosal fibre FA may also mediate success of inter-manual learning transfer in primates (Phillips et al., 2013). Currently, human data on structural differences associated with long-term fine manual training of the non-dominant hand are sparse; further studies

are needed to explore training-related myelin differences at hand area and across hemispheres.

**3.4.2. Cortical myelination, expertise and development.** Our left-lateralised auditory core group  $R_1$  difference holds implications for the expression of expertise effects in cortex. Previous morphometric and VBM studies have shown increased left-ward asymmetry and left-lateralised increases of planum temporale volume in musicians compared to non-musicians, and in AP musicians compared to non-AP musicians ( Schlaug et al., 1995; Zatorre et al., 1998; Luders et al., 2004; cf. Bermudez et al., 2009). Further, one VBM study also indicated a left-ward asymmetry of grey matter volume at Heschl's gyrus in musicians without AP (Luders et al., 2004).

Our data extend these results, and suggest a pattern of moderately increased cortical myelination at left auditory core in expert violinists compared to closely matched non-musicians. A left-lateralised asymmetry in mean white matter volume underlying the extent of Heschl's gyrus has previously been reported in healthy adults (Penhune et al., 1996). Such asymmetry was interpreted as facilitating increased conduction velocities to left hemisphere Heschl's gyrus (via greater myelination), promoting enhanced fine temporal processing by left primary auditory cortex (Penhune et al., 1996; see also Warrier et al., 2009; Zatorre & Belin, 2001). However, further studies have questioned the validity of such conclusions (Rademacher et al., 2001): the subfields (Te 1.0 and Te 1.1) of primary auditory cortex occupy only the medial two-thirds of Heschl's gyrus (and are best isolated by profiles of cortical rather than subcortical myelination; Dick et al., 2012; see also Morosan et al., 2001). More recent data suggest that cortical myelin profiles at auditory core show a lesser degree of left-ward asymmetry than adjacent temporal cortex (planum

temporale and superior temporal sulcus; Sigalovsky et al., 2006) (note however that the small sample size in that study meant these measurements were inherently noisy). Nevertheless, studies of expert phoneticians (who spend considerable time training to discriminate and transcribe speech sounds) have shown increased white matter volumes at Heschl's gyri bilaterally compared to controls (with left-lateralised effects lying just posterior to primary auditory cortex) (Golestani et al., 2011).

Our current findings from expert musicians and matched non-musicians do not support a general left-lateralised increase in myelin proxies at auditory core in adults (we did not find a hemisphere main effect) (cf. Penhune et al., 1996; Warrier et al., 2009). Instead, we find support for greater  $R_1$  at left core in expert violinists (further to Sigalovsky et al., 2006). Moreover, our results suggest that this effect is less robust across the full extent of Heschl's gyrus (cf. Penhune et al., 1996; Golestani et al., 2011), and is focused within the more medial region of higher  $R_1$  corresponding with auditory core.

Indeed, the quantitative nature of  $R_1$  allows us to appraise the extent of difference between our cohorts, by comparing our results to existing parcellations of cortex that used this method. Sereno et al. (2012) reported  $R_1$  differences of approximately 15 ms between V1 and a probabilistically-defined region having high correspondence with area MT; further, Dick et al. (2012) reported an  $R_1$  difference of up to 25 ms between the most medial and antero-lateral aspects of Heschl's gyrus. By comparison, the extent of average  $R_1$  differences we found between our violinists and non-musicians at left auditory core were on the order of 15 ms (scaled  $R_1$  mean  $\pm$  SD; 0.4 depth: violinists -  $668.31 \pm 16.96$ ; non-musicians -  $654.64 \pm 22.99$ ; 0.5 depth: violinists -  $657.03 \pm 16.1$ ; non-musicians -  $644.08 \pm 23.95$ ). These results suggest that the

difference between experts and non-experts at left auditory core reflects a similar extent of  $R_1$  difference between (for instance) primary and higher visual cortical regions (e.g., V1 and MT as measured by Sereno et al., 2012), but less than the differences measured between medial and antero-lateral Heschl's gyrus (Dick et al., 2012).

In line with our expectations, that we observed differences at middle (rather than superficial or inferior) cortical depths over left auditory core may suggest that violinists achieve relatively greater myelination of thalamocortical afferents to left core (likely arising from ventral medial geniculate nucleus; see Hackett, 2011; Kaas & Hackett, 2000). However, such an account is speculative. Allying the present results with tractography analyses may allow for more detailed explanation of the sources of the observed  $R_1$  differences at left core (an analysis program we intend to pursue).

In addition to expertise effects at left auditory core, we found a cluster of vertices at right superior temporal gyrus (rSTG) that showed a linear relationship with violinists' years of training (violinists with greater training showed higher  $R_1$  values). Moreover, the effect remained significant (albeit at a slightly lower threshold) when controlling for age. Notably, we did not find evidence of a significant group difference at the cluster. However, we did observe a weak trend toward a relationship between age and  $R_1$  values at the peak vertex for non-musicians ( $p = 0.09$ ); no such trend was evident for violinists. This might suggest that  $R_1$  values at the cluster reflect myelination in line with broader ageing and development in non-musicians, but mirror profiles of training experience in violinists. We also found that violinists' years of training predicted cortical  $R_1$  across left (but not right) Heschl's gyrus ROI (at 0.5 cortical

depth), although the effect was not robust when controlling for age (see Golestani et al., 2011; cf. Sluming et al., 2002).

The location of the rSTG cluster is consistent with previous imaging studies showing selectivity of right anterior STG towards: greater spectral than temporal acoustic complexity (Zatorre & Belin, 2001; PET); fixed pitch versus noise information, and diatonic or random melodies versus a fixed pitch (Patterson et al., 2002; fMRI). Bilateral responses to many of these acoustic properties were found in these studies (see also Griffiths et al., 1998); notably, the melodic versus fixed pitch contrast shown in Patterson et al. (2002) had a right-lateralised bias with a peak at a similar location to our cluster. We did not evaluate sensitivity to complex melodies alone in our behavioural experiments. However, it is possible that violinists' expertise with perception of complex acoustic spectra and analysis of melodic structures could lead to enhanced cortical myelination at anterior rSTG. Recent data from mouse models has shown that functional activity can directly influence myelin precursors: excitation of layer V pre-motor neurons that stimulated relevant forelimb movement in mice yielded increased proliferation of pre-myelin astrocytes (oligodendrocyte progenitor cells) at pre-motor layer V (Gibson et al., 2014). While the imaging evidence discussed above is functional in nature, our data may offer insight into related structural change arising from behaviourally-relevant auditory processing in expert violinists.

**3.4.3. Cortical myelination: behavioural implications.** A further goal of the present study was to relate cortical myelin proxies within auditory ROIs to behavioural indices of fine-grained auditory perception, and auditory attention, measured in the same participants. In particular, we explored whether the differences in acuity for instrument-relevant acoustical properties (onset rise

time, FM depth and AM depth) that we previously found between the cohorts would account for variance in cortical  $R_1$ .

We found that fine perception of AM depth (but neither FM depth, onset rise time, nor sustained auditory attention ability) accounted for significant variance in  $R_1$  at RH auditory core (with a marginal trend at LH core). For both core ROIs, a negative trend emerged: those with lower (i.e., more fine-grained) AM depth thresholds tended to have higher cortical  $R_1$  values. However, when we investigated these effects further at whole-brain level, we found relatively weak evidence of the negative trends noted for the ROI  $R_1$  means. Notably, the ROI effect spanned both groups; we found several non-musicians with low AM depth thresholds and correspondingly high mean  $R_1$  at core (see Figure 3.6).

Such effects might suggest a less expertise-specific relationship between fine perception and cortical myelination at auditory core. Given that variation in temporal envelope occurs in many auditory domains (e.g., speech; Shannon et al., 1995; Drullman et al., 1994), it is highly likely that the ability to perceive fine differences in temporal envelope extends to non-musicians also (e.g., Huss et al., 2013).

Previous fMRI studies have shown enhanced processing of fine temporal differences to be left rather than right-lateralised at anterior superior temporal plane in healthy adults (Zatorre & Belin, 2001; Schönwiesner et al., 2005). A similar functional asymmetry has been linked to the structure of Heschl's gyri: gross volumes of left (but not right) Heschl's gyrus correlated positively with the extent of BOLD activation for increased temporal complexity, whilst gross volumes of right (but not left) Heschl's gyrus correlated positively with activation for increased spectral complexity (Warrier et al., 2009).

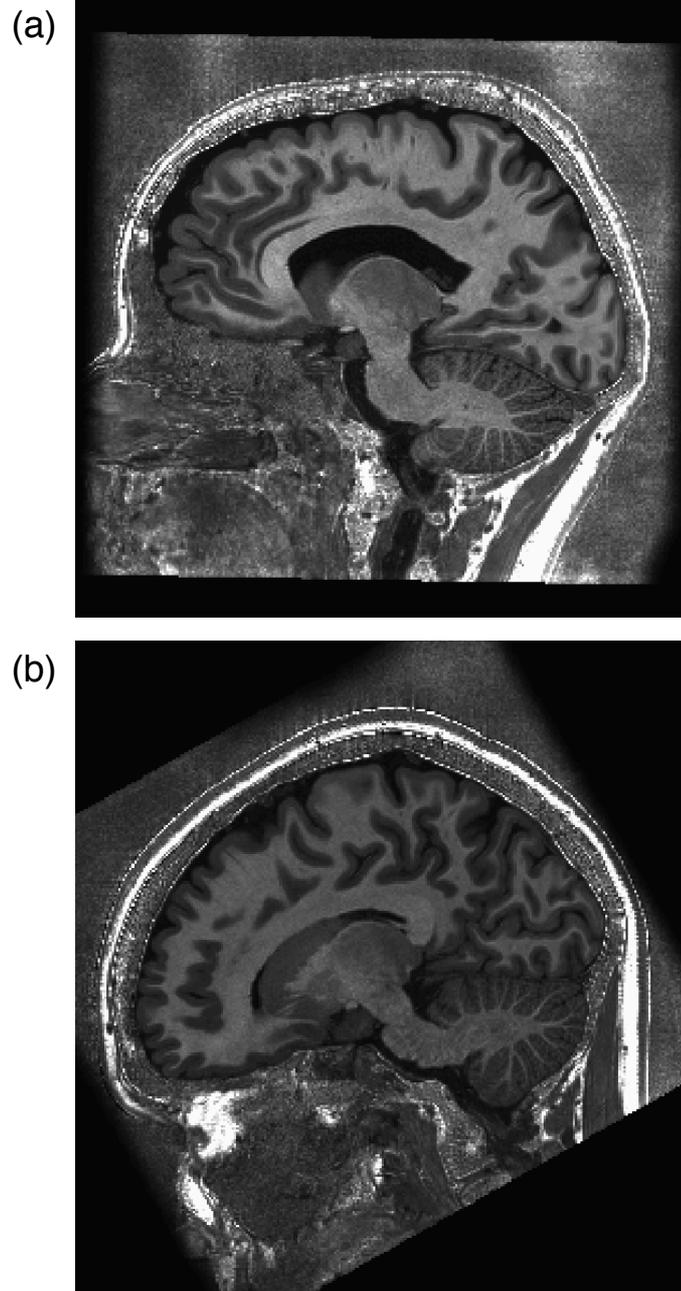
Nevertheless, our results suggest that left-lateralised increases in cortical myelination at auditory core (particularly in expert violinists) may only partly account for enhancements in behavioural measures of fine temporal perception (cf. Penhune et al., 1996; Warrier et al., 2009; see also Herdener et al., 2013). Indeed, improved melodic and rhythmic processing in both musicians and non-musicians has previously been shown to predict right-lateralised increases in cortical volume of Heschl's gyrus (Hyde et al., 2009), and increases in the mean volume of Heschl's gyri combined over hemispheres (Schneider et al., 2002; 2005; see also Seither-Preisler et al., 2014).

Taken together, the present results suggest that average cortical myelin proxies at auditory core can be predicted by fine temporal envelope perception metrics. However, these findings do not show a clear-cut agreement with previous suggestions of left hemisphere dominance for temporal processing, or agreement with right-lateralised musical training-related results. Given that our effects also manifested weakly at whole-brain level, these  $R_1$ -AM depth threshold relationships should be interpreted cautiously; future replications of these relationships may shed light on the robustness of cortical myelin and temporal envelope perception relationships.

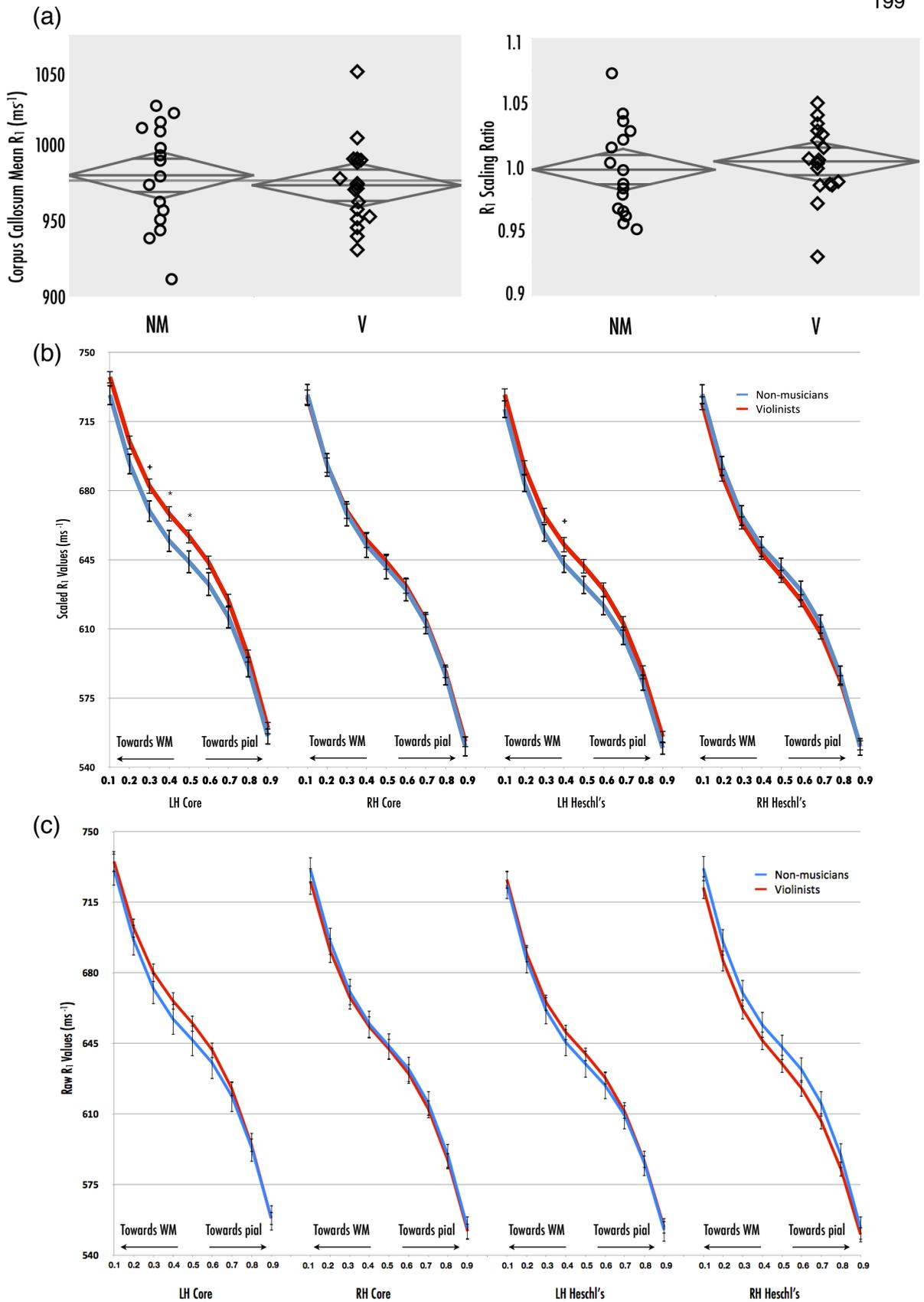
**3.4.4. Conclusions.** The present results suggest some evidence for expertise-based cortical structure differences within left auditory core, reflecting increases in quantitative proxies for cortical myelination. Further, we find that expert violinists' training extents show a linear relationship with myelin proxies at right-lateralised superior temporal gyrus; this may suggest a role for experience in shaping increases in cortical myelination within non-primary auditory cortical areas. Nevertheless, quantitative myelin metrics did not differ between primary motor hand area regions across hemispheres in experts and

non-experts. Such a result may indicate a role for bilateral transfer in shaping motor cortex myelin profiles in experts, revealing differences in cortical adaptations between auditory and motor networks. Finally, we demonstrate that fine-grained perception of temporal envelope cues may relate to cortical myelin proxies; however these results do not adhere strongly to the left-lateralised pattern we find for profiles of expertise. These cortical myelin and perception relationships are best interpreted cautiously.

Thus far, we have explored experience, learning and plasticity with respect to long-term models of expertise (via cohorts of expertly trained musicians and closely matched non-musicians). In the coming chapters, we explore these issues over short-term periods, using novel, interactive tasks to examine learning within the auditory modality and across modalities. In particular, we explore whether complex auditory cues may be combined and learned, and moreover, whether learning success may be mediated by development.



Supplemental Figure 3.1:  $R_1$  volumes for two participants; (a) shows participant with data acquired in canonical orientation; (b) shows participant with data acquired with slab rotated at 30 degrees.



Supplemental Figure 3.2: Effects of scaling  $R_1$  values by corpus callosum (CC)  $R_1$  mean. (a) left: corpus callosum  $R_1$  mean for each cohort (no significant difference,  $z < 0.9$ ,  $p > 0.3$ ; horizontal line: cohort grand mean); right:  $R_1$  CC scaling factor for each cohort (no significant difference,  $z < 0.9$ ,  $p > 0.3$ ). (b) and (c): auditory ROI depth fraction  $R_1$  data with (b) versus without (c) corpus callosum scaling applied; note that group means remain highly similar for each ROI with scaling applied compared to no scaling; SEM is reduced by scaling, notably at left auditory core.

Chapter 4: Short-term Auditory Learning  
within a Multi-modal Environment: Theory  
and Normative Data

## 4.1 Introduction

We are regularly faced with highly complex environments that pose major challenges to behaviour, attention, and learning. Imagine a family walking along a busy street at evening rush hour. Amidst a dynamic visual landscape of passing pedestrians, traffic, street lights and shop fronts, the family encounters a vast auditory scene. Tyre and brake noises, footsteps, sirens, conversation, bicycle bells and café music meld together. Our observers are thus faced with the non-trivial task of perceiving, attending and navigating within an incredibly dense sensory space. Moreover, the disparate lifetime experience our adult and child family members have with these acoustic and visual events (and the related behavioural and attentional demands) may differently impact their performance within this environment. Given such challenges, how can each of our observers begin to overcome these demands and use the available information in a way that optimises their behaviour and performance?

One possible strategy is to combine available cues within the auditory scene to guide behaviour with respect to relevant events. Indeed, sound may be critical to guiding both attention and behaviour, particularly given limitations upon visual processing (e.g., the forward position of the eyes and limited peripheral acuity; see Spence & Driver, 1997). Bregman (1990) described the auditory scene as a composite that arises from on-going sound events within the environment. In a manner analogous to vision, sounds fuse to produce a scene composed of streams that must be segregated by the listener; once segregated, distinct auditory streams or events may then be processed, attended to, and used as cues to inform behavioural decisions (Bregman, 1990; Gygi & Shafiro, 2011; see also Pressnitzer et al., 2011).

**4.1.1. Cue combination.** Of central importance is the manner in which auditory cues are combined within such complex auditory scenes. Indeed, while cues may be abundant within auditory environments, the relative usefulness of given cues can vary widely. Moreover, the utility of a combination of particular cues may be greater than the utility of a single cue alone. The Competition Model of Bates and MacWhinney (1987) offers an elegant account of cue competition and use within spoken language learning and development. Broadly, the model uses a connectionist framework to suggest that language learners infer both the validity and strength of multiple linguistic cues. Cue validity refers to the information value of a given cue, whereas cue strength refers to the relative weight that an organism attaches to a cue (vis à vis its validity) (Bates & MacWhinney, 1987). Central to mechanisms of cue validity are the relative statistics of the cue's occurrence: thus, validity arises as the product of a cue's availability (i.e., its overall proportion of occurrence with respect to all other cues) and its reliability (i.e., a proportion reflecting the number of times the cue is available and leads to a correct outcome, divided by the total availability of that cue) (Bates & MacWhinney, 1987; MacWhinney, Pleh & Bates, 1985). Hence, in learning to employ cues effectively, listeners may track the relative environmental statistics of cues (see 4.1.3), leading to weighting of a linguistic or auditory cue (or combination of cues) based on relative usefulness. Such a mechanism holds clear implications for cases in which auditory cue structure changes or becomes less reliable: if learners seek to use optimally reliable cue combinations, then violation of a cue combination should yield decrements in performance.

**4.1.2. Auditory cues: saliency, context and expectation.** Thus, a central question in the problem of sound cue use is how different combinations

of cues may facilitate or hinder processing of relevant events, in order to optimise behaviour. Listeners may rely on both the properties of incoming sounds and information about the on-going sound context, to attend to and form expectations about events (Kayser et al., 2005; Tsuchida & Cottrell, 2012; Niessen, 2008; Bendixen et al., 2009; see also Sohoglu et al., 2012). For instance, how we represent and store complex sound information reflects both the temporal properties of the sound, and the distribution of long-term spectral and envelope statistics (McDermott et al., 2013). Furthermore, bottom-up acoustical properties such as intensity, spectral content and temporal dynamics can guide attention to relevant sound events, by increasing a sound's relative saliency compared to other on-going sounds (i.e., acoustic context) (Kayser et al., 2005; Kalinli & Narayanan, 2007; Cusack & Carlyon, 2003). These bottom-up acoustical properties of sounds may further influence the status of a sound as an auditory 'object'; that is, a sound defined by specific acoustical features, which allow it to be processed, abstracted and categorised as a distinct perceptual entity (Griffiths & Warren, 2004). In addition, the incidence of sound events within auditory scenes may influence their perceived salience; models that account for the lifetime frequency of occurrence of a sound and its frequency of occurrence in the present on-going scene, suggest that less frequent sound events will typically be perceived as most salient (Tsuchida & Cottrell, 2012).

Moreover, listeners may form on-going semantic expectations about likely upcoming sounds in the auditory scene; for instance, combinations of sound cues can help the listener to determine further sound events that are most probable within a given context (Niessen, 2008; Keller & Stevens, 2004). Such expectations likely emerge across development as a result of extensive

exposure to and interaction with natural acoustic environments (e.g., Krishnan et al., 2013). Previous investigations of environmental auditory scene analysis have shown that contextual expectation can guide detection accuracy for auditory objects. When sounds are unexpected or incongruent (e.g., a dog bark) with the on-going acoustic context (e.g., an office scene), detection accuracy is improved relative to when sounds are congruent (e.g., a telephone in the same office scene) (Leech et al., 2009a; Gygi & Shafiro, 2011). Such findings have been replicated in school-aged cohorts also (Krishnan et al., 2013). Yet how we learn and acquire expectations based on sound events that occur as complex auditory contexts and objects is less well-understood (e.g., Keller & Stevens, 2004).

Indeed, while the above studies suggest that listeners use auditory contexts to form expectations about likely sound combinations ahead of upcoming events, the paradigms nevertheless rely on previously learned representations (that arise from experience with natural corpora of environmental sounds). Further, pre-existing differences in children's and adult's relative lifetime exposure to such corpora limit the extent to which sound context and sound object expectation can be investigated developmentally. Therefore, a learning paradigm where novel auditory objects combine with novel auditory contexts to form cues, may offer highly useful insight into the emergence of experience-driven expectancies (and perhaps expertise) within a task (see 4.1.6). Crucially, the novelty of such a paradigm would overcome developmental limitations in comparing children's and adult's relative prior exposure to sound combinations, whilst allowing further control over the acoustic properties of interest: saliency and on-going (i.e., temporally extended) acoustic context.

**4.1.3. Developmental learning and cue competition.** The issues discussed above have strong implications with respect to development. Although children may display relatively mature basic auditory processes early in development (see Boothroyd, 1996, and Werner, 2007, for review), issues of auditory scene analysis and selection of events within the scene to guide behaviour pose non-trivial demands for children. Children are more susceptible to competing auditory information (e.g., informational masking), both in naturalistic scenes (Krishnan et al., 2013) and for tone bursts presented with non-overlapping frequency maskers (Leibold & Bonino, 2009). Further, increases in the uncertainty concerning masker identity pose even greater difficulties for children compared to adults (Leibold & Neff, 2007).

The implications of such findings for children's performance are considerable. For instance, compared to adults, children require greater distinction (e.g., frequency separation) between incoming sounds to allow for selection of a relevant auditory stream (and performance of a behavioural task on sounds within that stream; Sussman et al., 2007). Moreover, models of children's language development suggest both cue detectability (i.e., the ease with which a cue can be perceived) and cue validity are essential for cue acquisition (MacWhinney et al., 1985).

While listening environments pose considerable challenges for children, studies nevertheless have shown that relatively young children can extract and learn relevant information from on-going auditory stimuli. After passively listening to a concatenated stream of synthetic, unstressed speech syllables (structured so that word boundaries were marked only by low transitional probabilities), 8 month old infants and 6-7 year old children (and adults) can discriminate previously heard tri-syllabic 'words' from tri-syllabic part-words (i.e.,

words that cross the low transitional probability boundary) (Saffran et al., 1996; 1997; cf. McNealy et al., 2011). Such evidence suggests that children can segment and learn specific ‘words’ in the stream, even in the absence of obvious segmental acoustic cues to word boundaries (e.g., stress or prosody; but see Hay & Saffran, 2012). Such effects have also been shown to be non-specific to speech; similar findings have been reported for sequences of pure tones (Saffran et al., 1999), and also visual stimuli (Kirkham et al., 2002; Saffran et al., 2007; Arciuli & Simpson, 2011; but see Conway & Christensen, 2006).

Yet despite this evidence of apparently sophisticated auditory statistical learning in even very young children, laboratory-based training studies have pointed to relative immaturity of auditory perceptual performance and learning in children. For example, cross-sectional data have shown significant reductions in children’s pure tone frequency difference limens as age increases (i.e., for 6-7 year olds vs. 10-11 year olds, but not 6-7 year olds vs. 8-9 year olds), with adults achieving significantly lower thresholds than children (Halliday et al., 2008; see also Banai & Yuval-Weiss, 2013). Moreover, training can improve thresholds across ages (although the extent of improvement in thresholds with training does not vary as a function of age; Halliday et al., 2008). Moreover, the effects of such training are largely specific to the trained task (e.g., frequency discrimination) and do not transfer to other auditory or speech-based tasks (e.g., non-word repetition, rhyme judgement) (Halliday et al., 2012) (although note that some studies suggest children trained on discrimination of speech phoneme continua transfer this learning to improved phonological awareness and word-in-noise discrimination abilities; Moore, Rosenberg & Coleman, 2005; see also Merzenich et al., 1996; Hayes et al., 2003; cf. Halliday, 2014).

Importantly, children's performance and learning on training tasks can vary widely, with group means often masking true performance patterns (Moore et al., 2005; 2008; 2011; Moore, Halliday & Amitay, 2009). In some cases, children can display adult-like levels of performance early in testing; however, a proportion of children may also display generally poorer thresholds, whilst others (~50%) may fail to comply with the training procedure (notably where adaptive protocols are used; Moore et al., 2008; 2009). Indeed, the relative difficulty younger cohorts have with respect to intensive auditory training and learning has been documented recently. In two studies, many adolescents failed to show improvements in backward masking performance (Huyck & Wright, 2013) and temporal interval discrimination (Huyck & Wright, 2011) after 10 days of training; in fact, a subset showed significantly worse performance. Adults and a further subset of adolescents did improve, although adolescents learned more slowly than adults (Huyck & Wright, 2013; 2011), whilst younger (11 year old) subjects did not learn (Huyck & Wright, 2011). Moreover, adult and adolescent learners transferred their learning to related conditions (e.g., backward masking with 10 ms pre-noise gap or a different bandpass noise) but not other conditions (e.g., forward masking). Yet adolescents who did not learn generalised their poorer performance to other conditions (i.e., forward masking) (Huyck & Wright, 2013). It is unclear however whether the consistently poorer performance of younger groups across training in these studies was due to inattention, fatigue or low motivation (training sessions were spread over 10 consecutive days).

Given the above findings of the difficulties complex auditory demands pose for children, few studies to date have explored children's in-task cue learning under conditions where multiple auditory cues of varying salience are present. Moreover, even less research has considered whether children can

indeed learn relevant auditory cue combinations, particularly where auditory objects can be distinguished from an on-going acoustic context. Such questions are critical to understanding how children are influenced by on-going auditory context, stimulus salience, and occurrence of sound events with respect to a task goal.

Further to the discussion above, the ability to combine information optimally within and across modalities remains immature up to early adolescence (Nardini, Bedford & Mareschal, 2012), showing relatively rapid development during late childhood (rather than a gradual improvement with increasing age; Barutchu, Crewther & Crewther, 2009). For instance, 4-5 and 7-8 year old children fail to combine haptic and visual information during spatial navigation, showing greater reliance on a single modality than even a non-optimal combination of information from both modalities (Nardini et al., 2008). Investigation of cross-modal processes also suggests that children differ from adults in the temporal window over which integration occurs: when auditory stimuli occur before visual, 10 year old children are more likely to report the stimuli as simultaneous at inter-onset intervals of up to 300 ms (differing significantly from adults, who show a decline in simultaneity judgements at inter-onset intervals  $> 100$  ms); yet the reverse asymmetry (i.e., visual before auditory) does not reflect the same increased temporal window for binding in children (i.e., children appear more adult-like) (Hillock et al., 2011). The findings of these studies hold implications for the extent to which children might make use of relevant information in one modality (e.g., audition), to inform behavioural decisions in another (e.g., vision, proprioception).

**4.1.4. Learning mechanisms and goal-directed behaviour.** In exploring online sound learning and developmental differences in learning of

cue combinations, task-specific demands are of critical importance. Animal models related to task-dependent goals and attention have helped to inform accounts of learning mechanisms. Animal studies have shown robust effects of sound learning that persist far beyond initial training (see Weinberger, 2004, for review). For instance, learning of tone pairs in rats is associated with enhancement of cortical maps at the frequency and expected onset time of the second tone of a pair (Zhou et al., 2010). Moreover, animal models have emphasised the role of task engagement versus passive exposure. Rats that learned to find locations in a cage that matched specific sound intensities (receiving food rewards upon finding the locations) showed a higher proportion of neurons that had best responses over a range of intensities presented during training; yoked controls showed best responses only for isolated intensities that were food-reinforced regardless of behaviour (Polley et al., 2004; see also Whitton et al., 2014). Similarly, ferrets aversively conditioned to inhibit a licking behaviour for target sinusoids presented within complex broadband sounds (temporally orthogonal ripple combinations) showed adaptations to neuronal spectro-temporal receptive fields (and profiles of lateral inhibition) close to the target sinusoid frequencies; control animals that did not perform the behavioural task (but passively encountered the stimuli) showed no such adaptations (Fritz et al., 2003; 2005).

In humans, task engagement serves as one key mechanism that can enhance learning of acoustic cues. Holt and colleagues (e.g., Wade & Holt, 2005; Lim & Holt, 2011) have shown that learning of variable, spectrally-complex sound cue categories that may be used to facilitate a relevant task (capturing versus shooting visual characters) is greatly enhanced in adults that progress to the furthest (i.e., most difficult) stages of the task. Moreover, in a

pre-post training fMRI study using this paradigm (Leech et al., 2009b), adults who became 'experts' over the course of the study (classifying the complex sound categories most accurately at post-test), also showed greatest increases in pre-post training BOLD signal change at left superior temporal sulcus (a region previously thought to respond selectively to intelligible speech; Leech et al., 2009b; Leaver & Rauschecker, 2010; Shultz et al., 2012; Agnew et al., 2011; see Scott & McGettigan, 2013). Mechanisms related to task reward and feedback also have been shown to play a critical role in building learned associations, tied to dopaminergic signalling of expected and unexpected reward onset (i.e., reward prediction error; see O'Doherty et al., 2003; McClure et al., 2003; Shohamy et al., 2004).

Based on the preceding discussion, active task engagement may be one means of facilitating auditory cue learning. An open question is whether adults and children might differ in their learning of useful auditory cue combinations during an active task. If the same active task demands pose a greater challenge for children than adults in general, then little to no cue learning and generally poorer task performance (e.g., longer reaction times and lower accuracy) might be expected in children compared to adults. However, if the active demands of the task pose a similar challenge for children and adults, then differences in overall learning outcomes may be attributable to other sources (e.g., difficulty in combining cues, or in attending to auditory information).

**4.1.5. Attention, cross-modal cuing and learning.** Modulation of attentional focus has been suggested as a key component arising from task learning. As outlined above, task engagement and learning may establish an initial bottom-up target for attention; this may then be selected and focused on via top-down mechanisms (Fritz et al., 2003; 2007; see also Posner, 1980).

However, one key issue that arises from questions concerning auditory learning is how auditory information is combined or integrated with visual information, since as discussed above (see 4.1), auditory cues may provide critical support to performance in the visual modality.

Studies of contextual cuing have shown robust contextual learning both for visual arrays alone (Chun & Jiang, 1998; Chun, 2000) and for auditory cues that signal the location of an upcoming visual stimulus (Kawahara, 2007; but see Brown et al., 1989). Further, animal findings suggest spatial navigation is optimised by combining multiple auditory cues with an isolated visual cue (Rossier et al., 2000). However, relatively little research in humans has investigated whether complex auditory cue learning can be influenced by attention, either within a single modality or across modalities.

Those studies that have explored attentional cuing in humans have shown that spatial cuing using overt (i.e., non-symbolic) cues can yield priming of visual targets by auditory cues but not vice versa (when eye movements are prevented or stimulus onset is rapid; Spence & Driver, 1997). However, where attention is guided covertly (i.e., by symbolic cues like arrows), cuing the probable location of an auditory target can also enhance detection of less frequent visual targets at that same location (Spence & Driver, 1996). Such findings question whether attentional process related to cue learning may be considered cross- or 'supramodal' (e.g., Farah et al., 1989), or whether both modalities operate independently during attentional deployment (Spence & Driver, 1997; Braga et al., 2013).

Despite the evidence above, few studies have explored whether learning of combinations of auditory contexts and auditory objects might be influenced by attentional allocation and performance. Although auditory scene analysis and

streaming can be influenced by attentional load (e.g., Alain & Izenberg, 2003; Cusack et al., 2004), it is not clear how attention influences learning of auditory cue combinations, whether that learning might be susceptible to attentional interference (e.g., Watkins et al., 2007), or whether individual and developmental differences in attentional performance might account for differences in learning auditory cue combinations (see Gomes et al. 2000).

Moreover, if complex auditory cue combinations can be learned, one further question is whether the combined auditory representations may be mapped cross-modally to cue a visual event. Although such cuing might be attentionally demanding, it would offer a means of using available cues in an optimal fashion. Finally, given that the ability to combine cues cross-modally develops relatively later in childhood (see 4.1.3), it is possible that adults may be more effective than children in learning cross-modal combinations of cues.

**4.1.6. The current studies.** The major goal of the present studies was to explore how fruitful associations might be built between combinations of complex auditory cues and visuo-spatial events, where use of the auditory cues was beneficial to behavioural performance in an active visual task.

Our main aim was to explore the ability to learn a combination of acoustic cues that formed novel auditory ‘scenes’. Each ‘scene’ comprised an on-going, broadband sound (loosely analogous to environmental broadband sound cues, e.g., McDermott et al., 2013; sound synthesis differed from McDermott et al. – see 4.2.3.1). Amidst the broadband sound, a punctate sound also occurred that was spectrally swept and temporally modulated so as to be relatively more salient than the broadband sound (based on Kayser et al., 2005). Thus, the broadband sound served to provide an auditory context on each trial; that is, it was temporally longer, not systematically modulated (compared to the punctate

sounds), and provided a period during which expectancy for events could be built (Niessen, 2008). In contrast, the punctate sounds were discrete auditory 'objects' (i.e., distinct, shorter, spectrally swept and temporally modulated sounds, independent of the context; see Griffiths & Warren, 2004). Our goal was to determine whether these two distinct sources of auditory information (context and punctate sounds) could be optimally combined and learned, based on their high cue validity with respect to an on-going visual task. Thus, within the visual task, the combination of both auditory cues provided more information than either cue alone. Learning was indexed via online improvements in task performance (and decrements when cue combination contingencies were switched), and post-task sound-visual location identification accuracy (4AFC).

A second major goal was to understand whether developmental constraints on auditory cue combination lead to clear differences in learning or task performance between children and adults. If children fail to associate auditory contexts and auditory objects with visual events (or simply attend to the visual modality), then no learning should be observed at 4AFC. Further, if children struggle to learn combinations of cues, then changes in the audio-visual contingencies within the task should not yield any significant effect on their in-task performance. However, if children attempt to combine cues yet experienced difficulty in learning mappings due to the complexity of the auditory 'scenes', then performance at 4AFC might reflect consistent confusion over auditory cues (particularly less salient ones).

A final aim was to examine whether differences in attentional performance could account for task performance or learning outcomes. The goal in this respect was to use metrics of sustained auditory attention (SAART) collected in adults, and sustained auditory (SAART) and selective visual (TEA-

Ch) attention collected in children, to explore relationships between learning and attentional performance.

In this chapter, we report normative data demonstrating that auditory contexts (experiment 4a) and auditory objects (experiment 4b) presented in isolation could be learned within a simple version of the visual task. This was important, as it laid the foundation for our subsequent experiments, by showing that each sound class could be learned. In chapter 5, we explore learning of combinations of these auditory cues in adults (experiment 5a), a paradigm we later extend to children (experiment 6). We further investigate whether properties such as acoustical saliency can account for learning outcomes and cue cost associated with cue contingency violation (experiment 5b).

## 4.2 Method

### 4.2.1 Participants

All subjects were right-handed, had normal or corrected to normal vision, and reported no history of hearing difficulties, hearing loss, or neurological insult or trauma.

**4.2.1.1. Experiment 4a.** Participants were 13 undergraduates recruited from the Birkbeck Department of Psychological Sciences participant pool (6 male, 7 female; mean age  $\pm$  SD: 26.9  $\pm$  5.1; range: 18-40). Participants received course credit or payment of £5 for participation.

**4.2.1.2. Experiment 4b.** Participants were 20 adults (8 male, 12 female; mean age  $\pm$  SD: 23.9  $\pm$  5.3; range: 18-44) recruited from Birkbeck Psychology undergraduate classes or from other courses within the University of London; a subset were working professionals. Participation was voluntary; no reimbursement or other incentive was provided.

### 4.2.2 Materials

All stimuli were created using Adobe Audition CS5.5 and Matlab 2010a (32-bit), running on MacBook Pro laptop computer (OS 10.7), and Adobe Audition 2.0 running on HP Pavilion dv2000 laptop computer (Windows XP). Stimuli were presented on MacBook Pro (OS 10.6 or 10.7) 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 by USB. All sounds were presented at a comfortable level fixed for all participants (50% sound card output or 50% iPad volume). Responses were logged in Matlab via USB using a Logitech Precision Gamepad.

### 4.2.3 Stimuli

Two classes of auditory stimuli were designed. The first class comprised four distinct complex broadband sound cues, each with duration of 900 ms (experiment 4a) (see footnote 1). Each sound was composed of three separate bands of filtered white noise; bands were combined to produce a single complex broadband cue (hereafter referred to as a contextual sound [CS]) in each case. The second class consisted of four distinct frequency swept, amplitude modulated noise segments, each of 300 ms duration (experiment 4b).

**4.2.3.1. Experiment 4a.** Complex, broadband sounds were synthesised such that each sound comprised three filtered noise bands, yielding structured broadband sound cues analogous to environmental noise (e.g., McDermott et al., 2013). Each band was produced by filtering 12 instances of the same initial 900 ms segment of white noise using a 4th order Butterworth filter (Audition 2.0), with varying filter centre frequencies (CFs) (see table 4.1). Individual noise bands were separated from each other by more than half the equivalent rectangular bandwidth (ERB) of their CF (such that the upper and lower ERB bound for a given CF did not overlap with the upper and lower ERB bound of the nearest adjacent CFs). ERBs were estimated using the formula of Moore and Glasberg (1983) ( $6.23f^2 + 93.39f + 28.52$ , where  $f$  is filter CF) and then rounded down to the nearest whole number in Hz.

Next, bands were grouped based on arbitrary subdivision of the CFs into low (< 1000 Hz), medium (>1000 Hz & < 2000 Hz) and high (> 2000 Hz) groupings. The bands that formed each contextual sound (CS) were then chosen, such that each CS consisted of one low, one medium and one high band. Close harmonic relationships (e.g., octaves) between bands within the same sound were avoided as much as was possible. Care was also taken to

ensure that listeners did not consistently rate the CSs as belonging to a specific category (e.g., speech phones, environmental sounds, etc.). Initial pilot testing suggested that listeners tended to rate the CSs as distinct from one another, and as sounding ‘unlike speech’ or ‘alien-like’. Table 4.2 displays the CFs of the bands that each CS was composed of. Individual bands were saved out as separate audio files (.wav format). Sound files for individual bands of each CS were added to a multi-track session in Audition (CS5.5) and then mixed out as a single mono audio track for each CS (44.1 kHz sampling rate, 16 bit quantisation). Mixed CS files were linearly tapered with 20 ms onset and offset

**Table 4.1:** Filter CFs and ERBs (both in Hz) for noise bands

<b>Band</b>	<b>CF</b>	<b>ERB</b>
1	253	52
2	473	74
3	693	96
4	913	118
5	1133	142
6	1353	166
7	1573	190
8	1883	227
9	2203	264
10	2563	308
11	2933	356
12	3353	411

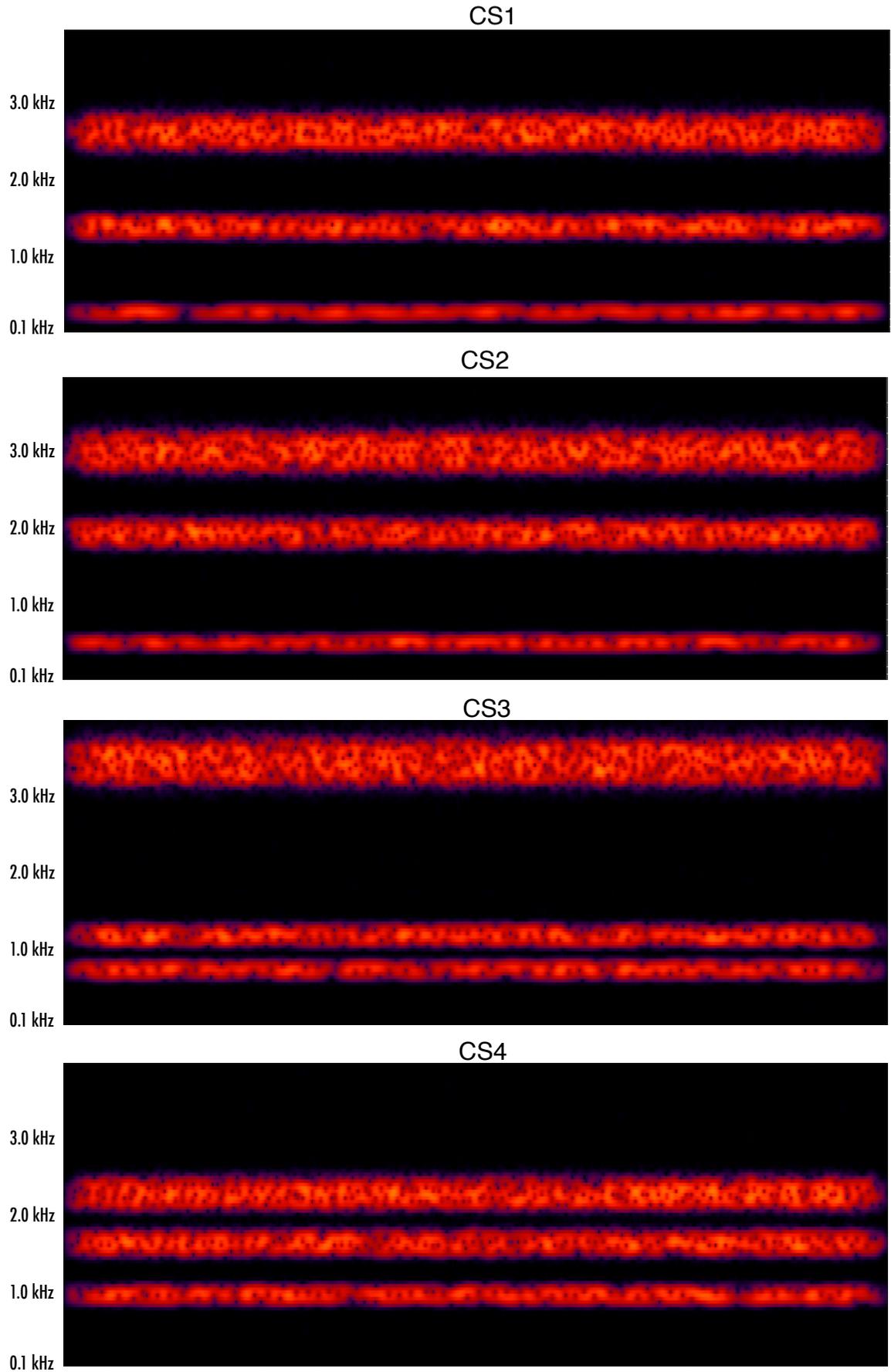


Figure 4.1: Spectrograms of CS stimuli used in experiment 4a (900 ms durations shown).

**Table 4.2:** CFs (in Hz) of filtered noise bands that comprised each CS

CS	Low Band	Mid Band	High Band
1	253	1353	2563
2	473	1883	2933
3	693	1133	3353
4	913	1573	2203

ramps, using a custom Matlab function. Finally, each of the four tapered CS files were normalised to equal total RMS amplitude (using Audition CS5.5). Spectrograms of each CS are displayed in Figure 4.1.

**4.2.3.2. Experiment 4b.** Stimuli for experiment 4b differed qualitatively from those used for experiment 4a. Experiment 4b stimuli were frequency swept noise bands, with varying rates of amplitude modulation applied.

Four instances of the same 300 ms white noise segment (excised from the beginning of the 900 ms noise sample used for experiment 4a stimuli) were filtered using dynamic equalisation in Audition 2.0. Four separate stimuli were produced (hereafter referred to as punctate sounds [PS]), frequency swept over a 2-4 kHz CF range according to the parameters outlined in table 4.3. Stimuli were narrow band in nature, determined using Q ratio (i.e., ratio of filter width to CF) values of 13 at CF of 4 kHz, decreasing to 10 at CF of 2 kHz. Next, each PS was amplitude modulated (AM; 100% depth) with a different modulation rate using the amplitude envelope processing tool in Audition 2.0. AM rates for each PS are also presented in table 4.3. Manipulation of frequency sweep and AM rate parameters were selected for stimuli, since within the acoustical saliency model of Kayser et al. (2005), such parameters have been shown to yield increases in detection performance and perceived salience (in humans and

primates), agreeing with the predictions of ‘saliency maps’ that code for these (and other) acoustic features. Separate sound files were created for each PS (.wav format; 44.1 kHz sampling rate, 16 bit quantisation). Sound files were linearly tapered with 20 ms onset and offset ramps, using a custom Matlab function. Finally, tapered PS files were normalised to equal total RMS amplitude (using Audition CS5.5). Spectrograms of PS stimuli are presented in Figure 4.2.

**Table 4.3:** Frequency sweep (CF kHz) and AM rate (Hz) parameters for each 300 ms PS

<b>PS</b>	<b>0-150 ms</b>	<b>150-300 ms</b>	<b>AM rate</b>
1	swept 2-4 kHz	swept 4-2 kHz	26.67
2	swept 4-2 kHz	swept 2-4 kHz	40
3	steady-state 2 kHz	swept 2-4 kHz	13.33
4	swept 4-2 kHz	steady-state 2 kHz	6.67

#### **4.2.4 Procedure**

All participants provided signed informed consent prior to beginning experimental sessions.

**4.2.4.1. Experiment 4a.** Participants sat in a softly lit, acoustically dampened booth in front of a laptop computer. Prior to beginning the experiment, participants were briefed on the nature of the task. Participants were informed that a series of visual ‘alien’ characters would appear on-screen against an outer space background. However, participants were not told anything specific about the sounds in the task or their function.

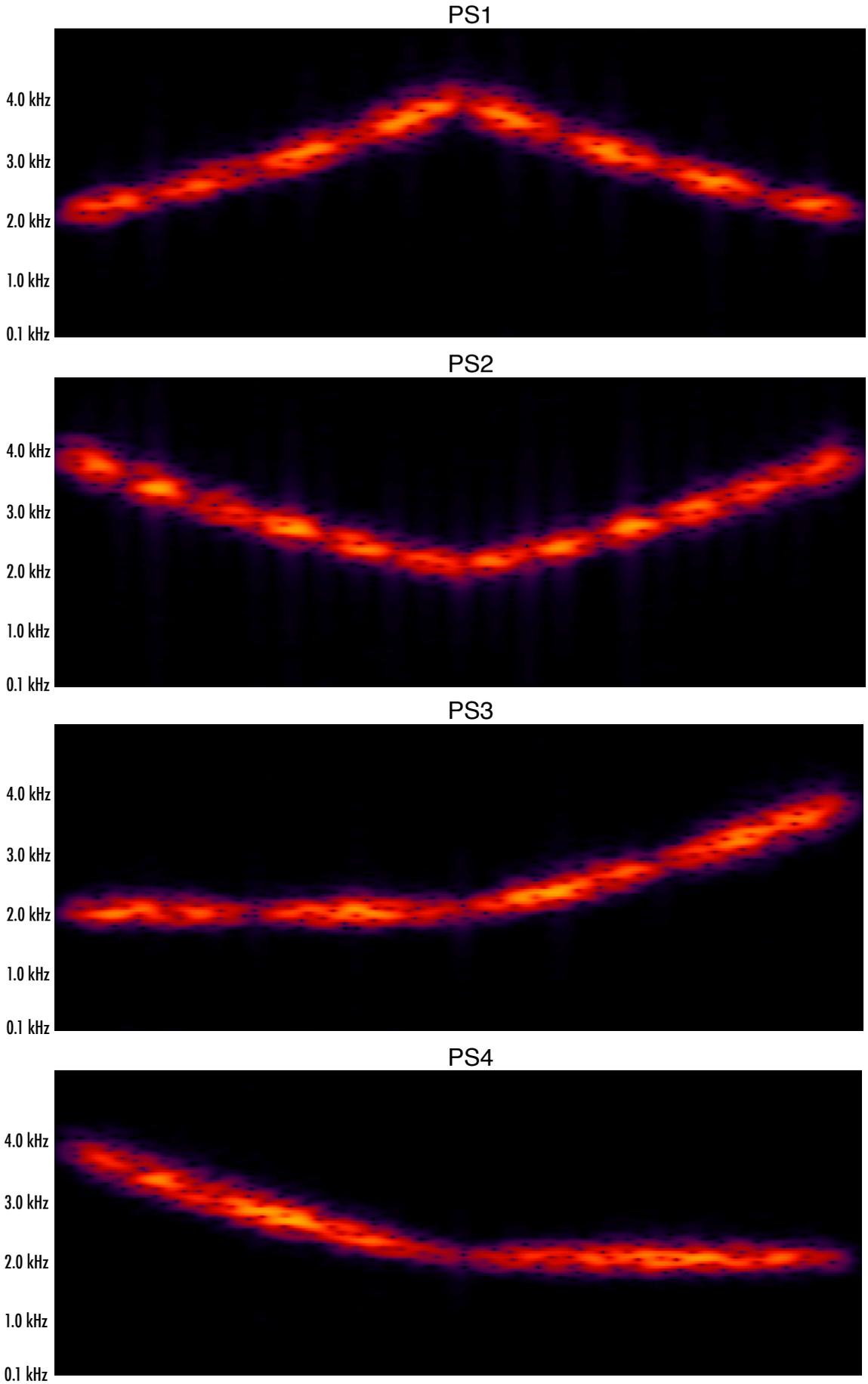
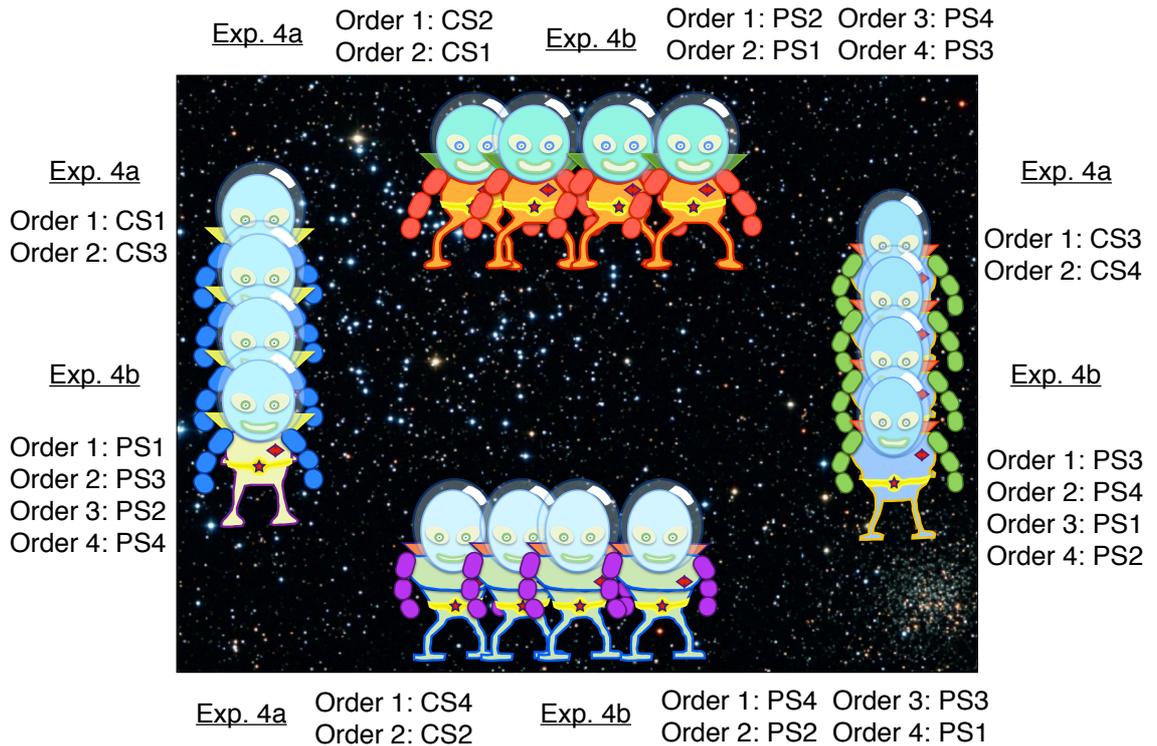


Figure 4.2: Spectrograms of PS stimuli used in experiment 4b (300 ms duration).

Each alien character appeared at its own distinct on-screen section (either top, bottom, left, or right of the space background). In addition, the exact position at which each particular alien appeared was jittered; thus, there were four particular positions in a given section on the background where a specific alien would appear (see Figure 4.3).

Onset of each alien character on-screen was preceded by one of the contextual sounds (CS), that provided an invariant cue to the section on the background image the upcoming alien would appear in. Each trial began with a 900 ms CS in tandem with presentation of the space background, followed by 300 ms of silence. After the 300 ms of silence had elapsed, a single alien character appeared at one of its four positions, within its particular section of the background image. Participants were instructed to 'catch' each alien character as soon as it appeared, by pressing the Logitech Precision Gamepad button that approximately matched the alien's on-screen section. If a participant pressed the correct gamepad button, a green crosshairs appeared over the character and the alien then disappeared. If participants pressed an incorrect button, the alien disappeared without the crosshairs appearing. If no response was made within 6 s, the trial timed out. Each trial ended with a black screen showing white text, displaying whether the participant had caught the alien, their current score, and a prompt to start the next trial (trial start was self-paced, by press of the 'c' key). Participants completed 8 blocks of 16 trials each (128 trials total). CSs were never repeated on consecutive trials (i.e., the same alien never appeared twice in a row). Matching of aliens and each CS was counterbalanced over two pseudorandom orders; pilot testing showed that two pseudorandom orders were sufficient to balance out any biases in responses and learning due to spatial position and/or sound-alien matchings.



**Figure 4.3:** Alien characters and on-image positions for experiments 4a and 4b (see below). On each trial, a single alien appeared in a particular section of the background image, at one of its four possible positions (note that position was jittered over trials). Sound-alien matchings for each counterbalanced order in experiments 4a & 4b are indicated at edges.

Immediately after completion of the task trials, participants completed a 24-trial 2AFC. The 2AFC tested participants' ability to match the alien characters with the CSs they had heard in the preceding task phase, thereby indexing learning. Participants were naïve to the fact that a 2AFC would be presented. On each 2AFC trial, a pair of the alien characters that had appeared during the task were presented (each alien at one of their specific positions from the task, held constant over 2AFC trials; i.e., visual position was not jittered). After 800 ms, one CS was presented. Participants then had 6 s to press the appropriate gamepad button to select the alien they believed matched with that CS.

**4.2.4.2. Experiment 4b.** Procedure for experiment 4b was identical to experiment 4a, with exception of stimuli; PSs were used instead of CSs. Note that PSs were shorter (300 ms) in duration than CSs (900 ms); in-task timing of events was hence the same as in experiment 4a, except for the difference in stimulus duration. In addition, two further counterbalanced orders of sound-alien matchings were used; thus, across the four orders, each sound was matched with every alien (see Figure 4.3). This change to the procedure for experiment 4a was made due to concerns that the frequency swept nature of PS stimuli might lead to response biases depending on whether a cue swept up or down, relative to the screen location it cued (for instance, a perceived incongruence between a downswept stimulus cuing the top of the screen may have biased participants' responses).

**4.2.4.3. SAART.** In addition to the learning tasks described above, all participants completed a response inhibition version of the SAART task as described previously (see chapter 2). The measure was collected as an index of participants' sustained attentiveness to sounds and as an index of executive control (specifically, inhibition of a prepotent response). Two fixed pseudorandom orders of sounds were presented, with the condition that target sounds never occurred consecutively. Participants fixated a central cross against a white background and heard 18 presentations of 9 short environmental sounds (162 trials total; sounds ranging in duration from 545-678 ms). Participants were required to respond as quickly as possible to all sounds except the target (a bird call). Upon hearing the target, participants inhibited the press response, and waited for the next sound (presented 2100 ms after the onset of the bird).

## 4.3 Results

### 4.3.1 Experiment 4a

Data from one participant could not be used due to experimenter error. Data from a further participant were not analysed due to high error rates (errors of omission on > 10% non-target trials) on a measure of sustained auditory attention (SAART) collected during the same session. Data from 11 participants are presented. Where the sphericity assumption was violated in ANOVAs (run in SPSS), Greenhouse-Geisser corrected degrees of freedom are reported.

#### 4.3.1.1 2AFC results

Figure 4.4 displays mean proportion correct for each CS. Mean proportion correct data were submitted to one-sample Wilcoxon signed rank (WSR) tests, testing distributions against a chance level of 50% (all tests two-tailed). Testing 2AFC distributions relative to chance revealed that the proportion of correct responses for CS3 and CS4 were significantly greater than chance (CS3: WSR = 20,  $p = 0.02$ ; CS4: WSR = 10.5,  $p = 0.031$ ). The distribution of proportion correct for CS1 was marginally different from chance (WSR = 16.5,  $p = 0.066$ ). Unexpectedly, the distribution of proportion correct for CS2 did not differ significantly from chance (WSR = 14,  $p = 0.25$ ). However, inspection of the distribution (see Figure 4.4) suggested that the lack of difference from chance may have driven by an outlier (a single participant that showed floor 2AFC performance for CS2). Re-running the analysis for CS2 removing the participant with floor performance showed that performance was close to significance for the rest of the cohort (WSR = 18.5,  $p = 0.068$ ). A repeated measures ANOVA with CS as factor showed no significant differences

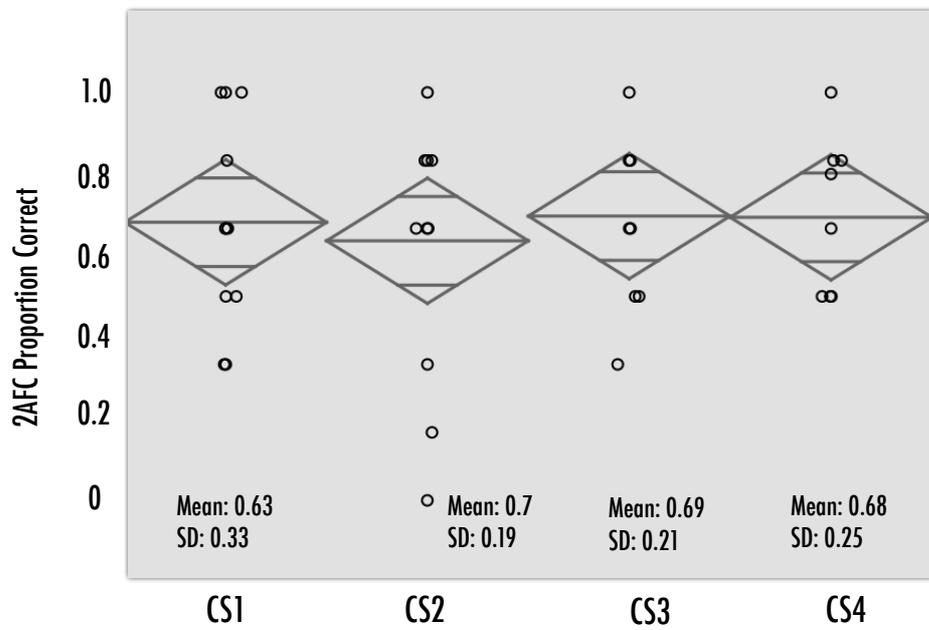


Figure 4.4: Proportion correct at 2AFC for each CS. Small diamonds show proportion correct for individual participants. Cross-subject means for proportion correct are shown as horizontal line within large diamond (upper and lower tips of large diamond show upper and lower bounds of 95% confidence interval; means and SDs are displayed below).

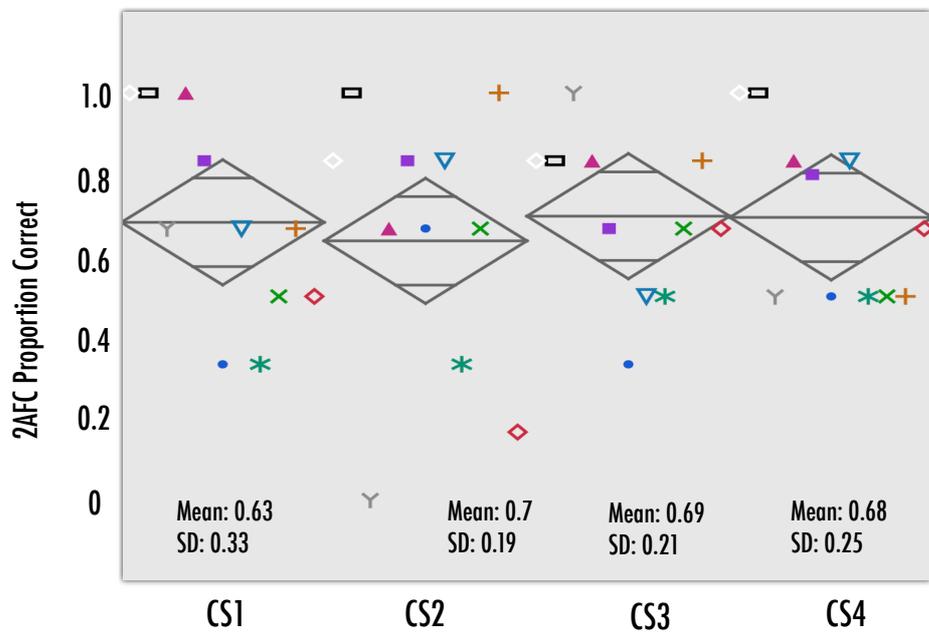


Figure 4.5: Data from Figure 4.4 replotted to highlight individual differences in 2AFC performance over each CS. Each participant is highlighted with a unique shape and colour (all other attributes as per Figure 4.4).

in learning performance across any of the sounds for the 11 participants [ $F(3, 30) = 0.23, p = 0.88$ ].

While statistical analysis suggested effects of learning indexed at 2AFC were less robust for CS1 and CS2, it is notable that for each CS, a subset of participants achieved ceiling or close to ceiling performance levels (see Figure 4.4 and 4.5). The profile of results suggests individual differences in the extent to which participants could identify each CS, and in turn map the CS to a particular alien exemplar.

To determine whether performance at 2AFC was biased by systematic confusion over multiple sound pairs, Kappa ( $\kappa$ ) was calculated over subjects for

**Table 4.4:** Rates of hits/misses and false alarms/correct rejections, with Kappa ( $\kappa$ ) data for each CS pair presented at 2AFC

	Response: CS1	Response: CS2		Response: CS1	Response: CS3
Presented: CS1	16 (0.9412)	1 (0.0588)	Presented: CS1	18 (0.8182)	4 (0.1818)
Presented: CS2	3 (0.1364)	19 (0.8636)	Presented: CS3	1 (0.0556)	17 (0.9444)
$\kappa = 0.794$			$\kappa = 0.751$		
	Response: CS1	Response: CS4		Response: CS2	Response: CS3
Presented: CS1	18 (0.9)	2 (0.1)	Presented: CS2	12 (0.75)	4 (0.25)
Presented: CS4	2 (0.1333)	13 (0.8667)	Presented: CS3	5 (0.2632)	14 (0.7368)
$\kappa = 0.767$			$\kappa = 0.485$		
	Response: CS2	Response: CS4		Response: CS3	Response: CS4
Presented: CS2	15 (0.8824)	2 (0.1176)	Presented: CS3	17 (0.7727)	5 (0.2273)
Presented: CS4	1 (0.0476)	20 (0.9524)	Presented: CS4	3 (0.1667)	15 (0.8333)
$\kappa = 0.839$			$\kappa = 0.6$		

each CS pair. As described by Cohen (1968),  $\kappa$  was calculated for each pair as the ratio of correct classifications occurring above chance relative to misclassifications occurring due to chance (see footnote 2). As displayed in table 4.4,  $\kappa$  and frequency data indicated relatively low confusion across most CS pairs. Notably, CS2 and CS3 showed lower  $\kappa$  compared to other CS pairs.

#### 4.3.1.2 In-task results

**4.3.1.2.1. Task accuracy.** Mean proportion correct responses to alien characters (i.e., alien ‘catch’) were high ( $> 0.93$ ) across each CS, and over all blocks. A 4 (CS)  $\times$  8 (block) ANOVA revealed no significant differences in accuracies due to CS or block; there was no significant CS  $\times$  block interaction [CS:  $F(2.1, 20.97) = 1.75, p = 0.2$ ; Block:  $F(7, 70) = 0.51, p = 0.82$ ; CS  $\times$  Block:  $F(3.09, 30.86) = 1.3, p = 0.29$ ].

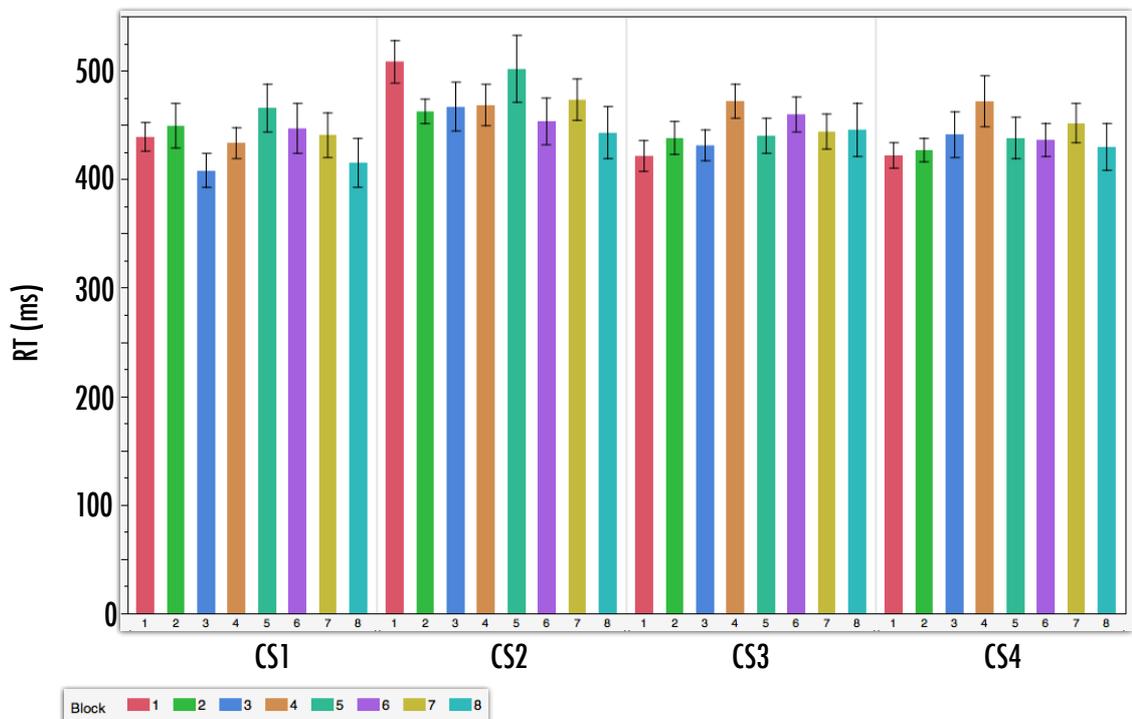


Figure 4.6: Mean RTs (ms) ( $\pm 1$  std. error) for each CS across blocks

**4.3.1.2.2. Task RTs.** Reaction times to each alien character were submitted to a 4 (CS) x 8 (block) ANOVA. A very marginal trend toward a main effect of CS was found,  $F(3, 30) = 2.3$ ,  $p = 0.1$ ,  $\eta_p^2 = 0.187$ ; there was no significant main effect of block [ $F(2.63, 26.3) = 0.92$ ,  $p = 0.43$ ], and no CS x block interaction [ $F(5.7, 57.04) = 1.6$ ,  $p = 0.17$ ] (see Figure 4.6).

### 4.3.2 Experiment 4b

Data from one participant were not analysed; at debrief, the participant reported difficulty in executing responses with the gamepad and inadvertent pressing of buttons that did not match either of the on-screen aliens during the 2AFC (resulting in near floor performance for all sounds at 2AFC). Data from 19 participants are presented. Where the sphericity assumption was violated in ANOVAs (run in SPSS), Greenhouse-Geisser corrected degrees of freedom are reported.

#### 4.3.2.1 2AFC results

Figure 4.7 displays 2AFC proportion correct data for each PS. Testing 2AFC distributions relative to chance revealed that proportions of correct responses for each PS were significantly different from chance (PS1: WSR = 41.5,  $p = 0.016$ ; PS2: WSR = 57,  $p = 0.002$ ; PS3: WSR = 38.5,  $p = 0.014$ ; PS4: WSR = 55.5,  $p = 0.012$ ) (all tests two-tailed). A repeated measures ANOVA with PS as a factor showed no significant differences in 2AFC accuracy over PS [ $F(3, 54) < 0.01$ ,  $p > 0.99$ ].

While statistical analyses suggested learning was significantly different from chance for each PS, it is notable that for each sound, a subset of participants achieved performance levels at or below chance (see Figure 4.7).

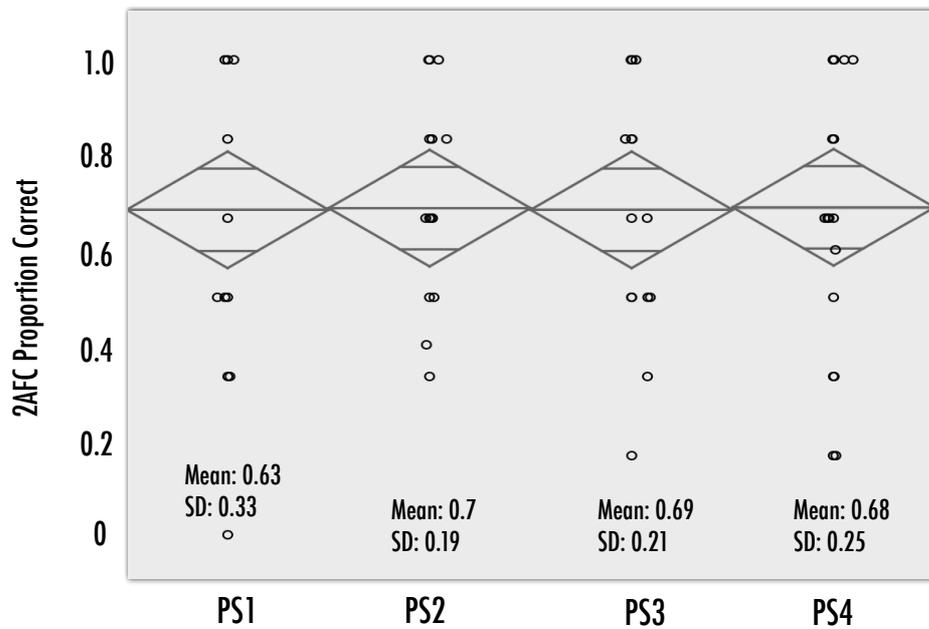


Figure 4.7: Proportion correct at 2AFC for each PS. Small circles show proportion correct for individual participants. Figure attributes as per Figure 4.4.

Table 4.5: Rates of hits/misses and false alarms/correct rejections, with Kappa ( $\kappa$ ) data for each PS pair presented at 2AFC

	Response: PS1	Response: PS2		Response: PS1	Response: PS3
Presented: PS1	26 (0.7027)	11 (0.2973)	Presented: PS1	22 (0.6111)	14 (0.3889)
Presented: PS2	10 (0.2973)	26 (0.7222)	Presented: PS3	12 (0.3243)	25 (0.6757)
$\kappa = 0.425$			$\kappa = 0.287$		
	Response: PS1	Response: PS4		Response: PS2	Response: PS3
Presented: PS1	28 (0.7778)	8 (0.2222)	Presented: PS2	24 (0.6857)	11 (0.3143)
Presented: PS4	14 (0.3784)	23 (0.6216)	Presented: PS3	9 (0.25)	27 (0.75)
$\kappa = 0.399$			$\kappa = 0.436$		
	Response: PS2	Response: PS4		Response: PS3	Response: PS4
Presented: PS2	28 (0.7368)	10 (0.2632)	Presented: PS3	23 (0.6216)	14 (0.3784)
Presented: PS4	10 (0.2703)	27 (0.7297)	Presented: PS4	10 (0.2632)	28 (0.7368)
$\kappa = 0.467$			$\kappa = 0.359$		

Thus, as with experiment 4a, data suggested profiles of individual differences with respect to learning each PS and mapping it to its respective alien character.

To determine whether performance at 2AFC was biased by systematic confusion over multiple sound pairs, Kappa ( $\kappa$ ) was calculated over subjects for each PS pair (see 4.3.1.1). As displayed in Table 4.5,  $\kappa$  and frequency data showed relative consistency across PS pairs; note however that  $\kappa$  for PS1 and PS3 was reduced relative to other PS pairs. Somewhat surprisingly, the Kappa data for each PS pair show lower values than were observed for the CS pairs (compare tables 4.4 and 4.5). This suggests that when participants attempted to map PS stimuli to their respective aliens/locations at 2AFC, relatively greater confusion occurred than for the cohort who mapped CS stimuli in a similar fashion. This might indicate that the particular PS exemplars were more difficult to distinguish from each other than the CS exemplars, despite the PS exemplars showing relatively distinct frequency sweep parameters (and showing robust patterns of learning at 2AFC). Critically however, the Kappa data in both instances enabled us to isolate CS and PS stimuli that showed the lowest extents of confusion; these CS and PS stimuli were then selected for use in experiment 5 (see 5.2.3).

#### **4.3.2.2 In-task results**

**4.4.2.2.1. Task accuracy.** Mean proportions of correct responses to alien characters (i.e., alien 'catch') were high ( $> 0.94$ ) across each PS, and over all blocks. A 4 (PS) x 8 (block) ANOVA revealed no significant differences in accuracies across PS or block; the PS x block interaction was not significant

[PS:  $F(3, 54) = 0.36, p = 0.78$ ; Block:  $F(3.77, 67.85) = 1.04, p = 0.39$ ; PS x Block:  $F(5.22, 93.96) = 1.89, p = 0.1$ ].

**4.4.2.2.2. Task RTs.** Analysis of reaction times to each alien character were submitted to a 4 (PS) x 8 (block) ANOVA. There was no significant main effect of PS, a very marginal main effect of block, and no significant interaction between the two factors [PS:  $F(3, 54) = 0.27, p = 0.85$ ; Block:  $F(3.59, 64.55) = 2.13, p = 0.094$ ; PS x Block:  $F(6.13, 110.34) = 1.16, p = 0.33$ ]. Mean RTs across PS and Block are displayed in Figure 4.8.

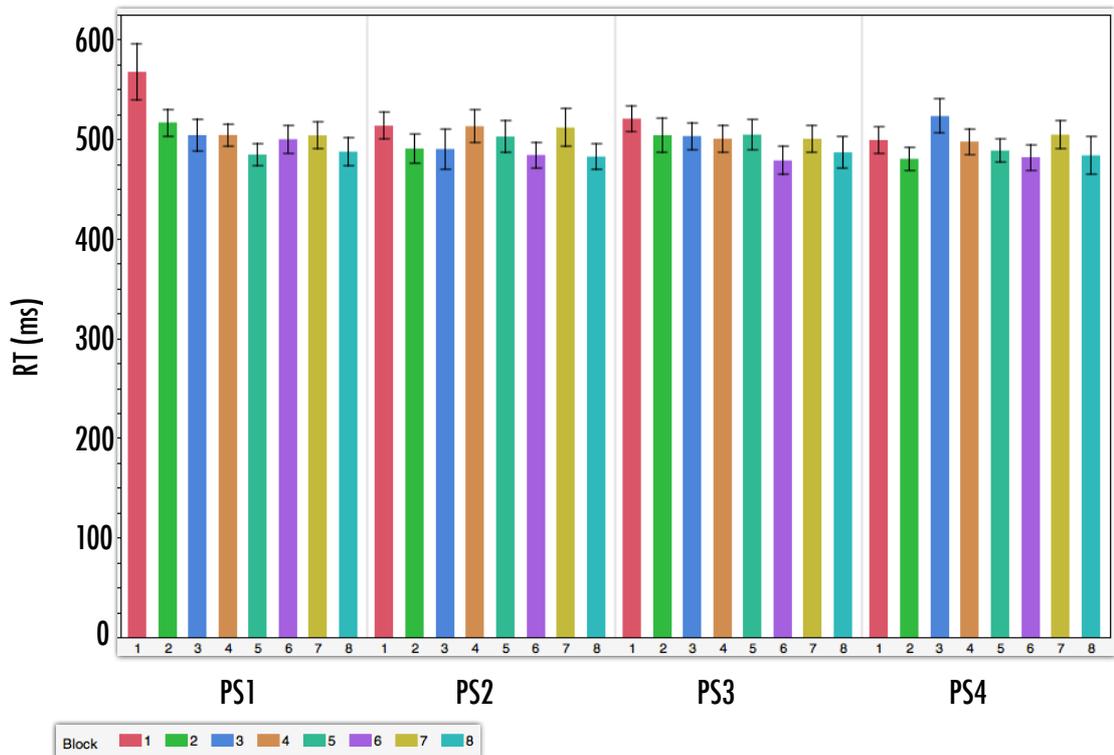


Figure 4.8: Mean RTs (ms) ( $\pm 1$  std. error) for each PS across blocks

#### 4.4 Chapter summary and discussion

In experiments 4a and 4b, we explored learning of fixed exemplars of particular auditory stimulus classes (broadband sound contexts and punctate sound objects), within an on-going visuo-spatial task. As expected, participants could learn to associate contextual and punctate auditory stimuli with visual targets at specific locations. These results are unsurprising based on extensive evidence of associative learning between auditory (and visual) stimuli and other environmental events (e.g., Molchan et al., 1994; Polley et al., 2004; Gallagher et al., 1999; see also Fanselow & Poulos, 2005). Nevertheless, it is important to highlight the differences in extent of learning success. In both experiments, while we found participants that achieved near ceiling levels of accuracy at 2AFC, we also observed participants that performed at close to floor levels. Indeed, this profile of individual differences in learning is a theme we will return to in the next chapter. Additionally, differences in broader abilities to sustain attention toward sounds did not account for differences in 2AFC learning outcomes (for the sake of brevity, data are not presented here).

The results presented in this chapter provide the foundation for developing more complex learning paradigms. In chapter 5, we explore learning of combinations of these cue stimuli (experiments 5a and 5b). Experiment 5a extended the present paradigm, introducing auditory contexts and objects on each trial; critically, learning a combination of both sound types provided the optimally informative cue to spatial events. This allowed us to probe questions regarding auditory learning abilities in adults; in particular, whether learning of complex sound combinations may occur following short periods of experience within an on-going active visual task. We further explored whether violations of cue contingencies would yield costs to task performance that would depend

upon success in learning to combine the auditory cue types. In chapter 6 (experiment 6), we extend this paradigm to a cohort of school-aged (i.e., 8-9 year old) children. This allowed us to address whether adults and children would differ in their abilities to combine, learn and use the cues in-task (given that children begin to develop the ability to combine cues within and across modalities at this age). Moreover, we also explored whether violation of both auditory cue types would yield differences in performance decrements for adults versus children. Experiment 5b extended experiment 5a, by introducing separate violations of contingencies for each cue type during the task; in this way, we could explore whether contingency violation for relatively more salient auditory objects would yield greater performance decrements than contingency violation for less salient auditory contexts, affording a test of mechanisms of attentional capture associated with cue saliency (further to Kayser et al., 2005).

## Chapter 4 Footnotes

1. The respective 900 ms and 300 ms white noise segments used to produce the CS stimuli for experiment 4a and PS stimuli for experiment 4b were excised from the beginning of the exact same 2100 ms white noise segment used to produce the stimuli for experiments 5a, 5b and 6. Thus, the 900 ms CSs were exactly the first 900 ms of the longer 2100 ms CSs.
2. Descriptive statistics for reliability of classification of sounds at post-game test are described throughout the results sections in chapters 4 and 5. Following the descriptive framework common to machine learning algorithms (as described in Kohavi & Provost, 1998), the following terms are used to describe classification accuracy and misclassification.

From Kohavi & Provost (1998), for a 2 x 2 confusion matrix of the form:

	Negative	Positive
Negative	a	b
Positive	c	d

Total accuracy:  $(a + d)/(a + b + c + d)$

True positive rate:  $d/(c + d)$

False positive rate:  $b/(a + b)$

Precision:  $d/(b + d)$

False negative rate:  $c/(c + d)$

True negative rate:  $a/(a + b)$

In addition, we utilise Kappa statistics as a metric of the extent of correct classification (i.e., agreement) that occurs over and above correct classification

arising as a result of chance (i.e., the baseline constraint) (Cohen, 1968; Landis & Koch, 1977). Kappa ( $\kappa$ ) is expressed as follows:

$$\kappa = \frac{\pi_0 - \pi_e}{1 - \pi_e}$$

where  $\pi_0$  is the total accuracy proportion  $[(a + d)/(a + b + c + d)]$  and  $\pi_e$  is the total proportion of accurate classifications arising due to chance  $[(((a + c)*(a + b)) + ((b + d)*(c + d)))/((a + b + c + d)^2)]$ .

Chapter 5: Short-term Auditory Learning  
within a Multi-modal Environment: Cue  
Combination and Saliency

## 5.1 Introduction

In chapter 4, we explored learning of isolated examples of CS (experiment 4a) and PS (experiment 4b) stimuli, based on their usefulness as cues within a visuo-spatial task. In this chapter, we expand upon the paradigm, examining whether CS and PS stimuli can be learned as a combination (rather than either stimulus in isolation), such that the combination of both yields a more effective cue than either stimulus alone. In experiment 5a, we explored cue combination learning in adults and examined whether violation of cue contingencies during the task would lead to performance costs. Experiment 5b further investigated these questions with adults by violating cue contingencies independently for CS and PS stimuli; this allowed us to test whether the more salient PS stimuli might be associated with greater cue cost than the CS stimuli.

## 5.2 Method

### 5.2.1 Participants

**5.2.1.1. Experiment 5a.** Participants were 24 right-handed adults (15 female, 9 male; mean age  $\pm$  SD: 26.5  $\pm$  6.1; range: 19–43) with normal or corrected to normal vision, and no history of hearing difficulties, hearing loss, or neurological insult or trauma. Participants were recruited from the Birkbeck College Department of Psychological Sciences participant pool, or from the working population.

**5.2.1.2. Experiment 5b.** Participants were 24 right-handed adults (17 female, 7 male; mean age  $\pm$  SD: 26.2  $\pm$  8.8; range: 19–54) with normal or corrected to normal vision, with no history of hearing difficulties, hearing loss, or neurological insult or trauma. Participants were recruited from the Birkbeck

College Department of Psychological Sciences participant pool, or from the working population.

### **5.2.2 Materials**

All stimuli were created using Adobe Audition CS5.5 running on MacBook Pro (OS 10.7). For the learning tasks, stimuli were presented and responses collected using a custom tablet application running on iPad 4 (iOS 7). A measure of sustained attention (SAART) was presented using Psychophysics Toolbox (version 3; Kleiner, Brainard & Pelli, 2007) running in Matlab (2010a; 32-bit) on MacBook Pro laptop computers (OS 10.6 and 10.7). All sounds were presented through Sennheiser HD-380 pro headphones for all devices and experiments.

### **5.2.3 Stimuli**

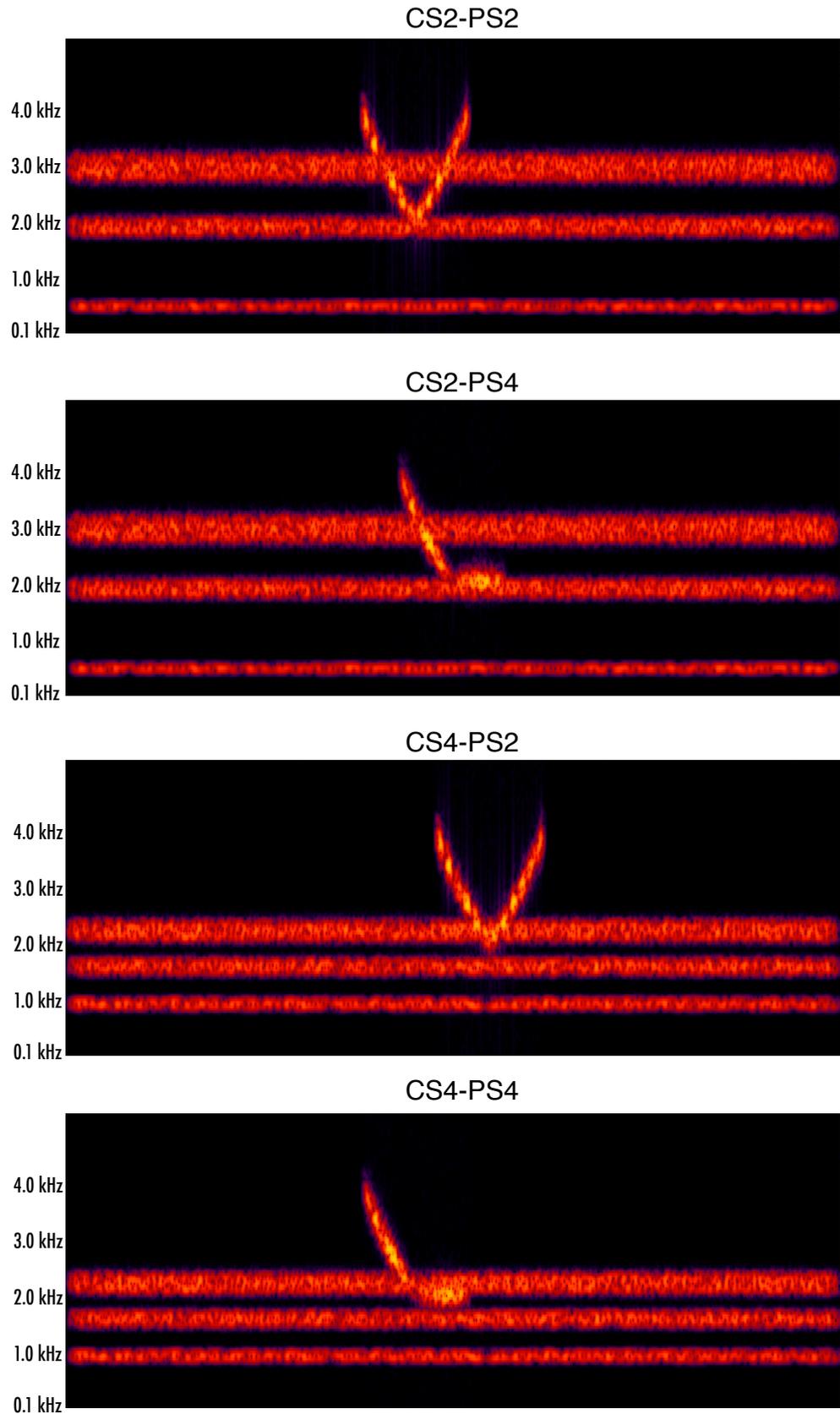
Stimuli were complex, novel auditory ‘scenes’, that comprised a combination of one CS and one PS, as used in experiments 4a and 4b, respectively (see Figure 5.1). From the results of experiment 4a and 4b, two examples of each of the CS and PS stimuli were selected, based on the response agreement data (i.e., Kappa [ $\kappa$ ]) calculated from participants’ 2AFC responses. From experiment 4a, CS2 and CS4 were selected, since response frequency and  $\kappa$  data indicated relatively low confusion of those sounds when paired (relative to other CS pairs; see table 4.4. and 4.3.1.1). Similarly, from experiment 4b, PS2 and PS4 were selected, as response frequency and  $\kappa$  data suggested low confusion of those sounds when paired (versus other PS pairs; see table 4.5 and 4.3.2.1).

CS and PS stimuli were combined using Audition CS5.5, to produce complex auditory ‘scenes’. CS and PS stimuli had respective durations of 2100 ms and 300 ms (see footnote 1, chapter 4). Each CS was combined with each PS, yielding four combinations of CS and PS cues (CS2-PS2; CS2-PS4; CS4-PS2; CS4-PS4; see Figure 5.1). Temporal onset of each PS within each CS was jittered according to three levels, to reduce the likelihood of participants learning a single fixed temporal point at which each PS would occur (see Figure 5.1). Thus, onset of the PS occurred at 800, 900 or 1000 ms post-onset of the CS. For each CS and PS combination, three examples were produced, reflecting the PS onset times just described. To ensure sufficient salience of the PS stimuli relative to the CS (and reduce effects of energetic masking as much as possible), both PS stimuli were scaled to a total RMS of 6 dB greater than the CS stimuli. CS and PS stimuli were then inserted into multi-track sessions in Audition CS5.5, and for each CS-PS combination (and for each extent of PS temporal jitter), were mixed out as a single mono track audio file (44.1 kHz sampling rate, 16 bit quantisation, .wav format).

#### **5.2.4 Procedure**

All adult participants provided signed informed consent prior to beginning the experiment.

**5.2.4.1 Experiment 5a.** Blocks 1-7 presented the cue contingencies as shown in the upper four panels of Figure 5.2. The locations cued by each CS and PS were orthogonal; thus, each CS cued locations at the left or right half of the screen, and each PS cued locations at the top or bottom half. During blocks 8 and 9, the sound-alien location contingencies were violated, by switching the half of the screen that each CS and PS sound cued. Thus, each CS and PS



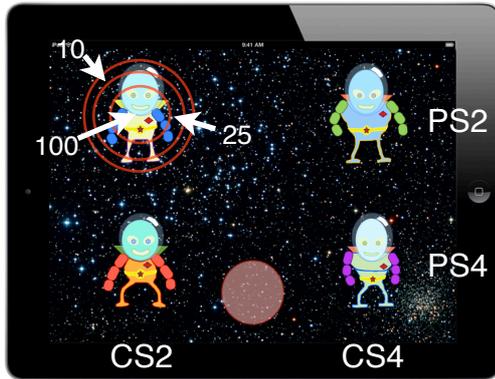
**Figure 5.1:** Spectrograms of CS and PS stimulus combinations used in experiments 5a & 5b. Note the illustration of PS temporal jitter; CS2-PS2: 800 ms PS onset; CS2-PS4: 900 ms PS onset; CS4-PS2: 1000 ms PS onset; CS4-PS4: 800 ms PS onset.

now cued locations on the opposite half of the screen relative to the first 7 blocks (e.g., if CS2 cued locations at the left half of the screen in blocks 1-7, it cued locations at the right half of the screen in blocks 8-9; the same applied to the PS stimuli, for the top and bottom halves of the screen; see Figure 5.2). The effect of this switch meant each CS-PS combination in blocks 8-9 cued the location diagonally opposite the location that it had cued during blocks 1-7. In blocks 10-11, the contingencies as presented in blocks 1-7 were re-established. As the game progressed, the time during which each alien remained on-screen decreased (blocks 1-3: 900 ms; blocks 4-5: 800 ms; blocks 6-11: 700 ms), thereby providing a gradual but small increase in task difficulty. If a response was not made within the allotted on-screen time for the alien in a particular block, the alien disappeared from the screen and the response was recorded as null. The harmonic mean of RTs for each condition was calculated for all subjects [i.e.,  $n / (1/a) + (1/b) + (1/c) \dots$ , where a, b, c are RTs to correct trials and  $n$  the number of correct observations for the condition]. [Ratcliff (1993) found that analysis power for simulated RT (ex-Gaussian) distributions was robust to outliers where an inverse transformation was applied (relative to trimmed means or SD cutoffs)].

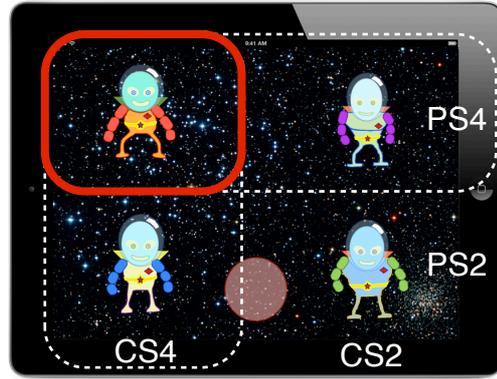
Adult participants sat in a softly lit, acoustically dampened booth and completed the touchpad game. Recall that in experiments 4a and 4b, participants were naïve to the fact that a test phase would follow the task (and in some cases showed relatively poor learning). To rule out the possibility that participants might not attend to the sounds during the task (focusing instead on the visual game), we informed participants before testing that listening to the

Contingencies (Blocks 1-7 & 10-11)

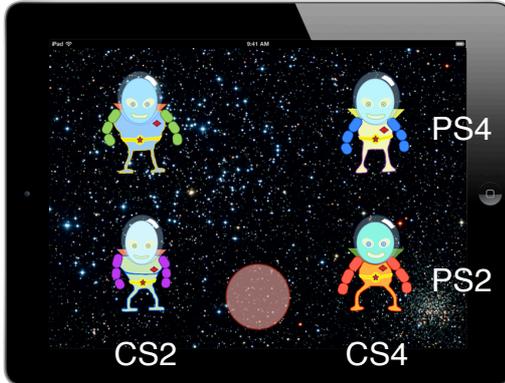
Order 1



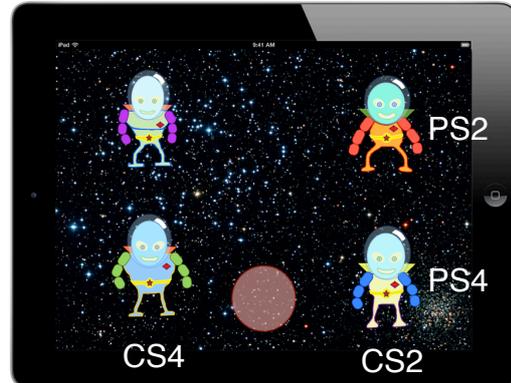
Order 2



Order 3

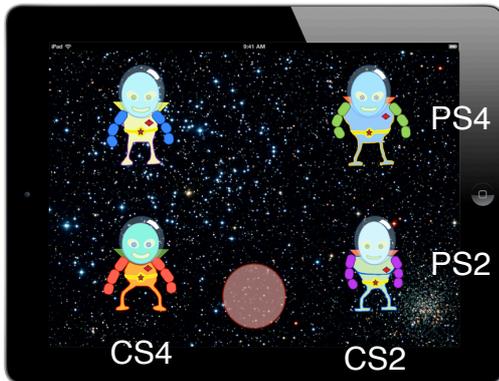


Order 4

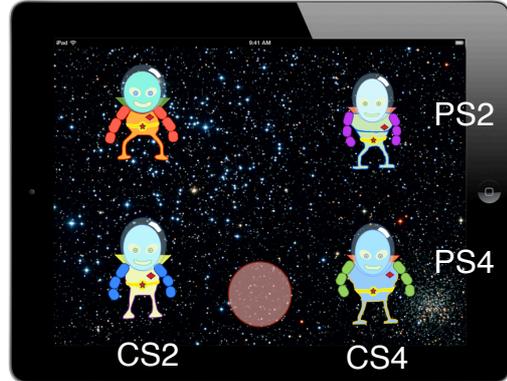


Contingency violations (Blocks 8-9)

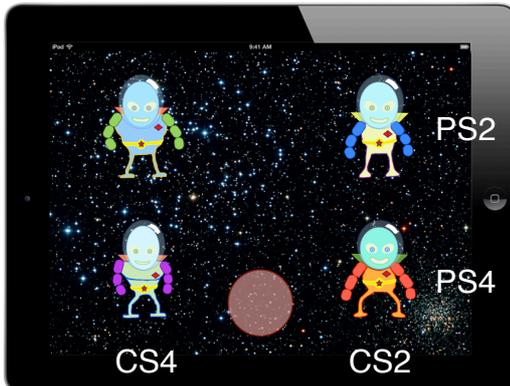
Order 1



Order 2



Order 3



Order 4

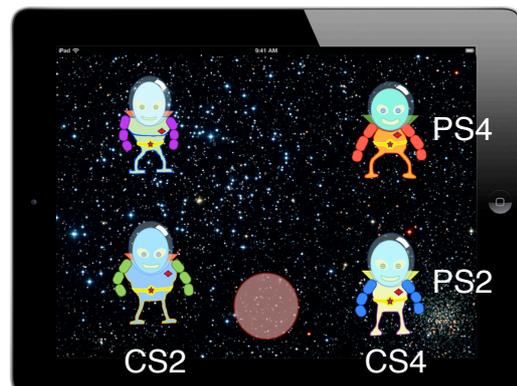


Figure 5.2: Design for learning task within experiment 5a. Each alien's location was cued by a combination of CS and PS stimuli. Across each counterbalanced order, CS stimuli always cued the left and right of the screen, and PS stimuli the top and bottom. Thus, combining both cue types provided a reliable cue to the alien's location. The upper right panel provides an example: CS4 cues the left side of the screen, whilst PS4 cues the top of the screen; thus, each sound alone cues two possible locations for the upcoming alien (dotted lines). Learning the combination of both (CS4-PS4) provides a reliable cue to a single alien at the upper left (red box; lines shown here are by way of example and were not presented in the task). In the lower half of the figure, violations (presented in blocks 8 and 9) of the contingencies set up over the first 7 blocks are depicted. The top left panel displays the game points scoring convention (note that the red circles were not visible during the task).

sounds could offer a useful cue as to which alien would appear and where it would appear, helping them to win more points. Participants began each trial with their finger resting on the red 'home' spot (see Figure 5.2), which was always displayed on-screen. Participants were instructed to tap each alien that appeared on-screen as quickly as possible as soon as it appeared. After tapping, participants then returned their finger to the 'home' spot before the next trial began. Participants completed 132 in-game trials (11 blocks of 12 trials), with the first block treated as practice. Each trial began with 1000 ms of silence, followed by onset of one of the CS-PS cue combinations. The alien character cued by the CS-PS combination always appeared on-screen 300 ms after the offset of the PS; this was the case in each of the PS temporal jitter conditions (thus, the alien appeared at 1100, 1200 or 1300 ms following the onset of the CS). Hence, the alien was presented with relatively close temporal proximity to the PS and whilst the CS was still being presented.

A 36 trial 4AFC task followed the 132 game trials. On each trial, all aliens were presented on-screen at the same time. After 500 ms of silence, one of the CS-PS combinations heard during the learned contingency phase of the game (Blocks 1-7 & 10-11) was played. Participants were instructed to wait until the sound had finished playing, and to then tap the alien that they believed matched the sounds they had just heard. Participants had 3.4 seconds from offset of the sounds to respond, after which the trial timed out.

Following the game and 4AFC, participants completed the same SAART task as the participants in experiments 4a and 4b. Details and procedure were identical to those experiments (see 4.2.4.3).

**5.2.4.2 Experiment 5b.** Experiment 5b followed a very similar structure and procedure to experiment 5a. The task differed primarily according to the nature of cue violations (see Figure 5.3). In experiment 5a, both CS and PS stimuli were violated simultaneously. In order to establish whether the relatively more salient PS cues modulated RT cost effects related to cue violation to a greater extent than the CS cues, we violated these cue types independently. Thus, CS (but not PS) violation occurred at block 5 and PS (but not CS) violation at block 9, across all experimental orders.

In designing the experiment, we opted to violate the PS stimuli at the latter point in the task since we expected that listeners might be less attentive to the less salient CS stimuli at latter blocks, particularly due to fatigue or inattention. Moreover, as discussed in this chapter, since we observed relatively rapid learning of cue combinations during the first three blocks of the task, we expected that optimal learners would still detect the CS violation at block 5. We were also concerned that counterbalancing the position at which the CS (but not PS) and PS (but not CS) violations occurred within a single experiment

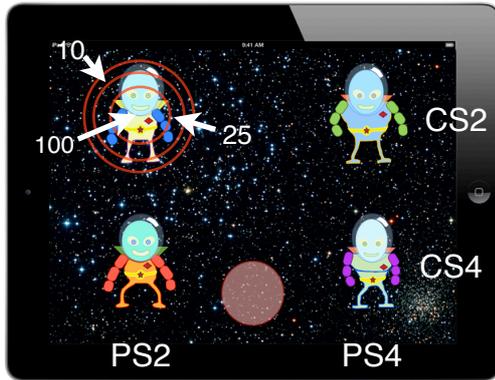
might lead to biases in performance that confounded order effects with the acoustic attributes of the sounds. Nevertheless, it is possible that violating the PS stimuli in the latter stages of experiment 5b (block 9) meant participants deemed the PS cues to be more reliable, since they remained constant over the initial 8 blocks. As this was a confound in the present design, we ran a control experiment with a small sample ( $N = 12$ ) in which the reverse order of contingency violations was presented; thus, PS (but not CS) violations occurred at block 5, while CS (but not PS) violations occurred at block 9. Results of this control experiment are presented in supplemental analyses 5.2.

In experiment 5b, cue combination contingencies were established from blocks 1-4 (block 1 was treated as practice); CS (but not PS) violation occurred at block 5; contingencies were re-established from blocks 6-8; PS (but not CS) violation occurred at block 9; contingencies were re-established at blocks 10 and 11. Finally, to rule out possible spatial effects related to the locations cued by each stimulus type, the locations cued by CS and PS stimuli were switched relative to experiments 5a and 6 (in experiments 5a and 6, each CS cued the left or right side of the screen, and each PS the top or bottom). In experiment 5b, CS stimuli cued the top or bottom of the screen, and PS stimuli the left or right side.

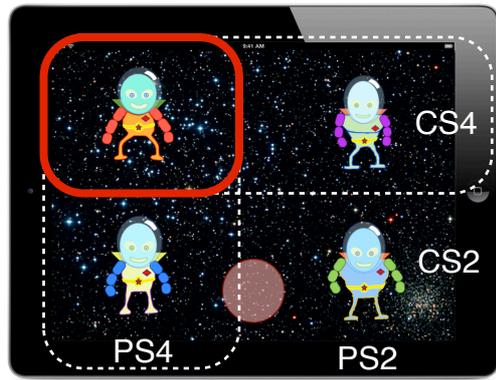
As in experiments 4a, 4b and 5a, all participants completed the same version of the SAART response inhibition task. Details and procedure were identical to the previous experiments.

Contingencies (blocks 1-4, 6-8 & 10-11)

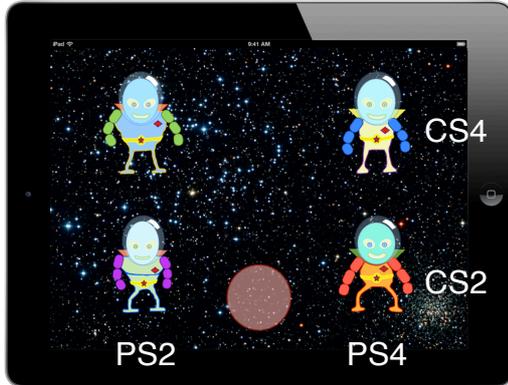
Order 1



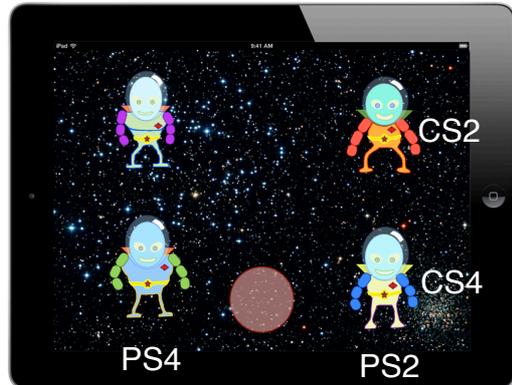
Order 2



Order 3

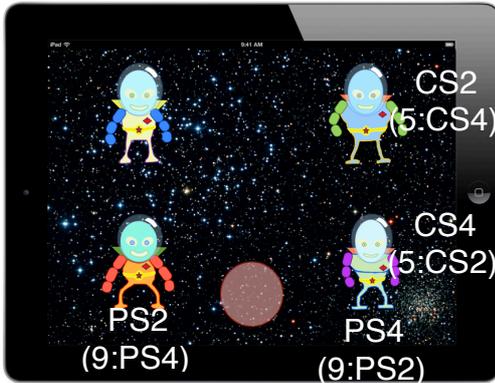


Order 4

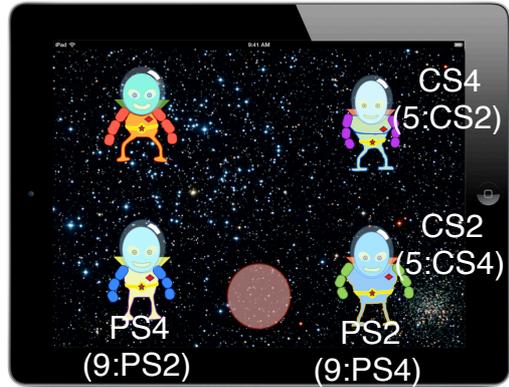


Contingency violations (CS: block 5; PS: block 9)

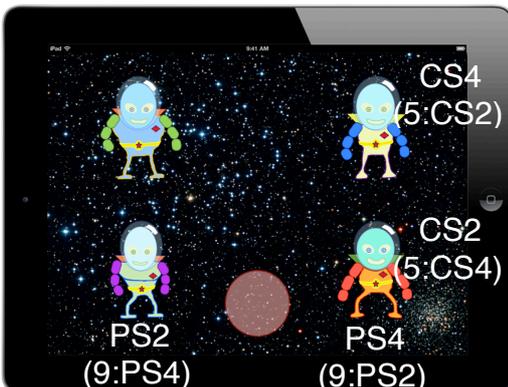
Order 1



Order 2



Order 3



Order 4

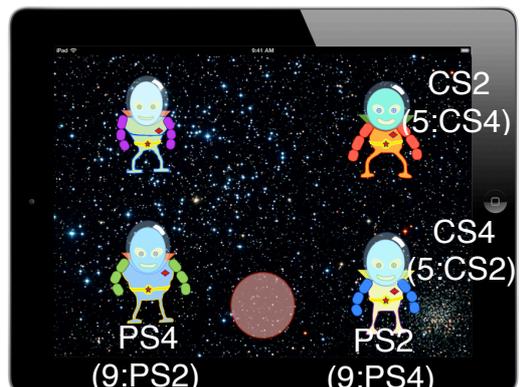


Figure 5.3: Design for learning task within experiment 5b. Task and cue structure differed from experiment 5a (see Figure 5.2). Violations of CS and PS stimuli occurred independently at separate blocks; CS (but not PS) violation at block 5, and PS (but not CS) violation at block 9 (violation block cue stimuli and the relevant block are indicated in parentheses in lower half of figure). In addition, locations cued by CS and PS stimuli were orthogonal to those of experiments 5a and 6; CS stimuli now cued the top or bottom of the screen, and PS stimuli the left or right side.

### 5.3 Results - Experiment 5a

Data from 24 subjects were analysed (except where noted; see below). In this experiment and those that follow, where the sphericity assumption was violated in ANOVA analyses (SPSS), Greenhouse-Geisser corrected degrees of freedom are reported.

#### 5.3.1 4AFC results

To determine whether the adult cohort had learned the mapping between each sound pair and the corresponding alien it cued, we first analysed proportion correct data for the 4AFC task across all subjects. One-sample Wilcoxon signed rank (WSR) tests of the proportion correct distribution for each CS-PS sound pair revealed that none of the distributions were significantly different from chance level (chance: 0.25) (see Figure 5.4a; CS2-PS2: WSR = 21.5,  $p = 0.52$ ; CS2-PS4: WSR = -20,  $p = 0.55$ ; CS4-PS2: WSR = -16,  $p = 0.61$ ; CS4-PS4: WSR = 41,  $p = 0.22$ ). Further, a repeated-measures ANOVA with sound pair as a within-subjects factor showed that proportion correct performance did not differ significantly across any of the four CS-PS pairs [ $F(3, 66) = 1.9, p = 0.14$ ].

Thus when considered as a whole, the adult cohort's performance in matching the sound pair cues with the respective aliens did not differ significantly from chance levels. Nevertheless, a major aim of the experiment was to evaluate differences in learning performance; in particular, we sought to account for patterns of in-task performance that reflected the learning outcomes as indexed at 4AFC. We therefore calculated a mean of 4AFC proportion correct performance for each subject collapsed across sounds. We then used the median of this distribution as a cut-off and performed a median split of the

cohort (12 participants above and 12 below). This was approach was used in Experiments 5b and 6 also (see footnote 1).

A mixed ANOVA with sound pair as the within-subjects factor and 4AFC median split as the between-subjects factor showed the expected highly significant main effect of 4AFC median split [ $F(1, 22) = 32.74, p < 0.0001, \eta_p^2 = 0.598$ ], with no significant main effect of sound pair or any significant interaction [both  $F(3, 66) < 2.0, p > 0.13$ ]. Indeed, one-sample Wilcoxon signed rank tests (two-tailed) showed that for those adults above the median split, performance was significantly different from chance for two sound pairs (CS2-PS2: WSR: = 35.5,  $p = 0.003$ ; CS4-PS4: WSR: = 32,  $p = 0.002$ ), and marginally different from chance for one further sound pair (CS4-PS2: WSR: = 17.5,  $p = 0.08$ ); performance was not significantly different from chance for one of the sound pairs (CS2-PS4: WSR: = 20,  $p = 0.12$ ). In contrast, those adults below the cohort median split performed significantly below chance levels across all sound pairs [CS2-PS2: WSR: = -33,  $p = 0.001$ ; CS2-PS4: WSR: = -33,  $p = 0.001$ ; CS4-PS2: WSR: = -34,  $p = 0.004$ ; CS4-PS4: WSR: = -27,  $p = 0.03$  (all one-sample WSR, two-tailed) (see Figure 5.4b)]. The median split of the cohort based on 4AFC performance was used as a between-subjects variable in subsequent analyses of data generated during the game stage.

Our 4AFC data also let us explore patterns of responses across sound pairs, offering further insight into stimulus dimensions that participants had learned or indeed, confused. Table 5.1 displays confusion matrices (count frequencies) for those above and below the cohort 4AFC median split.

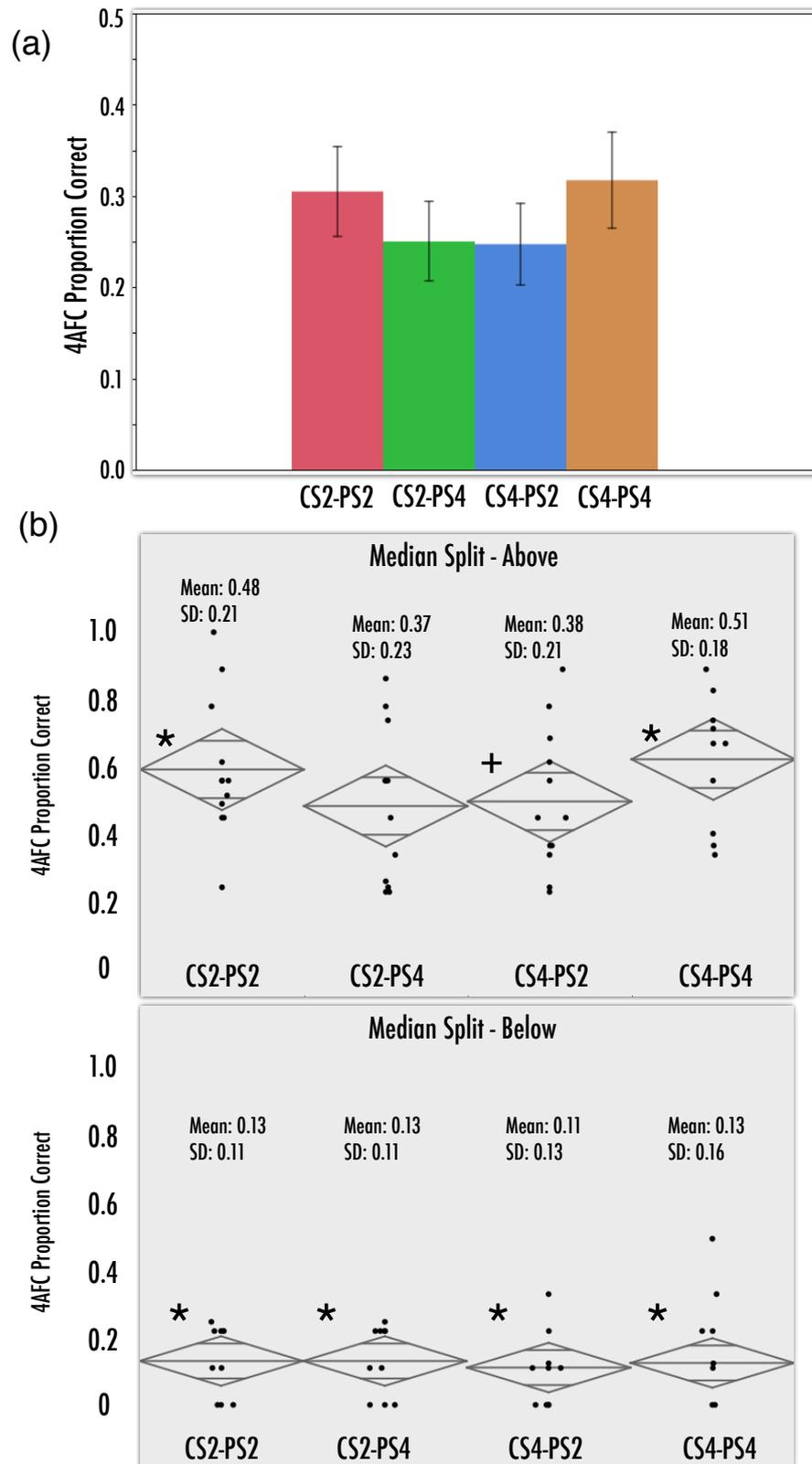


Figure 5.4: 4AFC proportion correct data for experiment 5a. (a) Mean ( $\pm 1$  std. err) proportion correct for each CS-PS pair across full cohort ( $n = 24$ ). (b) 4AFC proportion correct by median split of cohort (upper: above median; lower: below median); horizontal line within each diamond shows mean for that CS-PS, and upper and lower diamond tips the upper and lower bounds of 95% confidence intervals, respectively. WSR (two-tailed) tests: \*  $p < 0.05$  +  $p = 0.08$

Table 5.1: Confusion matrices for adults above and below 4AFC median

		Sound presented				Row total
		CS2-PS2	CS2-PS4	CS4-PS2	CS4-PS4	
Response	Above					
	CS2-PS2	<b>48</b>	17	17	17	99
	CS2-PS4	21	<b>38</b>	14	30	103
	CS4-PS2	21	12	<b>38</b>	27	98
	CS4-PS4	14	17	17	<b>49</b>	97
	Col. total	104	84	86	123	397

		Sound presented				Row total
		CS2-PS2	CS2-PS4	CS4-PS2	CS4-PS4	
Response	Below					
	CS2-PS2	<b>14</b>	57	17	19	107
	CS2-PS4	45	<b>14</b>	29	18	106
	CS4-PS2	13	33	<b>12</b>	49	107
	CS4-PS4	34	11	44	<b>13</b>	102
	Col. total	106	115	102	99	422

Since a major goal was to establish patterns of response agreement (and indeed, confusion) across CS-PS pairs, we report descriptive statistics for 2 x 2 tables comparing each possible cue-response pair (see Kohavi & Provost, 1998) (analysis of the full 4 x 4 confusion matrix for each group using Chi square models was precluded by the within-subject design; McNemar-Bowker analyses were also non-optimal, since comparison of symmetry above and below each diagonal was not of core interest to the present analyses).

Table 5.2 presents for all possible sound presented/response pairs, Kappa values and accuracy, true positive, false positive, true negative, false negative and precision proportions, for those above and below the cohort 4AFC median (see footnote 2).

**Table 5.2:** Kappa coefficients and proportion data for accuracy, true positive, false positive, true negative, false negative and precision data for adults above and below the cohort 4AFC median

Above						
Positive	CS2-PS2	CS2-PS2	CS2-PS2	CS2-PS4	CS2-PS4	CS4-PS2
Negative	vs. CS2-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS4
Kappa ( $\kappa$ )	0.384	0.384	0.516	0.490	0.300	0.327
Accuracy	0.694	0.694	0.758	0.745	0.649	0.664
True Pos.	0.738	0.738	0.738	0.731	0.559	0.585
False Pos.	0.262	0.262	0.262	0.269	0.441	0.415
True Neg.	0.644	0.644	0.778	0.760	0.742	0.742
False Neg.	0.356	0.356	0.222	0.240	0.258	0.258
Precision	0.696	0.696	0.774	0.760	0.691	0.691
Below						
Positive	CS2-PS2	CS2-PS2	CS2-PS2	CS2-PS4	CS2-PS4	CS4-PS2
Negative	vs. CS2-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS4
Kappa ( $\kappa$ )	-0.556	-0.067	-0.280	-0.407	-0.020	-0.574
Accuracy	0.215	0.464	0.338	0.295	0.482	0.212
True Pos.	0.197	0.452	0.424	0.326	0.438	0.197
False Pos.	0.803	0.548	0.576	0.674	0.563	0.803
True Neg.	0.237	0.480	0.277	0.267	0.542	0.228
False Neg.	0.763	0.520	0.723	0.733	0.458	0.772
Precision	0.237	0.519	0.292	0.298	0.560	0.214

As expected given the median split, Kappa, total accuracy and precision were markedly higher across all possible cue sound-response pairs for those above the median split versus those below (table 5.2; compare rows 3 & 11, rows 4 & 13 and rows 8 & 16). Inspection of table 5.2 indicates that for those above the 4AFC median, Kappa, accuracy and precision were highest, and error metrics (false positive and false negative) lowest when considering cue

sound-response pairs where both the CS and PS differed (table 5.2 columns 4 & 5); for those below the median, moderately negative Kappa was observed. As expected, this indicates that those adults who tended to perform better at the 4AFC task were less likely to confuse across both sound classes and more frequently confused across one of the two (i.e., CS or PS), whereas those below the median split tended to confuse across both sound types (i.e., CS and PS).

We hypothesised that adults might find the less salient CS stimuli relatively similar, and therefore might confuse across these sounds to a greater extent than the PS stimuli. We found limited support for this hypothesis. For instance, for those above the median split, we found a relative reduction of Kappa (0.3) and elevation of false positives (0.441) for CS2-PS4 vs. CS4-PS4 (suggesting a tendency to confuse across the CS, as expected); however, those above the median split also showed reduced Kappa (0.327) and elevated false positives (0.415) for CS4-PS2 vs. CS4-PS4 (thus indicating confusion across the PS) (see table 5.2, columns 6 & 7).

We also noted an unexpected profile of sound pair confusion for those below the median split. Inspection of table 5.2 shows that for instances of the same CS but different PS (i.e., CS2-PS2 vs. CS2-PS4; CS4-PS2 vs. CS4-PS4; see table 5.2 columns 2 & 7), Kappa was considerably more negative than for instances of the same PS but different CS (i.e., CS2-PS2 vs. CS4-PS2; CS2-PS4 vs. CS4-PS4; see table 5.2, columns 3 & 6). This suggests that those below the 4AFC median tended to confuse the more salient sound class when matching sound pairs to the respective aliens (perhaps accounting for their significantly below chance 4AFC proportion correct distributions).

### 5.3.2 In-task results

We used several dependent measures to index learning for each condition during the alien catching game: reaction times to alien onsets; accuracy of tap responses to alien onsets; means and standard deviations of Euclidean distances from the response location co-ordinates to the co-ordinates at the centre of each alien. Mixed ANOVAs were used to model the (within-subject) effects of sound pair cue (4 levels) and block (various levels) on game performance metrics. The 4AFC median split was used as a between-subjects factor, to explore whether participants above the 4AFC median split would show greater RT cost (together with reduced accuracy and/or increased spatial variability of responses), due to violation of the sound-pair alien contingencies.

**5.3.2.1. RTs to alien onsets.** RTs for valid responses only (i.e., responses occurring within the allowed on-screen time for the aliens at each block; see 5.2.4.1) were analysed. Note that block 1 was treated as practice and was not analysed; full omnibus analyses of blocks 2-11 are presented in supplemental analyses 5.1.

Central to our hypothesis was the effect of the 4AFC median split grouping in accounting for the cost to RTs when violations of the established cue structure were introduced within the task. If those participants above the 4AFC median had learned to map the cue combinations to the specific locations (or indeed, had abstracted the cue combination rules), then introduction of the cue violations should yield an increase in RTs during the cue violation blocks, relative to the preceding blocks.

We therefore predicted an interaction between 4AFC median split grouping, and block, which would be isolated to greater RTs for those above the 4AFC median split (rather than those below) at blocks where the cue

combinations were violated (compared to the preceding blocks). To model these effects, we first grouped block according to three levels: pre-violation (blocks 6 & 7), violation (blocks 8 & 9) and post-violation (blocks 10 & 11) (see Cohen et al., 1990), and calculated arithmetic means of the harmonic mean RTs for each condition per block pair. A 4 (sound pair) x 3 (block grouping) x 2 (4AFC median split) mixed ANOVA failed to show any significant interaction of 4AFC median split and block grouping,  $F(1.33, 29.25) = 0.81, p = 0.4$ . This was unexpected, particularly given the trend toward increased RTs at block 8 for those above the 4AFC median split (see Figure 5.5a, right side). There was no main effect of block grouping, 4AFC median split, sound pair, nor any significant interactions between these factors (all  $F < 0.9, p > 0.4$ ).

Given that the pattern of results in Figure 5.5a also suggested that the increase in RTs for those above the 4AFC median occurred at block 8 and was followed by a decline in RTs at block 9, we considered whether modelling the last block before the violation (block 7), the first violation block (8) and the first block after the cue combinations were re-introduced (block 10) would yield the expected interaction. Again however, a 4 (sound pair) x 3 (block: 7, 9 & 10) x 2 (4AFC median split) ANOVA did not show any significant main effects or interactions between any of these factors (all  $F < 1.1, p > 0.34$ ). We next collapsed across sound pair and calculated the difference in RTs between blocks 7 and 8 (i.e., block 8 - block 7). However, we found no significant difference between those above and below the median for this difference measure ( $z = 0.4, p = 0.7$ ), suggesting no significant difference in RT cost due to the cue violation.

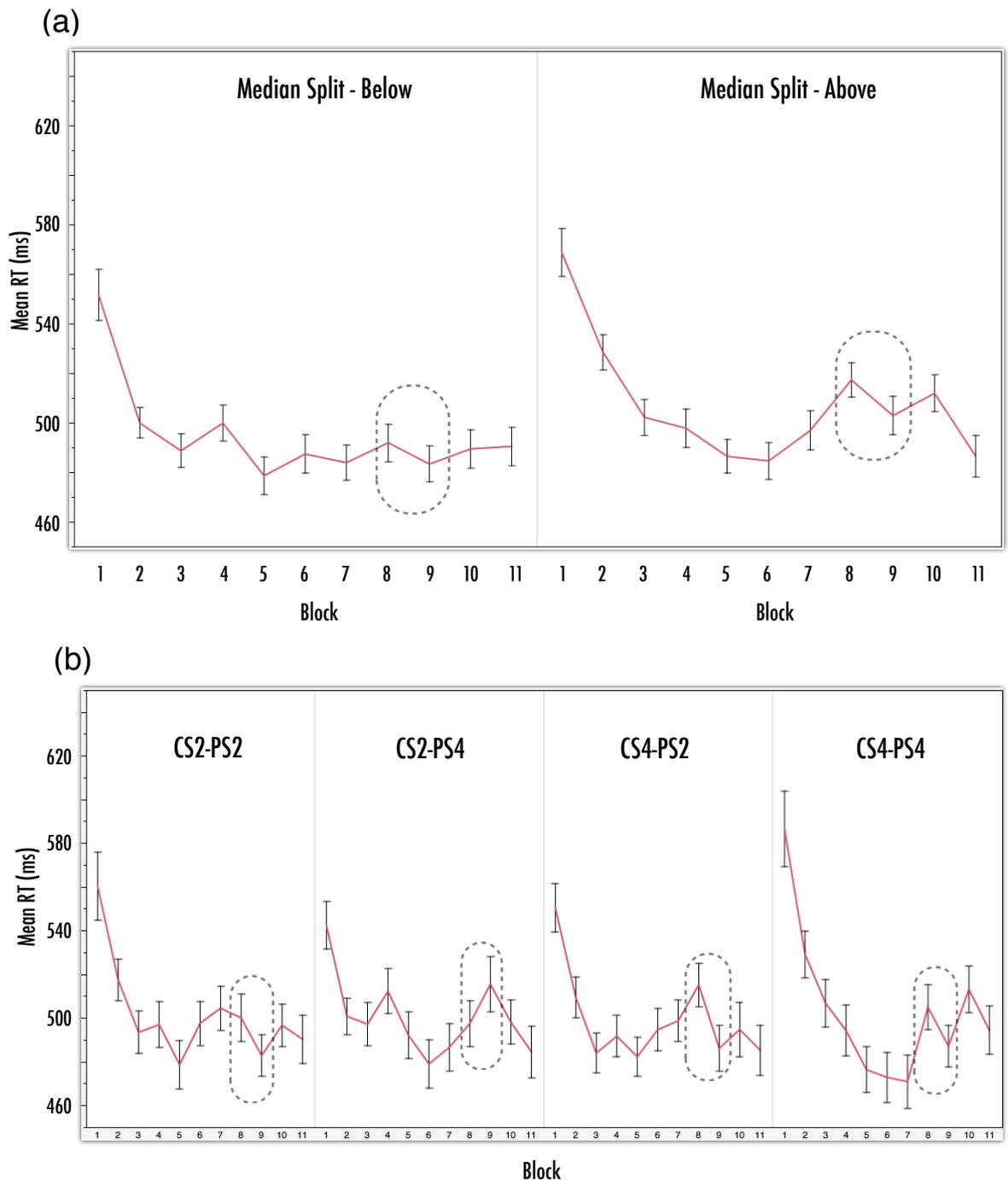


Figure 5.5: (a) arithmetic (i.e., non-harmonic) mean RTs across blocks 1-11, for those adults below (left side) and above (right side) the cohort 4AFC median. Note the elevation in RTs at block 8 (i.e., first cue violation block) for those above the 4AFC median; no increase is noted for those below the 4AFC median. Elevation at block 1 reflects initial practice on the task (block 1 was not analysed). (b) arithmetic mean RTs across blocks, split by CS-PS pair (4AFC cohorts collapsed) (see supplemental analyses 5.1). In (a) and (b), contingency violation blocks are highlighted with dashed ovals. Error bars:  $\pm 1$  std. error of mean.

Finally, we attempted to model the effects more fully across the latter 6 blocks, and included blocks 6 to 11 as a single factor (i.e., harmonic means not collapsed into pairs). This model [4 (sound pair) x 6 (block: 6, 7, 8, 9, 10 & 11) x 2 (4AFC median split)] again did not show any significant main effects or interactions between any factors (all  $F < 1.45$ ,  $p > 0.24$ ).

**5.3.2.2. Response accuracies.** As the introduction of cue violations did not lead to any significant RT cost for those above the cohort 4AFC median, we also considered whether response accuracy varied as a function of the cue violation. However, inspection of accuracy data across sound pair and block conditions revealed performance at or near ceiling levels for many of these conditions; therefore a full ANOVA model of sound pair, block grouping and 4AFC median split was not appropriate (due to violations of the assumptions of normality and homogeneity of variance for several conditions). To test whether those above and below the median split differed in accuracy during the cue violation stages of the game (blocks 8 & 9) relative to the preceding blocks (blocks 6 & 7), we averaged accuracy over sounds for each of these block pairs, and tested the difference in accuracy between blocks 6 & 7 and blocks 8 & 9, over the 4AFC median split. A non-parametric Wilcoxon signed rank test on the difference of the block pairs showed no significant difference between those above and below the 4AFC median ( $z = 0.7$ ,  $p = 0.46$ ). Indeed, mean accuracies were very high for both cohorts across both block pairs ([mean  $\pm$  SD] 4AFC below - blocks 6 & 7: 99%  $\pm$  0.01; blocks 8 & 9: 99%  $\pm$  0.02; 4AFC above - blocks 6 & 7: 98%  $\pm$  0.03; blocks 8 & 9: 97%  $\pm$  0.03). Hence, cue violation did not have a significant effect on task accuracy.

**5.3.2.3. Response spatial distance and variability.** While accuracy to alien character onsets did not differ significantly between the cue violation and

preceding blocks across the 4AFC median split of the cohort, we also considered whether metrics of spatial performance would indicate cohort differences. Over correct trials, we analysed mean and SD of Euclidean distance (in pixel co-ordinates) between the centre of the alien character and the location that the participant tapped. This allowed us to explore the effects of learning on spatial performance. Arithmetic means and means of standard deviations of Euclidean distances from alien centre to location tapped were calculated, and averaged over pre-violation (blocks 6 & 7), violation (blocks 8 & 9) and post-violation (blocks 10 & 11) blocks. We also initially conducted an omnibus model, with block modelled over 10 levels (i.e., blocks 2-11).

In line with the results of RT and accuracy data, we did not find any evidence of significant differences in mean or SD of response spatial distance when modelled over blocks, sound pairs and 4AFC median split. The initial omnibus models showed no significant effects of any of these factors on mean (all  $F < 1.66$ ,  $p > 0.21$ ) or SD (all  $F < 1.65$ ,  $p > 0.21$ ) of spatial distances. Similarly, modelling the effects across pre-violation, violation and post-violation blocks also yielded no significant effects of any factors on mean (all  $F < 1.85$ ,  $p > 0.18$ ) or SD (all  $F < 2.24$ ,  $p > 0.14$ ) of spatial distances.

### **5.3.3 Cross-task analyses**

Finally, we considered whether any of the differences in learning outcomes might relate to individual differences in attentional performance to sounds more generally. We therefore used measures from the SAART task (target inhibition accuracy and SD of RTs to non-targets) to predict learning metrics from the cue learning task (i.e., 4AFC proportion correct and in-game RT cost), using linear regression models. SAART RT SD data for one

participant indicated they were more than two standard deviations above the group mean SD; the participant was therefore not included in cross task analyses. Neither SAART non-target RT SDs nor number of correct target inhibitions differed significantly between those above and below the 4AFC median (both  $z < 1.26$ ,  $p > 0.2$ ).

We found only a weak marginal trend toward a predictive relationship between SAART target inhibition accuracy and 4AFC proportion correct collapsed across sounds [ $F(1, 21) = 3.7$ ,  $p = 0.068$ , adj.  $R^2 = 0.109$ ]; FDR-correction across the set of models showed this effect was non-robust ( $p > 0.2$ ). SAART RT SDs did not significantly predict 4AFC proportion correct, nor did either SAART metric significantly predict RT cost collapsed across sounds [all  $F(1, 21) < 2.7$ ,  $p > 0.11$ , uncorrected].

#### **5.3.4 Interim summary**

In experiment 5a, we investigated learning of combinations of CS and PS stimuli in adults. Results indicated a profile of differences in learning outcomes: while a median split showed that approximately half of the cohort were successful in mapping some of the sound combinations to the aliens' spatial locations at 4AFC, about half of the cohort performed at significantly below chance levels. Notably, even adults whose performance was at or above the 4AFC median split of the cohort showed difficulty in mapping one sound combination (CS2-PS4) to its respective location. This suggests that sound cue combination performance may be a non-trivial task even for adults, despite the quite limited set of available auditory cues. We also observed a slight increase in RTs when cue contingencies were violated, but only for those above the cohort 4AFC median. To our surprise, the effect was statistically non-robust,

and did not yield the expected block grouping x 4AFC median split interaction. We return to the learning of cue combinations (and costs associated with cue violations) in chapter 6, where we investigated these questions in a developing cohort. In the next experiment, we ask whether independent violations of CS and PS stimuli yield costs to performance for effective versus less effective cue learners. We expected that the PS violations would be associated with greater cue cost than the CS violations, due to the relatively greater salience of the PS stimuli.

## 5.4 Results - experiment 5b

In this experiment, we explored whether violating CS and PS stimuli independently would yield differences in cue cost that reflected the differences in relative stimulus salience (in experiment 5a, CS and PS stimuli were violated simultaneously). One participant responded on less than half of the post-game 4AFC trials and was excluded from all analyses; data from 23 participants were analysed (except where noted).

### 5.4.1 4AFC results

As in experiment 5a, we first considered whether adults had learned the mapping between the cue sound pairs and the corresponding aliens. 4AFC proportion correct data for each sound were analysed across all subjects. One-sample Wilcoxon signed rank tests of the distribution for each CS-PS pair relative to chance (0.25) showed that one pair was learned at significantly above chance levels (CS4-PS2: WSR = 84.5,  $p = 0.007$ ); however, performance for other CS-PS pairs did not differ significantly from chance levels (although a trend was noted for CS4-PS4: WSR = 54,  $p = 0.08$ ; CS2-PS2: WSR = 25.5,  $p = 0.45$ ; CS2-PS4: WSR = 51,  $p = 0.1$ ). Nevertheless, a repeated measures ANOVA showed that levels of learning did not differ significantly over the CS-PS pairs [ $F(3, 66) = 1.09$ ,  $p = 0.36$ ] (see Figure 5.6a).

Thus, in contrast to experiment 5a, participants did show some evidence of significant learning when considering the cohort as a whole. However, inspection of distributions suggested that the significant effect for CS4-PS2 was strongly driven by a subset of good performers. As in the previous experiment, a major goal was to explore differences in profiles of post-game learning, and to relate this to preceding in-game performance. We therefore calculated a mean

for each participant across all sound pairs at 4AFC, determined the median of this distribution, and performed a median split of the cohort. The split yielded 11 participants above the median and 12 below.

Tests of 4AFC proportion correct distributions for those above the median split showed that all sounds were matched with alien characters at significantly above chance levels (CS2-PS2: WSR = 33,  $p = 0.001$ ; CS2-PS4: WSR = 23,  $p = 0.018$ ; CS4-PS2: WSR = 33,  $p = 0.001$ ; CS4-PS4: WSR = 26.5,  $p = 0.004$ ). For those below the median split, tests of the 4AFC proportion correct distributions indicated that performance was significantly below chance for one sound (CS2-PS2: WSR = -36,  $p = 0.002$ ), and not significantly different from chance for others (CS2-PS4: WSR = -3.5,  $p = 0.8$ ; CS4-PS2: WSR = -13,  $p = 0.3$ ; CS4-PS4: WSR = -17,  $p = 0.2$ ) (see Figure 5.6b). As expected, a mixed ANOVA showed a highly significant main effect of median split [ $F(1, 21) = 36.95$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.638$ ], with no significant main effect of sound pair or any significant interaction [both  $F(3, 63) < 1.6$ ,  $p > 0.2$ ]. As in the previous experiment, 4AFC median split of the cohort was used as a between-subjects variable in analyses of in-game data.

As explored in experiment 5a, we also examined patterns of sound pair to alien mapping agreement (and confusion) at 4AFC. Table 5.3 presents confusion matrices (frequency counts) for those above and below the 4AFC median; table 5.4 presents confusion matrix descriptive statistics (Kappa, and proportions for accuracy, true positive, false positive, true negative, false negative and precision) over all possible cue sound-response pairs, for those above and below the cohort 4AFC median.

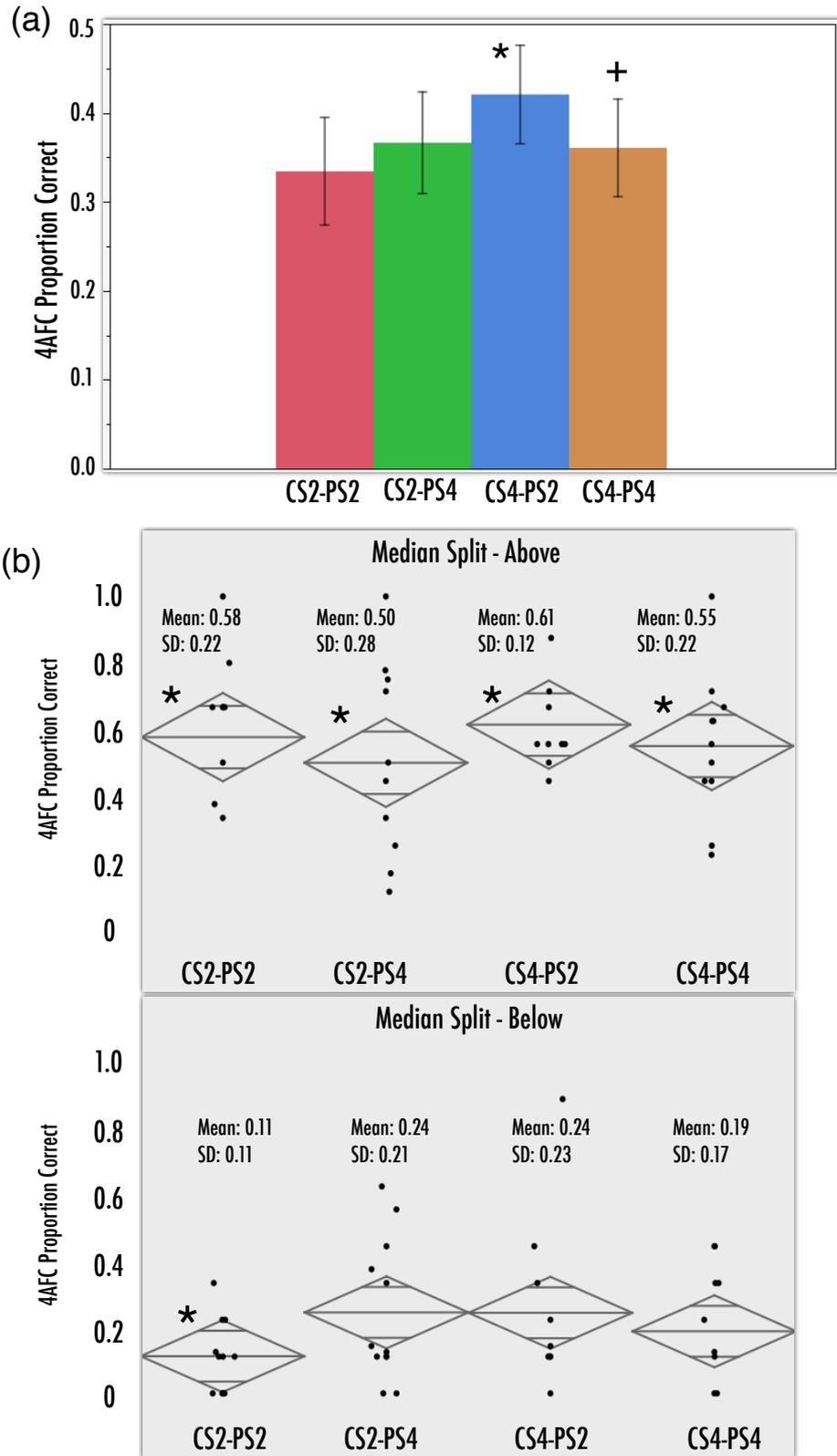


Figure 5.6: 4AFC proportion correct data for experiment 5b. (a) Mean ( $\pm 1$  std. err) proportion correct for each CS-PS pair across full cohort ( $n = 23$ ). (b) 4AFC proportion correct by median split of cohort (upper: above median; lower: below median); horizontal line within each diamond shows mean for that CS-PS, and upper and lower diamond tips the upper and lower bounds of 95% confidence intervals, respectively. WSR (two-tailed) tests: \*  $p < 0.05$  +  $p = 0.08$

Table 5.3: Confusion matrices for adults above and below 4AFC median

		Sound presented				Row total	
		CS2-PS2	CS2-PS4	CS4-PS2	CS4-PS4		
Response	Above	CS2-PS2	<b>51</b>	15	19	5	90
	CS2-PS4	23	<b>45</b>	2	19	89	
	CS4-PS2	16	1	<b>54</b>	18	89	
	CS4-PS4	6	15	21	<b>50</b>	92	
	Col. total	96	76	96	92	360	

		Sound presented				Row total	
		CS2-PS2	CS2-PS4	CS4-PS2	CS4-PS4		
Response	Below	CS2-PS2	<b>12</b>	15	52	25	104
	CS2-PS4	19	<b>25</b>	19	39	102	
	CS4-PS2	29	28	<b>26</b>	23	106	
	CS4-PS4	18	44	22	<b>30</b>	114	
	Col. total	78	112	119	117	426	

As expected, Kappa, accuracy and precision proportions were notably higher for those above the median split compared to those below (table 5.4; compare rows 3 & 11, and rows 8 & 16). Similar to the previous experiment, we observed very high Kappa, accuracy and precision, and low error proportions (false positives and false negatives) for those above the median split when considering cue sound-response pairs where both the CS and PS differed (see table 5.4, columns 4 & 5); this trend was not observed for those below the median split. Again, this suggests the tendency for those above the median split to confuse across one but not both sound types (with no such trend noted for those below the median split).

**Table 5.4:** Kappa coefficients and proportion data for accuracy, true positive, false positive, true negative, false negative and precision data for adults above and below the cohort 4AFC median

Above						
Positive	CS2-PS2	CS2-PS2	CS2-PS2	CS2-PS4	CS2-PS4	CS4-PS2
Negative	vs. CS2-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS4
Kappa ( $\kappa$ )	0.434	0.500	0.804	0.941	0.473	0.454
Accuracy	0.716	0.750	0.902	0.971	0.736	0.727
True Pos.	0.773	0.729	0.911	0.957	0.703	0.750
False Pos.	0.227	0.271	0.089	0.043	0.297	0.250
True Neg.	0.662	0.771	0.893	0.982	0.769	0.704
False Neg.	0.338	0.229	0.107	0.018	0.231	0.296
Precision	0.689	0.761	0.895	0.978	0.750	0.720
Below						
Positive	CS2-PS2	CS2-PS2	CS2-PS2	CS2-PS4	CS2-PS4	CS4-PS2
Negative	vs. CS2-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS4
Kappa ( $\kappa$ )	0.012	-0.330	-0.052	0.049	-0.203	0.108
Accuracy	0.521	0.319	0.494	0.520	0.399	0.554
True Pos.	0.444	0.188	0.324	0.568	0.391	0.531
False Pos.	0.556	0.813	0.676	0.432	0.609	0.469
True Neg.	0.568	0.473	0.625	0.481	0.405	0.577
False Neg.	0.432	0.527	0.375	0.519	0.595	0.423
Precision	0.387	0.293	0.400	0.472	0.362	0.542

As in the previous experiment, we hypothesised that confusion across the less salient CS would be more common than confusion across the more salient PS. In contrast to those adults below the 4AFC median in experiment 5a, here adults below the 4AFC median showed negative Kappa (-0.33), as well as increased false positive (0.813), and false negative (0.527) proportions for CS2-

PS2 vs. CS4-PS2. Adults below the median also showed negative Kappa (-0.203), and increased false positive (0.609) and false negative (0.595) proportions for CS2-PS4 vs. CS4-PS4 (in both instances, compared to when PS was fixed and CS varied; compare table 5.4, columns 2 & 3, and columns 6 & 7). This suggests that as expected, those below the 4AFC median more frequently confused (i.e., showed reduced agreement) over the less salient sound class. Indeed, Kappa values were relatively similar across pairs that shared a sound class for those above the 4AFC median (range: 0.434-0.5), suggesting relatively better agreement (and reduced confusion) at 4AFC.

#### **5.4.2 In-Game results**

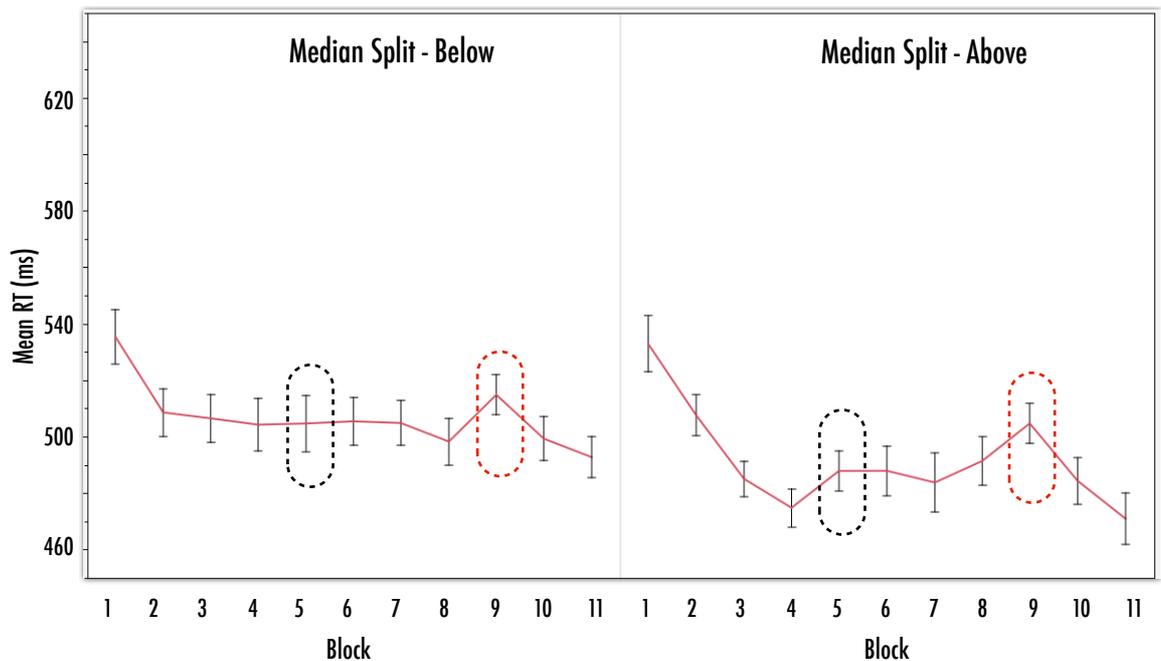
As in experiment 5a, data analysed were: in-game RTs to alien onsets; response accuracies to alien onsets; means and standard deviations of Euclidean distances from the alien centre co-ordinates to the tapped location co-ordinates. As in previous analyses, mixed ANOVAs were used, modelling effects of sound pair (4 levels) and block (various levels), along with 4AFC median split as a between-subjects factor. We predicted that those above the 4AFC median would show greater RT cost, and potential changes in accuracy and/or spatial variability as a result of separate CS and PS cue violation; however, we expected any effects to be reduced for CS relative to PS violation. No cue violation effects were expected for those below the median.

**5.4.2.1. RTs to alien onsets.** As in experiment 5a, harmonic means of RT data were analysed for each condition; block 1 was again treated as practice and was not analysed; omnibus analyses for blocks 2-11 are presented in supplemental analyses 5.2.

Critically, we tested the effects of separate CS and PS cue violations on RTs. For each sound pair, we calculated the difference in harmonic mean RTs between block 4 and block 5 (CS violation block) (block 5 - block 4) and between block 8 and block 9 (PS violation block) (block 9 - block 8). We then submitted the differences between these pairs of blocks to a 2 (block difference) x 4 (sound pair) x 2 (median split) ANOVA. There was no significant main effect of median split, block difference, or sound (all  $F < 1.46$ ,  $p > 0.24$ ). Unexpectedly, we also did not find a significant block difference x median split interaction [ $F(1, 21) = 0.94$ ,  $p = 0.34$ ,  $\eta_p^2 = 0.04$ ]. To further explore the predicted effect, we collapsed over sounds, and ran separate Wilcoxon tests for the differences over each pair of blocks. However, we again found no significant difference between those above and below the median split either for the difference of blocks 4 and 5 (CS cue violation;  $z = 0.83$ ,  $p = 0.41$ ), or the difference of blocks 8 and 9 (PS cue violation;  $z = 0.52$ ,  $p = 0.6$ ).

To explore the expected increase in RTs with cue violation, we also modelled the cue violation blocks alongside the immediately preceding and following cue contingency blocks. Considering first the CS cue violation (block 5), a 4 (sound pair) x 3 (Block: 4, 5, 6) x 2 (median split) ANOVA showed no significant main effects nor any significant interactions involving these factors (all  $F < 1.24$ ,  $p > 0.3$ ). Next, for the PS cue violation (block 9), a 4 (sound pair) x 3 (Block: 8, 9, 10) x 2 (median split) ANOVA showed a marginally significant main effect of block [ $F(2, 42) = 2.92$ ,  $p = 0.065$ ,  $\eta_p^2 = 0.122$ ], and surprisingly, no significant block x median split interaction [ $F(2, 42) = 0.3$ ,  $p = 0.74$ ]. No other main effects or interactions were significant ( $F < 1.42$ ,  $p > 0.23$ ).

In line with the marginal block main effect, pairwise post-hoc tests (FDR-corrected) showed marginally longer RTs at block 9 compared to block 8 and



**Figure 5.7:** arithmetic (i.e., non-harmonic) mean RTs across blocks 1-11, for adults below (left side) and above (right side) the cohort 4AFC median. Note the elevation in RTs at block 9 for both those above and below the 4AFC median. Elevation at block 1 reflects initial practice on the task (block 1 not analysed). Dashed ovals highlight the CS violation (black) and PS violation (red) blocks. Error bars:  $\pm 1$  std. error of mean.

compared to block 10 (both  $p = 0.08$ ); no significant difference emerged between blocks 8 and 10 ( $p = 0.94$ ). Thus, unexpectedly, the PS cue violation appeared to yield a marginal increase in RTs for participants above and below the 4AFC median split (see Figure 5.7). Notably however, we did not observe an increase in RTs at PS violation blocks for those below the 4AFC median when the PS violations occurred at block 5 (as was the case in our control experiment; see supplemental analyses 5.2). This may suggest that the effect noted for those below the median split (see Figure 5.7) arose (at least in part) due to the PS stimuli being relatively more stable and reliable than the CS stimuli over the first 8 blocks of the experiment.

**5.4.2.2. Response accuracies.** We also explored whether the introduction of cue violations influenced participants' accuracy during the task.

However, as noted for experiment 5a, performance accuracy was at or near ceiling levels for many conditions (precluding a full ANOVA model, due to violations of the assumptions of normality and homogeneity of variance). We therefore collapsed across sound pairs, and calculated accuracy difference scores between block 4 and block 5 (CS cue violation) and between block 8 and block 9 (PS cue violation). Using 4AFC median split of the cohort as a between subjects factor, we found no evidence of significant differences in accuracy for either difference score, between those above and below the median split (both  $z < 1.1$ ,  $p > 0.28$ ). We also analysed accuracy data for cue violation blocks alongside accuracy data for the immediately preceding and following blocks (collapsed over sounds). For the PS violation, a 3 (block: 8, 9, 10) x 2 (median split) ANOVA showed no significant main effects and no interaction of the two factors (all  $F < 1.92$ ,  $p > 0.16$ ). For the CS violation, a 3 (block: 4, 5, 6) x 2 (median split) ANOVA showed a main effect of block that was close to significance [ $F(2,42) = 3.09$ ,  $p = 0.056$ ,  $\eta_p^2 = 0.128$ ], with no other significant main effects or interactions (both  $F < 0.7$ ,  $p > 0.5$ ). However, post-hoc tests (FDR corrected) showed the block main effect was not robust (all  $p > 0.1$ ). Indeed, accuracies were high across blocks 4 to 6 for those above (mean  $\pm$  SD - block 4: 94.7%  $\pm$  0.07; block 5: 97.7%  $\pm$  0.04; block 6: 97%  $\pm$  0.04) and below (mean  $\pm$  SD - block 4: 95.8%  $\pm$  0.08; block 5: 97.9%  $\pm$  0.05; block 6: 95.8%  $\pm$  0.06) the 4AFC median split.

**5.4.2.3. Response spatial distance and variability.** Although response accuracies to alien character onsets were high (and not significantly different between those above and below the median split), we also explored whether the mean and variability of Euclidean distance between the alien centre and

location tapped varied as a function of cue violation, or indeed, 4AFC median split. Omnibus analyses are presented in supplemental analyses 5.2.

We modelled mean spatial distances across cue violation blocks and the immediately preceding and following blocks. Considering CS violation, a 4 (sound pair) x 3 (block: 4, 5, 6) x 2 (median split) ANOVA showed no significant effects of any factor (all  $F < 1.25$ ,  $p > 0.3$ ). For PS violation, a 4 (sound pair) x 3 (block: 8, 9, 10) x 2 (median split) ANOVA showed a main effect of sound that reached significance [ $F(3, 63) = 2.75$ ,  $p = 0.05$ ,  $\eta_p^2 = 0.116$ ], but no other significant effects ( $F < 1.19$ ,  $p > 0.31$ ). Post-hoc pairwise tests (FDR-corrected) showed a marginal difference in mean spatial distance between aliens cued by CS2-PS4 and CS4-PS4 ( $p = 0.06$ ); means ( $\pm$  SD) suggested spatial distances were increased for CS2-PS4 ( $31.5 \pm 10.1$ ) compared to CS4-PS4 ( $26.5 \pm 7.0$ ).

Finally, analyses of SD of spatial distances across omnibus, CS violation and PS violation blocks (all models as above) showed no significant effects of any factor within any model (all  $F < 2.0$ ,  $p > 0.15$ ).

#### **5.4.3 Cross task analyses**

Finally, we examined whether differences in learning outcomes might be predicted by individual differences in attentional performance as indexed by the SAART task (using SD of RTs to non-target sounds and number of correct inhibitions to target sounds). Two participants had SAART target inhibition accuracies more than two SDs below the cohort mean and were excluded from cross-task analyses (analysis  $n = 21$ ). Neither SAART non-target RT SDs nor number of correct target inhibitions differed significantly between those above and below the 4AFC median (all  $z < 1.65$ ,  $p > 0.1$ ).

SAART metrics did not significantly predict 4AFC proportion correct collapsed across sounds (both  $F < 0.8$ ,  $p > 0.4$ , uncorrected).

We further considered whether any RT cost due to CS cue violation would relate to SAART performance. However, neither SAART metric significantly predicted the difference in RTs between blocks 4 and 5 (i.e., block 5 - block 4) collapsed across sounds (all  $F < 0.25$ ,  $p > 0.6$ , uncorrected). Finally, we asked whether RT cost due to PS cue violation would be predicted by SAART performance. SAART RT SDs were a significant predictor of the difference in RTs between blocks 8 and 9 (i.e., block 9 - block 8) collapsed over sounds [ $F(1,19) = 5.60$ ,  $p = 0.029$ , adj.  $R^2 = 0.187$ ]; however, FDR-correction over all models showed the relationship to be non-robust ( $p > 0.17$ ). SAART target accuracies did not significantly predict the RT difference between blocks 8 and 9 (all  $F < 0.1$ ,  $p > 0.85$ , uncorrected).

#### 5.4.4 Interim summary

In experiment 5b, we asked whether violation of auditory cue contingencies independently of one another would lead to differences in cue cost, based on the relative differences in saliency between both sound types. We found no evidence of any significant cost associated with CS (but not PS) cue violation; this suggests that even for participants who showed significant learning at 4AFC (i.e., those above the cohort median), violation of the less salient cue type did not have a significant effect on performance. Interestingly however, we also did not find that those above the 4AFC median showed a significant performance cost when PS (but not CS) contingencies were violated. Instead, we observed a trend for both those above and below the 4AFC median to show increased RTs when PS (but not CS) contingencies were violated at

block 9. In partial support of our hypothesis, this suggests that the salient cue violation appears to have been detected by those above and below the 4AFC median, and appeared to influence RT performance for both cohorts (although the effect was only statistically marginal across the full cohort). However, our control experiment (see supplemental analyses 5.2) showed no increase in RTs for those below the median split when PS violations occurred earlier in the experiment (block 5), with a very small increase in RTs at block 5 for those who performed at near ceiling levels at 4AFC. As expected, this suggests that the timing of violations with respect to the number of learning blocks completed can influence the relative extent of cue costs arising from contingency violations.

## **5.5 Chapter summary**

In this chapter we have explored learning of complex sound combinations within the context of an on-going visuo-spatial task. Our results suggest strong variation over individuals in the extent of success in combining auditory cues (on-going contexts and shorter, punctate objects), and mapping these combinations to specific locations. Contrary to our hypothesis, we found limited evidence that simultaneous violation of the contingencies for both cue types yielded robust decrements in RT performance (experiment 5a). However, we found that when the relatively more salient cue type (PS) was violated, listeners showed a marginal increase in RTs regardless of their subsequent 4AFC learning performance (experiment 5b). Nevertheless, data from our control experiment appear to suggest that the order in which the violations occurred may have led to those below the 4AFC median split also showing a slight RT increase in experiment 5b.

In the next chapter, we investigate learning of the cue combinations from experiment 5a in a developing cohort (experiment 6). Of central interest was whether children at an age (8-9 years) at which cue combination abilities begin to emerge (within and across modalities; see Nardini et al., 2008; 2012) would show learning of complex auditory cue combinations within an active visuo-spatial task. Moreover, we explored whether violations of cue contingencies (as in experiment 5a) would yield decrements in children's task performance, in line with the expected costs arising from reduced cue reliability (Bates & MacWhinney, 1987). We compared the performance of the adults from experiment 5a to that of the children in experiment 6, asking whether development mediates success of cue combination learning (or indeed, costs associated with cue contingency violations).

## Chapter 5 Footnotes

1. A median split was performed for each of the cohorts in experiments 5a, 5b and 6 using 4AFC proportion correct. We opted to use the 4AFC median to split the cohorts (as opposed to upper versus lower quartiles, for instance) in order to include all of the data, and thus maximise statistical power within the RT, accuracy and spatial distance analyses of in-game data. Utilising the median entails the disadvantage of including some subjects who perform at close to chance levels within the 'above median' side of the split. However, running the analyses reported in this chapter with a 4AFC split based on the third of subjects performing the best versus the worst at 4AFC did not improve the significance of RT cost effects (over and above the analyses presented here).
2. As detailed in footnote 2 of chapter 4, descriptive statistics for classification of sound combinations at post-game test are described throughout the results sections in this chapter. The following terms are used to describe classification accuracy (and misclassification).

From Kohavi & Provost (1998), for a 2 x 2 confusion matrix of the form:

	Negative	Positive
Negative	a	b
Positive	c	d

Total accuracy:  $(a + d)/(a + b + c + d)$

True positive rate:  $d/(c + d)$

False positive rate:  $b/(a + b)$

Precision:  $d/(b + d)$

False negative rate:  $c/(c + d)$

True negative rate:  $a/(a + b)$

In addition, we utilise Kappa statistics as a metric of the extent of correct classification (i.e., agreement) that occurs over and above correct classification arising as a result of chance (i.e., the baseline constraint) (Cohen, 1968; Landis & Koch, 1977). Kappa ( $\kappa$ ) is expressed as follows:

$$\kappa = \frac{\pi_0 - \pi_e}{1 - \pi_e}$$

where  $\pi_0$  is the total accuracy proportion  $[(a + d)/(a + b + c + d)]$  and  $\pi_e$  is the total proportion of accurate classifications arising due to chance  $[(((a + c)*(a + b)) + ((b + d)*(c + d)))/((a + b + c + d)^2)]$ . Note that although Kappa typically indexes reliability of agreement (and hence positive values would be expected), it is also possible for Kappa to reflect reliable disagreement (hence yielding negative values) (Jurlink & Detsky, 2005).

## Supplemental Analyses

### Supplemental analyses 5.1: experiment 5a

**RTs.** As an initial model of performance across all stages of the task, we ran an omnibus mixed ANOVA with sound pair (each CS-PS pair; 4 levels) and block (2-11; 10 levels) as within subject factors, and 4AFC median split as a between-subject factor. The 4 x 10 x 2 ANOVA showed no significant main effects or significant interactions between any of the factors (all  $F < 2.1$ ,  $p > 0.1$ ).

To establish whether RT performance varied as a function of these factors during the learning stages of the task, a similar ANOVA treating block over 6 levels (i.e., blocks 2-7, during which sound-alien location contingencies were established) was performed. The 4 (sound pair) x 6 (block; 2-7) x 2 (4AFC median split) ANOVA revealed a significant main effect of block, [ $F(3.04, 66.94) = 4.04$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.155$ ] and a significant sound pair x block interaction [ $F(7.81, 171.7) = 3.72$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.145$ ]. However, there was no significant main effect of 4AFC median split grouping, nor any interactions involving this factor (all  $F < 0.8$ ,  $p > 0.5$ ), suggesting that those who were above the cohort median at 4AFC did not show any clear RT performance differences during learning stages compared to those below the median split. Post-hoc pairwise comparisons (FDR-corrected) for the block main effect showed close to significant differences (all  $p = 0.05$ ) reflecting facilitated RTs at blocks 3, 5 and 6 relative to block 2. Decomposing the sound x block interaction revealed marginally significant effects of block for CS2-PS2 [ $F(3.43, 78.94) = 2.38$ ,  $p = 0.068$ ,  $\eta_p^2 = 0.094$ ], and weak trends for CS4-PS2 [ $F(3.23, 74.35) = 2.29$ ,  $p = 0.08$ ,  $\eta_p^2 = 0.091$ ], and for CS2-PS4 [ $F(3.1, 71.32) = 2.17$ ,  $p = 0.097$ ,  $\eta_p^2 = 0.086$ ]. The effect of block was highly significant for CS4-PS4 [ $F(2.76, 63.39) =$

8.38,  $p < 0.0001$ ,  $\eta_p^2 = 0.267$ ]. Post-hoc pairwise comparisons (FDR-corrected) showed that for CS4-PS4, RTs at block 2 were significantly longer compared to blocks 4, 5, 6 and 7 (all  $p < 0.05$ ); RTs at block 3 were also significantly longer compared to blocks 5, 6 and 7 (all  $p < 0.05$ ). This suggests that RTs to aliens cued by CS4-PS4 declined as learning progressed, and to a relatively greater degree than noted for other CS-PS pairs (see Figure 5.5b and discussion).

### **Supplemental analyses 5.2: experiment 5b**

**RTs.** An initial omnibus ANOVA model of sound pair (4 levels), block (2-11; 10 levels) and 4AFC median split (2 levels) showed no significant main effects, nor any significant interactions between these factors (all  $F < 1.6$ ,  $p > 0.19$ ).

Next, we explored whether performance varied when considering only those early blocks over which the cue combinations were established (i.e., blocks 2-4 and blocks 6-8). A 4 (sound pair) x 6 (block: 2, 3, 4, 6, 7, 8) x 2 (median split) ANOVA showed no significant main effect of median split nor any significant interactions involving this factor (all  $F < 1.1$ ,  $p > 0.37$ ). There was no significant main effect of block [ $F(2.12, 44.49) = 0.99$ ,  $p = 0.39$ ,  $\eta_p^2 = 0.045$ ]. A sound pair x block interaction reached significance [ $F(6.26, 131.4) = 2.36$ ,  $p = 0.032$ ,  $\eta_p^2 = 0.101$ ]; however, post-hoc ANOVAs (splitting by sound or by block) failed to yield any significant effects (all  $F < 2.17$ ,  $p > 0.11$ ).

**Spatial Distance.** An omnibus 4 (sound pair) x 10 (block; 2-11) x 2 (median split) ANOVA on arithmetic mean Euclidean distances yielded a significant main effect of block [ $F(5.25, 110.34) = 2.65$ ,  $p = 0.024$ ,  $\eta_p^2 = 0.112$ ]; no other effects were significant (all  $F < 1.0$ ,  $p > 0.46$ ). Post-hoc pairwise tests (FDR-corrected) showed a weak trend toward a difference between blocks 2

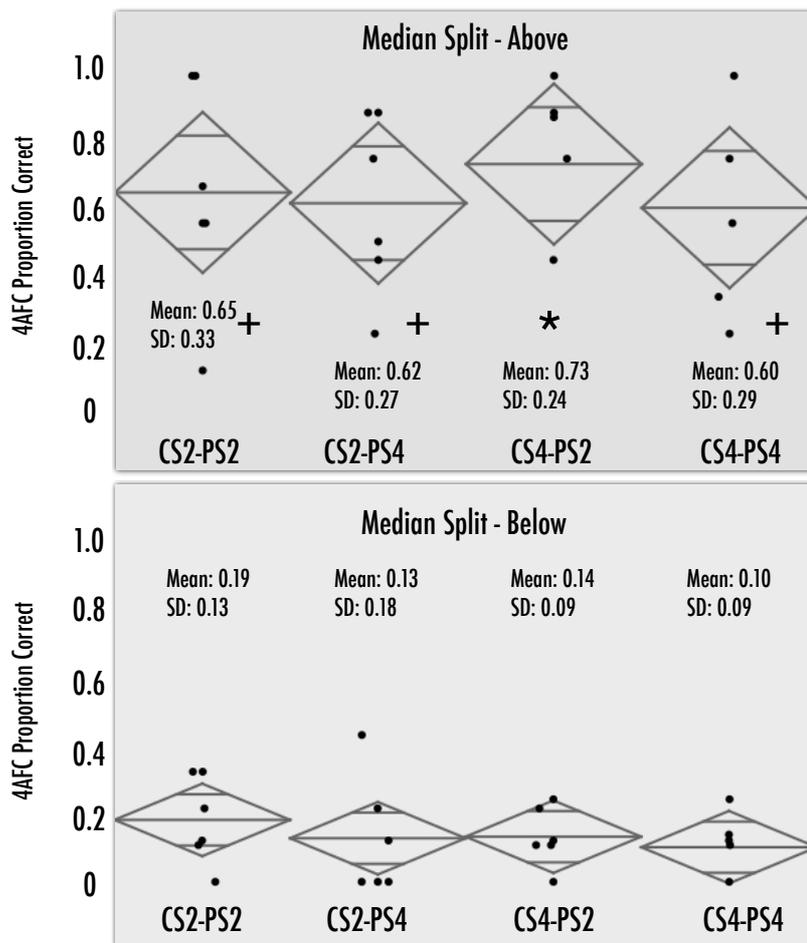
and 8 ( $p = 0.1$ ); means ( $\pm$  SD) suggested reduced distance between alien centre and tapped location at block 8 ( $27.7 \pm 8.0$ ) compared to block 2 ( $34.2 \pm 10.9$ ).

**Control experiment.** As outlined in methods (see 5.2.4.2), because the fixed order of contingency violations for CS and PS stimuli in experiment 5b was a confound within the design (i.e., the order was not counterbalanced across subjects), we also ran a control experiment in which we used the reverse order of contingency violations. Participants were 12 healthy right handed adults (9 female, 3 male; mean age  $\pm$  SD:  $26.2 \pm 5.2$ ; range: 20-38) recruited from the same participant pool as participants in experiments 5a and 5b; all met the same inclusionary criteria as per experiments 5a and 5b.

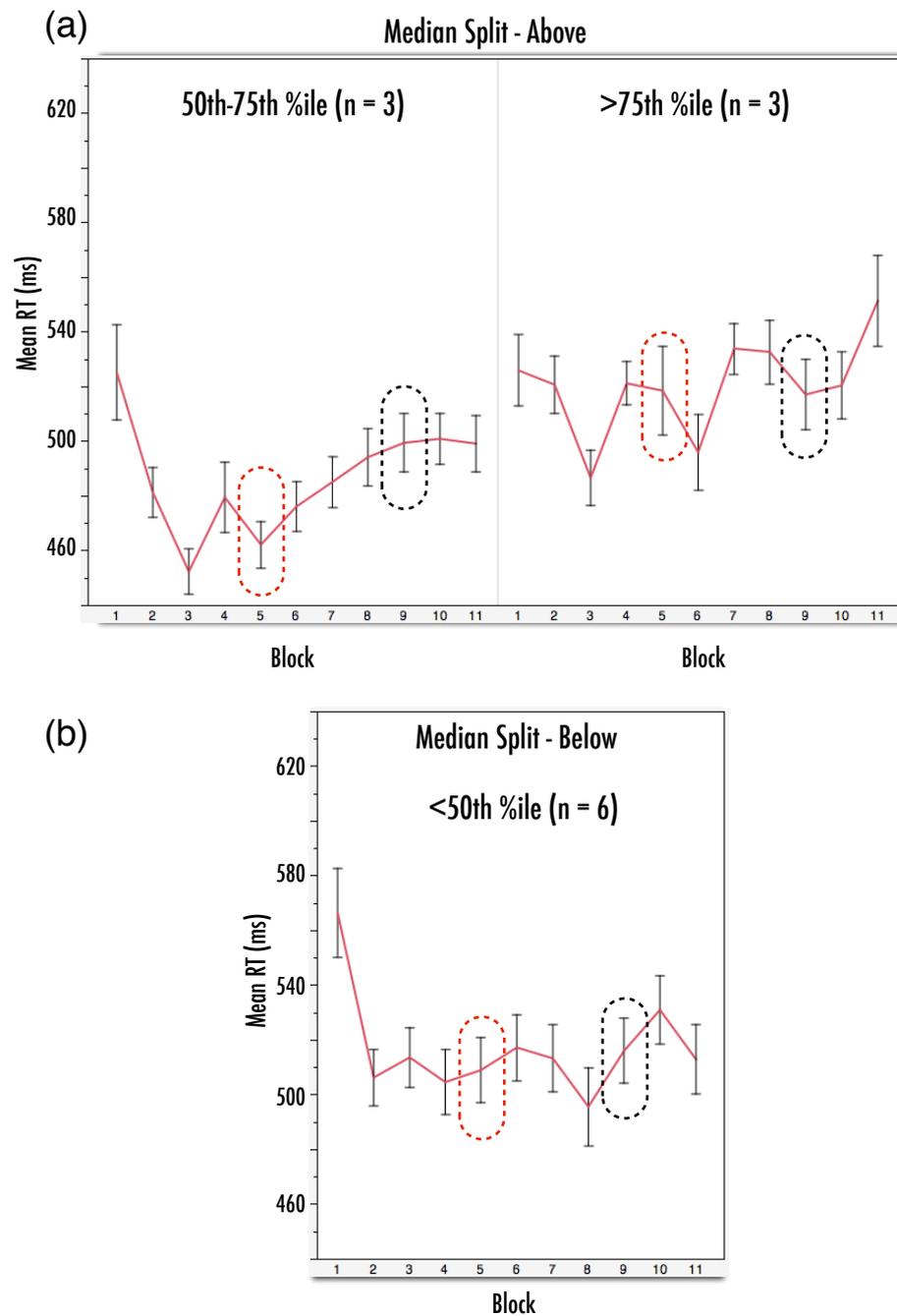
We again found evidence of individual differences in learning success at 4AFC (see Supplemental Figure [SF] 5.1). A median split of the cohort showed that those above the median identified one sound pair at significantly above chance level (CS4-PS2; WSR = 10.5,  $p = 0.031$ ), and three of the sound pairs at marginally above chance level (CS2-PS2; CS2-PS4; CS4-PS4; all WSR = 9.5,  $p = 0.063$ ) (all tests two-tailed). Those below the median identified two sound pairs at marginally below chance level (CS4-PS2 & CS4-PS4; both WSR = 7.5,  $p = 0.063$ ); the other sound pairs were not identified at significantly below chance level (both  $p > 0.15$ ).

Importantly, analyses of task harmonic mean RTs showed no significant costs associated with PS or CS contingency violations for those above or below the 4AFC median; there were no significant main effects nor interactions [PS violation: 4 (sound) x 3 (Block: 4, 5, 6) x 2 (median split), all  $F < 2.1$ ,  $p > 0.16$ ; CS violation: 4 (sound) x 3 (Block: 8, 9, 10) x 2 (median split), all  $F < 1.9$ ,  $p > 0.18$ ]. While the small sample size likely reduced statistical power, we

nevertheless did not see any RT increase at block 5 for those below the 4AFC median split, as noted at block 9 in experiment 5b (compare left side of Figure 5.7 and SF 5.2). We also noted an unexpected increase in RTs for those above the 4AFC median at block 4 – one block before the PS violations (see SF 5.2a). The exact source of this increase is unclear, although it likely reflects slowed responses following the very rapid decrease in RTs over the initial three blocks. Note that for those participants achieving the highest performance at 4AFC (see SF 5.2a, right side), the increase in RTs at block 4 was followed by elevated RTs at block 5, perhaps reflecting a cost due to the PS cue violation at block 5. These results are preliminary in nature and should be viewed cautiously.



Supplemental Figure 5.1: 4AFC proportion correct by median split of control experiment cohort (upper: above median; lower: below median); all attributes as per Figure 5.6b; WSR (two-tailed) tests: \*  $p < 0.05$  +  $p < 0.07$



**Supplemental Figure 5.2:** Arithmetic mean RTs across blocks for those (a) above and (b) below 4AFC median split in control experiment. (a) Left panel shows mean RTs across blocks for 3 participants between 50th and 75th percentiles for 4AFC proportion correct (sounds collapsed); right panel shows mean RTs for 3 participants above 75th percentile for 4AFC proportion correct (sounds collapsed). (b) Mean RTs for 6 participants below 4AFC median (i.e., 50th percentile). Dashed ovals highlight the PS violation (red) and CS violation (black) blocks. Error bars:  $\pm 1$  std. error of mean.

Chapter 6: Short-term Auditory Learning  
within a Multi-modal Environment:  
Development

## **6.1 Introduction**

In this chapter, we extend the experiment 5a paradigm to a cohort of school-aged children (experiment 6). We sought to explore whether children could learn combinations of complex auditory cues encountered within an on-going visuo-spatial task. Moreover, we aimed to investigate whether success of cue learning in children would be associated with greater cue costs when cue contingencies were violated. In latter sections, we compare children and adults directly. This allowed us to examine whether cue combination performance differs between children and adults, and whether cue cost associated with contingency violation was greater for children than adults. Finally, we provide a general discussion of findings from experiments presented in chapters 5 and 6.

## **6.2 Method**

### **6.2.1 Participants**

Participants were 24 school-aged children (11 female, 13 male; mean age [months]  $\pm$  SD:  $108.8 \pm 3.75$ ; range [months]: 103.3–113.9; 22 right-handed, 2 left-handed), with normal or corrected to normal vision. Children were recruited from and tested at a primary school in South London. All except two children had no history of hearing difficulties or hearing loss. One child had a history of ear infections with mild hearing loss; the condition had resolved by the time of testing according to parental report. A further child had suffered extensive hearing damage requiring surgery, and was not included in analyses.

### **6.2.2 Stimuli and materials**

All stimuli and materials were identical to experiments 5a and 5b (see 5.2.2 & 5.2.3).

### 6.2.3 Procedure

Parents of children provided signed informed consent in advance of experimental testing; children also provided verbal assent before beginning the experiment. Prior to testing, parents completed an 11-item questionnaire on their child's development, including questions on hearing and language issues, history of developmental disorders (if any), attentional difficulties (if any) and scholastic achievement.

Children were tested in pairs by two experimenters (i.e., one experimenter per child) in a quiet, familiar environment at their school. Children completed a series of tasks, measuring verbal ability, learning of cue combinations, selective visual attention, and sustained auditory attention. Tasks were framed with an earth and space narrative, with pictures used to explain the tasks. Description of each follows below (tasks were run in the order described for all participants, and took 30 minutes to complete).

**6.2.3.1. Test of Word Reading Efficiency (TOWRE).** Children completed the sight word reading efficiency subtest for familiar words from the TOWRE (Torgesen, Wagner & Rashotte, 1999). The subtest was included to provide an assay of children's general reading and broader cognitive development, and to detect potential language difficulties (further to parent report). The task measures verbal reading ability, requiring participants to read aloud a list of words that increases in difficulty (final score is the number of words correctly read aloud in 45 seconds; raw scores were used since age-standardised scores from UK-based samples of children were unavailable). Children listened to a recording of task instructions, and completed a short practice list. Children then read aloud one of the TOWRE lists (counterbalanced

across participants). Their reading was recorded as a sound file using Matlab and saved for scoring offline.

**6.2.3.2. Learning task (cue combination).** Children completed the same iPad learning task as described in experiment 5a (see 5.2.4.1). The procedure was similar, with the exception that an audio recording of the task instructions was played to children before the task. The task was also adapted so as to be child-friendly; after every two blocks, a 'UFO' picture appeared on-screen. Children could tap on-screen images to apply colourful stickers to the UFO.

**6.2.3.3. Test of Everyday Attention in children (TEA-Ch): Sky Search subtest.** Children completed the Sky Search subscale from the TEA-Ch battery (Manly et al., 2001). The Sky Search task measures visual selective attention to complex arrays, using a visual search task. The task requires children to find pairs of matching spaceship pictures presented side by side in vertical columns, as quickly as possible; more frequent, non-matching spaceship pairs are also presented in each column, requiring selective search for the matching targets. Children first completed an A4 sized practice sheet. Children circled matching pairs as quickly as possible, ticking a box to indicate when they had finished. Following this, children completed a larger (A3) test sheet (A or B, counterbalanced over participants), presenting 20 matching and 108 non-matching pairs. Children were timed as they completed the task, and circled as many of the matching pairs as they could find (the timer was stopped when children ticked to signal completion). Finally, children completed the motor control component of the task. The same matching target pairs as presented on the test sheet were shown without any non-matching pairs; children circled the matching items as quickly as possible (the timer was again stopped when they

indicated they had finished). Children's final score on the task was the time per target (i.e., time taken, divided by the number correctly circled) for the test sheet, minus the time per target for the motor control sheet. This provided a measure of visual selective attention that removed some of the variance associated with motor performance.

**6.2.3.4. SAART.** A shortened and adapted version of the response inhibition SAART as presented to adults was completed by children. The task provided a metric of both sustained attentiveness to sounds, and of broader executive function (indexed by response inhibition). Children listened to a recording of task instructions, and completed a brief pseudorandom practice with 18 sounds (4 target sounds). Children then completed 81 trials, with 17 target sounds (bird calls) and 64 non-targets. Note that the proportion of targets (no-go trials) relative to non-targets was increased relative to the adult version (162 trials, 18 targets), in order to provide a sufficient number of target trials so as to allow inhibitory performance to be measured over this shortened experiment.

### 6.3 Results

Data from one child were not analysed; parental report indicated a history of hearing damage requiring surgery that led to hearing loss in that child.

#### 6.3.1 4AFC results

First, 4AFC proportion correct data were analysed across the full child cohort. One-sample Wilcoxon signed rank tests (two-tailed) showed that the distribution of proportion correct for each CS-PS combination did not differ significantly from chance levels (0.25) (see Figure 6.1a; CS2-PS2: WSR = 23.5,  $p = 0.46$ ; CS2-PS4: WSR = 43.5,  $p = 0.19$ ; CS4-PS2: WSR = 26.5,  $p = 0.4$ ; CS4-PS4: WSR = -13.5,  $p = 0.65$ ).

Thus, the full cohort of children showed post-test performance that did not differ significantly from chance levels. Since a major aim of the experiment was to characterise profiles of learning at post-test and to use this to further explore learning in the preceding game phase, a median split was performed on 4AFC data. Children's mean 4AFC performance across all sounds was calculated, and a split of the group was performed above and below the cohort median of across-sound means. The median split left 11 children (6 girls and 5 boys) above the median and 12 below (5 girls and 7 boys); there were no significant differences in gender [ $\chi^2(1, n = 23) = 0.4, p = 0.54$ ] or age in months ( $z = 0.46, p = 0.64$ ) between children above and below the median. Indeed, based on this split, the children above the cohort median showed proportion correct levels that were significantly above chance for two sounds (CS2-PS2: WSR = 24.5,  $p = 0.024$ ; CS2-PS4: WSR = 32,  $p = 0.002$ ), and that were marginally above chance for the remaining two sounds (CS4-PS2: WSR = 18.5,  $p = 0.063$ ; CS4-PS4: WSR = 18.5,  $p = 0.06$ ) (all tests two-tailed).

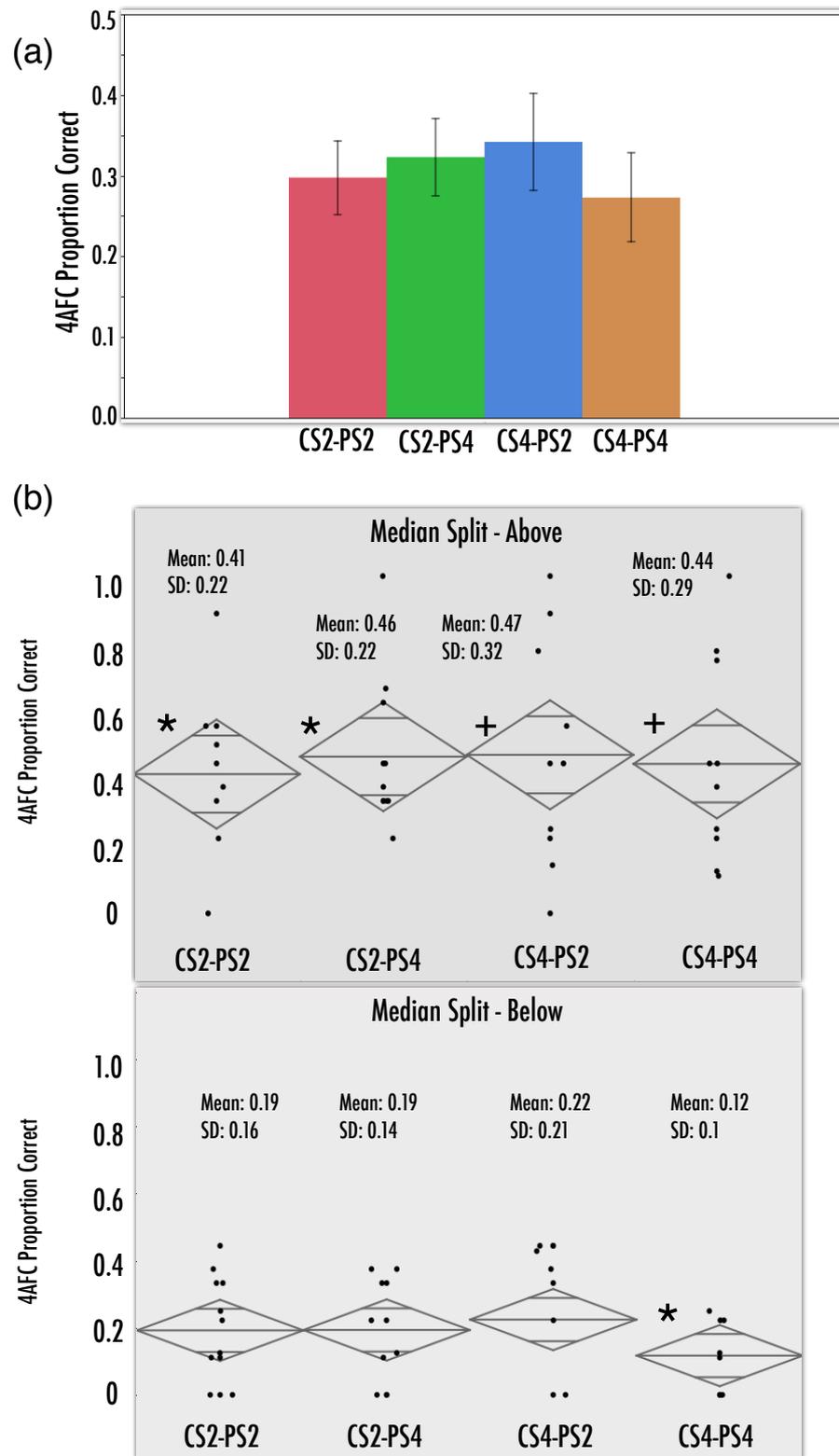


Figure 6.1: 4AFC proportion correct data for experiment 6. (a) Mean ( $\pm 1$  std. err) proportion correct for each CS-PS pair across full cohort ( $n = 23$ ). (b) 4AFC proportion correct by median split of cohort (left: below median; right: above median); horizontal line within each diamond shows mean for that CS-PS, and upper and lower diamond tips the upper and lower bounds of 95% confidence intervals, respectively. WSR (two-tailed) tests: \*  $p < 0.05$  +  $p < 0.07$

In contrast, the children below the cohort median did not differ significantly from chance levels in proportion correct for three of the sounds (CS2-PS2: WSR = -15.5,  $p = 0.18$ ; CS2-PS4: WSR = -16,  $p = 0.23$ ; CS4-PS2: WSR = -12,  $p = 0.4$ ), and were significantly below chance for one sound (CS4-PS4: WSR = -33,  $p = 0.001$ ). A mixed measures ANOVA showed a strong main effect of the median split grouping on 4AFC proportion correct, as expected [ $F(1, 21) = 15.15$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.419$ ], with no significant main effect of sound pair nor any interaction [both  $F(2.13, 44.75) < 0.8$ ,  $p > 0.4$ ].

In addition to 4AFC proportion correct data, we also explored agreement (and confusion) in sound pair to alien mapping at 4AFC. Confusion matrices presenting count frequencies for all children above and below the cohort 4AFC median are displayed in table 6.1. Table 6.2 presents confusion matrix descriptive statistics (Kappa, and proportions for accuracy, true positive, false positive, true negative, false negative and precision) over all possible cue sound-response pairs for those above and below the cohort 4AFC median.

Indeed, Kappa, accuracy and precision across all sound pairs were considerably higher for those above the median split compared to those below (see table 6.2; compare rows 3 & 11, rows 4 & 13 and rows 8 & 16). Similar to the adults above the median split in experiment 5a, the children above the median split also showed highest Kappa, accuracies and precision, and lowest error metrics (false positives and false negatives) when considering cue sound-response pairs where both the CS and PS differed (see table 6.2, column 4 and 5); children below the median split showed modest Kappa (0.352) for CS2-PS2 vs. CS4-PS4, but negative Kappa for CS2-PS4 vs. CS4-PS2 (-0.15).

**Table 6.1:** Confusion matrices for children above and below 4AFC median

		Sound presented				Row total
		CS2-PS2	CS2-PS4	CS4-PS2	CS4-PS4	
Response	Above					
	CS2-PS2	<b>40</b>	19	23	15	97
	CS2-PS4	23	<b>44</b>	7	22	96
	CS4-PS2	24	11	<b>46</b>	15	96
	CS4-PS4	12	16	24	<b>41</b>	93
	Col. total	99	90	100	93	382

		Sound presented				Row total
		CS2-PS2	CS2-PS4	CS4-PS2	CS4-PS4	
Response	Below					
	CS2-PS2	<b>20</b>	16	36	32	104
	CS2-PS4	27	<b>20</b>	29	29	105
	CS4-PS2	26	29	<b>23</b>	27	105
	CS4-PS4	27	39	24	<b>12</b>	102
	Col. total	100	104	112	100	416

As expected, this indicates that children above the median split more commonly confused across one – but not both – of the sound types (and respective spatial locations), whereas children below the median split performed in a less consistent fashion.

We also hypothesised that children above and below the median split would show greater confusion across the less salient of the two sound classes (i.e., the CS stimuli). Thus, we expected children to select – and confuse – aliens that shared the same PS but not the same CS (CS2-PS2 vs. CS4-PS2; CS2-PS4 & CS4-PS4; see table 6.2, columns 3 & 6, respectively), to a relatively greater extent than the reverse (i.e., CS2-PS2 vs. CS2-PS4; CS4-PS2 & CS4-PS4; see table 6.2, columns 2 & 7, respectively). However, we found limited

**Table 6.2:** Kappa coefficients and proportion data for accuracy, true positive, false positive, true negative, false negative and precision data for children above and below the cohort 4AFC median

Above						
Positive	CS2-PS2	CS2-PS2	CS2-PS2	CS2-PS4	CS2-PS4	CS4-PS2
Negative	vs. CS2-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS4
Kappa ( $\kappa$ )	0.333	0.291	0.500	0.667	0.383	0.383
Accuracy	0.667	0.647	0.750	0.833	0.691	0.690
True Pos.	0.678	0.635	0.727	0.863	0.667	0.754
False Pos.	0.322	0.365	0.273	0.137	0.333	0.246
True Neg.	0.657	0.657	0.774	0.807	0.719	0.631
False Neg.	0.343	0.343	0.226	0.193	0.281	0.369
Precision	0.635	0.625	0.769	0.800	0.733	0.657
Below						
Positive	CS2-PS2	CS2-PS2	CS2-PS2	CS2-PS4	CS2-PS4	CS4-PS2
Negative	vs. CS2-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS4
Kappa ( $\kappa$ )	-0.018	-0.171	0.352	-0.150	-0.355	-0.204
Accuracy	0.482	0.410	0.352	0.426	0.320	0.407
True Pos.	0.556	0.357	0.385	0.408	0.408	0.460
False Pos.	0.444	0.643	0.615	0.592	0.592	0.540
True Neg.	0.426	0.469	0.308	0.442	0.235	0.333
False Neg.	0.574	0.531	0.692	0.558	0.765	0.667
Precision	0.426	0.435	0.426	0.408	0.339	0.489

support for this hypothesis. For instance, for children above the median, Kappa was modest (0.383) for CS2-PS4 vs. CS4-PS4 (different CS); however, we also found modest Kappa (0.333) when considering CS2-PS2 vs. CS2-PS4 (different PS), and slightly lower Kappa (0.291) for CS2-PS2 vs. CS4-PS2 (different CS).

Contrary to our hypothesis, this suggests that children above the median split did not necessarily confuse (i.e., show reduced agreement) across CS stimuli to a greater extent than across PS stimuli.

For children below the median split, we found limited evidence of systematic confusion (i.e., disagreement) across CS relative to PS stimuli; indeed, negative Kappa values emerged for CS2-PS4 vs. CS4-PS4 (-0.355) and CS2-PS2 vs. CS4-PS2 (-0.171) (both different CS); however, Kappa for CS4-PS2 vs. CS4-PS4 was also negative (-0.204) (different PS). In each case, this indicates some consistent confusion or 'disagreement' concerning which sound pair matched with which alien; however, the occurrence of negative Kappa across both CS and PS sounds suggests children below the median did not show systematic confusion of just one sound class.

### **6.3.2 In-task results**

As in experiments 5a and 5b, in-game performance was assessed across each condition using: reaction time to alien onsets; accuracy of responses to alien character onsets; mean and standard deviation of Euclidean distance of response co-ordinate (i.e., location tapped), relative to the centre of each alien. Mixed ANOVAs were used to model within-subjects effects of sound pair and block, together with the between-subject effect of 4AFC median split (above/below). We expected that children above the 4AFC median split would show greater RT cost (with potentially lower accuracies and/or increased spatial variability of responses) compared to the children below the median split, when the in-game sound cues were violated (blocks 8 and 9) relative to the preceding blocks.

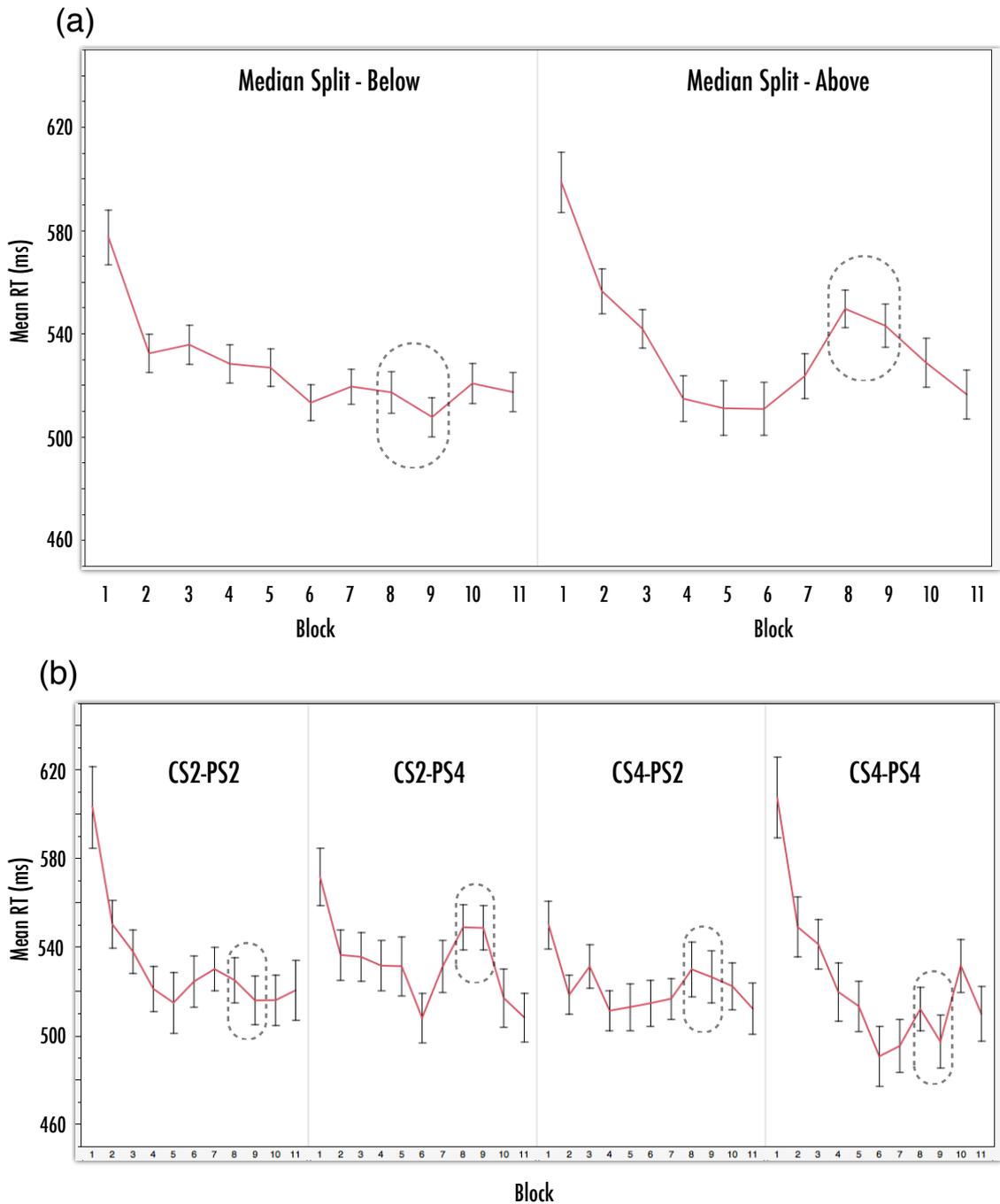
**6.3.2.1. RT to alien onsets.** As in experiment 5a and 5b, the harmonic mean of RTs for each condition was calculated for all subjects. Block 1 was again treated as practice; omnibus analysis of blocks 2 to 11 are reported in supplemental analyses 6.1.

We investigated whether the introduction of violations between the cue sounds and the locations they cued led to increases in harmonic mean RTs, and most importantly, whether the 4AFC median split of the cohort explained differences in any task RT cost effects. As in experiment 5a, for each sound pair, across-block arithmetic averages were calculated for blocks 6 and 7 (i.e., the two pre-cue violation blocks), blocks 8 and 9 (the cue violation blocks), and blocks 10 and 11 (the two post-cue violation blocks) (see Cohen et al., 1990).

A 4 (sound pair) x 3 (block grouping: pre-violation, violation, post-violation) x 2 (4AFC median split) mixed ANOVA showed a significant main effect of sound pair [ $F(2.37, 49.74) = 4.06, p = 0.018, \eta_p^2 = 0.162$ ] that did not interact with block grouping or 4AFC median split. Crucially, the expected interaction of block grouping and 4AFC median split was significant [ $F(2, 42) = 3.42, p = 0.042, \eta_p^2 = 0.14$ ].

Planned ANOVAs for the 4AFC median split showed that for the children above the median, the effect of block grouping was marginally significant [ $F(1.21, 12.05) = 3.67, p = 0.074, \eta_p^2 = 0.269$ ], whereas for the children below the median, there was no significant effect of block grouping [ $F(2, 22) = 0.31, p = 0.74, \eta_p^2 = 0.027$ ]. Figure 6.2a displays mean raw RT data across blocks, by 4AFC median split.

As expected, planned comparisons showed a close to significant ( $p = 0.061$ , two-tailed) increase in harmonic mean RTs averaged across blocks 8 and 9 (cue violation) compared to harmonic mean RTs averaged over blocks



**Figure 6.2:** (a) arithmetic (i.e., non-harmonic) mean RTs across blocks 1-11, for those children below (left side) and above (right side) the cohort 4AFC median. Note the elevation in RTs at blocks 8 and 9 (i.e., cue violation blocks) for those children above the 4AFC median; no increase is noted for those children below the 4AFC median. Elevation at block 1 reflects initial practice on the task (block 1 was not analysed). (b) arithmetic mean RTs across blocks, split by CS-PS pair (4AFC cohorts collapsed). Note the lower RTs for block 7 relative to blocks 2 and 3 for locations cued by CS4-PS4 (see supplemental analyses 6.1). In (a) and (b), contingency violation blocks are highlighted with dashed ovals. Error bars:  $\pm 1$  std. error of mean.

6 and 7, for the children above the 4AFC median. In contrast, the children below the 4AFC median showed no significant difference in harmonic mean RTs averaged over blocks 8 and 9, relative to blocks 6 and 7 ( $p = 0.51$ ; see Figure 6.2a and Table 6.3). This indicates that for children above the cohort median at 4AFC, there was a close to significant cost to RTs associated with violation of learned sound cue combinations; however, for children below the median, no significant RT cost occurred. Notably, the average of harmonic mean RTs over blocks 10 and 11 did not differ significantly from those at blocks 8 and 9 for the children above the 4AFC median ( $p = 0.1$ ), suggesting that RTs associated with learned relationships had not fully returned to pre-violation levels after the cues were re-established at block 10 (see Figure 6.2a, right side). Nevertheless, the general decline in RTs over blocks 10 and 11 agrees with the expected facilitatory effect of re-establishing the learned sound-alien location cues.

**Table 6.3:** Arithmetic means ( $\pm$  SD) of harmonic mean RTs (s) (and pairwise test significances) across block groupings for the cohort 4AFC median split.

	Below Median	Above Median
	<i>Mean (SD)</i>	<i>Mean (SD)</i>
Pre-violation (Blocks 6 & 7)	0.512 (0.05)	0.513 (0.08)
Violation (Blocks 8 & 9)	0.506 (0.05)	0.541 (0.06)
Post-violation (Blocks 10 & 11)	0.513 (0.06)	0.516 (0.08)
	<i>p value</i>	<i>p value</i>
Blocks 6 & 7 vs. Blocks 8 & 9	0.51	0.061 +
Blocks 6 & 7 vs. Blocks 10 & 11	0.95	0.5
Blocks 8 & 9 vs. Blocks 10 & 11	0.37	0.1

As noted above, a significant main effect of sound pair also emerged in the 4 x 3 x 2 ANOVA. Although not hypothesised a priori, post-hoc pairwise tests (FDR-corrected) showed that RTs at locations cued by CS4-PS4 were significantly faster than RTs at locations cued by CS2-PS4 ( $p = 0.01$ ). Inspection of Figure 6.2b suggests that across the full cohort, an increase in RTs occurred over latter blocks (particularly blocks 8 and 9) to aliens at locations cued by CS2-PS4, whereas relatively smaller RT increases were noted over the latter blocks for aliens at locations cued by CS4-PS4 (with exception of block 10).

**6.3.2.2. Response accuracies.** Since the introduction of cue violations led to a significant cost to RTs for children above the cohort 4AFC median, we asked whether the cue violations also led to a reduction of in-task accuracies. Accuracies to alien targets were submitted to a 4 (sound pair) x 3 (block grouping) x 2 (4AFC median split) ANOVA. There was no significant main effect of 4AFC median split [ $F(1, 21) = 2.85, p = 0.11, \eta_p^2 = 0.119$ ], nor any other significant main effects or interactions between factors (all  $F < 1.83, p > 0.1$ ). Accuracy levels across conditions were generally high, with mean accuracies not falling below 92% for any sound pair or block grouping condition across either 4AFC cohort. Importantly, this suggests that the RT cost did not reflect a significant reduction in accuracy during the violation blocks (which might have increased noise in RT measures, due to fewer correct trials to calculate RTs from).

**6.3.2.3. Response spatial distance and variability.** Although accuracy to alien targets was generally high, we also examined whether response spatial precision improved across correct trials, or if the variability of response spatial precision changed as a function of learning or cue violation. As in experiment 5a

and 5b, we analysed (for correct trials) the Euclidean distance (in pixel coordinates) between the centre of the alien character and the location tapped. Omnibus results across blocks 2-11 are presented in supplemental analyses 6.1.

Investigating the effects of cue violation on spatial performance, a 4 (sound pair) x 3 (block grouping) x 2 (4AFC median split) ANOVA showed no significant effects of any of these factors on mean distances between alien centre and tapped location (all  $F < 1.41$ ,  $p > 0.24$ ). A model with the same factors also showed only a weak marginal trend toward a sound pair x block grouping interaction [ $F(6, 126) = 1.91$ ,  $p = 0.084$ ,  $\eta_p^2 = 0.083$ ] for mean of standard deviations of Euclidean distances between alien centre and location tapped. Given the very marginal significance, the interaction was not investigated further. No other main effects or interactions reached significance.

Thus, as with the accuracy results, analysis of spatial performance suggested no significant learning or cue violation related effects within the game task.

### **6.3.3 Cross-task analyses**

As in previous experiments, we sought to examine whether individual differences in sustained auditory attention (and inhibitory control), as well as selective visual attention might predict children's learning performance – both in-game (RT cost) and at post-test (4AFC proportion correct). Sustained auditory attention was indexed using the modified SAART task (RT SDs over all non-target sounds and number of correct target inhibitions). Selective visual attention was indexed using the overall score from the TEA-Ch Sky Search subscale.

Linear regression analyses showed that only one SAART metric reached significance as a predictor of RT cost: number of correct target inhibitions predicted RT cost (blocks 8 & 9 - blocks 6 & 7) collapsed across sound pairs [ $F(1, 21) = 5.1, p = 0.035, \text{adj. } R^2 = 0.157$ ]; however, correcting for multiple comparisons (FDR-correction) across the set of models reduced the effect to non-significance ( $p > 0.2$ ). SAART RT SDs did not significantly predict in-game RT costs (all  $F < 1.0, p > 0.33$ , uncorrected). Moreover, neither SAART metric predicted learning performance at 4AFC collapsed across the four sound pairs (both  $F < 0.9, p > 0.38$ , uncorrected).

Similarly, TEA-Ch Sky Search subscale scores did not significantly account for either in-game RT cost effects or performance at 4AFC, collapsed across the sound pairs (all  $F < 0.8, p > 0.38$ , uncorrected).

Finally, we tested for any differences in attentional performance across the 4AFC median split of the cohort. Wilcoxon signed rank tests showed no significant differences between children above and below the 4AFC median, across any of the SAART measures, the TEA-Ch Sky Search score, or the TOWRE (all  $z < 1.3, p > 0.19$ ).

#### **6.3.4. Interim summary**

Further to the results with adults from experiment 5a, experiment 6 showed that children could also learn combinations of complex cues (on-going contexts and punctate objects), and in turn could map these to visuo-spatial locations. As in experiment 5a however, we again observed clear patterns of individual differences in learning success; the median split of the cohort based on 4AFC data showed near ceiling effects for some children, contrasting with floor effects observed for other children. In line with our hypothesis, we found

that children who showed evidence of cue combination learning at 4AFC also showed performance costs associated with the cue contingency violations during the task. In contrast to adults, these effects were statistically more robust (although two-tailed planned comparisons were marginal). Notably, differences in children's broader attentional skills also did not account for differences in learning success either within the game or at 4AFC (in agreement with results from experiment 5a). In the following section, we compare children's performance to that of adults.

#### **6.4 Adult-child comparisons: experiments 5a & 6**

As a major aim of the present experiments was to explore potential differences in cue combination and learning performance over development, we compared adult (experiment 5a) and child (experiment 6) cohorts across game and 4AFC data. We aimed to explore possible developmental effects that might account for any differences in learning outcomes. We therefore compared adults and children who were above their respective cohort 4AFC medians to each other; we also compared adults and children that were below their cohort 4AFC medians. Because previous analyses of accuracy and spatial performance data suggested no evidence of learning effects for either cohort, we did not analyse these data further.

##### **6.4.1 4AFC results**

We first explored whether adults and children who performed above the 4AFC median split of their respective cohorts differed significantly from each other for 4AFC proportion correct scores. A 4 (sound pair) x 2 (adults vs.

children) ANOVA showed no significant main effects nor any significant interactions (all  $F < 1.7$ ,  $p > 0.17$ ). Thus, adults and children above the 4AFC median split of their cohorts did not differ significantly from each other based on their proportion correct at 4AFC. Similarly, we considered whether adults and children below the 4AFC median split of their respective cohorts would differ from one another based on 4AFC proportion correct. Again, a 4 (sound pair) x 2 (adults vs. child) ANOVA showed no significant main effects nor any significant interactions (all  $F < 2.0$ ,  $p > 0.15$ ).

Thus, we did not find evidence to suggest that adults and children who were more effective at learning the cue combinations (as indexed by 4AFC proportion correct) showed any significant differences in their performance accuracy. Neither did we see any differences in proportion correct between adults and children that showed less effective learning at 4AFC (as expected).

We further compared relative confusion across cue sound/response pairings for adults and children above and below the 4AFC medians (Kappa coefficients displayed in table 6.4). Comparison of adults and children above their cohort 4AFC medians (table 6.4, row 3 & 4) indicated relatively similar metrics of 'moderate agreement', suggesting (in line with 4AFC proportion correct data) similar extents of correct classification of aliens with respect to cue sound pairs. In contrast, adults and children below their respective 4AFC medians had notably different profiles; those adults below their 4AFC median split showed a marked degree of 'disagreement' when considering CS-PS response combinations where PS varied but CS was constant (thus signifying confusion across the more salient acoustic cue; see table 6.4, row 7).

**Table 6.4:** Kappa coefficients for adults and children above and below their respective cohort 4AFC medians

Above						
Positive	CS2-PS2	CS2-PS2	CS2-PS2	CS2-PS4	CS2-PS4	CS4-PS2
Negative	vs. CS2-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS4
Adults	0.384	0.384	0.516	0.490	0.300	0.327
Children	0.333	0.291	0.500	0.667	0.383	0.383
Below						
Positive	CS2-PS2	CS2-PS2	CS2-PS2	CS2-PS4	CS2-PS4	CS4-PS2
Negative	vs. CS2-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS2	vs. CS4-PS4	vs. CS4-PS4
Adults	-0.556	-0.067	-0.280	-0.407	-0.020	-0.574
Children	-0.018	-0.171	0.352	-0.150	-0.355	-0.204

Although children below the 4AFC median showed some evidence of the same (compare adults and children below median, CS4-PS2 vs. CS4-PS4; table 6.4, row 7 & 8), children also displayed confusion across the less salient cue (e.g., CS2-PS4 vs. CS4-PS4; see table 6.4, row 8). Thus, further to the 4AFC proportion correct data, these values suggest that adults and children who performed more poorly at 4AFC did differ: adults appeared to do so based on consistently greater confusion of one class of acoustic cue; however, children appeared to show less selectivity and confused over multiple acoustic cues.

#### 6.4.2 In-task RTs

We first explored whether RT cost effects for those above their cohort 4AFC median differed between adults and children. We therefore modelled pre-violation (blocks 6 & 7), cue violation (blocks 8 & 9) and post-violation (blocks 10 & 11) blocks (see 5.3.2.1 & 6.3.2.1), across adults and children above their cohort 4AFC medians, collapsed over sounds. A 3 (block pair) x 2 (adult vs.

child) ANOVA showed a significant main effect of block grouping [ $F(2, 42) = 4.59, p = 0.016, \eta_p^2 = 0.179$ ], but no other significant main effect nor significant interaction (both  $F < 1.4, p > 0.25$ ). As expected, planned comparisons showed a significant difference in RTs at violation blocks (blocks 8 & 9) compared to pre-violation blocks (blocks 6 & 7) ( $p = 0.023$ ) and at violation blocks compared to post-violation blocks ( $p = 0.026$ ); pre- and post-violation blocks did not differ from each other ( $p = 0.53$ ). As shown in table 6.5, both adults and children above their cohort 4AFC medians showed increased RTs in response to cue combination violations.

Next, for adults and children above their 4AFC medians, we calculated the difference between the mean of blocks 6 and 7 (pre-cue violation) and the mean of blocks 8 and 9 (cue violation) (i.e., blocks 8 & 9 - blocks 6 & 7). However, these difference scores did not differ significantly between cohorts ( $z = 0.77, p = 0.42$ ). Thus, we did not find evidence to suggest significant developmental effects that mediated the extent of performance change (i.e., RT cost) associated with violating learned cue combinations.

Table 6.5: Arithmetic means ( $\pm$  SD) of harmonic mean RTs (s) across block groupings for adults and children above cohort 4AFC median split.

	Adults ( $n = 12$ )	Children ( $n = 11$ )
	<i>Mean</i>	<i>Mean</i>
Pre-violation (Blocks 6 & 7)	0.488 (0.07)	0.513 (0.08)
Violation (Blocks 8 & 9)	0.503 (0.05)	0.541 (0.06)
Post-violation (Blocks 10 & 11)	0.493 (0.06)	0.516 (0.08)

### 6.4.3 Interim summary

As in the analyses for experiments 5a and 6, we found a similar profile of learning effects based on 4AFC and in-task RT data. We found that adults and children above their cohort 4AFC medians did not differ from each other with respect to proportion correct identification at 4AFC. Similarly, adults and children below their cohort 4AFC medians also did not differ from each other in terms of 4AFC accuracy. Notably, violation of cue combinations led to an RT cost only for adults and children who were above their cohort 4AFC medians. The lack of interaction of this effect with adult/child status suggested relatively little difference in the extent of the cue violation cost that could be explained by developmental status. Similar to our previous analyses, we also observed that adults and children above and below their cohort 4AFC medians showed a relatively greater difference in RTs between early and latter learning blocks for CS4-PS4 than for other cue sound pairs (see supplemental analyses 6.2).

## 6.5 General Discussion

The aim of the present studies was to explore learning of novel auditory cue combinations in the context of an active visuo-spatial task. Firstly, we explored whether acoustic cues comprising an on-going context together with a relatively more salient auditory object, could be learned as a combination; that is, whether listeners might learn that the combination of both sounds afforded the most robust cue within the visuo-spatial task (versus either cue alone). We therefore explored performance improvements and learning outcomes associated with game play; we also indexed performance decrements (i.e., related cue cost) associated with changes to the cue contingencies. We further examined whether the relative salience of the cues might also play a role in influencing performance decrements (by violating each cue type separately in experiment 5b).

Secondly, and most importantly, we sought to account for developmental differences in the ability to map combinations of auditory cues to visual events. Given that children continue to develop the ability to combine cues within and across modalities even up to early adolescence, we expected that children might be less effective in learning cue combinations than adults; we also expected that if children did show evidence of learning, they might also demonstrate consistent confusion over the less salient acoustic contexts.

Finally, we aimed to explore possible contributions of broader attentional abilities toward task performance and learning. We reasoned that if general attentional mechanisms are involved in learning, then individual differences in auditory and/or visual attention might account for variance in learning outcomes.

**6.5.1. Cue combination and learning.** The present studies show that auditory cues that differ in both their temporal properties and acoustical saliency

can be combined and learned over relatively short timescales. Following no more than 120 in-game trials (excluding practice), we found evidence that participants could learn to map a combination of cues to an alien character at a visuo-spatial location. Notably, participants could achieve this where one cue was an extended temporal context and the other a temporally shorter yet relatively more salient auditory object (see Kayser et al., 2005; Griffiths & Warren, 2004). Such complexity of cue combination suggests that further to naturalistic auditory scene analysis paradigms (e.g., Leech et al., 2009a; Gygi & Shafiro, 2011), participants can combine and learn novel auditory objects and contexts, in a way that may have helped them to form expectations about impending visual events (see Niessen, 2008).

Notably, these mappings were achieved where the acoustical parameters were arbitrary with respect to the events they cued; that is, the present sounds had no physical relationship to the visual events (see also Kawahara, 2007). This suggests that associative relationships between complex, novel auditory cues and relevant events may be built even in the absence of any iconicity vis à vis the cue and event (further to Ho & Spence, 2008; Keller & Stevens, 2004). Listeners may have used broader semantic or metaphorical relationships in learning the mappings (see Keller & Stevens, 2004); for instance, during piloting, listeners reported the sounds to be 'alien-like' or from 'outer space'. Nevertheless, this would afford little direct benefit to learning the exact mapping of a given combination of sounds onto any particular location or alien.

While patterns of learning were apparent in all experiments, it is important to emphasise the clear differences in success of cue combination learning at 4AFC. At 4AFC, we did not find evidence of significant differences from chance levels in mapping sound pairs to aliens when testing across entire

cohorts, and only found evidence of learning when we split each cohort above and below the 4AFC medians.

This indicates strong individual differences in the ability to map particular auditory cue combinations to spatial locations. Notably, these individual differences appeared to be preserved even in adult cohorts. Similar individual differences have previously been shown in auditory learning of repeated noise segments in adults: while around one third of listeners tended toward ceiling accuracy in detection of repetition of noise segments, two thirds showed accuracy levels at approximately 50% (Agus et al., 2010). Despite the highly limited cue set here (just two context sounds and two punctate sounds), and knowledge that the sounds would be useful to visuo-spatial task performance, about half of the participants in each experiment showed relatively poorer mapping of the sound combinations to a given location. Why might this be?

One possibility is that participants found the acoustical properties of one or both cue types difficult to encode and/or retrieve (e.g., McDermott et al., 2013). The relative extent to which explicit versus implicit retrieval mechanisms are involved during post-task 4AFC performance is unclear (see Reber, 1993, for discussion). However, our data did allow us to probe whether a particular type of cue (i.e., CS or PS) was matched in a manner that showed consistent agreement with the combinations presented in the game, or alternatively, matched in a way that suggested confusion (or disagreement) with respect to the in-game cue combinations.

Indeed, we found that adults who performed below chance at 4AFC in experiment 5a tended to show relatively higher levels of confusion across the more salient PS stimuli. This was an unexpected result, given the relative distinctiveness of the features imposed by the frequency sweeps and AM rates

for each PS, and suggestions that such acoustic features are encoded by modulation-selective cells within the auditory system (see Cusack & Carlyon, 2003; Husain et al., 2004; Shamma, 2001). Indeed, based on evidence of relatively selective cortical processing mechanisms involved in perception of sounds deemed as ‘object-like’ (compared to ‘scene-like’; Lewis et al., 2012), it is perhaps counterintuitive that the adults in our study would tend to misidentify the PS stimuli more commonly.

The data may indicate that these participants attempted to focus on a single sound cue. For instance, these participants might have tried to learn the less salient (and typically more difficult) contextual sound types, yet in so doing, failed to combine them with the correct PS. We should note however that adults below their cohort 4AFC median in experiment 5b did not show this same pattern of results, and as expected, confused across the less salient context sounds to a greater extent (see 6.5.3, below).

Alternatively, it may be that the participants in experiment 5a tended to group the PS stimuli into a type of pseudo-category that was distinct from the CS stimuli (although such a ‘category’ would be highly limited in its scope, given that the stimulus sets comprised only two exemplars of each class). Previous studies of complex category learning for novel noise stimuli have shown that training listeners to identify exemplars within a particular category yields significantly reduced sensitivity ( $d'$ ) to the specific CFs of within-category exemplars; in contrast, when listeners are trained to discriminate between within-category exemplar CFs, sensitivity is improved (Guenther et al., 1999). While our paradigm did not directly manipulate the specific learning strategy listeners applied (learning was unsupervised), it may be that listeners below the median split attempted to identify the particular PS stimuli as belonging to an

object-like 'category'. In so doing, these listeners may have failed to individuate the PS stimuli, thus leading to unexpectedly high confusion at 4AFC. This is a largely speculative account however and will require testing in future experiments using true categories of stimuli.

One possible alternative is that some participants did not rely on the auditory cues and simply performed the visual task instead. Adopting a purely visual task strategy was sufficient to complete the game: those below the median split were no slower or any less accurate than those above, across any experiment. Moreover, participants below the 4AFC median split showed virtually no cost to RTs due to the sound contingency switches (see figures 5.5a and 6.2a). Indeed, Bayesian models of cross-modal spatial performance have shown that observers can use a combination of cue weights across the auditory and visual modalities based on the inverse of noise estimates within each modality (weighting the less noisy modality more heavily; Alais & Burr, 2004; Knill & Pouget, 2005). However, observers may also afford proportionally greater weight to the visual modality even when noise is greater in the visual than the auditory modality (Battaglia et al., 2004). Since the visual information in the present studies was relatively constant in its reliability (albeit with some spatial jitter), whereas auditory cue contingencies varied, it is plausible that some observers might have weighted visual information more heavily (further to Battaglia et al., 2004; Knill & Pouget, 2005). Furthermore, models of visual attention predict that high perceptual loads lead to lesser interference from distractors (when cognitive load is also low; Lavie et al., 2004; Lavie, 2005; see also Forster & Lavie, 2011). One possibility therefore is that combinations of complex auditory cues occurring within a visual task might present a high perceptual load, serving to guide attention toward the visual stimuli only.

However, this remains an open question and will be explored in future experiments.

Previous studies with a similar paradigm to ours required the use of auditory cues in order to facilitate in-task progress (i.e., to anticipate visuo-spatial events; Wade & Holt, 2005; Lim & Holt, 2011). However, this was not feasible in the present paradigm since we intended to compare adults and children on the same task (imposing limitations on task duration and difficulty). Future versions of the present paradigm with adults could extend the incentive to learn and use the auditory cues by increasing task difficulty (e.g., shortening alien on-screen time via an adaptive procedure) thereby enhancing the value of the sound combinations as necessary and useful cues.

**6.5.2. Cue combination, competition and development.** A major aim of the present experiments was to explore developmental differences in the ability to combine auditory cues and apply those cues to a subsequent visual task. Indeed, such a question entails both combination of available cues within a single modality and mapping of cues cross-modally. While adults can achieve such mappings with relative ease (e.g., Kawahara, 2007), school-aged children often do not combine over multiple features even within a single modality (Nardini et al., 2012), and may fail to combine available cues cross-modally at all (Nardini et al., 2008).

An issue central to the use of auditory cues was whether children would combine auditory cue information across both contextual sounds and punctate objects. Indeed, as with our adult data, we found that a subset of children showed a high degree of success in mapping the combination of both cues to relevant locations/alien characters at 4AFC. Interestingly however, we found limited support for our hypothesis that children would tend to confuse the more

salient of these sound classes less consistently. In fact, our data suggest that children who were less successful in mapping the sound combinations to specific locations tended to make responses that suggested confusion across the more salient sound cues as well as the less salient cues. Moreover, our data also indicate that children who showed success in learning the cue combination mappings displayed RT costs when the auditory cue contingencies were violated. Further to data showing that children can use naturalistic auditory information at contextual levels to form expectancies about sound object occurrence (Krishnan et al., 2013), the present study shows that children can learn complex combinations of entirely novel context and object sound classes that arise in a visuo-spatial task, and can map these learned sound combinations to specific locations/characters.

As outlined in chapter 4 (4.1.1), the Competition Model predicts that cues showing high validity (the product of cue availability and reliability; see Bates & MacWhinney, 1987) will be more likely to be weighted as strong cues. Previous applications of the model's predictions to language learning have shown that cue validity accounts for Hungarian children's reliance on accusative case marking in perception of spoken sentences (owing to its high validity compared to cues such as word order in Hungarian; see MacWhinney et al., 1985).

Indeed, the present data suggest that combinations of novel, non-linguistic auditory cues that display relatively high validity may be combined and weighted by child learners. Moreover, in line with the Competition Model's predictions of cue cost where a cue is less reliable (Bates & MacWhinney, 1987), we found that children above the 4AFC median showed an increase in RTs when auditory cue contingencies were violated. Taken together, these results support the Competition Model's predictions within a general framework

that can account for children's cue use and performance both in linguistic and non-linguistic auditory domains (MacWhinney et al., 1985; Bates & MacWhinney, 1987).

Indeed, our results show that similar to adults, by 8-9 years, some children show reliable learning of auditory cues, together with application of those cues to a visual task. These results hold implications for studies of cross-modal cue combination. Knowland et al. (2013) found that across later childhood and early adolescence, the amplitude differences between auditory-only and audio-visual speech perception conditions for early ERP (N1 & P2) components showed a positive increase with age. If (as suggested by Knowland et al.) such findings indicate a developmental increase in the competition between sensory inputs, then it is possible that children may differentially weight information within the auditory and visual modalities across development (see also Burr & Alais, 2004). However, related studies have suggested that such effects appear specific to audio-visual speech perception; where illusory percepts are evoked by non-speech audio-visual cues (e.g., two auditory tones presented with a single visual circle flash, leading to the percept of two visual flashes; Shams et al., 2000), data indicate no developmental changes in the magnitude of the percept (Tremblay et al., 2007).

While our study differs in that we used complex auditory contexts and salient auditory objects as cues to later visual events, the above issues are relevant with respect to development of cross-modal cue combination. As with our adult data, approximately half of our child cohort showed no evidence of significant learning at 4AFC; moreover, those children also showed no evidence of any in-task cue cost (i.e., RT increase) when auditory cue contingencies were violated. This may suggest that some children afforded little weight to the

auditory cues (see Battaglia et al., 2004), perhaps due to a difficulty in combining the auditory and visual information, or a lack of focus on the auditory cues in general.

One question that arises from such evidence is whether these effects reflect stable individual differences (e.g., Bates et al., 1995); that is, do children who tend not to learn auditory cues within an on-going visual task also show a lack of such cue learning into adulthood? Indeed, the narrow age range of our child cohort did not allow us to fully probe age-effects on learning (we found no significant age differences between those above and below the 4AFC median split). If (as with the flash illusion described above) some aspects of non-speech cross-modal cue perception are developmentally stable, then it is possible that poorer cross-modal learning may persist into adulthood. The present study cannot address this issue; however, future longitudinal work with the same child cohort could offer insight into these questions.

**6.5.3. Auditory cues and learning.** As outlined above, an important consideration with respect to cue combination was the extent to which listeners would weight the combined cue arising from both sound types, rather than relying on a single cue alone. Central to this was the nature of the more salient punctate objects, and the likely greater capture of attention associated with their occurrence compared to the context sounds. Based on the predictions of the model of Kayser et al. (2005) (see also Kalinli & Narayanan, 2007), we designed punctate stimuli that varied across many acoustical parameters (i.e., frequency swept, with varying AM rates, and scaled to +6 dB SNR compared to contexts), yielding relatively distinct and salient cues when compared to the spectrally invariant context stimuli. In experiment 5b, we predicted that when

violating each cue type in turn, we should find a relatively greater cue cost associated with violation of the more salient punctate cues.

Although our analyses showed marginal significance, we did find a trend toward increased in-task RTs when punctate cues (but not contextual cues) were violated; however, when contextual cues (but not punctate cues) were violated, we saw little if any increase in RTs. Most notably, we found that the punctate cue violation yielded an increase in RTs for those above and below their cohort 4AFC median. This was an unexpected finding, given that profiles of RT costs were noted only for those above the median split in experiments 5a and 6 (the RT cost noted for adults in experiment 5a was statistically non-significant, versus the close to significant effect found for children in experiment 6, suggesting that adults may have recovered more effectively from the cue violation). However, since the PS (but not CS) cue contingencies were always violated at a late stage in experiment 5b (block 9), it appeared that the RT cost observed for both halves of the cohort might simply have reflected the relatively more reliable cuing by the PS stimuli up to block 9. Probing this possibility in more detail, our control experiment (see supplemental analyses 5.2) showed that when the PS (but not CS) cue contingencies were violated earlier in the experiment (block 5), we did not observe any increase in RTs for those below the 4AFC median. As expected, this may suggest that the extent of cost associated with cue violation depends at least in part upon the point in the learning task at which the violations occur. Moreover, our control experiment also showed little if any evidence of increases in RTs for those above the 4AFC median when the contingencies for the CS (but not PS) cues were violated later (block 9) in the experiment; this agrees with the lack of RT cost associated with contingency violations for the less salient CS stimuli in experiment 5b.

Agreeing with the predictions of the Kayser et al. (2005) model, the results from experiment 5b may suggest that punctate, salient cues can influence performance to a greater extent than the on-going acoustic context. The cue cost demonstrated by both effective and less effective cue combination learners in experiment 5b (i.e., those above versus below 4AFC median) indicates that both halves of the cohort showed some sensitivity to the locations cued by the salient sounds. Moreover, the patterns of confusion across the context sounds at 4AFC by those below the median split suggests some evidence of learning of the punctate sounds (although note that this contrasted with the confusion across the PS stimuli we observed for those below the median in experiment 5a; see 6.5.1).

Taken together, these findings hold implications for the kinds of non-semantic and spatially invariant acoustic features that learners can acquire as cues. A spectrally variable sound with dynamic amplitude envelope and greater intensity (compared to an on-going context), may provide a more readily learnable cue than variations in an on-going broadband sound (further to Ho & Spence, 2008; see also Fritz et al., 2003; 2005).

Finally, we also noted an unexpected pattern of learning for one sound combination. In experiments 5a and 6, CS4-PS4 showed significantly greater extents of RT facilitation relative to other sound combinations across initial versus latter learning blocks (although we did not replicate this effect in experiment 5b). It is unclear why this particular sound pair would show a relatively greater extent of RT decrease as learning proceeded; one possibility is that children and adults may have found its particular acoustical attributes easier to encode (or indeed, more salient) than the other cue combinations, making it a more effective cue to use. Notably however, when cue combinations

were violated, we did not find that the increase in RTs for CS4-PS4 was any greater than for other sound pairs (i.e., there was no group x median split x sound interaction, although we did see a main effect for that sound; see experiment 6).

**6.5.4. Attention, learning and generalisation.** A final goal of the present studies was to explore the role of attentional mechanisms in relation to complex cue combination and learning. Effects of auditory stimuli in capturing attention with respect to visual events (in the absence of receptor surface shifts) have been documented extensively (Spence & Driver, 1997; Ho & Spence, 2008; see also Driver & Spence, 1998). Moreover, developmental accounts of attention and learning have posited that sustained attention to sound may be one key mechanism underlying task engagement and related learning success (see Gomes et al., 2000). Yet no study yet had tested whether general auditory or visual attentional performance might underlie complex cross-modal cue combination learning.

We therefore tested whether individual differences in attentional performance (particularly amongst children; further to Gale & Lynn, 1972) would relate to learning outcomes. Those with poorer general attentional performance might have been expected to encounter greater difficulty in attending to sounds, which could limit access to lower-level auditory stimulus features, hindering further cue combination and learning success.

Contrary to predictions, we found no evidence across experiments that individual differences in sustained auditory attention performance (response inhibition accuracy and variability of non-target RTs) showed robust relationships with either 4AFC learning outcomes or in-task RT cost (when correcting for multiple comparisons). Neither did we find that measures of

sustained auditory attention or selective visual attention accounted for any significant variance in children's learning performance. Differences in children's learning performance were also not explained by more general profiles of language development; those above and below the 4AFC median showed no evidence of any differences in TOWRE scores. Previous studies have found that composite metrics of auditory attention can predict children's performance on lower-level auditory tasks (such as backward masking, simultaneous masking, and frequency discrimination; Moore et al., 2010). In contrast, our results suggest that more general attentional skills show limited relation to success of learning auditory cue combinations, and mapping of those combinations cross-modally (further to Miller & Weiss, 1981).

These findings suggest that general attentional metrics account for little variance in learning outcomes for children and adults. This raises the question as to whether task-specific performance might better explain learning success. Indeed, Halliday et al. (2012) showed that children trained on auditory frequency discrimination, visual frequency grating discrimination or phoneme discrimination tasks improved on the particular task they were trained on, with no generalisation to language-related measures. Further, animal models have shown relatively selective adaptations of neuronal spectro-temporal receptive fields, that are specific to the tone frequencies (and adjacent sidebands) encountered during on-task auditory learning (Fritz et al., 2003; 2005).

Taken together, these results suggest that success in learning to combine auditory cues may reflect a qualitatively different set of mechanisms than those indexed by the attentional metrics used here. One possible avenue for future experiments is a measure of divided attention (e.g., Schiff & Knopf, 1985). If the ability to split attention between both auditory cues and visual task

goals is a key mechanism at play in the present study, then related indices of divided cross-modal attention (e.g., the Sky Search Dual Task measure; Manly et al., 2001) may serve as a useful predictor of cue combination learning outcomes.

**6.5.5. Conclusions.** In a busy, complex environment with many competing sources of information, it can be beneficial to combine cues from the auditory modality and use them to inform broader behaviour. The present studies show that auditory cues – both complex contexts and punctate objects – can be combined and learned over short timeframes, and further applied to a related visuo-spatial task. Moreover, such abilities may be present even in school-aged children, further suggesting that the mechanisms enabling cue combination within audition and across modalities can develop by later childhood. As predicted by the Competition Model, violations of cue contingencies (i.e., reductions in cue validity) can lead to performance decrements for children and adults alike (although the extent of these effects may be specific to learning success, and statistically more robust in children).

Nevertheless, we observed clear patterns of differences in cue combination and learning success. While both children and adults can learn abstract auditory cue combinations as they map to spatial events, subsets of children and adults may not learn such mappings. This might arise through systematic confusion of a given auditory cue type, or attentional capture by one cue (limiting the ability to focus on and learn another cue type). Alternatively, subsets of both adults and children may weight their behaviour specifically toward the visual modality, without learning or using the available auditory cues. Whether such trends reflect stable individual differences, or behavioural processes that are malleable (perhaps through development or training)

remains to be addressed. However, present data suggest that general sustained attentiveness to sound accounts for little variance when exploring underlying mechanisms.

In sum, the ability to learn and combine auditory cues may occur within busy multi-modal task environments; however, auditory cue combination, learning and cue mapping are complex processes. Although potentially beneficial to performance, combinations of auditory cues may not necessarily be learned as a result of active task completion. Variation over individuals appears to characterise success of learning outcomes, both for children and adults.

## Supplemental Analyses

### Supplemental analyses 6.1: experiment 6

**RTs.** First, we considered if there was an overall effect of sound pair (each CS-PS pair; 4 levels), block (10 levels; blocks 2-11) or 4AFC median split grouping on harmonic mean RTs. A 4 x 10 x 2 mixed model ANOVA showed a marginal main effect of block on RTs [ $F(3.95, 82.89) = 2.41, p = 0.056, \eta_p^2 = 0.103$ ]; however, pairwise comparisons across blocks revealed only marginal trends toward significant differences after correction for multiple comparisons (FDR; all  $p > 0.09$ ). There were no significant main effects of sound pair or 4AFC median split, nor any significant interactions (all  $F < 1.98, p > 0.1$ ).

A similar mixed model ANOVA treating block over 6 levels (i.e., blocks 2-7, during which contingencies were established) showed no significant main effect or interactions involving median split grouping (all  $F < 1.47, p > 0.23$ ). A significant main effect of block emerged [ $F(2.53, 53.07) = 3.8, p = 0.02, \eta_p^2 = 0.153$ ], as did a sound pair x block interaction [ $F(6.27, 131.6) = 2.16, p = 0.048, \eta_p^2 = 0.093$ ]. Post-hoc tests (FDR-corrected) on the block main effect showed marginal differences in RTs between blocks 2 and 4, and between blocks 2 and 7 (facilitated RTs at blocks 4 and 7 versus block 2; both  $p = 0.06$ ); marginal differences were also noted between blocks 3 and 4, blocks 3 and 6, and blocks 3 and 7 (facilitated RTs at blocks 4, 6 and 7 versus block 3 in each case; all  $p = 0.06$ ). The sound pair x block interaction was isolated to significantly longer RTs at block 2 and at block 3 when compared to blocks 4, 5, 6 and 7 (all  $p < 0.05$ ), for CS4-PS4 (all tests FDR-corrected) (there were no significant differences between blocks 2 and 3, nor between blocks 4, 5, 6 or 7; all  $p > 0.1$ ). This indicates the decrease in RTs to aliens cued by CS4-PS4 between the initial

and latter learning blocks was significantly greater than for aliens cued by other sound pairs (see Figure 5.9b and discussion).

**Spatial distance.** A 4 (sound pair) x 10 (block; 2-11) x 2 (median split grouping) omnibus ANOVA model showed no significant effects of any of these factors on mean or SD of Euclidean distance between response co-ordinate and alien centre (all  $F < 1.97$ ,  $p > 0.12$ ).

### Supplemental analyses 6.2: adult-child comparisons

**RTs.** We further investigated whether profiles of RT performance during the earlier stages of the task (i.e., where cue combinations were established) differed between adults and children, both for those above and below their respective cohort 4AFC medians.

For children and adults above their cohort 4AFC medians, we performed a 4 (sound pair) x 6 (block: 2-7) x 2 (adult vs. child) ANOVA. The model showed a highly significant main effect of block [ $F(2.57, 54.02) = 5.92$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.220$ ] and sound x block interaction [ $F(6.42, 134.8) = 3.02$ ,  $p = 0.007$ ,  $\eta_p^2 = 0.126$ ], but no significant main effect of adult vs. child nor any significant interactions with this factor (all  $F < 1.0$ ,  $p > 0.34$ ). For the main effect of block, pairwise comparisons (FDR-corrected) indicated marginally significant differences between both block 3 and block 7 compared to block 2 (both  $p = 0.06$ ); blocks 4, 5, and 6 each differed from block 2 (all  $p = 0.05$ ), whilst each was also marginally different from block 3 (all  $p = 0.06$ ). As in previous analyses, the sound x block interaction was isolated to CS4-PS4 [effect of block:  $F(2.58, 56.66) = 8.77$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.285$ ]; post-hoc pairwise comparisons (FDR-corrected) showed significant differences between block 2 and each of blocks 4, 5, 6 and 7 (all  $p < 0.05$ ) (with no difference between block

2 and block 3;  $p > 0.1$ ); there were significant differences between block 3 and each of blocks 4, 5, 6 and 7 (all  $p < 0.05$ ); no other comparisons reached significance.

For adults and children below their cohort 4AFC medians, we performed the same analysis. The 4 (sound pair) x 6 (block: 2-7) x 2 (adult vs. child) ANOVA also showed a significant sound x block interaction [ $F(15, 330) = 2.92$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.117$ ] and very weak trend toward a block main effect [ $F(3.03, 66.7) = 2.23$ ,  $p = 0.092$ ,  $\eta_p^2 = 0.092$ ]; again, adult vs. child was neither significant as a main effect nor in any interactions (all  $F < 2.4$ ,  $p > 0.13$ ). The significant sound x block interaction was again isolated to CS4-PS4 [effect of block:  $F(3.15, 72.41) = 5.60$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.196$ ]. Post-hoc pairwise comparisons (FDR-corrected) showed a significant difference between block 2 and block 7 ( $p = 0.03$ ); there was a significant difference between block 2 and block 6 ( $p = 0.05$ ), and for block 6 and block 7 when each was compared to block 3 (both  $p = 0.05$ ).

## Chapter 7: General Discussion

## 7.1 General Discussion

In this thesis we have investigated experience, development, learning and plasticity within audition. Targeting both long- and short-term timeframes, we have explored how experience across a lifetime or a laboratory session can influence auditory perception and cognition. Further, we have investigated how long-term expertise relates to adaptations to cortical regions involved in expert performance. Together, these experiments have helped to guide our understanding of auditory plasticity. We briefly review the aims and findings of the experiments conducted.

In chapter 2, we explored fine-grained and contextually-mediated auditory perception in two cohorts of musicians trained on instruments with very different acoustic constraints (i.e., violin vs. piano). We asked whether musicians' expertise would yield instrument-specific perceptual advantages (i.e., comparing musician cohorts to each other), but also considered perceptual advantages of musicianship more generally (i.e., comparing musicians to a closely matched cohort of non-musicians). Moreover, we investigated whether musical expertise transferred to cognitive tasks that present demands analogous to the area of expertise; in particular, we explored sustained auditory attention, auditory scene analysis and multi-modal sequencing tasks. We further asked whether any advantages on higher cognitive tasks (such as attentional metrics) might account for differences in perceptual skill, both for experts and non-experts. Our results showed that expert musicians perceived low-level sound features more finely than non-experts, but we found no evidence of instrument-specific advantages in this regard. Instrument-specific perceptual advantages did emerge however within a contextually-relevant task (musical tuning perception). We found little evidence to suggest musical expertise

transferred to non-musical cognitive skills. Nevertheless, our results showed that attentional performance could account for fine perception of certain auditory temporal envelope cues (AM depth), to a greater extent than could expertise.

In chapter 3, we expanded these results and asked whether one of our expert cohorts (violinists) would differ from the non-expert cohort with respect to indices of cortical structure. In particular, we completed the first investigation of cortical myelin proxies ( $R_1$ ) in experts musicians and non-experts, with a focus on auditory and motor cortical regions. We found that experts had increased  $R_1$  at a medial region of left Heschl's gyrus corresponding with auditory core; however, these increases were subtle. To our surprise, we did not support a hypothesised reduction in motor cortical  $R_1$  asymmetry at hand area (despite violinists' intensive training of the non-dominant hand). We nevertheless found some support for training-related plasticity: greater years of training in violinists were associated with higher  $R_1$  at right superior temporal gyrus (an area well outside of primary auditory regions).

In chapters 4, 5 and 6, we investigated learning over short timescales, and examined plasticity with respect to building novel auditory representations. Using an active, multi-modal task, we found that some listeners could learn abstract sound cues that invariantly predicted the location of an upcoming visual target, even without prior knowledge of the utility of these cues. Following this, we showed that both adults and children could learn combinations of sounds that formed a novel auditory 'scene' (comprising contextual and object-like cues). Yet in each experiment, we found clear differences in learning outcomes: while some adults and children could achieve expert-like learning of sound combinations, we also found that others showed little if any learning of the

sound combination to visual location mapping. Moreover, measures of attentional ability in adults and children did not account for inter-individual variation in learning success.

In the following sections, we integrate the above findings with existing literature on issues including (where relevant) plasticity, expertise, development, environment, learning and generalisation. We begin with consideration of long-term expertise.

## **7.2 Long-term expertise: development, environment and plasticity**

Central to the developmental perspective within this thesis is the exploration of plasticity that occurs through experience accrued over extended periods of time. Indeed, assignment of individuals to protracted periods of instruction is costly and presents extensive logistical challenges (although it is possible; see Schellenberg, 2004; Moreno et al., 2008; 2011; Norton et al., 2005; Hyde et al., 2009). Nevertheless, one solution is the study of expert groups: individuals who train over periods of years, often within well-defined domains. Indeed, study of a single expert group can yield self-selection confounds (e.g., pre-dispositions ahead of training that confer an advantage leading to pursuit of a given field; Corrigan et al., 2013). Thus, an elegant solution is to compare experts trained within the same domain yet with clear differences in the nature and demands of their training. By comparing such experts, many of the developmental considerations of motivation to pursue and persist with training (as well as constraints on resources and effort) can be mediated (see Ericsson et al., 1993). This allows for more controlled investigation – and critically, testable predictions – concerning the particular perceptual and cognitive outcomes we might expect in expert groups.

In chapter 2, we aimed to use this design with respect to expert musicianship. Drawing on expert violinists and pianists (closely matched for years of training), we explored whether the different perceptual and cognitive demands of their training might relate to differences in expert outcomes (further to Strait et al., 2010; 2012a). To our surprise, we did not find support for instrument-specific perceptual advantages for low-level acoustic features: violinists and pianists perceived differences in onset rise time, AM depth and FM depth cues more finely than non-musicians, but did not differ significantly from each other. However, we did find that perception of very subtle tuning differences yielded an expertise-driven distinction between both cohorts. While violinists tended to make in-tune judgements that adhered closely to just intonation (the tuning system typical to violin), pianists' judgements only tended to adhere to equal temperament (the tuning system used for pianos) when that system was paired with a tuning deviance (see Loosen, 1994; 1995). Moreover, pianists' tuning ratings in this instance were closely related to their FM depth thresholds, whereas we did not find any such relationship for violinists. Further, we found a trend for violinists who began training early in life (aged 3-4 years) to rate just intonation as more in-tune than a very fine tuning discrepancy (less than 10% of a semitone), when compared to peers who began violin training at later ages.

Taken together, these results suggest that both environmental and developmental factors may interact and spur plasticity of auditory perceptual representations. For two expert cohorts, we show clear distinctions in perceptual performance within a musical context, likely reflecting the differential demands of their respective training. Moreover, we argue that within a given expert violinist cohort where training environments were more similar, early

development may play a role in mediating adaptations to very fine perceptual skill, but only in musical contexts. Such results perhaps suggest that certain environmental and development effects act within relatively selective confines, and yield perceptual outcomes that are specific to the expert's realm.

### **7.3 Long-term expertise: specificity versus generality of transfer**

One of the major implications of the results from chapter 2 is the nature of transfer of expertise with respect to perceptual and cognitive skills.

**7.3.1. Perceptual transfer.** As discussed above (7.2), we found support for improved fine perceptual discrimination of low-level auditory features in expert musicians, compared to non-experts. These results agree with the predictions of reverse hierarchy theory (RHT) as proposed by Ahissar and colleagues (Ahissar et al., 2009; see also Hochstein & Ahissar, 2002).

Briefly, RHT posits that perceptual discrimination will initially be dominated by higher-level perceptual skills; that is, based on a relatively superficial or surface-level of discrimination. However, as learning proceeds, access to low-level and finely detailed perceptual acuity improves; as a result, these low-level representations may then begin to feedforward to higher levels, and further influence many related higher level perceptual representations (Ahissar et al., 2009). Indeed, this framework mirrors some of the hallmarks of expertise as discussed in chapter 1 (1.3.1): experts can access finely-detailed subordinate levels of representation with relative ease (contrasting with non-experts' focus at the basic or superordinate level) (Tanaka & Taylor, 1991; Tanaka et al., 2005; Palmeri et al., 2004).

Thus, our expert violinists and pianists could very finely discriminate acoustic features like AM depth, FM depth and onset rise time (presented with a

non-musical timbre, via a paradigm they had not encountered before) and to a more fine-grained extent than our non-musicians. This suggests that musical expertise affords access to representation of low-levels of acoustic detail, in line with RHT. However, it is important to acknowledge the rapid pace of perceptual learning that occurred within the tasks, even for non-musicians. In particular, onset rise time and AM depth cues were learned by all cohorts; moreover, we found non-musicians' last runs for each measure did not differ significantly from musicians' first runs. Such a result also agrees with the predictions of RHT: as learning of a stimulus class proceeds, access to fine levels of detail improves, leading to further reductions in perceptual thresholds (Ahissar et al., 2009).

Nevertheless, we also found some support for the role of more general top-down attentional mechanisms with respect to perceptual acuity. We found that variability of sustained attention performance (i.e., RT SDs) could account for more variance in AM depth thresholds than expert musician status (somewhat surprisingly). Indeed, further to the rapid learning for AM depth stimuli discussed above, these results suggest that individual differences in attentional performance may partly explain fine perceptual acuity for temporal envelope (further to Strait et al., 2010; 2012b; Tervaniemi et al., 2009). Although the predictions of RHT fit with our perceptual findings, RHT proposes feedforward and feedback mechanisms between low and higher levels of perceptual representation. Our results suggest that broader executive attentional mechanisms may partly explain fine perceptual acuity, further to the proposals of Ahissar and colleagues.

While Ahissar et al. (2009) also argue that the testing protocol itself plays a role in perceptual learning (see Figure 4, Ahissar et al., 2009), further studies have suggested that rapid perceptual learning may occur within the first few

hundred trials of perceptual experiments, independent of improvements arising simply due to procedural or task-based factors (Hawkey, Amitay & Moore, 2004). Indeed, it is notable that we found most robust learning for low-level temporal (AM depth and onset rise time) parameters, and moreover, that we saw learning for a further temporal property (AM depth) even after participants had become quite familiar with the testing protocol (having completed the onset rise time task first). However, we should acknowledge that learning for FM depth using the same paradigm was reduced somewhat. Thus, it appears that procedural factors alone could not fully explain the present results. Our results might suggest that low-level frequency or pitch cues are more robustly represented in experts, or perhaps are less easily trained over brief periods. Indeed, one previous study showed that several hours of training were needed before non-musicians' thresholds for pitch reached musician levels (Micheyl et al., 2006).

Taken together, our findings largely agree with the predictions of RHT and extend these to expert musician cohorts: regardless of training on a fixed or non-fixed pitch instrument, our experts had fine thresholds for low-level auditory cues. However, further to RHT, we also suggest that top-down attentional mechanisms may account for fine perceptual acuity for certain acoustic cues above effects of expertise.

**7.3.2. Broader transfer.** Our results with respect to generalisation of expertise to cognitive skills present a more complex picture. As outlined in chapter 1 (1.5), the conditions under which generalisation occurs may reflect some of the facets of the training environment and the trained skill itself. For instance, learning on laboratory-based perceptual tasks may yield improvements on the trained dimension, yet relatively limited transfer to

untrained perceptual cues (e.g., Wright et al., 1997; 2010; Halliday et al., 2012; Hervais-Adelman et al., 2011) or untrained higher-level lexical dimensions (Davis et al., 2005). However, training on broader, multi-modal and interactive tasks may yield robust generalisation of learning, both to lower-level perceptual skills and higher-level attentional abilities (e.g., Green & Bavelier, 2003; 2007; 2012; Dye et al., 2009).

Our present results suggest that despite the varied complexity of environments that musicians encounter over prolonged periods of time (practice rooms, chamber ensembles, orchestras, etc.), and the cognitive skills they likely hone within these environments (sustaining attention to output, sequencing, timing and co-ordinating responses, and auditory stream segregation), such skills appear not to transfer to laboratory measures that index some of these abilities (cf. Parbery-Clark et al., 2009a; 2009b; 2011; 2012; Strait et al., 2010; 2012b; see Ruggles et al., 2014). Why might this be? As discussed in chapter 2, the measures we employed reflected stimulus and task properties that both musicians and non-musicians were likely to have been familiar with (e.g., environmental sounds, everyday listening scenarios, multi-modal game-like interfaces). While we expected that the skills these measures indexed might have been more generally enhanced by musicians' training and experience, it is also possible that the familiarity of these contexts across all cohorts masked the potential for manifestation of transfer effects.

**7.3.3. Implications for transfer and plasticity.** This discussion then raises the question as to the environments and constraints that affect expertise transfer. One possibility is that transfer to low-level skills that are similar to the expert domain is likely (given the specificity of expert ability; Chi, 2006). Indeed, previous studies of expert visual skills have suggested that expert radiologists

could transfer their expertise to low-luminance contrast detection for dots presented on X-ray film, even where expert visual search and pattern recognition skills were of no benefit to the task (Sowden et al., 2012). In contrast, novices showed poor transfer of training across low-level stimulus variation (switch of image contrast to or from negative); notably however, novice training that involved real mammograms and entailed visual search and pattern recognition did partly transfer across image contrast switches (Sowden et al., 2000).

In line with our findings and the predictions of RHT, such results suggest that expertise can benefit low-level perceptual discrimination. However, these findings also suggest that some of the demands of the training environment may determine whether transfer occurs more broadly (Sowden et al., 2000). Based on previous studies of learning transfer (Logan, Lively & Pisoni, 1991; Lively et al., 1994; but see also Lim & Holt, 2011), it may be that variable training sets, task demands or multi-modal environments are more likely to yield plastic outcomes that generalise across a variety of domains (Ahissar et al., 2009; Green & Bavelier, 2012). Given these considerations, expertise appears to afford potential for fine acuity for low-level stimuli; however, as discussed above, the potential for far transfer to higher cognitive skills may be limited by the context in which those skills are both trained and tested (e.g., via novel game-like interfaces, using everyday sounds, or with multi-modal cues present; Green & Bavelier, 2012). Moreover, based on the framework of Lövdén and colleagues (see 1.2.3), if there is only partial overlap between the functional capacity of the expert's cognitive skills and the cognitive demands posed by another domain, then adaptations related to expertise may offer only limited benefit to that domain (see Lövdén et al., 2010).

## 7.4 Expertise and cortical plasticity

Related to the question of expertise and general versus specific outcomes is the issue of manifestation of cortical plasticity. Indeed, as discussed in chapter 1 (1.3.2), expertise may be associated with adaptations to cortical or sub-cortical structures that are necessarily involved in performance within the expert domain (e.g., Maguire et al., 2000; Teki et al., 2012; Roberts et al., 2013; Amunts et al., 1997; Gartner et al., 2013; Golestani et al., 2011).

**7.4.1. Cortical  $R_1$  in experts and non-experts.** Major questions arising from investigations of experts concern both the specificity of adaptations at a cortical level, along with the underlying source(s) of the structural adaptations under study (Zatorre et al., 2012). Indeed, previous studies have shown that expertise is associated with volumetric and/or thickness enhancements to specific cortical regions thought to be involved in the expert skill (e.g., Gaser & Schlaug, 2003; Dohn et al., 2013; Bermudez et al., 2009). However, such evidence suggests relatively coarse or macro-level adaptations to structure that may be accounted for by many cellular or tissue properties; for instance, both volumetric and cortical thickness measures can be influenced by synaptic density, glial cell processes, vasculature and myelin (see Zatorre et al., 2012a; Bermudez et al., 2009; Draganski & May, 2008). Hence, studies of experts using quantitative measures of tissue properties may allow for more specific conclusions regarding the mechanisms that underlie cortical plasticity in experts. Moreover, relating quantitative measures of cortical structure to training metrics offers a strong means of testing for the experience-dependent nature of plastic adaptations (see Zatorre et al., 2012a).

In probing these issues, we explored profiles of cortical structure adaptations in expert violinists and closely matched non-musicians, using

quantitative proxies for cortical myelin ( $R_1 = 1/T_1$ ; Dick et al., 2012; Sereno et al., 2012; Sigalovsky et al., 2006; Lutti et al., 2014). Our results revealed profiles of cortical structure differences between violinists and non-musicians that were limited to auditory regions. We found increases in cortical  $R_1$  in violinists versus non-musicians at left auditory core; however, these increases were subtle, focused at mid cortical depths, and were less robust over the full extent of left Heschl's gyrus (cf. Schneider et al., 2002). Notably, we also found that cortical thickness at right Heschl's gyrus in violinists was significantly greater than in non-musicians (see Bermudez et al., 2009; Dohn et al., 2013). Critically, we did not support a predicted increase in cortical  $R_1$  at right hemisphere hand area in violinists versus non-musicians (cf. Amunts et al., 1997; Bangert & Schlaug, 2006). This result was surprising and suggests a key difference in expert outcomes when comparing both auditory and motor regions in violinists (see 7.4.3).

Investigating behavioural metrics with respect to  $R_1$  provided some evidence of experience-dependent relationships. We found that independent of age effects, violinists with greater years of training had higher  $R_1$  at right anterior superior temporal gyrus (STG). Notably, we did not find any evidence of a group difference at this cluster (although we did see a trend toward an age effect in non-musicians, but not in violinists). Conversely, we found that while years of training accounted for variance in  $R_1$  at left Heschl's gyrus in violinists, the effect was not robust when controlling for age. In exploring behavioural relationships, we found that AM depth thresholds could account for variance in  $R_1$  at auditory core ROIs (particularly at right hemisphere); however, these effects only reached significance when considering both cohorts together and

were not robust at whole-brain level. We suggest this relationship should be viewed cautiously and will require further replication.

**7.4.2. Experience-dependent plasticity and expertise.** The above findings hold clear implications for the nature of experience-dependent plasticity with respect to expertise. In particular, we find only partial support for a model of experience-dependent cortical adaptation. Our finding of increased cortical  $R_1$  at right STG in violinists with greater years of training agrees with the predictions of an experience-dependent account (further to Bengtsson et al., 2005; Sluming et al., 2002; see also Seither-Preisler et al., 2014). However, that we find evidence of a cohort difference at left auditory core that is not readily accounted for by training metrics draws some doubt over this model.

In particular, it is important to stress that the correlational design used in the present study did not allow us to determine that cortical  $R_1$  differences are directly a result of violin training. Indeed, as other authors have noted (see Gaser & Schlaug, 2003; Herholz & Zatorre, 2012; Zatorre, 2013), pre-training brain structure enhancements that predispose individuals toward musical expertise cannot be ruled out entirely. Nevertheless, longitudinal studies have found no differences in brain structure when comparing children who intended to pursue music lessons to children who did not, before any training had begun (Norton et al., 2005). Moreover, follow-up with the same cohorts showed that a year of music lessons yielded increases in deformation-based metrics of cortical structure at right Heschl's gyrus and bilateral motor cortex in the musically trained children (Hyde et al., 2009).

Thus, although we cannot establish causation, existing literature suggests that many cortical adaptations following musical training are experience-dependent. Furthermore, although weak, we did observe a trend for

violinists who began training earlier in development (3-4 years) to show higher cortical  $R_1$  at left auditory core than violinists who began training at later ages. Indeed, such a result might suggest a role for early training or ‘sensitive period’ effects; that is, training at early periods of development may influence structural adaptation at auditory core to a greater extent than training at later ages. Indeed, this trend agrees with findings of greater adaptations to subcortical tracts (i.e., higher fractional anisotropy at isthmus of corpus callosum and left temporal lobe) in musicians who began training before age 7, compared to musicians who began training later (Steele et al., 2013; see also Imfeld et al., 2009). We should caution that our early training results were statistically weak (since dividing our cohort into those with and without early training necessarily reduced statistical power). However, future investigations of cortical myelin adaptations in expert musicians would do well to explore effects of age of training onset.

#### **7.4.3. Expertise and cortical adaptations: auditory and motor cortex.**

Contrary to our prediction, we did not find that the non-dominant hand training violinists engage in leads to increases in cortical  $R_1$  measured at the contralateral hemisphere (indeed, we found a hemisphere main effect only; LH > RH). Moreover, the lack of any group differences at motor cortex (either within ROI or whole-brain analyses) suggests a clear difference with auditory cortical areas with respect to expert cortical adaptations.

In chapter 3, we discussed mechanisms of bilateral transfer as one possible explanation for the lack of group  $R_1$  differences across cortical hand areas (see 3.4.1). One further account concerns the fundamental differences in organisation between the auditory and motor systems. While auditory core receives afferents from the medial geniculate of the thalamus, motor cortical

afferents of the cortico-spinal tract arrive from the spinal cord via the pyramidal decussation, pons and cerebral peduncles (Nolte, 2009). One possibility is that differences in myeloarchitecture between musicians and non-musicians may manifest within the cortico-spinal tract itself (Imfeld et al., 2009; Gartner et al., 2013; see also Han et al., 2009). Given that the cerebral peduncles and cortico-spinal tract myelinate relatively earlier in development (Baumann & Pham-Dinh, 2001; Nakagawa et al., 1998; Paus et al., 2001), this suggests training early in life (closer to the age at which myelination finishes) may be a candidate mechanism for structural plasticity within these tracts in musicians. Indeed, a recent diffusion tensor imaging study found that musicians who began training earlier in development had greater mean diffusivity within cortico-spinal tract than musicians who began training later (Imfeld et al., 2009). While such results are outside the scope of the present study, in the future we intend to pursue diffusion tensor methods with our cohort as a means of probing violin expertise effects within the cortico-spinal tract. Our present results suggest that myelin proxies within cortex of expert violinists are increased at left auditory core but not motor hand areas, compared to non-musicians.

### **7.5 Short-term learning, plasticity and expertise**

A core theme explored within this thesis has been the development of plastic adaptations and expertise through online learning. Indeed, following periods of relatively short term-training, individuals can develop detailed perceptual and cognitive representations of novel stimuli (e.g., Wade & Holt, 2005; Leech et al., 2009b; Gauthier & Tarr, 1997; 2002). In some respects, the features of the representations that develop can be regarded as expert-like. For instance, learners develop access to fine-grained subordinate levels of

description (and may access subordinate levels as efficiently as basic levels) (Bukach et al., 2012). Further, learners may also show sensitivity to particular feature configurations (and suffer disruptions to performance when those configural relations are violated) (Gauthier et al., 1998; Gauthier & Tarr, 1997; 2002). Such performance closely mirrors that observed in real-world expert domains (e.g., Gauthier et al., 2003), and suggests that short-term learning and plasticity for novel stimuli can lead to representations that bear certain hallmarks of expertise.

Similar arguments have been made for auditory stimuli encountered in the context of online tasks. Indeed, listeners can develop finely honed representations of complex auditory categories that serve as cues to visual events (Wade & Holt, 2005). Moreover, post-training discrimination performance for these complex categories may correlate with functional adaptations to superior temporal regions typically involved in speech perception (Leech et al., 2009b; this finding is mirrored by imaging studies of novel visual object learning and fusiform face area – see Wong et al., 2009).

**7.5.1. Auditory cue learning and ‘expertise’.** A major goal of the present studies was to explore short-term novel sound learning. We asked whether listeners could learn complex sounds that cued visual events, particularly where a combination of sounds provided the optimal cue. Our aim was to model some of the complexity of real-world performance, where listeners typically rely on learned contextual and object-based auditory cues (differing in duration, spectral content, temporal envelope and relative saliency) to inform expectations about likely events (Leech et al., 2009a; Krishnan et al., 2013; Gygi & Shaffiro, 2011; see also Niessen, 2008). In so doing, we aimed to show that complex, scene-like auditory stimuli could be learned online, to the benefit

of a relevant visual task. In the most basic instantiation, we expected that optimal learners would be able to map isolated instances of broadband (context-like) sounds or punctate (object-like) sounds to a specific location (given that the sounds invariantly cued the upcoming location of a visual target). Extending this paradigm, we predicted that optimal learners would be able to combine both broadband and punctate cues and in turn map this optimally informative combination to a given visual location.

We anticipated that if participants could learn the particular cue types, it might be possible to index expert-like outcomes. Specifically, we expected that ‘experts’ might achieve very high levels of performance accuracy when asked to match the cues to their corresponding locations. Moreover, we expected that where a combination of cues were violated during the learning task itself, optimal learners might show in-task costs associated with the disruption to cue structure (Kawahara, 2007) (in a broadly similar vein to the costs associated with disruptions to configural processing for learned novel visual objects; see Gauthier et al., 1998). Moreover, we sought to examine these abilities across development. Indeed, both children and adults can make use of auditory contextual information to inform expectations (Krishnan et al., 2013), and may weight optimally informative cues where available (MacWhinney et al., 1985); yet children may nevertheless struggle to combine available cues even within modalities (e.g., Nardini et al., 2012). Such contrasting findings suggested developmental differences in cue learning and use, hence motivating our study.

In support of our hypotheses, we found some evidence of cue learning that approached levels that might be considered as expert-like. We observed profiles of differences in learning success: in each experiment, a subset of participants tended toward very high levels of cue learning accuracy, whereas

others showed near floor performance. Indeed, in order to index learning effects thoroughly in latter cue combination experiments, we divided our cohorts by a median split of post-learning identification accuracy. In so doing, we found some evidence of ‘expert-like’ performance hallmarks: when we violated the spatial configuration of cues, we found that participants with better 4AFC learning outcomes also showed increases in task RTs (but no costs to accuracy) compared to those with poorer learning outcomes (further to Kawahara, 2007). Notably however, this effect was most robust in children (while mean RTs suggested the effect in adults in experiment 5a, we did not find it to be statistically significant).

As predicted, we also observed that the particular cue type could influence whether there was a cost related to violation of the cue structure; yet surprisingly, this effect did not depend on post-task learning success. Across better and poorer adult learners, we found a trend for increases in RTs (but no cost to accuracy) when we violated the configuration of the punctate cue structure (but not of the contextual cue structure). Indeed, costs associated with feature-specific configural switches have previously been shown for expertise with visual ‘greeble’ categories (Gauthier et al., 1998). Our stimuli do not reflect auditory categories per se (and were learned in a manner very different to the greebles of Gauthier et al.); yet it is notable that violations of the spatial configuration of a relatively salient auditory cue type tended to yield some cost to performance both for better and poorer learners. Such a result may suggest that features of cue stimuli can influence online performance, even if the ability to learn a combination of particular cues is limited (further to Kayser et al., 2005; see also Tsuchida & Cottrell, 2012). Nevertheless, it is important to note that a control experiment in which we reversed the order that the violations of

particular cues occurred in did not yield support for cue cost effects for more salient auditory stimuli, that were independent of cue combination learning success. In other words, it appears that the extent to which a salient cue is stable during learning blocks is also a factor in determining whether or not costs to performance occur where that salient cue alone is violated. In future experiments, it will be important to control for any potential order effects when investigating effects of cue structure violation for more versus less salient auditory cues, as such order effects appear to critically influence the extent of cue cost that occurs for a more salient stimulus type.

A remaining question is whether our 'expert' learners developed efficient access to subordinate levels of detail with respect to the auditory cues. Indeed, we found limited evidence that better learners showed a cost associated with violation of the less salient contextual cue (experiment 5b). Nevertheless, that better learners across cue combination experiments could achieve near ceiling accuracy (and show relatively low confusion) at 4AFC suggests some ability to distinguish the less salient contexts. While this perhaps reflects some access to subordinate levels of stimulus detail, we should acknowledge that the limited exemplars within the stimulus sets used here do not allow us to make firm conclusions on this point. Future experiments that use a broader array of contextual and punctate sound stimuli in the same experiment may allow us to determine whether listeners can disambiguate between multiple exemplars from the same stimulus category based on subordinate level features.

**7.5.2. Auditory cue learning and development.** As noted above, while we observed strong differences in success of learning within both adult and child cohorts, we also found that overall extents of learning among better child and adult learners were not significantly different. As discussed in chapter 6

(6.5.2), such findings agree with the predictions of the competition model (Bates & MacWhinney, 1987; MacWhinney et al., 1985), suggesting that children can learn to weight valid cues (i.e., where cue availability and reliability are relatively high) even in novel learning scenarios.

Indeed, the question then arises as to whether children's performance can be considered as 'expert-like'. For instance, in learning visual greebles, typically developing children achieve relatively poorer identification accuracy at subordinate category levels compared to adults (whereas basic level accuracy does not differ; Scherf et al., 2008). In the present study (as noted above), children and adults who performed above their cohort 4AFC median did not differ significantly in identification accuracy; yet we found that children showed a more statistically robust RT cost associated with violation of cue configural relations. Such a finding may be interpreted in two ways. On the one hand, it may suggest that children had developed a strong representation of spatial configural relations among cues such that violation reliably disrupted their performance speed – perhaps suggesting an expert-like understanding of the sound cues and their structure. However, on the other hand it is also possible that children's performance was less flexible than that of adults; thus when the useful and (up to that point) reliable sound cues were violated, children may have failed to adjust to the new cue relations or adopt a new strategy. Indeed, children's focus on task-relevant information improves during the school-age years (e.g., Schiff & Knopf, 1985), but children may also continue to develop the ability to select appropriate task strategies up to late childhood (Miller et al., 1986; DeMarie-Dreblow & Miller, 1988; Miller & Weiss, 1981). If children who performed better on the learning task tended to selectively focus on the sound cue relations, disruption of these relations may have limited their ability to

switch to another set of cue relations. As noted previously (see 6.5.4), differences in attentional ability did not account for differences in learning success in our child cohort, although it is possible that the ability to divide attention could account for differences in learning outcomes (see Schiff & Knopf, 1985; Miller & Weiss, 1981).

In future studies, we may be able to probe possible 'expert-like' outcomes in children by using broader stimulus sets (see above) and by violating cues independently. In particular, distinct cue costs associated with punctate and contextual sound violation might suggest access to subordinate levels of detail and perhaps expert-like outcomes in children.

## **7.6 Concluding remarks**

The experience we have with sounds has the potential to shape and guide our perceptual, cognitive and neural processing of the auditory world. Across timeframes of years and even minutes, we can develop and hone elaborate perceptual and cognitive skills with respect to sound. Moreover, the experience we have within complex environments across ontogeny may help to shape the structure of brain areas involved in sound perception, cognition and expert performance. In this thesis, we have argued that an expertise framework can offer a means of accounting for the parallel roles of environment and ontogeny in guiding auditory plasticity.

We have seen that expert musicians trained over periods of years under intensive and specific auditory constraints adapt tuning perception skills that appear specific to the demands they encounter during training. Yet expertise may also benefit more general abilities to finely perceive low-level signal features of auditory input, regardless of the nature of training demands. Further,

attentional performance may serve to mediate fine-grained auditory perception in some regards. However, broader generalisation and transfer of expertise to cognitive skills that reflect some of the demands of the training environment appears limited.

Moreover, we have found that expert violinists and non-experts show differences in the structure of auditory cortical regions (based on proxies for cortical myelin content); these differences can manifest at a between-cohort level (left auditory core) and within the expert cohort alone (reflecting extent of training; right STG). Nevertheless, both auditory and motor cortical regions show clear differences in the extent of expertise-related adaptation indexed using cortical myelin proxies. Such a difference suggests a key distinction between auditory and motor cortical areas involved in expert performance.

We have seen that experience over the course of a laboratory session can lead to adaptations to both auditory perceptual skill and the ability to combine and learn useful auditory cues. Across ontogeny, we have shown that auditory cue learning adaptations can occur for adults and children within timeframes of minutes and can yield outcomes that may approach performance considered expert-like. Nevertheless, experience in itself does not necessitate that learning and plasticity will occur; clear profiles of inter-individual variation in learning success may follow experience with complex auditory cues.

In sum, these experiments have helped to further understanding of auditory learning, expertise and plasticity. Yet many questions remain. Under what circumstances can musicians' expertise generalise to cognitive skills? Do expert motor skills lead to adaptations to subcortical fibre tracts? Can children show evidence of access to subordinate levels of complex auditory cues, suggesting expert-like abilities? These questions have not been answered fully

by the present experiments. Our future studies will aim to explore these issues, and further enhance our understanding of plasticity in the auditory system.

## References

- Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, *15*(7), 1063–1073.
- Alais, D., & Burr, D. (2004). The Ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.
- Agnew, Z.K., McGettigan, C., & Scott, S.K. (2011). Discriminating between auditory and motor cortical responses to speech and non-speech mouth sounds. *Journal of Cognitive Neuroscience*, *23*(12), 4038–4047.
- Agus, T.R., Thorpe, S.J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: insights from noise. *Neuron*, *66*, 610–618.
- Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 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.
- Amunts, K., Schlaug, G., Jancke, L., H, S., Schleicher, A., Dabringhaus, A., & Zilles, K. (1997). Motor cortex and hand motor skills: Structural compliance in the human brain. *Human Brain Mapping*, *5*, 206–215.
- Arciuli, J., & Simpson, I. C. (2011). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286–304.
- Ashida, G., & Carr, C.E. (2011). Sound localisation: Jeffress and beyond. *Current Opinion in Neurobiology*, *21*(5), 745–751.
- Aslin, R.N., Pisoni, D.B., Hennessy, B.L., Perey, A.J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development*, *52*, 1135–1145.

- 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.
- Banai, K., & Yuval-Weiss, N. (2013). Prolonged development of auditory skills: A role for perceptual anchoring *Cognitive Development*, *28*(3), 300–311.
- Bangert, M., & Schlaug, G. (2006). Specialization of the specialized in features of external human brain morphology. *European Journal of Neuroscience*, *24*(6), 1832–1834.
- Barbey, A.K., & Patterson, R. (2011). Architecture of explanatory inference in the human prefrontal cortex. *Frontiers in Psychology*, *2*, 1–9. doi: 10.3389/fpsyg.2011.00162.
- 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.
- Barkovich, A.J., Kjos, B.O., Jackson, D.E., & Norman, D. (1988). Normal maturation of the neonatal and infant brain: MR imaging at 1.5T. *Radiology*, *166*, 173–180.
- Barnea-Goraly, N., Menon, V., Eckert, M., Tamm, L., Bammer, R., Karchemskiy, A., Dant, C. C., et al. (2005). White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study. *Cerebral Cortex*, *15*(12), 1848–1854.

- Bartzokis, G. (2004). Age-related myelin breakdown: a developmental model of cognitive decline and Alzheimer's disease. *Neurobiology of Aging, 25*, 5–18.
- Barutchu, A., Crewther, D. P., & Crewther, S. G. (2009). The race that precedes coactivation: development of multisensory facilitation in children. *Developmental Science, 12*(3), 464–473.
- Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 157–193). Hillsdale, NJ: Lawrence Erlbaum.
- Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In Fletcher, P., & MacWhinney, B. (Eds.), *The Handbook of Child Language* (pp. 96–151). Oxford: Basil Blackwell.
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A: Optics, Image Science and Vision, 20*(7), 1391–1397.
- Baumann, N., & Pham-Dinh, D. (2001). Biology of oligodendrocyte and myelin in the mammalian central nervous system. *Physiological Reviews, 81*(2), 871–927.
- 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.

- Bendixen, A., Schroger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, *29*(26), 8447–8451.
- Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., & Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience*, *8*(9), 1148–1150.
- 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.
- Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology*, *17*(1-3), 35–54.
- Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex*, *19*(7), 1583–1596.
- 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.
- Boothroyd, A. (1996). Auditory development of the hearing child. *Scandinavian Audiology, Suppl 46*(26), 9–16.

- 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.
- Boyke, J., Driemeyer, J., Gaser, C., Büchel, C., & May, A. (2008). Training-induced brain structure changes in the elderly. *The Journal of Neuroscience*, *28*(28), 7031–7035.
- Bukach, C.M., Gauthier, I., & Tarr, M.J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159–166.
- Buonomano, D.V., & Merzenich, M.M. (1998). Cortical plasticity: from synapses to maps. *Annual Review of Neuroscience*, *21*, 149–186.
- 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.
- Brown, M. L., Newsome, S. L., & Glinert, E. P. (1989). An experiment into the use of auditory cues to reduce visual workload. *ACM SIGCHI Bulletin*, *20*(SI), 339–346.
- Callaghan, M.F., Helms, G., Lutti, A., Mohammadi, S., Weiskopf, N. (2014). A general linear relaxometry model of  $R_1$  using imaging data. *Magnetic Resonance in Medicine*, advanced access. doi: 10.1002/mrm.25210

- Chang, E. F., Bao, S., Imaizumi, K., Schreiner, C. E., & Merzenich, M. M. (2005). Development of spectral and temporal response selectivity in the auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(45), 16460–16465.
- Chi, M.T.H. (2006). Laboratory methods for assessing experts' and novices' knowledge. In Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 167–184). Cambridge: Cambridge University Press.
- Chobert, J., François, C., Velay, J.L., Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex*, *24*(4), 956–967.
- Chun, M. M. (2000). Contextual cuing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–177.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.
- 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.
- Coch, D., Sander, L.D., & Neville, H.J. (2005). An event-related potential study of selective auditory attention in children and adults. *Journal of Cognitive Neuroscience*, *17*(4), 605–622.
- Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 17–30.

- Cohen, J. (1968). Weighted Kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220.
- 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
- Conway, C.M., & Christiansen, M.H. (2006). Statistical learning within and between modalities. *Psychological Science*, 17(10), 905–912.
- 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, 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.
- Cusack, R., & Carlyon, R. P. (2003). Perceptual asymmetries in audition. *Journal of Experimental Psychology: Human Perception and Performance*, 29(3), 713–725.
- Cusack, R., Decks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 643–656.

- Davis, M. H., Johnsruide, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences *Journal of Experimental Psychology: General*, *134*(2), 222–241.
- DeCasper, A.J., & Spence, M.J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behaviour and Development*, *9*, 133–150.
- DeMarie-Dreblow, D., & Miller, P.H. (1988). The development of children's strategies for selective attention: evidence for a transitional period. *Child Development*, *59*(6), 1504–1513.
- 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.
- Dick, F., Taylor-Tierney, A., Lutti, A., Josephs, O., Sereno, M. I., & Weiskopf, N. (2012). In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *Journal of Neuroscience*, *32*(46), 16095–16105.
- Dohn, A., Garza-Villarreal, E.A., Chakravarty, M.M., Hansen, M., Lerch, J.P., & Vuust, P. (2013). Gray- and white-matter anatomy of absolute pitch possessors. *Cerebral Cortex*, early access, doi: 10.1093/cercor/bht334
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Changes in grey matter induced by training. *Nature*, *427*, 311–312.
- Draganski, B., & May, A. (2008). Training-induced structural changes in the adult human brain. *Behavioural Brain Research*, *192*(1), 137–142.
- Drullman, R., Festen, J.M., & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America*, *95*(2), 1053–1064.

- Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, 2(7), 254–261.
- 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.
- Dye, M.W.G., Green, C.S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science*, 18(6), 321–326.
- Ehrlé, N., & Samson, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain and Cognition*, 58, 133–147.
- Emery, B. (2010). Regulation of oligodendrocyte differentiation and myelination. *Science*, 330(6005), 779–782.
- Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance *Psychological Review*, 100(3), 363–406.
- 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.

- Faiss, J.H., Dähne, D., Baum, K., Deppe, R., Hoffman, F., Kohler, W., ... Zetl, U.K. (2014). Reduced magnetisation transfer ratio in cognitively impaired patients at the very early stage of multiple sclerosis: a prospective, multicenter, cross-sectional study. *British Medical Journal*, *4*, e004409.
- Fanselow, M. S., & Poulos, A. M. (2005). The neuroscience of mammalian associative learning. *Annual Review of Psychology*, *56*(1), 207–234.
- Farah, M. J., Wong, A. B., Monheit, M. A., & Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: modality-specific or supramodal. *Neuropsychologia*, *27*(4), 461–470.
- Flege, J.E., Munro, M.J., MacKay, I.R.A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*(5), 3125–3134.
- Fornari, E., Knyazeva, M. G., Meuli, R., & Maeder, P. (2007). Myelination shapes functional activity in the developing brain. *NeuroImage*, *38*(3), 511–518.
- Forster, S., & Lavie, N. (2011). Entirely irrelevant distractors can capture and captivate attention. *Psychonomic Bulletin & Review*, *18*(6), 1064–1070.
- François, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, *21*, 2357–2365.
- François, C., Jaillet, F., Takerkart, S., & Schön, D. (2014). Faster sound stream segmentation in musicians than in nonmusicians. *PloS one*, *9*(7), e101340.
- 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.

- Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, *6*(11), 1216–1223.
- Fritz, J. B., Elhilali, M., & Shamma, S. A. (2005). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *Journal of Neuroscience*, *25*(33), 7623–7635.
- Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hearing Research*, *229*(1-2), 186–203.
- 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.
- Gale, A., & Lynn, R. (1972). A developmental study of attention. *British Journal of Educational Psychology*, *42*(3), 260–266.
- Gallagher, M., McMahon, R. W., & Schoenbaum, G. (1999). Orbitofrontal cortex and representation of incentive value in associative learning. *The Journal of Neuroscience*, *19*(15), 6610–6614.
- Gartner, H., Minnerop, M., Pieperhoff, P., Schleicher, A., Zilles, K., Altenmüller, E., & Amunts, K. (2013). Brain morphometry shows effects of long-term musical practice in middle-aged keyboard players. *Frontiers in Psychology*, *4*(636), 1–13.
- Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, *23*(27), 9240–9245.
- Gauthier, I., & Tarr, M.J. (1997). Becoming a 'greeble' expert: exploring mechanisms for face recognition. *Vision Research*, *37*(12), 1673–1682.

- Gauthier, I., & Tarr, M.J. (2002). Unravelling mechanisms for object recognition: bridging brain activity and behaviour. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 431–446.
- Gauthier, I., Skudlarski, P., Gore, J.C., & Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197.
- Gibson, E. M., Purger, D., Mount, C. W., Goldstein, A. K., Lin, G. L., Wood, L. S., ... Monje, M. (2014). Neuronal activity promotes oligodendrogenesis and adaptive myelination in the mammalian brain. *Science*, *344*(6183), 1252304–1252304.
- Glasser, M.F., & Van Essen, D.C. (2011). Mapping human cortical areas in vivo based on myelin content as revealed by T<sub>1</sub>- and T<sub>2</sub>-weighted MRI. *The Journal of Neuroscience*, *31*(32), 11597–11616.
- Goebel, W. (2001). Melody lead in piano performance: expressive device or artifact? *Journal of the Acoustical Society of America*, *110*, 563–572.
- Goebel, 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.
- Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron*, *35*, 997–1010.
- Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, *17*(3), 575–582.

- Golestani, N., Price, C. J., & Scott, S. K. (2011). Born with an ear for dialects? structural plasticity in the expert phonetician brain. *Journal of Neuroscience*, *31*(11), 4213–4220.
- Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., & Cowan, N. (2000). The development of auditory attention in children. *Frontiers in Bioscience*, *5*, d108–d120.
- Gordon, J.W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America*, *82*, 88–105.
- Gore, J.C., & Kennan, R.P. (1999). Physical and physiological basis of magnetic relaxation. In Stark, D.D., & Bradley Jr., W.G. (Eds.), *Magnetic Resonance Imaging* (vol. 1) (pp. 33–42). St. Louis: Mosby.
- Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S.K. (2002). Amplitude envelope onsets and developmental dyslexia: a new hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(16), 10911–10916.
- Green, C.S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*, 534–537.
- Green, C.S., & Bavelier, D. (2007). Action video game experience alters the spatial resolution of vision. *Psychological Science*, *18*(1), 88–94.
- Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. *Current Biology*, *22*(6), R197–R206.
- Green, C. S., Pouget, A., & Bavelier, D. (2010). Improved probabilistic inference as a general learning mechanism with action video games. *Current Biology*, *20*(17), 1573–1579.

- 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.
- Griffiths, T. D., Buechel, C., Frackowiak, R. S. J., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, *1*(5), 422–427.
- Griffiths, T.D., & Warren, J.D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*, 887–892.
- Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–562.
- Grothe, B., & Sanes, D.H. (1993). Bilateral inhibition by glycinergic afferents in the medial superior olive. *Journal of Neurophysiology*, *69*(4), 1192–1196.
- Groussard, M., La Joie, R., Rauchs, G., Landeau, B., Chételat, G., Viader, F., ... Platel, H. (2010). When music and long-term memory interact: Effects of musical expertise on functional and structural plasticity in the hippocampus. *PloS one*, *5*(10), e13225. doi:10.1371/journal.pone.0013225.s002
- Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, *106*(5), 2900–2912.
- Gygi, B., & Shafiro, V. (2011). The incongruity advantage for natural sounds presented in natural auditory scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 551–565.

- Hackett, T.A. (2007). Organization and correspondence of the auditory cortex of humans and nonhuman primates. In Kaas, J.H. (Ed.) *Evolution of the Nervous System* (pp 109 –119). Oxford, UK: Elsevier.
- Hackett, T. A. (2011). Information flow in the auditory cortical network. *Hearing Research, 271*(1-2), 133–146.
- Hackett, T. A., Preuss, T. M., & Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *The Journal of Comparative Neurology, 441*(3), 197–222.
- Hagler, D.J., Saygin, A.P., & Sereno, M.I. (2006). Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage, 33*, 1093–1103.
- Hall, D.E., & Hess, J.T. (1984). Perception of musical interval tuning. *Music Perception, 2*, 166–195.
- Hall, J.W., Buss, E., & Grose, J.H. (2005). Informational masking release in children and adults. *Journal of the Acoustical Society of America, 118*(3), 1605–1613.
- Halliday, L. F. (2014). A Tale of Two Studies on Auditory Training in Children: A Response to the Claim that 'Discrimination Training of Phonemic Contrasts Enhances Phonological Processing in Mainstream School Children' by Moore, Rosenberg and Coleman (2005). *Dyslexia*. doi:10.1002/dys.1470.
- Halliday, L.F., Taylor, J.L., Edmonson-Jones, A.M., & Moore, D.R. (2008). Frequency discrimination learning in children. *Journal of the Acoustical Society of America, 123*(6), 4393–4402.
- Halliday, L. F., Moore, D. R., Taylor, J. L., & Amitay, S. (2011). Dimension-specific attention directs learning and listening on auditory training tasks. *Attention, Perception, & Psychophysics, 73*(5), 1329–1335.

- Halliday, L.F., Taylor, J.L., Millward, K.E., & Moore, D.R. (2012). Lack of generalization of auditory learning in typically developing children. *Journal of Speech, Language and Hearing Research, 55*, 168–181.
- Halwani, G. F., Loui, P., Rüber, T., & Schlaug, Gottfried. (2011). Effects of Practice and Experience on the Arcuate Fasciculus: Comparing Singers, Instrumentalists, and Non-Musicians. *Frontiers in Psychology, 2*, doi: 10.3389/fpsyg.2011.00156
- Han, Y., Yang, H., Lv, Y.-T., Zhu, C.-Z., He, Y., Tang, H.-H., Gong, Q.-Y., Luo, J.-Y., Zang, F.-Y, Dong, Q. (2009). Gray matter density and white matter integrity in pianists' brain: a combined structural and diffusion tensor MRI study. *Neuroscience Letters, 459*(1), 3–6.
- Hanajima, R., Ugawa, Y., Machii, K., Mochizuki, H., Terao, Y., Enomoto, H., ... Kanazawa, I. (2001). Interhemispheric facilitation in the hand motor area in humans. *Journal of Physiology, 531*(3), 849–859.
- Hannon, E.E., & Trainor, L.J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Sciences, 11*, 466–472.
- Hawkey, D. J. C., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience, 7*(10), 1055–1056.
- Hay, J.F., & Saffran, J.R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy, 17*(6), 610–641.
- Hayes, E.A., Warrier, C.M., Nicol, T.G., Zecker, S.G., & Kraus, N. (2003). Neural plasticity following auditory training in children with learning problems. *Clinical Neurophysiology, 114*, 673–684.

- 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.
- Helms, G., Dathe, H., & Dechent, P. (2008). Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. *Magnetic Resonance in Medicine, 59*, 667–672.
- Helms, G., Draganski, B., Frackowiak, R., Ashburner, J., & Weiskopf, N. (2009). Improved segmentation of deep brain grey matter structures using magnetization transfer (MT) parameter maps. *NeuroImage, 47*, 194–198.
- Herdener, M., Esposito, F., di Salle, F., Boller, C., Hilti, C. C., Habermeyer, B., Scheffler, K., et al. (2010). Musical training induces functional plasticity in human hippocampus. *Journal of Neuroscience, 30*(4), 1377–1384.
- Herdener, M., Esposito, F., Scheffer, K., Schneider, P., Logothetis, N. K., Uludag, K., & Kayser, C. (2013). Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex, 49*(10), 2822–2833.
- 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.
- Herholz, S.C., & Zatorre, R.J. (2012). Musical training as a framework for brain plasticity: behaviour, function and structure. *Neuron, 76*, 486–502.
- Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 283–295.

- Herzmann, G., & Curran, T. (2011). Experts' memory: an ERP study of perceptual expertise effects on encoding and recognition. *Memory & Cognition, 39*(3), 412–432.
- Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia, 49*(3), 461–467.
- Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied, 11*(3), 157–174.
- Hoffman, G., Mürbe, D., Kuhlisch, E., & Pabst, F. (1997). Unterschied des auditiven frequenzdiskriminationsvermögens bei musikern verschiedener fachbereiche. *Folia Phoniatica et Logopaedica, 49*, 21–25.
- Hoffman, R.R. (1996). How can expertise be defined? Implications of research from cognitive psychology. In R. Williams, W. Faulkner & J. Fleck (Eds.), *Exploring Expertise* (pp. 81–100). Edinburgh: University of Edinburgh Press.
- Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron, 36*, 791–804.
- Holtmaat, A., & Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience, 10*(9), 647–658.
- Hopkin, B. (1996). *Musical instrument design: Practical information for instrument making*. Tucson: Sharp Press.

- Huang, Z.J., Kirkwood, A., Pizzorusso, T., Porciatti, V., Morales, B., Bear, M.F., Maffei, L., & Tonegawa, S. (1999). BDNF regulates the maturation of inhibition and critical period of plasticity in mouse visual cortex. *Cell*, *98*, 739–755.
- Hurford, J.R. (1991). The evolution of the critical period for language acquisition. *Cognition*, *40*, 159–201.
- Husain, F.T., Tagamets, M.-A., Fromm, S.J., Braun, A.R., & Horwitz, B. (2004). Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modelling and an fMRI study. *NeuroImage*, *21*, 1701–1720.
- Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, *47*(6), 674–689.
- Hutchinson, S., Lee, L. H. L., Gaab, N., & Schlaug, G. (2003). Cerebellar volume of musicians. *Cerebral Cortex*, *13*, 943–949.
- Huttenlocher, P.R., & Dabholkar, A.S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *The Journal of Comparative Neurology*, *387*, 167–178.
- Huxley, A.F., & Stämpfli, R. (1949). Evidence for saltatory conduction in peripheral myelinated nerve fibres. *Journal of Physiology*, *108*, 315–339.
- Huyck, J. J., & Wright, B. A. (2011). Late maturation of auditory perceptual learning. *Developmental Science*, *14*(3), 614–621.
- Huyck, J.J., & Wright, B.A. (2013). Learning, worsening and generalization in response to auditory perceptual training during adolescence. *Journal of the Acoustical Society of America*, *134*(2), 1172–1182.

- 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.
- Imfeld, A., Oechslin, M. S., Meyer, M., Loenneker, T., & Jäncke, L. (2009). White matter plasticity in the corticospinal tract of musicians: A diffusion tensor imaging study. *NeuroImage*, *46*(3), 600–607.
- Irvine, D.R.F., Park, V.N., & McCormick, L. (2001). Mechanisms underlying the sensitivity of neurons in the lateral superior olive to interaural intensity differences. *Journal of Neurophysiology*, *86*, 2647–2666.
- James, T.W., & James, K.H. (2013). Expert individuation of objects increases activation in the fusiform face area of children. *NeuroImage*, *67*, 182–192.
- Jurlink, D.N., & Detsky, A.S. (2005). Kappa statistic. *Canadian Medical Association Journal*, *173*(1), 16.
- Justus, T., & List, A. (2005). Auditory attention to frequency and time: an analogy to visual local-global stimuli. *Cognition*, *98*, 31–51.
- Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 11793–11799.
- Kaas, J.H., Hackett, T.A., & Tramo, M.J. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology*, *9*, 164–170.
- Kaneko, T. (2013). Local connections of excitatory neurons in motor-associated cortical areas of the rat. *Frontiers in Neural Circuits*, *7*, 1–17. doi: 10.3389/fncir.2013.00075
- Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society: Biological Sciences*, *361*, 2109–2128.

- Karmiloff-Smith, A.K. (2012). From constructivism to neuroconstructivism: the activity-dependent structuring of the human brain. In Marti, E., & Rodrigues, C. (Eds.), *After Piaget* (pp. 1–14). New Jersey: Transaction Publishers.
- Karpicke, J.D., & Pisoni, D. B. (2004). Using immediate memory span to measure implicit learning. *Memory and Cognition*, *32*, 956–964.
- Kawahara, J. I. (2007). Auditory-visual contextual cuing effect. *Perception & Psychophysics*, *69*(8), 1399–1408.
- 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.
- Keller, P., & Stevens, C. (2004). Meaning from environmental sounds: Types of signal-referent relations and their effect on recognizing auditory icons. *Journal of Experimental Psychology: Applied*, *10*(1), 3–12.
- Kempermann, G., Kuhn, H. G., & Gage, F. H. (1997). More hippocampal neurons in adult mice living in an enriched environment. *Nature*, *386*, 493–495.
- Kempermann, G., Kuhn, H. G., & Gage, F. H. (1998). Experience induced neurogenesis in the senescent dentate gyrus. *The Journal of Neuroscience*, *18*(9), 3206–3212.
- Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.
- 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*, ECVF Abstract Supplement.
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712–719.
- Knowland, V.C.P., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M.S.C. (2013). Audio-visual speech perception: a developmental ERP investigation. *Developmental Science, 17*(1), 110–124.
- 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.
- Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning, 30*(2-3), 271–274.
- Kopiez, R. (2003). Intonation of harmonic intervals: Adaptability of expert musicians to equal temperament and just intonation. *Music Perception, 20*, 383–410.
- 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*(4), 331–359.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11*, 599–605.
- 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, 300*, 46–55.

- Kuhl, P.K., Tsao, F.M., & Liu, H.M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Science of the United States of America*, *100*(15), 9096–9101.
- Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.
- 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.
- Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load Theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339–354.
- Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82.
- Leaver, A.M., & Rauschecker, J.P. (2010). Cortical representation of natural complex sound categories: Effects of acoustic features and auditory object category. *The Journal of Neuroscience*, *30*(22), 7604–7612.
- Leech, R., Gygi, B., Aydelott, J., & Dick, F. (2009a). Informational factors in identifying environmental sounds in natural auditory scenes. *Journal of the Acoustical Society of America*, *126*, 3147–3155.

- Leech, R., Holt, L. L., Devlin, J. T., & Dick, F. (2009b). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *The Journal of Neuroscience*, *29*(16), 5234–5239.
- Leibold, L. J., & Bonino, A. Y. (2009). Release from informational masking in children: Effects of multiple signal bursts. *Journal of the Acoustical Society of America*, *125*(4), 2200–2208.
- Leibold, L. J., & Neff, D. F. (2007). Effects of masker spectral-variability and masker fringes in children and adults. *Journal of the Acoustical Society of America*, *121*(6), 3666–3676.
- Lenneberg, E. (1967). *Biological Foundations of Language*. Oxford: Wiley.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.
- Lewis, J.W., Talkington, W.J., Tallaksen, K.C., & Frum, C.A. (2012). Auditory object salience: human cortical processing of non-biological action sounds and their acoustic signal attributes. *Frontiers in Systems Neuroscience*, *6*, 1–15. doi:10.3389/fnsys.2012.00027.
- Lim, S.J., & Holt, L. L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*, *35*(7), 1390–1405.
- Lively, S.E., Pisoni, D.B., Yamada, R.A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*(4), 2076–2087
- Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America*, *89*(2), 874–886.

- 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.
- Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, *136*(4), 659–676.
- Luders, E., Gaser, C., Jaencke, L., & Schlaug, G. (2004). A voxel-based approach to gray matter asymmetries. *NeuroImage*, *22*, 656–664.
- Lutti, A., Hutton, C., Finsterbusch, J., Helms, G., & Weiskopf, N. (2010). Optimization and validation for mapping of the radiofrequency transmit field at 3T. *Magnetic Resonance in Medicine*, *64*(1), 229–238.
- Lutti, A., Dick, F., Sereno, M.I., & Weiskopf, N. (2014). Using high-resolution quantitative mapping of  $R_1$  as an index of cortical myelination. *NeuroImage*, *93*, 176–188.
- Lynch, M.P., & Eilers, R.E. (1991). Children's perception of native and nonnative musical scales. *Music Perception*, *9*(1), 121–132.
- MacWhinney, B., Pleh, C., & Bates, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology*, *17*, 178–209.
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 4398–4403.

- 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.
- Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., Robertson, I.H. (2001). The differential assessment of children's attention: the Test of Everyday Attention for Children (TEA-Ch), normative sample and ADHD performance. *Journal of Child Psychology and Psychiatry*, *42*(8), 1065–1081.
- Marie, C., Magne, C., & Besson, M. (2011). Musicians and the 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.
- May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends in Cognitive Sciences*, *15*(10), 475–482.
- McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*, 339–346.
- McDermott, J.H., Lehr, A.J., & Oxenham, A.J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, *20*, 1035–1041.
- McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, *16*(4), 493–498.
- McMullen Jonaitis, E., & Saffran, J. (2009). Learning harmony: the role of serial statistics. *Cognitive Science*, *33*, 951–968.

- McNealy, K., Mazziotta, J. C., & Dapretto, M. (2011). Age and experience shape developmental changes in the neural basis of language-related learning. *Developmental Science, 14*(6), 1261–1282.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143–178.
- 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.
- Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science, 271*(5245), 77–81.
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*(1-2), 36–47.
- Miller, D. J., Duka, T., Stimpson, C. D., Schapiro, S. J., Baze, W. B., MacArthur, M. J., ... Sherwood, C.C. (2012). Prolonged myelination in human neocortical evolution. *Proceedings of the National Academy of Sciences of the United States of America, 109*(41), 16480–16485.
- Miller, P. H., & Weiss, M. G. (1981). Children's attention allocation, understanding of attention, and performance on the incidental learning task. *Child Development, 52*(4), 1183–1190.
- Miller, P.H., Haynes, V.F., DeMarie-Dreblow, D., & Woody-Ramsey, J. (1986). Children's strategies for gathering information in three tasks. *Child Development, 57*(6), 1429–1439.

- Molchan, S. E., Sunderland, T., McIntosh, A. R., Herscovitch, P., & Schreurs, B. G. (1994). A functional anatomical study of associative learning in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 8122–8126.
- Moore, B.C.J., & Glasberg, B.R. (1983). Suggested formulae for calculating auditory filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, *74*(3), 750–753.
- 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.
- Moore, D. R., Rosenberg, J. F., & Coleman, J. S. (2005). Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children. *Brain and Language*, *94*(1), 72–85.
- Moore, D. R., Ferguson, M. A., Halliday, L. F., & Riley, A. (2008). Frequency discrimination in children: Perception, learning and attention. *Hearing Research*, *238*(1-2), 147–154.
- Moore, D.R., Halliday, L.F., & Amitay, S. (2009). Use of auditory learning to manage listening problems in children. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 409–420.
- Moore, D.R., Ferguson, M.A., Edmonson-Jones, A.M., Ratib, S., & Riley, A. (2010). Nature of auditory processing disorder in children. *Pediatrics*, *126*, e382–e390.

- Moore, D.R., Cowan, J.A., Riley, A., Edmonson-Jones, A.M., & Ferguson, M.A. (2011). Development of auditory processing in 6- to 11-year-old children. *Ear & Hearing, 32*(3), 269–285.
- Mondor, T.A., & Bregman, A.S. (1994). Allocating attention to frequency regions. *Perception and Psychophysics, 56*, 268–276.
- 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., 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.
- 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.
- Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage, 13*(4), 684–701.
- Morrongiello, B.A., Kulig, J.W., & Clifton, R.K. (1984). Developmental changes in auditory temporal perception. *Child Development, 55*(2), 461–471.
- Morrongiello, B.A., & Trehub, S.E. (1987). Age-related changes in auditory temporal perception. *Journal of Experimental Child Psychology, 44*, 413–426.
- 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.

- Münste, T.F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*, 473–478.
- Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., & Münste, T.F. (2003). The fate of sounds in conductors' brains: an ERP study. *Cognitive Brain Research*, *17*, 83–93.
- Nakagawa, H., Iwasaki, S., Kichikawa, K., Fukusumi, A., Taoka, T., Ohishi, H., & Uchida, H. (1998). Normal myelination of anatomic nerve fibre bundles: MR analysis. *American Journal of Neuroradiology*, *19*, 1129–1136.
- Nakahara, H., Zhang, L. L., & Merzenich, M. M. (2004). Specialization of primary auditory cortex processing by sound exposure in the 'critical period.' *Proceedings of the National Academy of Sciences of the United States of America*, *101*(18), 7170–7174.
- Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, *18*(9), 689–693.
- Nardini, M., Bedford, R., & Mareschal, D. (2012). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(39), 17041–17046.
- Niessen, M. E., van Maanen, L., & Andringa, T. C. (2008). Disambiguating sounds through context. *International Journal of Semantic Computing*, *2*(3), 327–341.
- 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.
- Nittrouer, S., Lowenstein, J.H., & Tarr, E. (2013). Amplitude rise time does not cue the /ba/-/wa/ contrast for adults or children. *Journal of Speech, Language and Hearing Research*, *56*(2), 427–440.

- Nolte, J. (2009). *The Human Brain: An Introduction to its Functional Anatomy* (6th ed.) Philadelphia, PA: Mosby.
- Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., & Schlaug, G. (2005). Are there pre-existing neural, cognitive, or motoric markers for musical ability. *Brain and Cognition*, *59*(2), 124–134.
- O'Doherty, P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *28*, 329–337.
- Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., & Jaencke, L. (2010). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. *Frontiers in Human Neuroscience*, *3*, doi:10.3389/neuro.09.076.2009
- 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*(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., 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.
- 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., 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., Hittner, E., & Kraus, N. (2013). Musical training enhances neural processing of binaural sounds. *The Journal of Neuroscience, 33*, 16741–16747.
- Partridge, S. C., Mukherjee, P., Henry, R. G., Miller, S. P., Berman, J. I., Jin, H., Lu, Y., et al. (2004). Diffusion tensor imaging: Serial quantitation of white matter tract maturity in premature newborns. *NeuroImage, 22*(3), 1302–1314.
- 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.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36, 767–776.
- Paus, T., Collins, D. L., Evans, A. C., Leonard, G., Pike, B., & Zijdenbos, A. (2001). Maturation of white matter in the human brain: a review of magnetic resonance studies. *Brain Research Bulletin*, 54(3), 255–266.
- Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, 9(2), 60–68.
- 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.
- Penhune, V. B., Zatorre, R. J., McDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, 6, 661–672.
- Phillips, K.A., Schaeffer, J.A., & Hopkins, W.D. (2013). Corpus callosum microstructure influences intermanual transfer in chimpanzees. *Frontiers in Systems Neuroscience*, 7, 1–6. doi:10.3389/fnsys.2013.00125
- 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.

- Platt, J. R., & Racine, R. J. (1985). Effect of frequency, timbre, experience, and feedback on musical tuning skills. *Perception & Psychophysics*, *38*(6), 543–553.
- Polley, D. B., Heiser, M. A., Blake, D. T., Schreiner, C. E., & Merzenich, M. M. (2004). Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(46), 16351–16356.
- Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.
- Pressnitzer, D., Suied, C., & Shamma, S.A. (2011). Auditory scene analysis: the sweet music of ambiguity. *Frontiers in Human Neuroscience*, *5*, doi: 10.3389/fnhum.2011.00158.
- Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., & Zilles, K. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage*, *13*(4), 669–683.
- Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and non-musicians. *Music Perception*, *24*, 37–48.
- Ramsden, S., Richardson, F.M., Josse, G., Thomas, M.C., Ellis, C., Shakeshaft, C., Seghier, M.L., & Price, C.J. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature*, *479*, 113–116.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.

- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of 'what' and "where" in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11800–11806.
- Reber, A.S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford: Oxford University Press.
- Rezlescu, C., Barton, J.J.S., Pitcher, D., & Duchaine, B. (2014). Normal acquisition of expertise with greebles in two cases of acquired prosopagnosia. *Proceedings of the National Academy of Sciences of United States of America*, early edition. doi: 10.1073/pnas.1317125111
- Roberts, L., & Mathews, M.V. (1984). Intonation sensitivity for traditional and non-traditional chords. *Journal of the Acoustical Society of America*, *75*, 952–959.
- Roberts, R.E., Bain, P. G., Day, B. L., & Husain, M. (2013). Individual differences in expert motor coordination associated with white matter microstructure in the cerebellum. *Cerebral Cortex*, *23*(10), 2282–2292.
- 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.
- Rossier, J., Haerberli, C., & Schenk, F. (2000). Auditory cues support place navigation in rats when associated with a visual cue. *Behavioural Brain Research*, *117*, 209–214.

- 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.
- Saffran, J., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105.
- Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.
- Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*(3), 669–680.
- Schellenberg, E.G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*, 511–514.
- Scherf, K.S., Behrmann, M., Minshew, N., & Luna, B. (2008). Atypical development of face and greeble recognition in autism. *The Journal of Child Psychology and Psychiatry*, *49*(8), 838–847.
- Schiff, A.R., & Knopf, I.J. (1985). The effect of task demands on attention allocation in children of different ages. *Child Development*, *56*(3), 621–630.
- Schlaug, G, Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*(5198), 699–701.

- Schmithorst, V. J., & Wilke, M. (2002). Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neuroscience Letters*, *321*, 57–60.
- 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.
- Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., Dosch, H. G., et al. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience*, *8*(9), 1241–1247.
- Schnupp, J., Nelken, I., & King, A. (2011). *Auditory Neuroscience: Making Sense of Sound*. Cambridge, MA & London, UK: The MIT Press.
- Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature Neuroscience*, *12*(11), 1370–1371.
- Schönwiesner, M., Rübsem, R., & Cramon, Von, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, *22*(6), 1521–1528.
- Scott, S.K., & McGettigan, C. (2013). The neural processing of masked speech. *Hearing Research*, *303*, 58–66.
- Scott, S.K., Blank, C.C., Rosen, S., & Wise, R.J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–2406.

- Seither-Preisler, A., Parncutt, R., & Schneider, P. (2014). Size and synchronization of auditory cortex promotes musical, literacy, and attentional skills in children. *The Journal of Neuroscience*, *34*(33), 10937–10949.
- 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.
- Sereno, M. I., Lutti, A., Weiskopf, N., & Dick, F. (2012). Mapping the human cortical surface by combining quantitative T<sub>1</sub> with retinotopy. *Cerebral Cortex*, *23*(9), 2261–2268.
- 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.
- Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, *5*(8), 340–348.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*, 788.
- Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304.

- Sherwood, C. C., Holloway, R. L., Erwin, J. M., Schleicher, A., Zilles, K., & Hof, P. R. (2004). Cortical orofacial motor representation in old world monkeys, great apes, and humans. *Brain, Behavior and Evolution*, *63*(2), 61–81.
- Shohamy, D., Myers, C.E., Grossman, S., Sage, J., Gluck, M.A., & Poldrack, R.A. (2004). Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. *Brain*, *127*, 851–859.
- 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.
- Shultz, S., Vouloumanos, A., & Pelphrey, K. (2012). The superior temporal sulcus differentiates communicative and non-communicative auditory signals. *Journal of Cognitive Neuroscience*, *24*(5), 1224–1232.
- Siegel, J.A., & Siegel, W. (1977). Absolute identification of notes and intervals by musicians. *Perception and Psychophysics*, *21*, 143–152.
- Siegel, L.S., & Ryan, E.B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, *60*(4), 973–980.
- Sigalovsky, I. S., Fischl, B., & Melcher, J. R. (2006). Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: a possible marker for primary cortex and hemispheric differences. *NeuroImage*, *32*(4), 1524–1537.
- 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.

- Slater, J., Tierney, A., & Kraus, N. (2013). At-risk elementary school children with one year of classroom music instruction are better at keeping a beat. *Plos One*, *e77250*, 1–9. doi:10.1371/journal.pone.0077250
- Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's Area in male symphony orchestra musicians. *NeuroImage*, *17*(3), 1613–1622.
- Sohoglu, E., Peelle, J.E., Carlyon, R.P., & Davis, M.H. (2012). Predictive top-down integration of prior knowledge during speech perception. *The Journal of Neuroscience*, *32*(25), 8443–8453.
- Sowden, P. T., Davies, I. R. L., & Roling, P. (2000). Perceptual learning in the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 379–390.
- Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature Neuroscience*, *6*(3), 309–315.
- Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 1005–1030.
- Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*(1), 1–22.
- 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.

- Steele, C.J., Scholz, J., Douaud, G., Johansen-Berg, H., Penhune, V.B. (2012). Structural correlates of skilled performance on a motor sequence task. *Frontiers in Human Neuroscience*, 6, doi: 10.3389/fnhum.2012.00289
- 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. *The Journal of Neuroscience*, 33(3), 1282–1290.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). New Jersey: Lawrence Erlbaum Associates.
- 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. (2014). Biological impact of auditory expertise across the lifespan: Musicians as a model of auditory learning. *Hearing Research*, 308, 109–121.
- 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.
- 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.

- Summers, M.M., Fisniku, L.K., Anderson, V.M., Miller, D.H., Cipelotti, L., & Ron, M.A. (2008). Cognitive impairment in relapsing-remitting multiple sclerosis can be predicted by imaging performed several years earlier. *Multiple Sclerosis, 14*, 197–204.
- Sussman, E., Wong, R., Horváth, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research, 225*(1-2), 117–127.
- Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology, 91*(3), 1282–1296.
- Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science, 12*(1), 43–47.
- Tanaka, J.W., & Taylor, M. (1991). Object categories and expertise: is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457–482.
- Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science, 16*(2), 145–151.
- Taylor, H.G., & Heilman, K.M. (1980). Left-hemisphere motor dominance in righthanders. *Cortex, 16*(4), 587–603.
- 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.
- Teixeira, L.A. (2000). Timing and force components in bilateral transfer of learning. *Brain and Cognition, 44*, 455–469.

- Teki, S., Kumar, S., Kriegstein, von, K., Stewart, L., Lyness, C. R., Moore, B. C. J., Capleton, B., et al. (2012). Navigating the auditory scene: an expert role for the hippocampus. *The Journal of Neuroscience*, *32*(35), 12251–12257.
- Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife*, *2*(0), e00699–e00699.
- 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.
- Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., & Friederici, A.D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *European Journal of Neuroscience*, *30*, 1636–1642.
- 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.
- Todorich, B., Pasquini, J.M., Garcia, C.I., Paez, P.M., & Connor, J.R. (2009). Oligodendrocytes and myelination: the role of iron. *Glia*, *57*, 467–478.
- Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (1999). *TOWRE, Test of Word Reading Efficiency*. PRO-ED: Austin, Texas.
- Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Welker, E., & Svoboda, K. (2002). Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, *420*, 788–794.

- 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.
- Tremblay, C., Champoux, F., Voss, P., Bacon, B.A., Lepore, F., & Theoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study. *Plos One*, e742, 1–5. doi:10.1371/journal.pone.0000742
- Tsuchida, T., & Cottrell, G. W. (2012). Auditory saliency using natural statistics. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 1048–1053.
- 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.
- Wade, T., & Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *The Journal of the Acoustical Society of America*, 118(4), 2618–2633.
- Wareham, A.C. (2005). Action potential: Generation and propagation. *Anaesthesia and Intensive Care Medicine*, 6(6), 200–203.
- Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., & Kraus, N. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *Journal of Neuroscience*, 29(1), 61–69.

- van Wassenhove, V., & Nagarajan, S. S. (2007). Auditory cortical plasticity in learning to discriminate modulation rate. *The Journal of Neuroscience*, *27*(10), 2663–2672.
- Watkins, S., Dalton, P., Lavie, N., & Rees, G. (2007). Brain mechanisms mediating auditory attentional capture in humans. *Cerebral Cortex*, *17*(7), 1694–1700.
- Weiskopf, N., Hutton, C., Josephs, O., & Deichmann, R. (2006). Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3T and 1.5T. *NeuroImage*, *33*, 493–504.
- Weiskopf, N., Lutti, A., Helms, G., Novak, M., Ashburner, J., & Hutton, C. (2011). Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *NeuroImage*, *54*(3), 2116–2124.
- Weiskopf, N., Suckling, J., Williams, G., Correia, M.M., Inkster, B., Tait, R., Ooi, C., Bullmore, E.T., & Lutti, A. (2013). Quantitative multi-parameter mapping of R1, PD\*, MT and R2\* at 3T: a multi-center validation. *Frontiers in Neuroscience*, *7*, doi: 10.3389/fnis.2013.00095
- Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience*, *5*(4), 279–290.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: evidence of perceptual reorganization during the first year of life. *Infant Behaviour and Development*, *7*, 49–63.
- Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, *46*, 233–251.
- Werker, J.F., & Yeung, H.H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, *9*(11), 519–527.

- Werner, L. A. (2007). Issues in human auditory development. *Journal of Communication Disorders, 40*(4), 275–283.
- Werner-Olsho, L., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982). Auditory frequency discrimination in infancy. *Developmental Psychology, 18*(5), 721–726.
- Wessel, D.L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3*, 45–52.
- Westermann, G., Mareschal, D., Johnson, M.H., Sirois, S., Spratling, M.W., Thomas, M.S.C. (2007). Neuroconstructivism. *Developmental Science, 10*(1), 75–83.
- 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
- Whitton, J. P., Hancock, K. E., & Polley, D. B. (2014). Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. *Proceedings of the National Academy of Sciences*. Advanced access; doi:10.1073/pnas.1322184111
- Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology, 26*(6), 1003–1017.
- Wilbrecht, L., Holtmaat, A., Wright, N., Fox, K., & Svoboda, K. (2010). Structural plasticity underlies experience-dependent functional plasticity of cortical circuits. *Journal of Neuroscience, 30*(14), 4927–4932.

- Wong, A. C. N., Palmeri, T. J., Rogers, B. P., Gore, J. C., & Gauthier, I. (2009). Beyond shape: How you learn about objects affects how they are represented in visual cortex. *PloS one*, *4*(12), e8405.
- Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., & Zatorre, R. J. (2008). Volume of left Heschl's gyrus and linguistic pitch learning. *Cerebral Cortex*, *18*(4), 828–836.
- Wright, B.A., & Fitzgerald, M.B. (2001). Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(21), 12307–12312.
- Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalisation of auditory temporal interval discrimination in humans. *The Journal of Neuroscience*, *17*(10), 3956–3963.
- Wright, B. A., Wilson, R. M., & Sabin, A. T. (2010). Generalization lags behind learning on an auditory perceptual task. *Journal of Neuroscience*, *30*(35), 11635–11639.
- Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, *15*, 1234–1242.
- Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science*, *342*(6158), 585–589.
- Zatorre, R.J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*, 946–953.
- Zatorre, R.J., Perry, D.W., Beckett, C.A., Westbury, C.F., & Evans, A.C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 3172–3177.

- Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012a). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature Neuroscience*, *15*(4), 528–536.
- Zatorre, R.J., Delhommeau, K., & Zarate, J.M. (2012b). Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Frontiers in Psychology*, *3*, 1–17. doi: 10.3389/fpsyg.2012.00544.
- 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.
- Zevin, J.D. (2012). A sensitive period for shibboleths: the long tail and changing goals of speech perception over the course of development. *Developmental Psychobiology*, *54*(6), 632–642.
- 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.
- Zhou, X., & Merzenich, M. M. (2007). Intensive training in adults refines A1 representations degraded in an early postnatal critical period. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(40), 15935–15940.
- Zhou, X., de Villers-Sidani, E., Panizzutti, R., & Merzenich, M. (2010). Successive-signal biasing for a learned sound sequence. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(33), 14839–14844.

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.