RESEARCH ARTICLE

Enhanced voice recognition in musicians

Allison J. Sletcher, Stefania S. Moro[ID], Jennifer K. E. Steeves*

Department of Psychology and Centre for Vision Research, York University, Toronto, Canada

* steeves@yorku.ca

Abstract

Musicians typically have extensive auditory experience and demonstrate better pitch, timbre, and tempo discrimination compared to non-musicians. Musical training is also correlated with earlier and more robust cortical and subcortical responses to linguistic stimuli. We asked whether musical expertise may contribute to other auditory tasks, namely person and object recognition when both auditory and visual cues to identity are available. Musicians and non-musicians learned face-voice and car-horn "identity" pairs. Using a forced choice, old/new paradigm, participants were tested for recognition of the learned stimuli presented among distractor stimuli under three stimulus conditions (auditory only, visual only, and bimodal audiovisual). Compared to non-musicians, musicians were more sensitive at recognizing voices but not object sounds. Further, voice recognition sensitivity was positively correlated with both years of musical training and hours of weekly practice suggesting an influence of experience on performance. This differential performance for people and object stimuli is consistent with distinct neural substrates for face and object processing. Overall, this study demonstrates that experience in a sensory domain can benefit aspects of that sensory ability, such as voice but not object sound recognition, likely due to plasticity in distinct neural processing pathways.

Introduction

Humans can easily recognize people in an instant when seeing a familiar face or hearing a familiar voice. Recognition skills begin even before birth, when a fetus responds preferentially to their mother's voice compared to a stranger's [1]. Shortly after birth, infants respond preferentially to their mother's face [2] and show a preference for faces and voices over other stimuli [3, for review see 4]. Person recognition skills continue to develop through childhood and adolescence until fully mature in adulthood [5,6]. Individual differences in person recognition become more apparent in adulthood when recognition skills reach maturity [7].

Face and voice recognition ability can vary across individuals. On one hand, an individual may be unable to recognize face identity, gender, and facial expression (prosopagnosia) [8–12] or may have the inability to recognize voices (phonagnosia)

[13,14]. On the other hand, some individuals have superior face [15–17] and voice recognition skills [18,19]. Superior voice recognition may be associated with accentuated auditory skills and specialization of auditory cortical pathways through training. For example, highly trained forensic voice experts have superior voice discrimination skills when asked to identify a target voice from a "lineup" of voices [20,21]. Linguists, speech pathologists, and musicians who spend more time evaluating auditory stimuli than the average person may also become auditory specialists [19]. Musicians have established skills in pitch, timbre, and tempo of music [22,23]. Compared to non-musicians, musicians have superior ability to discriminate pitch and fundamental frequencies [24–26], timbre differences [24–27], and tempo [24,25] in both linguistic and musical stimuli.

These cues are common to both linguistic and voice processing and could provide a strong foundation for musicians to develop superior voice recognition. Speech and language tasks, such as speech-in noise recognition (i.e., "cocktail party" scenarios) may be improved by musical training, by strengthening shared resources [for review see 28] through increasing the listening capacity in both ideal acoustic conditions and also difficult acoustic environments [for review see 29]. Training studies on speech-in noise perception [28] and experience dependent plasticity in the auditory system [30–32] suggest that musical training can provide long lasting benefits to auditory function including simple perceptual enhancements and factors impacting higher order cognition such as working memory and intelligence [for review see 33,34]. However, genetic [35] and epigenetic factors (such as behavioural traits like personality [36], motivation [37], and the interaction between factors [38]) are also likely to contribute to musical and speech-in noise recognition [35].

Musical training appears to be associated with plasticity in both cortical and subcortical brain regions that process pitch, duration, and onset time of voice stimuli [for review see 39]. For example, when presented with linguistic pitch patterns musicians have enhanced and more accurate frequency encoding in the inferior colliculus [40]. It has also been shown that musicians have modulated inter-regional neural communication compared to non-musicians [41,42]. For example, in music/speech categorical processing musicians had increased activation in early primary auditory cortex whereas less experienced non-musicians had increased activation in downstream, higher-order linguistic brain areas such as the inferior frontal gyrus [41]. These findings indicate that cortical and subcortical processes may be under different learning timescales or constraints whether in terms of short-term experiences or for long-term experiences such as musical training. Recent evidence suggests that the enhanced frequency-following response (a scalp-recorded neuroelectric brain recording that serves as a neural index of sound encoding in EEG) may be impacted by myogenic responses, such as a postauricular muscle artifact which may contribute to the musician related differences observed in frequency-following responses [43].

Cortical and subcortical responses have been correlated to years of musical training [39,40,44] establishing a relationship with length of musical training. A relationship between musical training and speech processing at a cortical level has also been demonstrated, where enhanced and earlier cortical responses to syllabic duration

and voice onset time has also been measured in children [45,46]. Finally, years of musical training is positively correlated to faster learning of voice identification in a non-native language [47]. English and Mandarin speaking musicians have an advantage for voice identity recognition for unfamiliar languages that may be attributed to superior pitch processing abilities [48]. While years of music training relationships are suggestive of a training effect, recent studies have demonstrated that inherent, genetic predispositions may also contribute to differences in neural responses [49], speech processing behaviours [50], voice emotion recognition [51], the desire to pursue musical activities [52], and to commit to training for longer than less musically inclined peers [53, for review].

The specialization of auditory processing in musicians may also extend to sound recognition more broadly. Musicians demonstrate better auditory memory for spoken words and environmental sounds [54, for review see 55], but they do not demonstrate better visual memory for object categories [35,37]. Cohen and colleagues (2011) found musicians had better auditory memory for spoken words and environmental sounds compared to non-musicians for an object category recognition task. Typically, in object recognition studies, objects are presented from a range of object categories followed by a memory recognition task to investigate the perceptual processing of general object recognition as opposed to recognition of a specific object representation [54,56]. To date, there have been no studies investigating whether musicians have enhanced auditory sensitivity for specific object identity recognition. The current study asks whether musical expertise contributes to person and object recognition when both auditory and visual cues to identity are available.

Methods

Participants

Participants reported normal hearing, normal or corrected-to-normal visual acuity, and no health issues related to vision or hearing. Visual acuity was assessed with an EDTRS eye chart (Precision Vision™, La Salle, IL). All participants identified English as their first language or were early bilingual with English learned before the age of five years. Musicians and non-musicians were recruited through the York University Undergraduate Research Participant Pool (URPP) and the surrounding community. The recruitment period began on July 18, 2022 and ended on February 29, 2024. The study was approved by the Office of Research Ethics at York University and all participants provided informed consent.

Musicians. Thirty-five participants [mean age = 21 years (SD = 6 years); mean years of music training = 13 years (SD = 3 years); mean hours of weekly practice = 14 hours/week (SD = 11 hours/week)] self-identified as professional or semi-professional musicians. Musicians reported 8 years or more of musical training beginning before the age of ten years and participated in a regular weekly music practice.

Non-musicians. Thirty-five non-musicians participated as control participants. Non-musicians [mean age = 21 years (SD = 8 years); mean years of music training = 0.06 years (SD = 0.04 years)] self-identified as non-musicians, with less than one year of music training, and no current music practice.

Stimuli

All stimuli have been used previously in similar person and object recognition studies in patients with visual agnosia or unilateral eye enucleation [57,58]. See Hoover et al. [57] and Moro et al. [58] for more detailed stimulus information.

Person recognition task. Visual stimuli consisted of 50 greyscale images of female faces, cropped into an oval to eliminate hairlines, with all identifiable markers such as moles and nose rings digitally removed. All face images pictured in this manuscript have given written informed consent (as outlined in PLOS consent form) for use in publication. Auditory stimuli were 50 female voices speaking the same ten second neutral phrase in English. Each auditory clip was controlled for amplitude, background noise, and distinguishing markers such as long pauses and mispronounced words (average sound pressure level (SPL) = 50.6dB (44.2dB − 57.8dB)).

Object recognition task. Visual stimuli consisted of 50 greyscale cars presented from the same angle, with all identifiable markers such as ornaments, markings, and license plates removed. Auditory stimuli were 50 unique horn sounds. Each auditory clip was ten seconds in length and controlled for amplitude (average SPL = 53.5 dB (43.9dB − 59.6dB)).
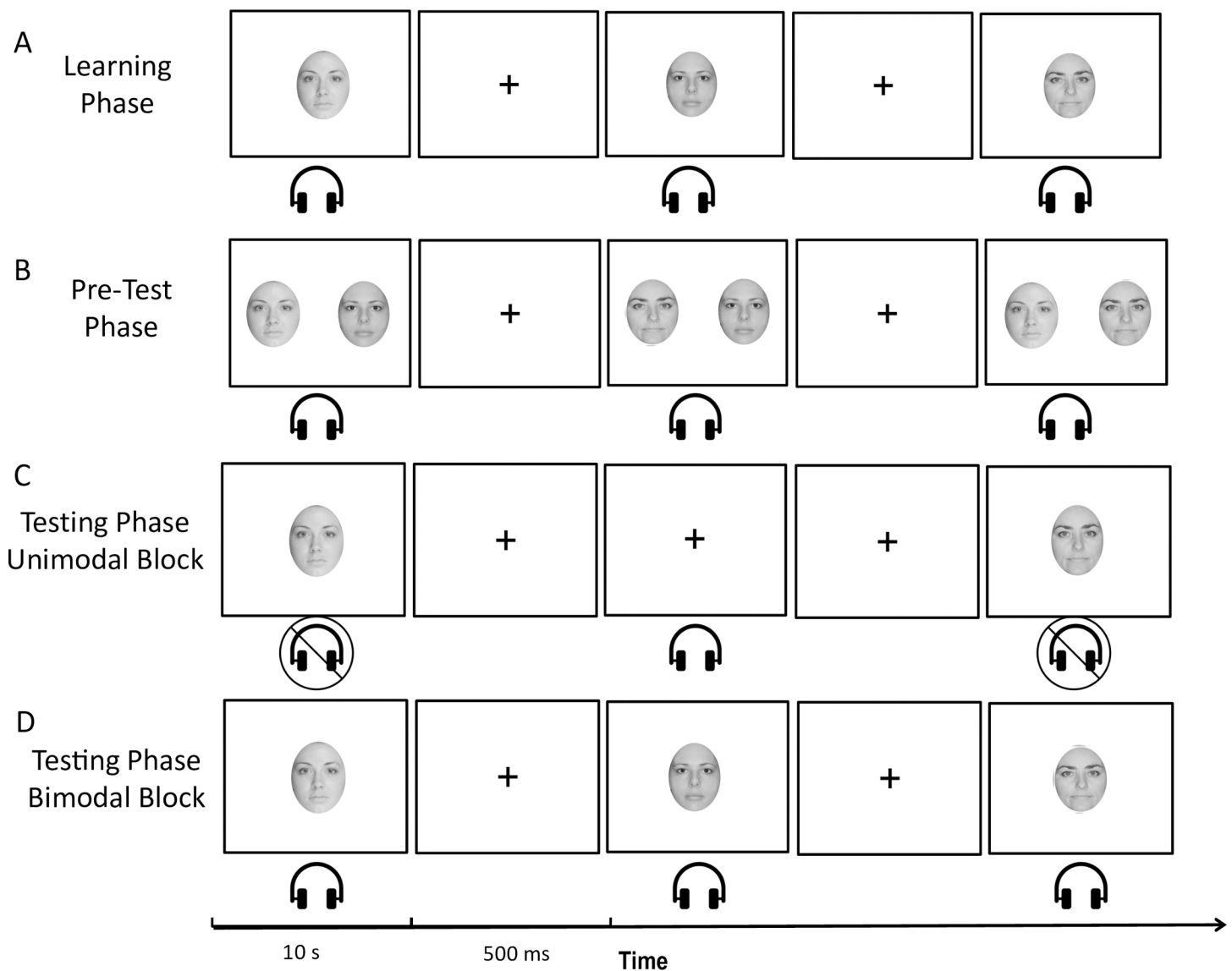
Procedure

Following the same paradigm as previous studies [57,58] stimuli were presented using Inquisit 6.6.0 [59] on a 23-inch computer display positioned 60 cm from the participant in a dimly lit room and using SONY noise cancelling headphones (Model #: MDR-ZX110NC). Participants responded with a choice of two designated keys on a computer keyboard. Person recognition trials consisted of ten faces paired with ten designated voices to be learned as an "identity" pair. Object recognition trials consisted of ten cars paired with ten designated horns to be learned as an "identity" pair. *Learning phase:* Participants were instructed that they would be tested on their ability to recognize the ten identities, made up of a unique face-voice or car-horn pair and that they had to learn to associate each specific face/car with its corresponding voice/horn, respectively. Identity pairs were presented for a total of four repetitions. Each presentation began with a fixation cross (500 ms) followed by the stimulus pair (10 s) (see Fig 1A and 2A). *Pre-test phase:* To practice their knowledge of recently learned identity pairs and to ensure an adequate level of learning of the paired identities, participants were presented with two learned visual stimuli on the screen and heard one learned auditory stimulus (see Fig 1B and 2B). They were instructed to press the corresponding key on the keyboard for the visual stimulus that was paired to the appropriate auditory stimulus that created an identity pair. Participants could not advance to the next trial until they responded with the correct key. Each identity pair was presented twice with an unrelated learned visual stimulus. *Testing phase:* Following the pre-test recognition was measured in two different blocks: 1. unimodal (visual only and auditory only) and 2. bimodal (face-voice or car-horn pairs). The unimodal block consisted of visual and auditory stimuli presented alone (see Fig 1C and 2C). Stimuli were presented in random order with learned stimuli presented twice and 20 new visual and 20 new auditory distractor stimuli for a total of 80 trials. Participants were asked to press a corresponding key on the keyboard if the stimulus was "learned" or "new". In the bimodal block, participants were presented with combined visual and auditory stimuli. This block consisted of two presentations of each learned identity pair and 30 distractor pairs. Four different stimuli combinations were presented: (1) a learned visual and learned auditory identity pair (congruent learned pair); (2) a new visual stimulus and a new auditory stimulus that are both unfamiliar (congruent new pair); (3) a learned visual stimulus with a new auditory stimulus (incongruent new pair); and (4) a learned auditory stimulus with a new visual stimulus (incongruent new pair) (see Fig 1D and 2D). Participants were asked to press a corresponding key on the keyboard if the stimulus pair was "learned" or "new". Key responses were counterbalanced across participants.

Results

Participant sensitivity scores were calculated for all conditions (visual only, auditory only, overall bimodal, congruent bimodal, and incongruent bimodal). Overall bimodal sensitivity scores were calculated based on both congruently and incongruently paired stimuli. Participants with perfect scores were adjusted based on recommendations by Macmillan and Creelman [61], by adjusting their score by a constant value (-0.025 in the case where participants scored 100). Approximately 10% of the data for the person identification task and 9% of the data for the object identification task were adjusted in this manner. All statistical analyses were conducted using IBM SPSS Statistics 29 and jamovi 2.3.28.

Person identity recognition performance

A 2x3 mixed-design analysis of variance (ANOVA) was performed on person identity recognition sensitivity scores with Group (Musicians vs Non-musicians) as the between-subject factor and the Stimulus Condition (Auditory vs Visual vs Bimodal) as the within-subject factor. Data met the assumptions of sphericity (W = 0.97, *p* = 0.37) and homogeneity of
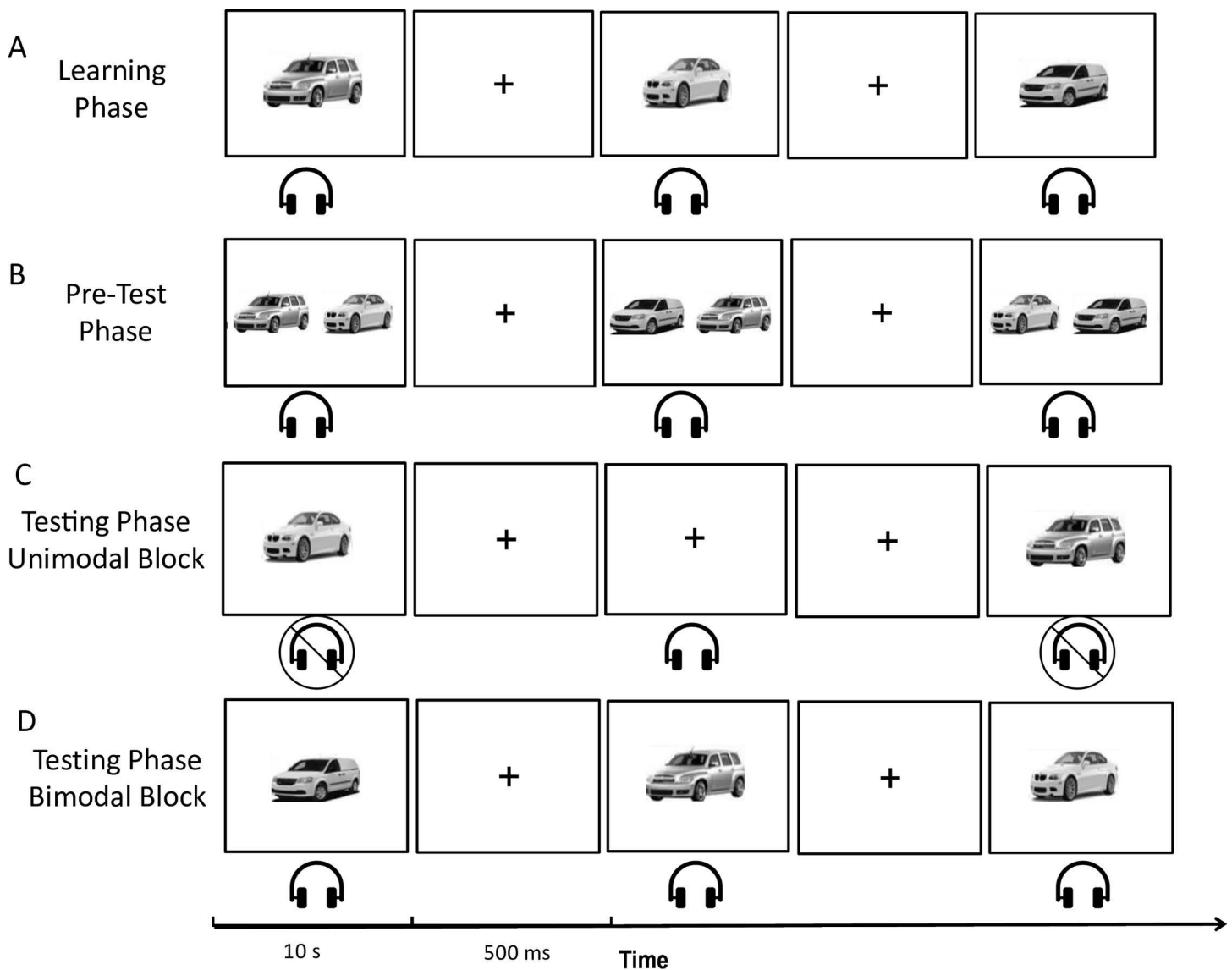
Fig 1. Schematic diagram representing the person identity recognition procedure adapted from Moro et al. [58] A. Learning phase. B. Pre-test phase. C. Testing phase, unimodal block. D. Testing phase, bimodal block.

https://doi.org/10.1371/journal.pone.0323604.g001

variances (Unimodal Visual: $F(1, 68) = 0.11$, $p = 0.74$; Unimodal Auditory: $F(1, 68) = 0.02$, $p = 0.89$; and Bimodal: $F(1,68) = 0.53$, $p = 0.47$) and therefore a parametric hypothesis test was conducted. There was a main effect of Group, $F(1, 68) = 5.40$, $p = 0.02$, $\eta_p^2 = 0.07$, where recognition sensitivity was higher for musicians compared to non-musicians (Fig 3). There was a main effect of Stimulus Condition, $F(2, 136) = 102.88$, $p < 0.001$, $\eta_p^2 = 0.60$. No significant interaction between Group and Stimulus Condition for person identity recognition was found, $F(2, 136) = 0.62$, $p = 0.54$, $\eta_p^2 = 0.009$.

To assess whether musicians are more sensitive to auditory, visual, or bimodal stimuli compared to non-musicians Bonferroni corrected post hoc comparisons were conducted between Groups for each Stimulus Condition. In the auditory condition, musicians ($M = 2.38$, $SD = 0.54$, 95% CI [2.20, 2.57]) had higher sensitivity compared to non-musicians ($M = 2.05$, $SD = 0.61$, 95% CI [1.84, 2.26]), $t(68) = 2.41$, $p = 0.02$, $d = 0.58$. There were no differences between groups in the bimodal

Fig 2. Schematic diagram representing the object identity recognition procedure adapted from Moro et al. [58] A. Learning phase. B. Pre-test phase. C. Testing phase, unimodal block. D. Testing phase, bimodal block. Car images in this Figure were taken from the open-access Stanford Cars Dataset [60] and are used for illustrative purposes.

condition (musicians: $M = 2.81$, SD = 0.62, 95% CI [2.60, 3.02]; non-musicians: $M = 2.62$, SD = 0.54, 95% CI [2.43, 2.80]), $t(68) = 1.40$, $p = 0.17$, $d = 0.33$, or the visual condition (musicians: $M = 3.44$, SD = 0.52, 95% CI [3.26, 3.62]; non-musicians: $M = 3.27$, SD = 0.57, 95% CI [3.08, 3.47]), $t(68) = 1.30$, $p = 0.20$, $d = 0.31$.

Bonferroni corrected post hoc comparisons were conducted to investigate the difference between Stimulus Conditions. Sensitivity was lower for auditory ($M = 2.21$, SD = 0.60, 95% CI [2.07, 2.36]) compared to visual stimuli ($M = 3.36$, SD = 0.55, 95% CI [3.22, 3.49]), $t(68) = -14.03$, $p < 0.001$, $d = -1.68$ and bimodal stimuli ($M = 2.71$, SD = 0.58, 95% CI [2.58, 2.85]), $t(68) = -5.93$, $p < 0.001$, $d = -0.71$. Visual sensitivity was higher compared to bimodal sensitivity, $t(68) = -8.85$, $p < 0.001$, $d = -1.06$. This indicates that performance was higher overall for the visual domain as participants were more sensitive to visual stimuli compared to auditory stimuli.
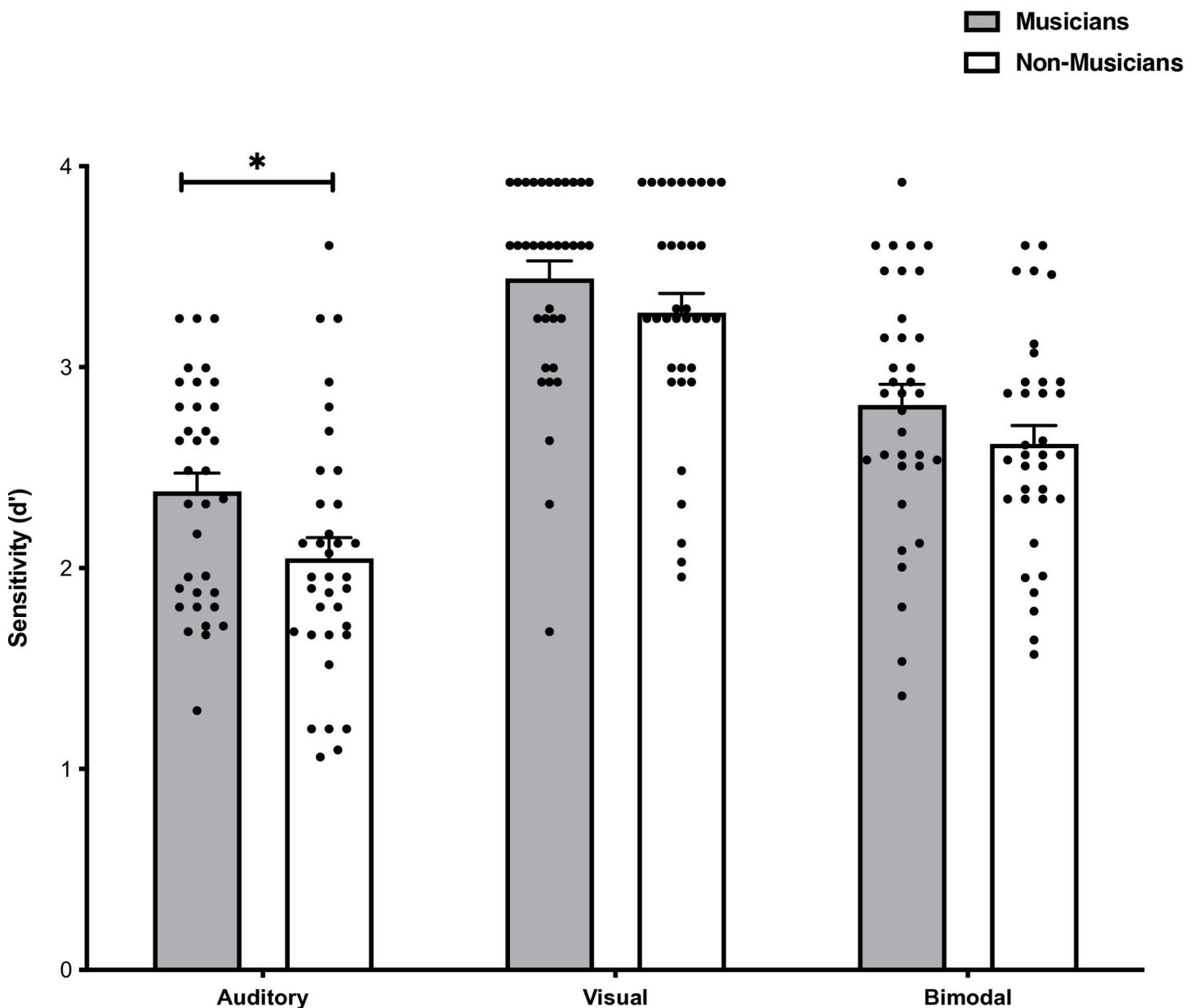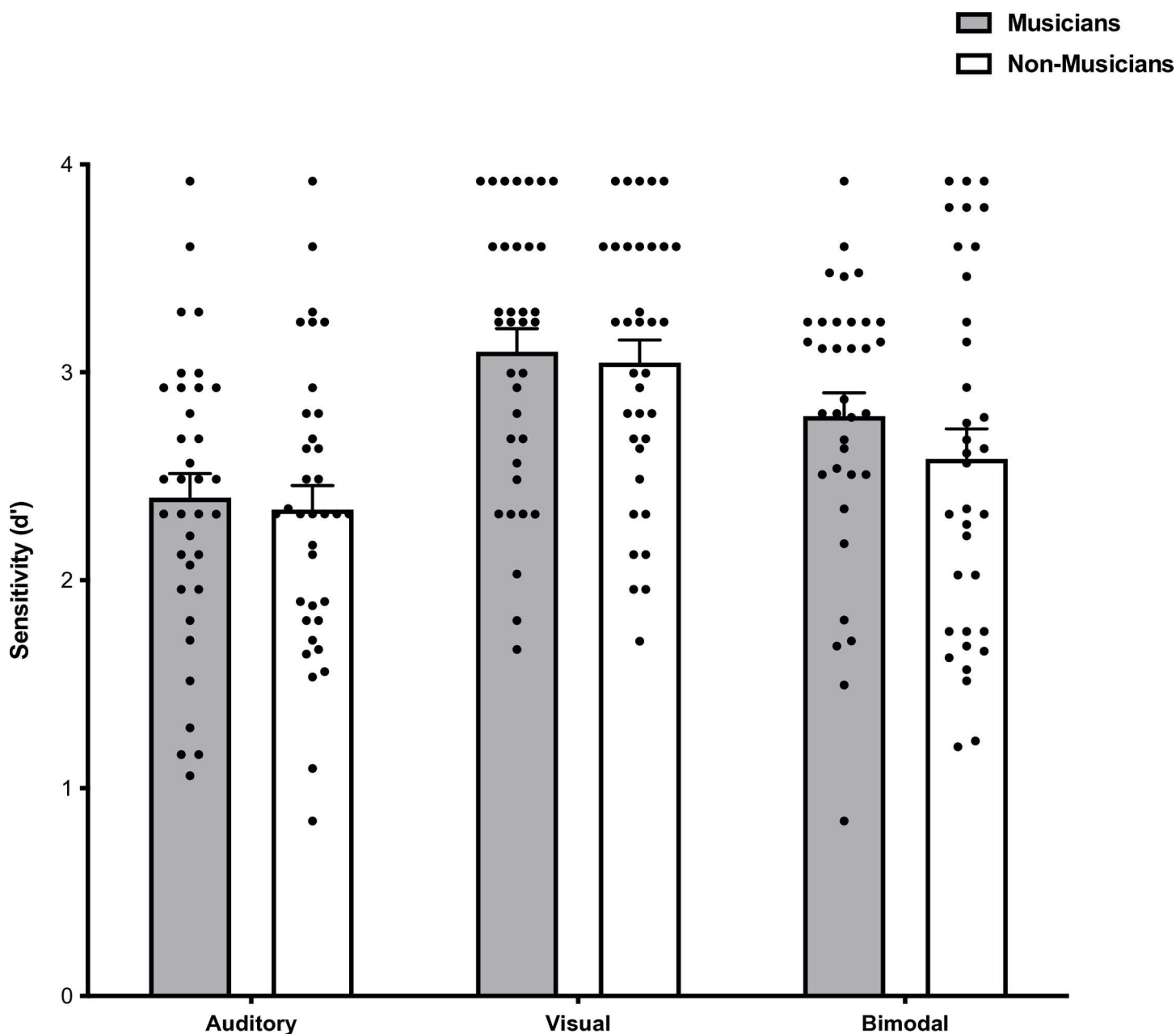
Fig 3. Person recognition sensitivity scores for each Stimulus Condition (Auditory vs Visual vs Bimodal) according to Group (Musicians vs Non-musicians). Musicians had better voice recognition sensitivity compared to non-musicians. Error bars represent the standard error of the mean.

Object identity recognition performance

A 2x3 mixed-design ANOVA was performed on object identity recognition sensitivity scores of between-subject Groups (Musicians vs Non-musicians) and within-subject Stimulus Condition (Visual vs Auditory vs Bimodal) (Fig 4). Data met the assumptions of sphericity (W = 0.986, p = 0.619) and homogeneity of variances (Unimodal Visual: $F(1, 68) = 0.006$, $p = 0.939$; Unimodal Auditory: $F(1, 68) = 0.003$, $p = 0.957$; and Bimodal: $F(1,68) = 4.427$, $p = 0.04$) and therefore a parametric hypothesis test was conducted. Unlike for person recognition, there were no Group differences for object recognition, $F(1, 68) = 0.61$, $p = 0.44$, $\eta_p^2 = 0.009$. Similar to person recognition, there was a main effect of Stimulus Condition,

Fig 4. Overall object recognition sensitivity scores for each Stimulus Condition (Auditory vs Visual vs Bimodal) according to Group (Musicians vs Non-musicians). No differences were found between groups in object recognition sensitivity across all modalities. Error bars represent standard error of the mean.

$F(2, 136) = 34.52$, $p < 0.001$, $\eta_p^2 = 0.34$, and no interaction between Group and Stimulus Condition, $F(2, 136) = 0.53$, $p = 0.59$, $\eta_p^2 = 0.008$.

Bonferroni corrected post hoc pairwise comparisons of Stimulus Condition indicated that sensitivity for the visual stimuli ($M = 3.07$, SD = 0.65, 95% CI [2.92, 3.23]), was higher compared to the auditory stimuli ($M = 2.37$, SD = 0.68, 95% CI [2.21, 2.53]), $t(68) = 8.41$, $p < 0.001$, $d = 1.01$, and the bimodal stimuli ($M = 2.69$, SD = 0.77, 95% CI [2.50, 2.87]), $t(68) = 4.77$,

$p < 0.001$, $d = 0.57$. Sensitivity for the bimodal stimuli was higher compared to the auditory stimuli, $t(68) = 3.56$, $p = 0.002$, $d = 0.42$. Similar to person recognition, performance was higher overall for the visual domain compared to the auditory and bimodal domains for object recognition across groups.

Relationship between musical experience and auditory sensitivity

Spearman rank-order correlation analyses were conducted to examine the relationship between musical experience and auditory sensitivity for musicians and non-musicians. The musical experience variable (Years of Musical Training) was not normally distributed therefore a non-parametric analysis was conducted (Non-musician: W = 0.25, $p < 0.001$; Musicians: W = 0.94, $p = 0.07$). There was a small positive correlation between years of training and performance sensitivity, where more training experience was related to higher sensitivity for voices in person recognition. Further, there was no relationship between years of training and object sound recognition across all stimulus condition modalities. See Table 1 for a breakdown of the correlation results.

Relationship between hours of weekly practice and auditory sensitivity

Spearman rank-order correlation analyses were conducted to examine the relationship between hours of weekly musical practice and auditory sensitivity for musicians and non-musicians. The hours of weekly musical practice variable was not normally distributed therefore a non-parametric correlation analysis was conducted (Non-musician: W < 0.001, $p < 0.001$; Musicians: W = 0.830, $p < 0.001$). There was a moderate positive correlation between hours of weekly practice and performance sensitivity, where increased hours of weekly practice was related to higher sensitivity for voices in person recognition. Further, there was no relationship between hours of weekly practice and object sound recognition across all stimulus condition modalities. See Table 1 for detailed Spearman correlation results.

Comparison across voices and horns

To determine whether there was an effect of the type of auditory stimulus (Voices or Horns) we conducted parametric paired samples t-tests to compare the auditory sensitivity scores of Musicians and Non-musicians across the person identity and object identity tasks (Fig 5). Musicians did not differ in their auditory sensitivity for voices ($M = 2.38$, SD = 0.54, 95% CI [2.20, 2.57]) compared to object sounds ($M = 2.40$, SD = 0.66, 95% CI [2.16, 2.63]), $t(34) = -0.125$, $p = 0.90$, $d = -0.02$. Non-musicians demonstrated a strong trend for enhanced auditory sensitivity for object sounds ($M = 2.34$, SD = 0.69, 95% CI [2.10, 2.58]) compared to voices ($M = 2.05$, SD = 0.61, 95% CI [1.84, 2.26]), $t(34) = -2.031$, $p = 0.05$, $d = -0.34$.

Table 1. Spearman rank-order correlations examining the relationship between years of training and hours of weekly practice for auditory person and object identity recognition sensitivity. Significant correlations are indicated with an asterisk (p-values are not adjusted for multiple comparisons).

	Auditory Sensitivity	
	ρ	p
	Years of Training	
Person Identity Recognition	0.26	0.03*
Object Identity Recognition	0.05	0.67
	Hours of Weekly Practice	
Person Identity Recognition	0.30	0.01*
Object Identity Recognition	0.07	0.57

https://doi.org/10.1371/journal.pone.0323604.t001
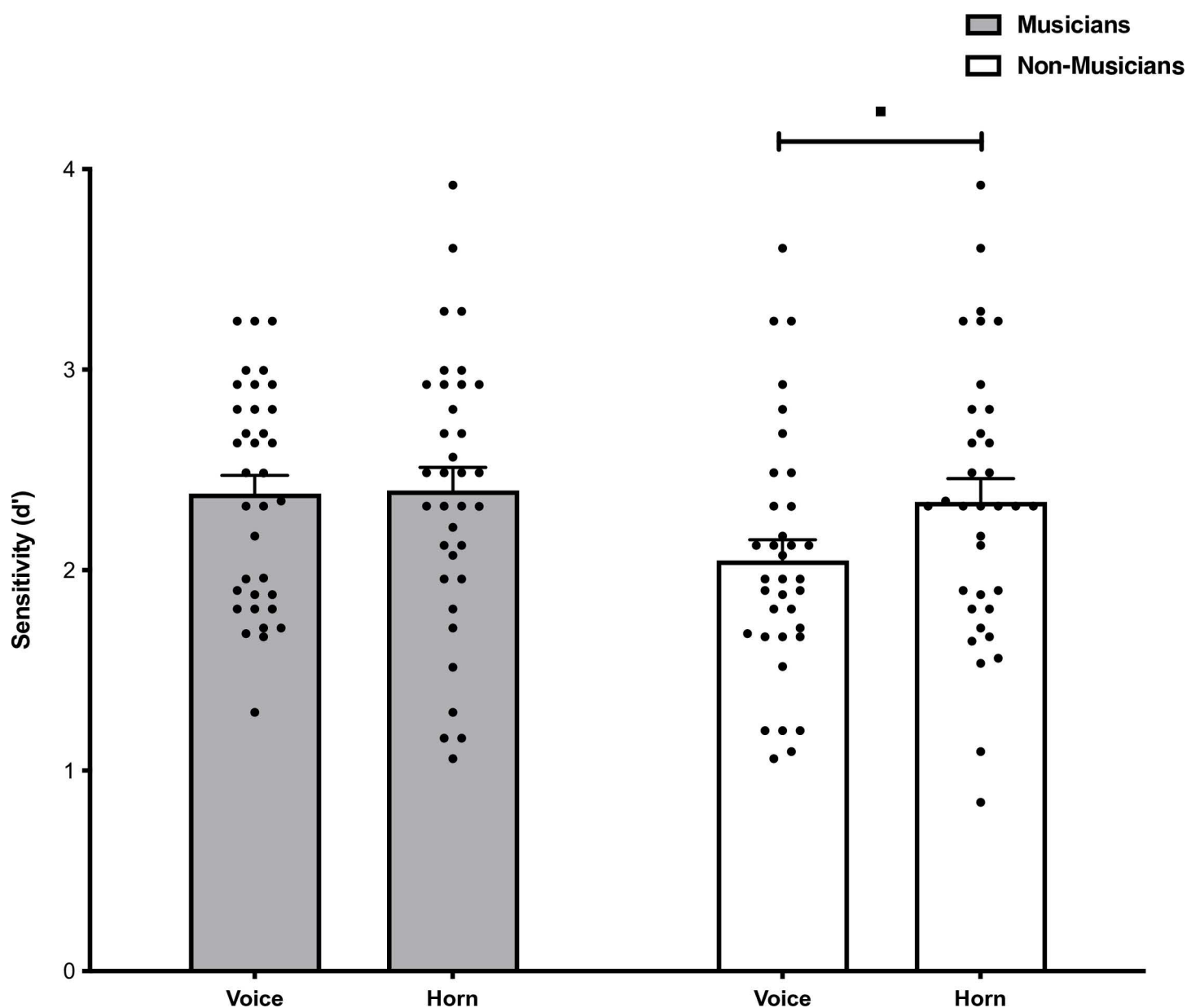
Fig 5. Auditory only stimulus sensitivity in the person (face/voice) and object (car/horn) identity tasks for Musicians (grey bars) and Non-musicians (white bars). Error bars represent standard error of the mean.

Discussion

We demonstrated that musicians from our sample have better sensitivity for recognition of voices compared to non-musicians from our sample and that this ability is related to the extent of experience as measured by hours of weekly practice and years of training. These findings suggest that learned skill in auditory processing through musical specialization contributes to driving group differences between musicians and non-musicians. Further, our study demonstrated that the enhancement of auditory sensitivity for voice perception is not replicated in auditory sensitivity for object sound

perception. This indicates that person identity processing is a unique process and that the auditory advantage of musical specialization may not broadly translate to other domains such as object recognition, specifically for car-horn identification. Finally, our study demonstrated that better unimodal auditory sensitivity for voices is positively correlated with years of musical experience, as well as hours of practice indicating that learned skill and expertise impact voice recognition for musicians.

These findings suggest that musicians have specialized auditory skills that transfer to voice recognition and are aligned with previous studies demonstrating that musicians have enhanced recognition of timbre, pitch, and fundamental frequencies [24–26,62], which may be supported by enhanced cortical and subcortical processing of auditory information [39,40,44]. On one hand, musicians enhanced auditory sensitivity was restricted to voices in the person recognition task and did not translate to enhanced auditory sensitivity in the object identity recognition task. On the other hand, non-musicians show a strong trend for better sensitivity to object sounds compared to voices, which aligns with previously reported data [57]. These results indicate that musical expertise may only benefit certain aspects of sound processing. This is consistent with the notion of distinct neural substrates for object and person processing [11,57,63–69]. Our findings are consistent with prior studies of patients demonstrating that person compared to object identity recognition can be differentially affected by clinical vision disorders [57,58,70]. Clinical case studies have shown that lesions to discrete regions within visual cortex can lead to specific deficits in face recognition (prosopagnosia) [10–12,57,70,71], and in object recognition (object agnosia) [72–74]. Similarly in the auditory domain, clinical case studies have shown lesions in auditory cortex can result in impaired recognition of melodic tunes or tone deafness (amusia) and voice recognition (phonagnosia) while sparing environmental sounds recognition [75]. Our data are consistent with such clinical cases [75,76].

Voice recognition sensitivity was positively correlated with both years of training and hours of weekly practice indicating an influence of experience through a learned skill. Both behaviour and brain response differences in musicians is related to years of musical training. For example, faster voice processing in unfamiliar languages and enhanced cortical and subcortical responses to auditory stimuli has also been correlated to years of musical training in musicians [39,40,44,47]. These relationships indicate that the development of auditory expertise through experience can influence neuroplastic changes in the brain. Future research should evaluate the neural correlates of enhanced auditory sensitivity for voices compared to object sounds in musicians.

In conclusion, compared to non-musicians, musicians have enhanced sensitivity for voice identity recognition but not object sound recognition. This auditory specialization is positively correlated with both years of musical training and hours of weekly practice indicating an influence of experience through a learned skill. Overall, this study demonstrates the contribution of musical experience towards enhancement of a specific sensory ability, such as voice but not object sound recognition likely due to plasticity in distinct neural processing pathways.

Author contributions

Conceptualization: Allison J. Sletcher, Stefania S. Moro, Jennifer K. E. Steeves.

Data curation: Allison J. Sletcher.

Formal analysis: Allison J. Sletcher.

Funding acquisition: Jennifer K. E. Steeves.

Investigation: Allison J. Sletcher.

Methodology: Stefania S. Moro, Jennifer K. E. Steeves.

Project administration: Stefania S. Moro, Jennifer K. E. Steeves.

Resources: Jennifer K. E. Steeves.

Supervision: Stefania S. Moro, Jennifer K. E. Steeves.

Validation: Stefania S. Moro, Jennifer K. E. Steeves.

Visualization: Stefania S. Moro.

Writing – original draft: Allison J. Sletcher.

Writing – review & editing: Stefania S. Moro, Jennifer K. E. Steeves.

References

1. Kisilevsky BS, Hains SMJ, Brown CA, Lee CT, Cowperthwaite B, Stutzman SS, et al. Fetal sensitivity to properties of maternal speech and language. Infant Behav Dev. 2009;32(1):59–71. https://doi.org/10.1016/j.infbeh.2008.10.002 PMID: 19058856

2. Pascalis O, de Schonen S, Morton J, Deruelle C, Fabre-Grenet M. Mother's face recognition by neonates: A replication and an extension. Infant Behavior and Development. 1995;18(1):79–85. https://doi.org/10.1016/0163-6383(95)90009-8

3. Johnson MH, Dziurawiec S, Ellis H, Morton J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition. 1991;40(1–2):1–19. https://doi.org/10.1016/0010-0277(91)90045-6 PMID: 1786670

4. Mercure E, Kischkel L. Social Perception in Infancy: An Integrative Perspective on the Development of Voice and Face Perception. In Frühholz S, Belin P, (Eds.). The Oxford Handbook of Voice Perception. Oxford University Press; 2018.

5. Bennetts RJ, Murray E, Boyce T, Bate S. Prevalence of face recognition deficits in middle childhood. Q J Exp Psychol (Hove). 2017;70(2):234–58. https://doi.org/10.1080/17470218.2016.1167924 PMID: 26999413

6. Germine LT, Duchaine B, Nakayama K. Where cognitive development and aging meet: face learning ability peaks after age 30. Cognition. 2011;118(2):201–10. https://doi.org/10.1016/j.cognition.2010.11.002 PMID: 21130422

7. Dalrymple KA, Garrido L, Duchaine B. Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. Dev Cogn Neurosci. 2014;10:10–20. https://doi.org/10.1016/j.dcn.2014.07.003 PMID: 25160676

8. Damasio AR, Damasio H, Van Hoesen GW. Prosopagnosia: anatomic basis and behavioral mechanisms. Neurology. 1982;32(4):331–41. https://doi.org/10.1212/wnl.32.4.331 PMID: 7199655

9. Damasio A, Tranel D, Damasio H. Face agnosia and the neural substrates of memory. Annu Rev Neurosci. 1990;13:89–109.

10. Ellis HD, Florence M. Bodamer's (1947) paper on prosopagnosia. Cogn Neuropsychol. 1990;7:81–105.

11. Steeves J, Dricot L, Goltz HC, Sorger B, Peters J, Milner AD, et al. Abnormal face identity coding in the middle fusiform gyrus of two brain-damaged prosopagnosic patients. Neuropsychologia. 2009;47(12):2584–92. https://doi.org/10.1016/j.neuropsychologia.2009.05.005 PMID: 19450613

12. Steeves JKE, Culham JC, Duchaine BC, Pratesi CC, Valyear KF, Schindler I, et al. The fusiform face area is not sufficient for face recognition: evidence from a patient with dense prosopagnosia and no occipital face area. Neuropsychologia. 2006;44(4):594–609. https://doi.org/10.1016/j.neuropsychologia.2005.06.013 PMID: 16125741

13. Garrido L, Eisner F, McGettigan C, Stewart L, Sauter D, Hanley JR, et al. Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia. 2009;47(1):123–31. https://doi.org/10.1016/j.neuropsychologia.2008.08.003 PMID: 18765243

14. Van Lancker DR, Cummings JL, Kreiman J, Dobkin BH. Phonagnosia: a dissociation between familiar and unfamiliar voices. Cortex. 1988;24(2):195–209. https://doi.org/10.1016/s0010-9452(88)80029-7 PMID: 3416603

15. Bobak AK, Bennetts RJ, Parris BA, Jansari A, Bate S. An in-depth cognitive examination of individuals with superior face recognition skills. Cortex. 2016;82:48–62. https://doi.org/10.1016/j.cortex.2016.05.003 PMID: 27344238

16. Robertson DJ, Noyes E, Dowsett AJ, Jenkins R, Burton AM. Face Recognition by Metropolitan Police Super-Recognisers. PLoS One. 2016;11(2):e0150036. https://doi.org/10.1371/journal.pone.0150036 PMID: 26918457

17. Russell R, Duchaine B, Nakayama K. Super-recognizers: people with extraordinary face recognition ability. Psychon Bull Rev. 2009;16(2):252–7. https://doi.org/10.3758/PBR.16.2.252 PMID: 19293090

18. Aglieri V, Watson R, Pernet C, Latinus M, Garrido L, Belin P. The Glasgow Voice Memory Test: Assessing the ability to memorize and recognize unfamiliar voices. Behav Res Methods. 2017;49(1):97–110. https://doi.org/10.3758/s13428-015-0689-6 PMID: 26822668

19. Schweinberger SR, Zäske R. Perceiving speaker identity from the voice. In Frühholz S, Belin P, (Eds.). The Oxford Handbook of Voice Perception. Oxford University Press; 2018.

20. Eladd E, Segev S, Tobin Y. Long-term working memory in voice identification. Psychology, Crime & Law. 1998;4(2):73–88. https://doi.org/10.1080/10683169808401750

21. Schiller NO, Köster O. The ability of expert witnesses to identify voices: a comparison between trained and untrained listeners. IJSLL. 1998;5(1):1–9. https://doi.org/10.1558/ijsll.v5i1.1

22. Baumann O, Belin P. Perceptual scaling of voice identity: common dimensions for different vowels and speakers. Psychol Res. 2008;74(1):110–20. https://doi.org/10.1007/s00426-008-0185-z PMID: 19034504

23. Latinus M, Belin P. Human voice perception. Curr Biol. 2011;21:143–5.

24. Deguchi C, Boureux M, Sarlo M, Besson M, Grassi M, Schön D, et al. Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: behavioral, electrophysiological and psychoacoustic study. Brain Res. 2012;1455:75–89. https://doi.org/10.1016/j.brainres.2012.03.034 PMID: 22498174

25. Hutka S, Bidelman GM, Moreno S. Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. Neuropsychologia. 2015;71:52–63. https://doi.org/10.1016/j.neuropsychologia.2015.03.019 PMID: 25797590

26. Tervaniemi M, Just V, Koelsch S, Widmann A, Schröger E. Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. Exp Brain Res. 2005;161(1):1–10. https://doi.org/10.1007/s00221-004-2044-5 PMID: 15551089

27. Rammsayer T, Altenmüller E. Temporal Information Processing in Musicians and Nonmusicians. Music Perception. 2006;24(1):37–48. https://doi.org/10.1525/mp.2006.24.1.37

28. Alain C, Zendel B, Hutka S, Bidelman G. Turning down the noise: the benefit of musical training on the aging auditory brain. Hear Res. 2014;308:162–73.

29. Coffey EBJ, Mogilever NB, Zatorre RJ. Speech-in-noise perception in musicians: A review. Hear Res. 2017;352:49–69. https://doi.org/10.1016/j.heares.2017.02.006 PMID: 28213134

30. Pantev C, Herholz SC. Plasticity of the human auditory cortex related to musical training. Neurosci Biobehav Rev. 2011;35(10):2140–54. https://doi.org/10.1016/j.neubiorev.2011.06.010 PMID: 21763342

31. de Villers-Sidani E, Simpson KL, Lu Y-F, Lin RCS, Merzenich MM. Manipulating critical period closure across different sectors of the primary auditory cortex. Nat Neurosci. 2008;11(8):957–65. https://doi.org/10.1038/nn.2144 PMID: 18604205

32. Bidelman GM, Alain C. Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. J Neurosci. 2014;35(3):1240–9. https://doi.org/10.1523/JNEUROSCI.3292-14.2015 PMID: 25609638

33. Moreno S, Bidelman GM. Examining neural plasticity and cognitive benefit through the unique lens of musical training. Hear Res. 2014;308:84–97. https://doi.org/10.1016/j.heares.2013.09.012 PMID: 24079993

34. Herholz SC, Zatorre RJ. Musical training as a framework for brain plasticity: behavior, function, and structure. Neuron. 2012;76(3):486–502. https://doi.org/10.1016/j.neuron.2012.10.011 PMID: 23141061

35. Schellenberg EG. Music training and speech perception: a gene-environment interaction. Ann N Y Acad Sci. 2015;1337:170–7. https://doi.org/10.1111/nyas.12627 PMID: 25773632

36. Corrigall KA, Schellenberg EG, Misura NM. Music training, cognition, and personality. Front Psychol. 2013;4:222. https://doi.org/10.3389/fpsyg.2013.00222 PMID: 23641225

37. McAuley JD, Henry MJ, Tuft S. Musician Advantages in Music Perception: An Issue of Motivation, Not Just Ability. Music Perception. 2011;28(5):505–18. https://doi.org/10.1525/mp.2011.28.5.505

38. Anderson S, Kraus N. Auditory Training: Evidence for Neural Plasticity in Older Adults. Perspect Hear Hear Disord Res Res Diagn. 2013;17:37–57. https://doi.org/10.1044/hhd17.1.37 PMID: 25485037

39. Kraus N, Chandrasekaran B. Music training for the development of auditory skills. Nat Rev Neurosci. 2010;11(8):599–605. https://doi.org/10.1038/nrn2882 PMID: 20648064

40. Wong PCM, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci. 2007;10(4):420–2. https://doi.org/10.1038/nn1872 PMID: 17351633

41. Bidelman GM, Walker B. Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network. Neuroimage. 2019;201:116022. https://doi.org/10.1016/j.neuroimage.2019.116022 PMID: 31310863

42. Coffey EBJ, Herholz SC, Chepesiuk AMP, Baillet S, Zatorre RJ. Cortical contributions to the auditory frequency-following response revealed by MEG. Nat Commun. 2016;7:11070. https://doi.org/10.1038/ncomms11070 PMID: 27009409

43. Bidelman GM, Sisson A, Rizzi R, MacLean J, Baer K. Myogenic artifacts masquerade as neuroplasticity in the auditory frequency-following response. Fron Neurosci. 2024; https://doi.org/10.3389/fnins.2024.1422903

44. Musacchia G, Strait D, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear Res. 2008;241(1–2):34–42. https://doi.org/10.1016/j.heares.2008.04.013 PMID: 18562137

45. Moreno S, Marques C, Santos A, Santos M, Castro SL, Besson M. Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. Cereb Cortex. 2009;19(3):712–23. https://doi.org/10.1093/cercor/bhn120 PMID: 18832336

46. Chobert J, François C, Velay J-L, Besson M. Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. Cereb Cortex. 2014;24(4):956–67. https://doi.org/10.1093/cercor/bhs377 PMID: 23236208

47. Bregman MR, Creel SC. Gradient language dominance affects talker learning. Cognition. 2014;130(1):85–95. https://doi.org/10.1016/j.cognition.2013.09.010 PMID: 24211437

48. Xie X, Myers E. The impact of musical training and tone language experience on talker identification. J Acoust Soc Am. 2015;137(1):419–32. https://doi.org/10.1121/1.4904699 PMID: 25618071

49. Mankel K, Bidelman GM. Inherent auditory skills rather than formal music training shape the neural encoding of speech. Proc Natl Acad Sci U S A. 2018;115(51):13129–34. https://doi.org/10.1073/pnas.1811793115 PMID: 30509989

50. Swaminathan S, Schellenberg EG. Musical competence and phoneme perception in a foreign language. Psychon Bull Rev. 2017;24(6):1929–34. https://doi.org/10.3758/s13423-017-1244-5 PMID: 28204984

51. Correia AI, Castro SL, MacGregor C, Müllensiefen D, Schellenberg EG, Lima CF. Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. Emotion. 2022;22(5):894–906. https://doi.org/10.1037/emo0000770 PMID: 32718172

52. Wesseldijk LW, Mosing MA, Ullén F. Why is an early start of training related to musical skills in adulthood? A genetically informative study. Psyc Sci. 2021; https://doi.org/10.1177/0956797620959014

53. Schellenberg EG, Lima CF. Music Training and Nonmusical Abilities. Annu Rev Psychol. 2024;75:87–128. https://doi.org/10.1146/annurev-psych-032323-051354 PMID: 37738514

54. Cohen MA, Evans KK, Horowitz TS, Wolfe JM. Auditory and visual memory in musicians and nonmusicians. Psychon Bull Rev. 2011;18(3):586–91. https://doi.org/10.3758/s13423-011-0074-0 PMID: 21374094

55. Talamini F, Altoè G, Carretti B, Grassi M. Musicians have better memory than nonmusicians: A meta-analysis. PLoS One. 2017;12(10):e0186773. https://doi.org/10.1371/journal.pone.0186773 PMID: 29049416

56. Rodrigues AC, Loureiro M, Caramelli P. Visual memory in musicians and non-musicians. Front Hum Neurosci. 2014;8:424. https://doi.org/10.3389/fnhum.2014.00424 PMID: 25018722

57. Hoover AEN, Démonet J-F, Steeves JKE. Superior voice recognition in a patient with acquired prosopagnosia and object agnosia. Neuropsychologia. 2010;48(13):3725–32. https://doi.org/10.1016/j.neuropsychologia.2010.09.008 PMID: 20850465

58. Moro SS, Hoover AEN, Steeves JKE. Short and long-term visual deprivation leads to adapted use of audiovisual information for face-voice recognition. Vision Res. 2019;157:274–81. https://doi.org/10.1016/j.visres.2018.01.009 PMID: 29567099

59. Millisecond. Inquisit (6.6.0.). Millisecond Software; 2022. https://www.millisecond.com

60. Krause J, Stark M, Deng J, Fei-Fei L. 4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013 (3dRR-13). Sydney, Australia; 2013.

61. Macmillan NA, Creelman CD. Detection theory: a user's guide. Taylor & Francis Group; 2004.

62. Chartrand J-P, Belin P. Superior voice timbre processing in musicians. Neurosci Lett. 2006;405(3):164–7. https://doi.org/10.1016/j.neulet.2006.06.053 PMID: 16860471

63. Solomon-Harris LM, Mullin CR, Steeves JKE. TMS to the "occipital face area" affects recognition but not categorization of faces. Brain Cogn. 2013;83(3):245–51. https://doi.org/10.1016/j.bandc.2013.08.007 PMID: 24077427

64. Duchaine B, Nakayama K. The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia. 2006;44(4):576–85. https://doi.org/10.1016/j.neuropsychologia.2005.07.001 PMID: 16169565

65. Hildebrandt A, Wilhelm O, Herzmann G, Sommer W. Face and object cognition across adult age. Psychol Aging. 2013;28(1):243–8. https://doi.org/10.1037/a0031490 PMID: 23527744

66. Shakeshaft N, Plomin R. Genetic specificity of face recognition. Proc Natl Acad Sci USA. 2015;112:12887–92.

67. Wilmer JB, Germine L, Chabris CF, Chatterjee G, Williams M, Loken E, et al. Human face recognition ability is specific and highly heritable. Proc Natl Acad Sci U S A. 2010;107(11):5238–41. https://doi.org/10.1073/pnas.0913053107 PMID: 20176944

68. Wilmer JB, Germine L, Chabris CF, Chatterjee G, Gerbasi M, Nakayama K. Capturing specific abilities as a window into human individuality: the example of face recognition. Cogn Neuropsychol. 2012;29(5–6):360–92. https://doi.org/10.1080/02643294.2012.753433 PMID: 23428079

69. Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 2001;293(5539):2425–30. https://doi.org/10.1126/science.1063736 PMID: 11577229

70. Föcker J, Best A, Hölig C, Röder B. The superiority in voice processing of the blind arises from neural plasticity at sensory processing stages. Neuropsychologia. 2012;50(8):2056–67. https://doi.org/10.1016/j.neuropsychologia.2012.05.006 PMID: 22588063

71. Duchaine B, Nakayama K. Dissociations of face and object recognition in developmental prosopagnosia. J Cogn Neurosci. 2005;17(2):249–61. https://doi.org/10.1162/0898929053124857 PMID: 15811237

72. Humphreys GW, Rumiati RI. Agnosia without prosopagnosia or alexia: evidence for stored visual memories specific to objects. Cogn Neuropsychol. 1998;15(3):243–77. https://doi.org/10.1080/026432998381177 PMID: 28657515

73. McMullen PA, Fisk JD, Phillips SJ, Maloney WJ. Apperceptive agnosia and face recognition. Neurocase. 2000;6(5):403–14. https://doi.org/10.1080/13554790008402711

74. Moscovitch M, Winocur G, Behrmann M. What Is Special about Face Recognition? Nineteen Experiments on a Person with Visual Object Agnosia and Dyslexia but Normal Face Recognition. J Cogn Neurosci. 1997;9(5):555–604. https://doi.org/10.1162/jocn.1997.9.5.555 PMID: 23965118

75. Peretz I, Kolinsky R, Tramo M, Labrecque R, Hublet C, Demeurisse G, et al. Functional dissociations following bilateral lesions of auditory cortex. Brain. 1994;117 (Pt 6):1283–301. https://doi.org/10.1093/brain/117.6.1283 PMID: 7820566

76. Xu X, Biederman I, Shilowich BE, Herald SB, Amir O, Allen NE. Developmental phonagnosia: Neural correlates and a behavioral marker. Brain Lang. 2015;149:106–117.