UNIVERSITY OF LISBON

FACULTY OF PSYCHOLOGY

EXPLORING ASSOCIATIONS BETWEEN MUSICAL SOPHISTICATION AND EMOTION RECOGNITION IN INDIVIDUALS DIFFERING IN MUSICAL ABILITIES

Alexandre Manuel Vieira dos Santos

Dissertação orientada pela Professora Doutora Ana P. Pinheiro

e coorientada pelo Professor Doutor César F. Lima

elaborada para a obtenção do grau de Mestre em Neuropsicologia Clínica e Experimental

INTERUNIVERSITY MASTER IN CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY (IMCEN)

2021

Agradecimentos

À Professora Doutora Ana Pinheiro, desde logo pela disponibilidade para orientar esta Tese de Mestrado e, sobretudo, pelo esforço de coordenação que mantivemos desde cedo e ao qual sempre correspondeu. A sua orientação foi determinante para ultrapassar certos obstáculos que fui encontrando ao longo da minha caminhada. Do mesmo modo, ao Professor Doutor César Lima, que me aceitou coorientar e partilhar a sua *expertise* sobre o tema em questão.

À Marta Pereira pelas tutorias e materiais fornecidos, que tanto me ajudaram a compreender em que consistia o Projeto MUSE e como deveria organizar o meu trabalho a partir daí, além do *know-how* tão necessário às recolhas EEG/comportamentais. Ressalto também o seu profissionalismo, sentido de missão e exemplar dedicação, que permitiram ao Projeto dar passos decisivos e atingir metas ambiciosas. Aos participantes, que tive oportunidade de conhecer, pela sua colaboração e perseverança. Aos meus colegas de laboratório agradeço todo o apoio prestado, particularmente àqueles que me possibilitaram participar em sessões de recolhas experimentais e contribuíram para as minhas aprendizagens no âmbito das atividades realizadas no VoicES Lab.

Aos meus pais, Margarida Vieira e Manuel Santos, que tanto estimo e amo profundamente, por me terem dado todas as condições de principiar e concluir este mestrado. Estou-lhes humildemente grato para todo o sempre pelas oportunidades que proporcionaram e continuam a proporcionar. Por me terem protegido e saído do caminho quando era preciso. Por me deixarem cometer erros e aprender com eles. Por me deixarem viver os meus sonhos e aspirações. Por serem o que são e às vezes muito mais. Por me amarem sem igual.

Entre amigos e amigas, um especial e emotivo abraço ao João Almeida e à Ana Sofia Patatas por sempre me terem dado forças para seguir de queixo levantado, principalmente nas alturas mais difíceis e em que não acreditava em mim mesmo. Por sinal, os nossos momentos juntos são alimento para a alma. *Pun intended*.

À Ana Rita, por ter vivido comigo algumas das horas mais felizes dos meus últimos dois anos. Um beijo de despedida.

Index

Abstract

Music training and musical sophistication have been shown to be associated with nonmusical domains of cognition. Many authors argue that these associations demonstrate experience-driven neuroplasticity. This study attempts to replicate and lend additional support to the idea that both trained and untrained individuals with higher self-reported musical abilities can achieve similar performance levels in three identical emotion recognition tasks. In the current study, participants (*N* = 31) with different scores on the Goldsmiths Musical Sophistication Index had to identify the emotion that best characterized each exemplar of faces, prosody or nonverbal vocalizations in separate tasks, using one of seven response options representing emotional labels. Contrary to the hypothesis, we did not find a significant association between musical sophistication and average accuracy scores in vocal emotion recognition tasks. Also, as predicted, we found no association between musical sophistication and recognition accuracy in the facial emotion task. An exploratory analysis revealed an association between musical sophistication and recognition of fear. These mixed findings are partly in line with the musical sophistication/musical training literature. Possible methodological pitfalls are discussed. Future studies are to continually improve on the current body of work with different techniques and research methods.

Keywords: Neuroplasticity; musical sophistication; emotion recognition; facial expressions; speech prosody; nonverbal vocalizations

Resumo Alargado

A formação musical ao longo da vida e as aprendizagens a ela associadas apresentam-se como bons modelos para estudar a plasticidade do cérebro humano, na medida em que a prática de um instrumento, por exemplo, requer um conjunto de faculdades cognitivas que, por sua vez, estão relacionadas com determinadas estruturas e funções cerebrais implicadas no processamento de conteúdo cognitivo-afetivo. A especificidade deste tipo de processamento (com treino vs. sem treino musical) tem sido analisada em estudos de imagiologia cerebral, eletrofisiológicos e comportamentais que revelam efeitos de facilitação numa variedade de processos cognitivos de âmbito geral (e.g., funções executivas e inteligência) e particular (e.g., processamento da fala, linguagem escrita, capacidades visuoespaciais, etc). Deste modo, os investigadores argumentam que os correlatos encontrados constituem casos onde a neuroplasticidade é orientada pela experiência, através de um mecanismo que denominam de *far transfer* ("transferência longa"), uma vez que ocorre normalmente entre um domínio musical (e.g., aulas de canto/treino vocal) e um domínio não-musical (e.g., cognitivo). Por oposição, a *near transfer* ("transferência curta") ocorre no mesmo domínio, no qual a aptidão aprendida tem aplicação imediata na tarefa experimental (e.g., aprender acordes e tocá-los no instrumento).

Ora, certos autores (Sala & Gobet, 2020) admitem que a formação musical não promove melhorias nas faculdades cognitivas ou na excelência académica, visto que a maioria dos estudos analisados são transversais e o impacto é nulo. Nestes casos, a validade ecológica pode ser posta em causa, uma vez que estes estudos usaram metodologias distintas e nem sempre as mais adequadas, variando quase sempre ao nível do delineamento da intervenção (e.g., individual ou grupal), das idades dos participantes e da operacionalização das variáveis representativas do treino musical e das faculdades cognitivas em foco. Além disso, a evidência de estudos longitudinais é escassa e apresenta igualmente um conjunto de limitações, entre as quais a não aleatorização dos participantes, a não utilização de condições controlo equivalentes ao treino musical e a falta de medidas de controlo adequadas (e.g., traços de personalidade, estatuto socioeconómico e indicadores de funcionamento cognitivo). Contudo, outros autores (Bigand & Tillmann, 2021) reanalisaram *esses* dados e demonstraram efeitos de transferência "modestos", mas estatisticamente significativos, quando compararam os resultados das condições experimentais e de controlo de estudos de *transferências longas*.

Assim sendo, a plasticidade induzida pelo treino poderá não explicar todas as diferenças entre músicos e indivíduos sem formação musical. De facto, a experiência musical ou os conhecimentos musicais são muitas vezes aferidos pela capacidade de tocar um instrumento musical e o nível de proficiência exibido, enquanto que outras capacidades que dizem respeito à relação do indivíduo com a música não são tidas em conta. Na literatura, esta relação é definida pelo conceito de sofisticação musical ("musical sophistication"; Mullensiefen et al., 2014), que engloba comportamentos e competências musicais numa diversidade de dimensões como o envolvimento ativo com a música (e.g., quanto tempo e recursos monetários são investidos em música), competências percetivas (e.g., audição musical), treino musical (e.g., treino musical formal recebido), competências de canto (e.g., performance vocal durante o canto), e envolvimento emocional com a música (e.g., capacidade de falar sobre emoções expressas pela música). Por conseguinte, é possível identificar estes comportamentos/competências na população em geral, e não só exclusivamente em músicos. Acresce que certos autores (e.g., Martins et al., 2021) reforçam que fatores preexistentes como predisposições genéticas podem favorecer o surgimento de competências musicais naturalmente boas em indivíduos sem treino musical. De facto, estudos recentes têm mostrado que boas competências de perceção musical em indivíduos sem treino musical estão relacionadas com melhores desempenhos em domínios não musicais (e.g., Mankel & Bidelman, 2018; Swaminathan & Schellenberg, 2017). Outros fatores, nomeadamente ambientais, como as faculdades cognitivas, a personalidade e o estatuto socioeconómico parecem, no entanto, predizer diferenças individuais no treino musical (Corrigall et al., 2013; Swaminathan & Schellenberg, 2018).

Nas últimas duas décadas, tem havido interesse crescente em investigar aspetos do processamento socioemocional, nomeadamente a capacidade de reconhecer emoções através da voz e faces. Embora a música esteja intimamente ligada com processos emocionais e sociais, o estudo de efeitos de transferência para estes domínios tem conhecido tímidos avanços, sobretudo no que toca à sofisticação musical.

O presente estudo teve como objetivo usar medidas explícitas de reconhecimento emocional para explorar associações entre a sofisticação musical, avaliada com recurso à adaptação portuguesa do Gold-MSI (Lima et al., 2020), e o reconhecimento de vozes e faces, com base na precisão das respostas dos participantes (em *Hu scores*; Wagner, 1993). Nesse sentido, três tarefas idênticas foram administradas a todos os participantes.

Para cada tarefa, os participantes ou viam ou ouviam exemplares pré-selecionados de estímulos de três categorias (vocalizações não verbais, prosódia de discurso e expressões faciais) e tinham de identificar a qualidade emocional do estímulo apresentado com um de sete rótulos que representavam as seis emoções (raiva, nojo, medo, alegria, tristeza e neutralidade) e a opção "nenhuma das anteriores", para quando a voz/face não expressasse nenhum dos estados emocionais atrás mencionados. Os participantes também responderam ao questionário Gold-MSI e a outros relativos a informação sociodemográfica e a sintomas de COVID-19, que, quando presentes, poderiam levar ao cancelamento da sessão experimental. Quanto às hipóteses, nós esperávamos que a sofisticação musical estivesse associada ao reconhecimento emocional de vozes, fosse para a prosódia de discurso ou para as vocalizações não verbais. Especificamente, em linha com Correia e colaboradores (2020), os participantes que reportassem maiores competências percetivas teriam um melhor desempenho no reconhecimento de vozes. Também esperávamos que a sofisticação musical não estivesse associada ao reconhecimento de expressões faciais. Ainda explorámos a possibilidade da sofisticação musical estar associada ao reconhecimento de emoções específicas para cada tarefa de reconhecimento.

As análises de correlação não revelaram nenhuma associação entre a sofisticação musical e o reconhecimento de vozes e faces. Além disso, a associação encontrada para as faces foi negativa, ao contrário da nossa predição, sendo apenas verificada parcialmente. Assim, apesar de sustentarem a ideia de que não existe uma vantagem clara no reconhecimento emocional de faces para indivíduos com valores elevados no Gold-MSI, estes resultados não mostraram a relação previamente identificada entre sofisticação musical e desempenho nas tarefas de reconhecimento de vozes (ver Correia et al., 2020). Por outro lado, a análise exploratória revelou uma correlação marginalmente significativa entre o *Envolvimento Ativo* (i.e., subescala do Gold-MSI) e o reconhecimento de medo. Houve uma correlação particularmente forte entre o reconhecimento do medo na tarefa da prosódia de discurso e o envolvimento ativo com a música que parece ter contribuído de forma decisiva para que a anterior estivesse mais saliente.

Em suma, esta investigação propôs-se explorar correlações entre os dados da escala de autoavaliação Gold-MSI e as médias da precisão de resposta (corrigida quanto a potenciais enviesamentos) em três tarefas de reconhecimento emocional variando apenas no que toca ao estímulo usado. Os resultados não mostraram melhorias de

desempenho nos indivíduos que pontuaram tendencialmente mais alto no Gold-MSI, tanto para as tarefas de voz como para a de faces. Poderá ser importante no futuro utilizar métodos complementares para se compreender as variáveis de interesse e as relações entre elas de forma extensiva. Por exemplo, técnicas de EEG poderão ajudar a descrever o decurso temporal do processamento emocional para cada tipo de estímulo, enquanto que as de fMRI poderão localizar as regiões ativas durante o reconhecimento emocional e como esses padrões de conectividade variam de acordo com o tipo de emoção. Posto isto, defendemos que este estudo tem relevância na literatura da sofisticação musical.

Palavras chaves: Neuroplasticidade; sofisticação musical; reconhecimento emocional; expressões faciais; prosódia de discurso; vocalizações não verbais

1. Introduction

1.1. The neuroplasticity hypothesis of musical training

Lifelong music training and extensive music experience provide unique opportunities for studying the plasticity of the human brain. Musical training as a multisensory experience involving a complexity of cognitive abilities and underlying neural networks relates to changes in brain function and structure that affect both cognitive and affective processing (Barrett et al., 2013). Changes have not only been linked to specific cognitive domains such as speech segmentation (François et al., 2013), pitch processing of spoken language (Schön et al., 2004), phonological awareness (Degé & Schwarzer, 2011), syntax processing (Jentschke & Koelsch, 2009), higher-level language abilities (e.g., reading; Anvari et al., 2002), verbal memory (Franklin et al., 2008), visual attention (Roden et al., 2014) and other visuospatial abilities (Hetland, 2000), mathematic (Vaughn, 2000) abilities, but also to domain-general cognitive abilities such as intelligence (Swaminathan, Schellenberg, & Khalil, 2017) and executive functions (Frischen, Schwarzer, & Degé, 2019).

These apparent advantages in nonmusical abilities leads authors to assume they happen as a result of far transfer, the assumption that training in one domain could generalize to other distantly related domains (e.g., between music and verbal memory; for a review, Miendlarzewska & Trost, 2014). This implies evidence of learning and cross domain plasticity (Herholz & Zatorre, 2012), even though most studies do not examine longitudinal training to establish causal effects. In fact, a recent meta-analysis by Sala and Gobet (2020) claims that music training does not cause improvements in any domain-general cognitive skill or academic achievement. However, results were obtained from children in educational programs, either receiving training vs. no training or training in other areas. Ecological validity is, thus, matter of discussion as it depends on the study design (i.e., individual vs group-based lessons), age-appropriate interventions and on the degree to which training reflects measures of the cognitive benefits (Swaminathan & Schellenberg, 2016).

Over the years researchers have been proposing distinct explanations for how the transfer occurs: (1) some suggest that training affects executive functions, namely working memory, which in turn improves general cognitive abilities (for a review, Criscuolo et al., 2019; Schellenberg & Peretz, 2008); (2) another view holds that musical training shapes auditory brainstem response to speech by generating high-fidelity copies of auditory stimuli (Kraus & Chandrasekaran, 2010); (3) OPERA hypothesis (Patel, 2011) posits that greater subcortical encoding of speech happens in musically trained individuals when there is a neural overlap (O) between a given speech and music skill, the music skill demands precise (P) auditory processing, the music training elicits positive emotion (E), and when the musical activities require repetition (R) and focused attention (A); and (4) theories of overlap in temporal processing for music and speech that favor rhythm-based music interventions for training language skills (Gosvvarni, 2011; Tallal & Gaab, 2006). All these explanatory theories have been used as frameworks to account for associations between musical training and higher-level cognitive abilities (e.g., Moreno & Bidelman, 2014; Miendlarzewska & Trost, 2014).

1.2. Individual differences in musical sophistication

So, training-induced plasticity might not fully explain differences between musicians and individuals without training (also termed 'nonmusicians') as most effect sizes found are small according to Hattie (2008)'s barometer of influence (Bigand & Tillmann, 2021). Musical sophistication is used to describe the multi-faceted nature of 'musical experience' or 'musical expertise', concept that is often reduced to the ability to play an instrument and the degree of proficiency of performing musicians, while other skills tending to musical predispositions and engagement are ignored (Levitin, 2012). Musical sophistication is comprised of "musical skills, expertise, achievements, and related behaviors" (Mullensiefen et al., 2014, p. 2) across multiple dimensions namely active musical engagement, self-reported perceptual abilities, musical training, self-reported singing abilities and sophisticated emotional engagement with music. Therefore, a variety of musical skills and behaviors are likely to be observed in the general population, and not exclusively in musicians. Manifestations of this repertoire include informal music listening, acquired functional uses of music (i.e., regulate arousal and mood, achieve self-awareness, and as an expression of social relatedness; Schäfer et al., 2013) or even humming and singing along with lyrical musical tunes (Radocy & Boyle, 2012). Musical behaviors and abilities are present in all human societies and may be predicted by several genetic (e.g., Tan et al., 2014) and environmental variables (e.g., cognitive abilities, personality and socioeconomic status; Corrigall et al., 2013; Swaminathan & Schellenberg, 2018).

Recently, a cross-sectional investigation explored the relationship between musical sophistication and cognitive performance in a balanced sample of 35 musicians and 35 nonmusicians (Porflitt & Rosas, 2020). Among other cognitive variables, musical sophistication correlated with executive functions such as working memory and cognitive inhibition, and fluid intelligence, an IQ estimate. The main finding was that musical sophistication explained more variance than other variables in the model (i.e., age, socioeconomic status and laterality) that are generally associated with cognitive performance and executive functions (e.g., Best et al., 2009; Hackman et al., 2015; Hirnstein et al., 2010). These results show that, like musical training, musical sophistication might contribute to domain-general skills via executive functions. Indeed, Bonetti and Costa (2017) found that fluid intelligence was associated with performance in music tasks in 4- to 6-year-old children with no musical training. Hence, musical sophistication and intelligence might share a relation independent of musical training.

Another correlational study probing musical sophistication in auditory processing (Dawson et al., 2017) reported enhanced behavioral pitch discrimination in musically oriented Finnish speakers, as yielded by Goldsmiths Musical Sophistication Index (Gold-MSI; Mullensiefen et al., 2014) generalized scores. McKenzie and collaborators (2019) further support and extend this finding in that musical sophistication correlated with pitch discrimination among English speakers. Meanwhile, research on a new Musical Emotion Discrimination Task used Gold-MSI subscales musical training, emotional music skills and active engagement with music, and showed a significant relationship between them and pitch discrimination (MacGregor & Müllensiefen, 2019). This collective evidence indicates that music sophistication might impact higher-order auditory processing, similarly to music training, by fine-tuning auditory perception abilities that subserve sensory aspects of voice extraction.

1.3.Emotion recognition in voice and faces

A different perspective is that the affective experience associated with music is the principal component for engaging in musical activities (Juslin and Laukka, 2004). In that sense, emotional skills might be responsible for maintaining a typical level of emotional engagement with music. Therefore, it is possible that perception of musical emotions relies on general measures of emotional ability. Akkermans and colleagues (2019) showed an influence of self-reported emotional engagement with music (Gold-MSI subscale) on accuracy rates in a musical emotion decoding task. As noted earlier, because pitch encoding in acoustical processing is fundamental for auditory perception, it seems likely that differences in auditory perception may also contribute to recognition ability in both speech and music. This possibility is reinforced by a study of individuals with pitch processing difficulties who exhibited lower sensitivity to vocal and facial expressions, thus affecting their judgements of perceived emotions (Lima et al., 2016).

Communication of basic emotions in both music and vocal expressions is processed by certain domain-specific modules that react autonomously to acoustic features present in the stimulus (e.g., speed, pitch range, and timbre; see Juslin & Laukka, 2004) (Fodor, 1983). As such, a verbal vocalization might be perceived as *angry* if it is reproduced at a fast rate, with loud intensity and harsh timbre, whereas a musical composition might impart the *same* emotion if it has those characteristics. So, emotional contagion might explain why listeners often become *moved* by musical and vocal expressions via similarities in the acoustic cues (e.g., Juslin, Harmat, & Eerola, 2014). Incidentally, there is mounting evidence of emotional contagion with both vocal (e.g., Hawk, Fischer, & Van Kleef, 2012; Hietanen, Surakka, & Linnankoski, 1998; Lopes, 2015) and facial expressions (e.g., Aldunate & González-Ibáñez, 2017; Egermann & McAdams, 2012; Manera, Grandi, & Colle, 2013), thus pointing to a putative mechanism of emotion induction and recognition in these modalities.

Most research in vocal expressions have focused on nonverbal emotional cues in the human voice (in terms of pitch, loudness, tempo, rhythm, and timbre; for a review, Grandjean, Bänziger, & Scherer, 2006), the so-called emotional speech prosody, and purely nonverbal vocalizations, such as laughter and crying. Because nonverbal vocalizations do not contain any linguistic information, they enable a more primitive type of communication. Hence, for this reason it is thought that emotional prosody and nonverbal vocalizations engage different articulatory and perceptual mechanisms (for reviews see Liebenthal, Silbersweig, & Stern, 2016; Scott, Sauter, & McGettigan, 2010). Still, several studies with adult listeners indicated that they can accurately identify a selection of positive and negative emotions from both emotional prosody and nonverbal vocalizations, heard in isolation and regardless of semantics (e.g., Castro & Lima, 2010; Lima, Castro, & Scott, 2013; Pell, Monetta, Paulmann, & Kotz, 2009). Furthermore, other studies asserted that emotion recognition is more accurate for nonverbal vocalizations than for emotional prosody (Hawk et al., 2009; Lausen & Hammerschmidt, 2020; Sauter, 2006).

According to a critical review by Martins, Pinheiro and Lima (2021), the state of art suggests an association between music training and emotion recognition abilities in vocal expressions including emotional prosody (e.g., Lima & Castro, 2011; Thompson, Schellenberg, & Husain, 2004) and nonverbal vocalizations (e.g., Correia et al., 2020; Parsons et al., 2014). In another sense, the authors emphasize that other pre-existing factors such genetic predispositions might endow untrained individuals with naturally good musical abilities. In a twin study, Mosing and colleagues (2014) showed variability in music practice is heritable up to 40%-70%, and that when genetic propensity was held constant, more practice did not translate into improvements in musical abilities. This might suggest that genetic and other environmental factors parallel to musical training can potentially account for differences between trained musicians and untrained individuals. That is apparent in lines of research that seek to reproduce the advantages of musicians in musical and non-musical domains in untrained individuals. Indeed, correlational studies found musician-like encoding of noise-degraded speech sounds in untrained adults with higher music perception abilities (Mankel & Bidelman, 2018), an association between natural musical abilities and phoneme perception consistent with benefits documented in musicians (see Swaminathan & Schellenberg, 2017; Swaminathan & Schellenberg, 2020), and a positive relationship between self-report measures of music perception and vocal emotion recognition, even when music training is controlled for (Correia et al., 2020). In addition, musicians and untrained participants with higher musical abilities were equally good at recognizing vocal expressions. Meanwhile, no advantage was reported for emotion recognition in facial expressions though only few studies attempted to capture this effect (e.g., Correia et al., 2020; Farmer, Jicol, & Petrini, 2020). Yet, cross-sectional examinations cannot ascertain causality and, thus, caution is advised when interpretating cross-sectional data. Besides, the shortcomings of the existing longitudinal evidence (i.e., no randomization of participants, non-active controls and lack of adequate control measures – personality traits, socioeconomic status and cognitive functioning indexes) did not allow to verify long-term effects linked to musical training and/or musical abilities. Also, the findings cited above highlight the need to consider informal music engagement and related factors in certain cognitive domains.

1.4. The current study and hypotheses

In the current study, we used an explicit measure of emotion recognition in which the stimuli presented (i.e., nonverbal vocalizations, speech prosody and facial expressions) and the different emotional states conveyed by them (i.e., anger, disgust, fear, happiness, sadness, neutral) were manipulated to investigate the relationship between musical sophistication, assessed using a Portuguese adaptation of Gold-MSI (Lima et al., 2020), and emotion recognition in voice and faces.

To this end, three identical tasks were conducted. For each task, participants either heard or saw pre-selected stimuli from three categories (nonverbal vocalizations, speech prosody and facial expressions) and had to categorize the emotional quality of the stimuli based on seven *emoticons* representing six emotional states (anger, disgust, fear, happiness, sadness, neutral) and the option "none of the above", for when the voice/face did not express any of the previous emotional states.

We predicted that musical sophistication would be associated with increased vocal emotion recognition, both for speech prosody and nonverbal vocalizations (Hypothesis 1). Specifically, in line with Correia and collaborators (2020), participants who reported greater perceptual abilities in Gold-MSI are expected to perform better at recognizing emotions in speech prosody and nonverbal vocalizations. Musical sophistication has been shown to correlate with auditory perception in musical beat perception (Müllensiefen et al., 2014), mistuning perception (Larrouy-Maestri, Harrison, & Müllensiefen, 2019) and, more recently, Timbre Perception Test (TPT; Lee & Müllensiefen, 2020) and Pitch Imagery Arrow Task (PIAT; see Gelding et al., 2021). Furthermore, some individuals with naturally good auditory abilities have superior neural processing and perception of speech (Mankel & Bidelman, 2018; Swaminathan & Schellenberg, 2017). Mankel and collaborators (2020) corroborate the assumption that, in the absence of music training, inherent auditory skills impart improvements to categorical speech processing, which may explain experience-driven neuroplasticity effects.

We also predicted musical sophistication would not translate into performance advantage in emotional face recognition (Hypothesis 2). A number of studies in which facial recognition accuracy did not differ between musicians and nonmusicians seem to suggest higher-order benefits of musical training might be restricted to the auditory domain (e.g., Correia et al., 2020; Farmer & Petrini, 2020; Twaite, 2016; Weijkamp &

Sadakata, 2017). It might owe to the fact that transfer effects of musical training on enhanced cognitive abilities are still relatively weak (Sala & Gobet, 2017). Therefore, we admit the possibility that natural musical abilities might influence vocal and facial emotion recognition in dissimilar ways.

An additional exploratory analysis addressed whether musical sophistication would be associated with the recognition of specific emotions.

2. Method

2.1. Participants

Twenty-one musicians (12 male), ranging from 18-54 years (*M* = 30.44, *SD* = 8.70), and 10 musically naïve controls (three male), ranging from 19-36 years (*M* = 23, *SD* = 2.45), participated in the current study. A female musician was excluded from subsequent analyses for not completing all experimental tasks. The musicians were recruited using a convenience sampling approach and word of mouth from professional orchestras, music schools and others (Orquestra Metropolitana de Lisboa, Academia Nacional Superior de Orquestra, Escola Profissional da Metropolitana, Conservatório de Música da Metropolitana, Orquestra Sinfónica Juvenil, Coro Gulbenkian; freelancer jobs such as specialist in music therapy, music teacher and private instrument teacher). The controls were college students and some of them received course credits for their participation. They had no music training aside from music classes within the school curriculum, while musicians had at least 4 years of formal music training (ranging from 4-24 years; *M* = 11, *SD* = 5.35). Musicians were instrumentalists who played guitar (n = 4), piano (n = 2), violin (n = 2), accordion (n = 1), cello (n = 1), clarinet (n = 1), drums (n = 1), and organ (n = 1); in addition to instrumental training, 11 of them had vocal training in classical singing. All the participants were native European Portuguese speakers, right-handed or ambidextrous, had no history of neurological illness and were not taking any medication for neurological indication, had not received electroconvulsive treatment, had normal hearing and normal (or corrected to normal) vision. Hand dominance was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). Ethical approval for the study was obtained from the Ethics Committee, Faculty of Psychology, University of Lisbon. All participants signed a written consent form before the experiments were conducted.

2.2. Stimuli

The stimuli used for the emotion recognition task consisted of exemplars of three types of stimuli – speech prosody, nonverbal vocalizations, and facial expressions. Twelve exemplars of six emotion categories (angry, disgust, fear, sadness, happiness and neutral) were selected for each stimulus type. The stimuli were retrieved from previously validated corpora (speech prosody, Castro & Lima, 2010; nonverbal vocalizations, Lima, Castro, & Scott, 2013; facial expressions, Karolinska Directed Emotional Faces database, Goeleven, De Raedt, Leyman, & Verschuere, 2008). Speech prosody stimuli containing one of nine emotionally neutral semantic contents (e.g., “O quadro está na parede”, The painting is on the wall) were short sentences (*M* = 1386 ms , *SD* = 234 ms) with different intonations, produced by two female speakers. Nonverbal vocalizations were short vocal bursts (*M* = 971 ms, *SD* = 285 ms) devoid of verbal content (e.g., laughs), produced by both male and female speakers. Faces were colored photographs of facial expressions by male and female actors posing with no beards, moustaches, earrings, eyeglasses, or visible make-up. Whenever possible, stimuli selection was informed by recognition accuracy in validation studies of the different corpora (speech prosody, Castro & Lima, 2010; nonverbal vocalizations, Lima, Castro, & Scott, 2013; facial expressions, Goeleven, De Raedt, Leyman, & Verschuere, 2008), to match recognition accuracy across stimulus type (85,17% for vocalizations, 75,28% for speech prosody and 76,32% for facial expressions). To decrease the likelihood of ceiling effects, stimuli with the highest accuracy were not selected. The auditory stimuli were normalized to 70 dB of intensity with a Praat script (Praat; Boersma, 2001).

2.3. Materials

2.3.1. Goldsmiths Musical Sophistication Index (Gold-MSI)

A Portuguese adaptation of the Gold-MSI was utilized (Gold-MSI-P; Lima et al., 2020). The Gold-MSI is a self-report measure of musical sophistication, "a psychometric construct that can refer to musical skills, expertise, achievements, and related behaviors across a range of facets that are measured on different subscales" (Mullensiefen et al., 2014).

The Gold-MSI-P encompasses 38 items. For 31, participants rate their level of agreement or disagreement with statements (e.g., “Consigo cantar ou tocar música de memória.”) using a Likert-type scale from 1 (completely disagree) to 7 (completely agree). For the remaining seven items, using ordinal scales with seven response alternatives (e.g., “Sei tocar [number from 0 to ‘6 ou mais’] instrumentos musicais”) participants respond according to the musical behavior assessed. The items are grouped into five subscales: active engagement, nine items; perceptual abilities, nine items; musical training, seven items; singing abilities, seven items; and emotions, six items. A general factor of musical sophistication is based on 18 items from all the five subscales. Hence, Gold-MSI can be used as a multidimensional instrument, covering different facets of musical behavior, or as a unidimensional one, tapping musical sophistication more broadly. An additional supplementary question assesses the instrument participants play best including voice. Internal consistency across all subscales is estimated to range from 0.82 to 0.91 (i.e., good or excellent reliability) and all test–retest correlations from 0.84 to 0.94 (i.e., very high).

2.4. Procedure

Participants filled out the Gold-MSI and other online questionnaires regarding demographic information and COVID-19 symptomatology on Qualtrics software, Version June 2021 (Qualtrics, Provo, UT, USA). Then, they were tested individually in a quiet room at the Voice, Emotion, and Speech (VoicES, University of Lisbon) Neuroscience Lab, following distancing and disinfection protocols to reduce risk of contagion of COVID-19. The test period occurred from September 2020 to December 2020.

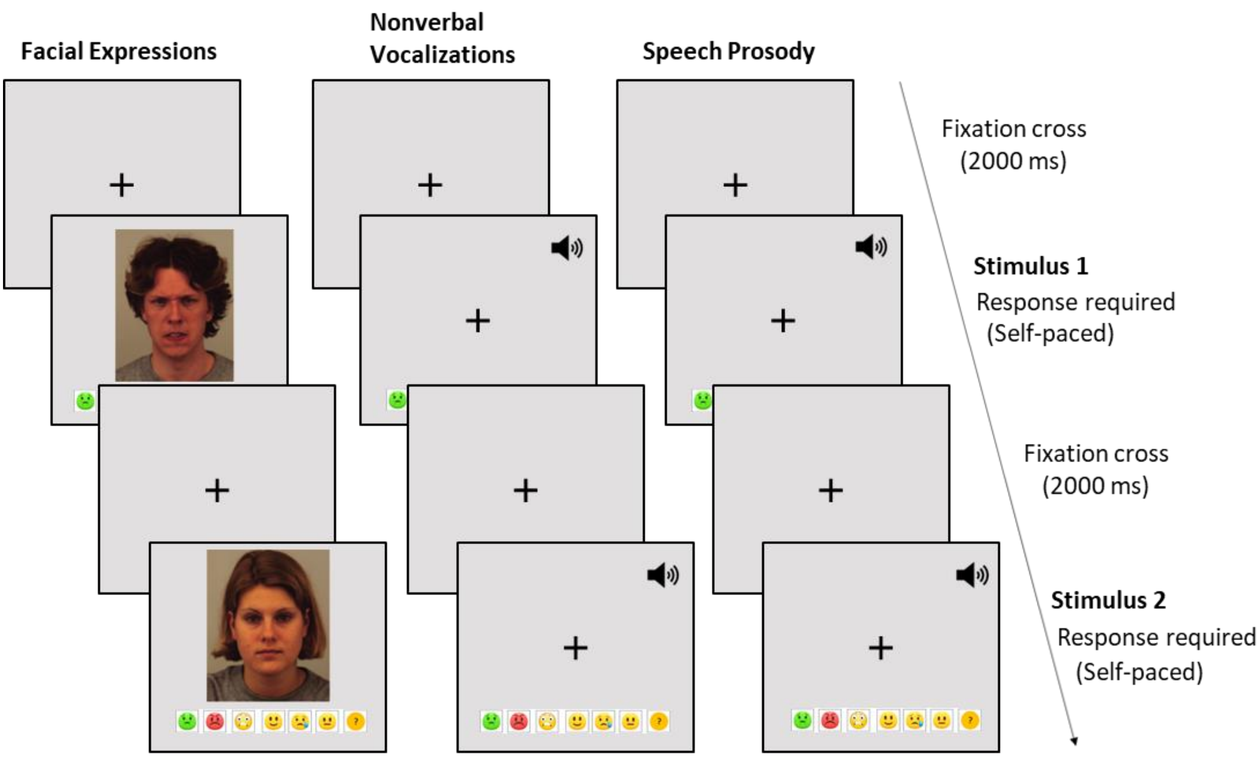
The experiment consisted of three identical emotion recognition tasks, only differing in the type of the stimuli presented to convey emotion – nonverbal vocalizations, speech prosody and facial expressions. The order of the tasks was randomized across participants, and each testing session lasted on average 1.5 hours. Short breaks were allowed between tasks. The auditory stimuli were presented via high-quality headphones (Sennheiser HD 62 TV), with the volume adjusted to a comfortable level for each participant. Facial expressions were presented centrally on the computer screen, with a black background.

In each task, the trials were presented in two blocks, preceded by a practice block with four trials, that included different stimuli from the experimental blocks. Twelve stimuli of each emotion category (i.e., anger, disgust, fear, happiness, sadness, neutral) were presented once, in a total of 72 experimental trials. Their order was pseudorandomized so that no more than two stimuli from the same category could appear consecutively.

Figure 1 shows the design of the experimental trials. In each trial, a fixation cross appeared centrally on the screen for 2000 ms, followed by stimulus presentation. The cross remained on the screen during stimulus presentation in the auditory blocks. Participants were asked to identify the emotion that best characterized each stimulus, by pressing one of 7 *emoticons* on the keyboard. The order of the *emoticons* on the keyboard was counterbalanced across participants as there were three possible combinations. All *emoticons* were displayed on the screen alongside the stimulus and accompanied by the respective labels (anger, disgust, fear, happiness, sadness, neutral and none of the previous). The task was self-paced, but participants were asked to respond as fast and as accurately as possible. No feedback was provided. The tasks were implemented with E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA, USA).

Figure 1

Illustration of experiment trials in the three emotion recognition tasks



2.5. Data Preparation and Analysis

Accuracy rates for Emotion and Stimulus in emotion recognition were corrected for potential response biases since percentage-correct scores do not account for cases in which participants only used one or a closed set of response options (e.g., only deciding between 'happiness' and 'sadness'). By expressing accuracy as a function of response and stimulus frequency, any Hu score is thereby insensitive to bias, to proportions of stimuli of different types, and to the number of categories. When all the stimuli from a category are correctly identified, and the target response category is always correctly used, the Hu score is equal to 1; when no categorial stimulus (e.g., disgust) is correctly identified, the Hu score is equal to 0. Hu scores were computed separately for each emotion category and stimulus type, and average scores were computed for Emotion and Stimulus.

Statistical analyses were performed with IBM SPSS Statistics for Windows, version 27 (IBM Corp., Armonk, N.Y., USA). To investigate the role of the experimental manipulations in the behavioral data, we used a repeated-measures ANOVA. To test our hypotheses, we explored correlations between emotion recognition and musical sophistication, as assessed by the subscales from the Gold-MSI. Outliers were not included in this analysis.

3. Results

3.1. Recognition Accuracy

A response was considered correct when it corresponded to the target stimuli. The proportion of correct identifications per emotion type was distributed across stimulus categories and computed individually. Figure 2 presents the mean distribution of correct identifications (expressed as Hu scores). Overall accuracy was around .75. Recognition rates ranged from .35, for disgust in speech prosody, to .94, for happiness in nonverbal vocalizations (see Table 1 for descriptive statistics).

A repeated-measures ANOVA was conducted with Emotion and Stimulus Type as within-subject factors resulting in a 6 (emotion type: anger vs. disgust vs. fear vs. happiness vs. sadness vs. neutral) x 3 (stimulus type: speech prosody vs. nonverbal vocalizations vs. facial expressions) design. There were two significant main effects of Emotion, $F\ (5, 145) = 15.38$, $p < .001$, partial $\eta 2 = .35$, and stimulus, $F\ (2, 58) = 48.49$, p

< .001, partial η2 = .63, on the Hu scores, showing that these variables have effectively been manipulated.

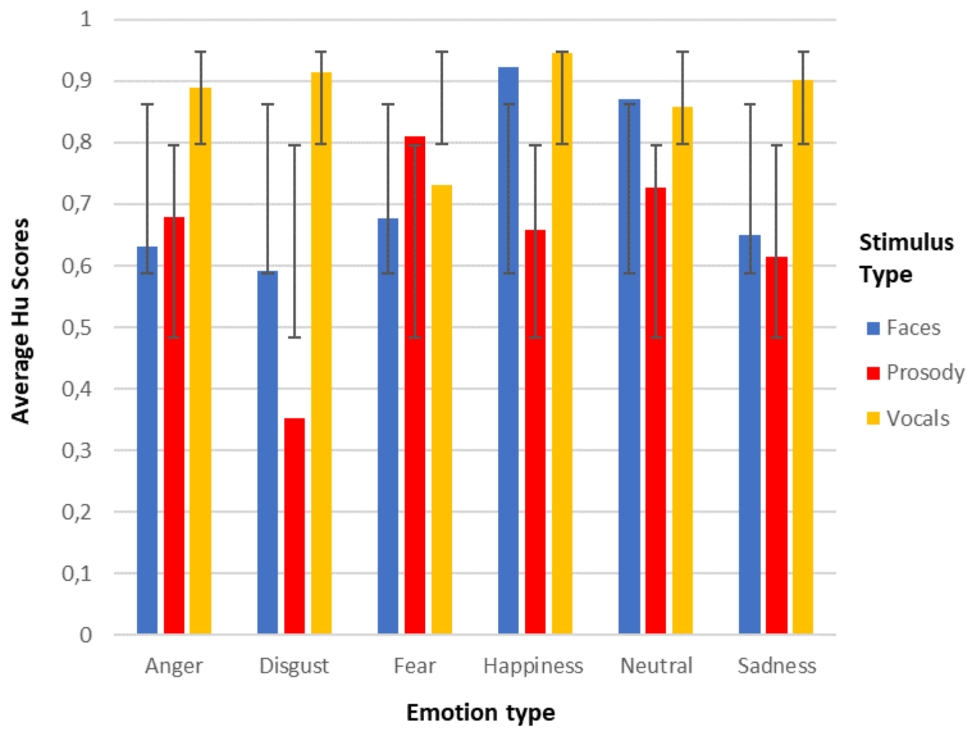


Figure 2. Average Hu scores for each emotion and stimulus type. Error bars indicate the standard deviations.

Table 1

Descriptive Statistics

	Full sample (*N* = 31)		
	M	SD	Range
Gold-MSI (general factor)	86.35	3.79	42-117
Active Engagement	39.06	1.55	17-51
Perceptual Abilities	51.45	1.45	27-63
Musical Training	30.84	2.50	7-48
Singing Abilities	33.74	1.61	18-48
Emotions	33.84	0.76	24-42
Emotion Recognition (average Hu scores)			
Faces	.74	.02	.55-.91
Prosody	.65	.02	.39-.87
Vocalizations	.88	.02	.65-1.0
Anger	.76	.02	.39-.92
Disgust	.63	.03	.20-.87
Fear	.75	.02	.46-.95
Happiness	.84	.01	.59-.95
Neutral	.82	.02	.50-.97
Sadness	.74	.02	.42-.95

Results of associations between emotion recognition and musical abilities, in the form of Gold-MSI's general factor and subscales, are exhibited as zero-order correlations in Table 2. Contrary to Hypothesis 1, musical sophistication did not correlate with emotion recognition accuracy in nonverbal vocalizations and speech prosody. On the other hand, in line with Hypothesis 2, musical sophistication did not correlate with emotion recognition accuracy in facial expressions. Nonetheless, the respective Pearson's coefficients indicated that lower accuracy was obtained by individuals with higher self-reported musical sophistication, and vice versa. The exploratory analysis regarding the correlations between Emotion in the Hu scores and Gold-MSI's measures did not yield any statistically significant relation apart from a marginally significant one with Fear ($p = 0.59$).

Table 2

Correlations Between Emotion Recognition (average Hu scores by Stimulus and Emotion) and Musical Sophistication (Gold-MSI General Factor and Subscales)

	Gold-MSI					
	AE	PA	MT	SA	E	GF
Hu Scores	*r*	*r*	*r*	*r*	*r*	*r*
Stimulus						
Faces	-.13	-.19	-.25	-.04	-.27	-.17
Prosody	-.10	.07	.10	-.03	.04	.05
Vocalizations	.11	.04	.11	.17	.01	.14
Emotion						
Anger	-.13	.00	-.08	-.03	-.17	-.04
Disgust	.00	-.04	-.02	.11	-.05	.06
Fear	-.35[a]	.06	-.07	.04	.10	-.06
Happiness	-.18	-.32	-.16	-.12	-.23	-.17
Neutral	.18	-.08	.10	-.24	.02	.35
Sadness	-.08	-.04	-.06	-.23	-.21	-.16

Note. AE – Active Engagement; PA – Perceptual Abilities; MT – Musical Training; SA – Singing Abilities; E – Emotions; GF – General Factor; *r* – Pearson’s coefficients

[a]$p = 0.59$

4. Discussion

The present study investigated the association between musical sophistication and the ability to recognize emotions in vocal and facial expressions. First, we evaluated the effectiveness of the manipulated variables (i.e., Stimulus and Emotion) on the recognition scores. Second, we examined correlations between emotion recognition accuracy and Gold-MSI scores to verify the hypotheses. The analysis did not support the idea that performance improvements from sophisticated musical engagement are limited to the auditory domain (Hypothesis 1). Higher performance in speech prosody and nonverbal vocalizations was not found for individuals with naturally good perceptual abilities, as expected. In contrast, Hypothesis 2 was corroborated since there was no advantage in emotional face recognition even though it was higher for individuals with lower self-reported scores of musical sophistication. This opposed to our prediction. Finally, the exploratory analysis revealed a (marginally significant) negative correlation between Gold-MSI's Active Engagement subscale and recognition of fear.

In addition to the hypotheses, these results contradicted original findings showing consistently significant positive associations (vs. non-significant and often negative associations) between musical sophistication and vocal emotion recognition, and positive associations (vs. negative associations) between musical sophistication and emotional face recognition (Correia et al., 2020; MacGregor & Müllensiefen, 2019). In musical training studies, the musician performance advantage with vocal expressions is supported to some degree (e.g., Correia et al., 2020; Lima & Castro, 2011; Parsons et al., 2014; Thompson, Schellenberg & Husain, 2004;), in contrast to null evidence for facial expressions, as already noted (e.g., Correia et al., 2020, Farmer & Petrini, 2020; Weijkamp & Sadakata, 2017).

4.1. Nonverbal vocalizations and speech prosody tasks

Recognition accuracy in the vocal expression tasks did not correlate significantly with any of the measures of musical sophistication. Note that there was no substantial departure from normality in the accuracy for vocalizations (skewness, range = -0.72 – 0.42; kurtosis, range = -0.11 – 0.82; Curran et al., 1996), even though it was tendentially higher than that for prosody (see Table 1), with one participant reaching the maximum score. This is, however, consistent with some studies on the recognition of anger, disgust, fear, sadness and happiness reporting increased detection in vocalizations, compared to speech prosody (Hawk et al. 2009; Sauter et al, 2013). Hence, the absence of an association between recognition of vocal expressions and musical sophistication cannot be ascribed to ceiling effects.

As previously held, some authors claim that the effect sizes are small and, therefore, these studies might require large samples of heterogeneous groups of both trained and untrained musicians for the effects to surface in the analyses (see Sala & Gobet, 2017). In musical training literature, certain studies including 16 musicians or less did not find any association with emotion recognition for nonverbal vocalizations and speech prosody (e.g., Park et al., 2015; Weijkamp & Sadakata, 2017), so it is possible that the sample size was not enough to permit observable effects on recognition. Moreover, Correia and colleagues (2020) showed that highly trained musicians and nonmusicians scoring higher in musical sophistication equated in emotion recognition of both nonverbal vocalizations and speech prosody. This might be evidence that at a behavioral level musical training does not provide a surplus over musical abilities. Compatible with this notion is the finding that musician-like advantages in phoneme perception and speech processing were observed in individuals without formal music training (Mankel & Bidelman, 2018). Finally, Thompson and colleagues (2004) suggested that singing lessons might have disrupted vocal emotion perception as trained vocal patterns could have conflicted with vocal emotion expression. In this study, a good number of participants (n = 11) had had lifelong singing lessons, so we do not reject the possibility that it tempered with the results.

4.2. Facial expressions task

As predicted, there was no association between recognition of facial expressions and the several facets of musical sophistication. Lima and colleagues (2016) revealed difficulties in the processing of facial expressions in congenital amusia (i.e., musical disorder linked to deficits in pitch perception). This raises the question of whether typically developing individuals varying in musical abilities/musical training can identify facial emotions differently. Indeed, Correia and colleagues (2020) found no significant correlations between four measures of musical sophistication and the average recognition rates of facial expressions, unlike for speech prosody and nonverbal vocalizations. Meanwhile, the musical training studies failed to report associations with facial emotion recognition (Correia et al., 2020; Farmer & Petrini, 2020; Weijkamp & Sadakata, 2017).

On the other hand, there are findings that link emotion decoding ability to the recognition of emotional face expressions (e.g., MacGregor & Müllensiefen, 2019; Thompson, 2010) and to self-reported emotional engagement with music (i.e., Gold MSI Emotions subscale) (Akkermans et al., 2019; MacGregor & Müllensiefen, 2019). In fact, when we analyzed Emotion scores by each task, three correlations turned up for facial expressions: a significant one in recognition of disgust and two marginally significant in recognition of anger and happiness (see Appendix A). These results provide further evidence of a potential relationship between facial emotion recognition and sophisticated emotional engagement with music.

4.3. Exploratory analysis

We found a marginally significant correlation between Active Engagement and recognition accuracy for fear. Incidentally, there was a significant correlation between recognition of fear in the prosody task and active engagement with music, which might elucidate the unpredicted association (see Appendix A).

Age can be a decisive link between the two variables. Lima and colleagues (2020) identified a particularly strong correlation between the participant's age and self-reported active engagement. Beyond the fact that older adults display greater difficulties in correctly decoding fear compared to younger adults (for a review, Ruffman et al., 2008), a study using six emotions (angry, disgust, fear, sadness, happiness and neutral) demonstrated only one significant difference between the younger and the older adults for fear recognition (Abbruzzese et al., 2019). Thus, we argue that active engagement and recognition of fear might be sensitive to fine differences in age-related factors. This might help explain the correlation found.

4.4. Limitations and future directions

One of the shortcomings of the current study is the relatively small sample size, which deterred between-group analyses with musical training and median-splits of musical sophistication due to concerns with insufficient statistical power. Ideally, the sample would be composed of 35 musicians and 35 musically naïve controls, as yielded by an initial calculation of the statistical power required. Thus, it is worth considering this situation might have given rise to type II errors since some results were unexpected, bringing into question the direction and strength of the relationships.

Another shortcoming respects to the use of other relevant control such as age, sex, socioeconomic status, personality and measures of cognitive abilities. A systematic exploration of these measures could have aided a more comprehensive view of their individual contributions to emotion recognition and how they vary across groups and conditions. Also, given that the tasks were self-paced, there might have been confounding effects accruing from longer response times (e.g., familiarity effect) that were not controlled for.

Future studies should also consider the inclusion of 3D models of emotion recognition for facial expressions to capture finer-grained sensitivities to dynamic facial features that static images (i.e., photographs) do not allow to tap into. Beyond adding to the ecological validity, these would possibly lead to increased detection rate in some participants and, thus, accentuate individual differences. Moreover, other levels of analyses with EEG and fMRI could, for example, help address *when* the emotional processing occurs for each stimulus category and *where* in the brain (e.g., temporal lobes and modality-independent regions responsible for socioemotional processing) emotion decoding shows connectivity patterns across a range of emotions, respectively. Together with refining the understanding of the putative common mechanisms underlying musical sophistication and emotion recognition, it would be interesting to compare singers and multiple classes of instrumentalists as to latent nuances in recognition accuracy.

4.5. Conclusion

In this study, we inspected correlações between Gold MSI self-report scores of musical sophistication and mean accuracy rates (in the form of Hu scores) in three stimuli-dependent emotion recognition tasks. The results presented novel evidence incompatible with our hypothesis that emotion recognition in the vocal expressions tasks would be linked to musical sophistication. Additionally, they demonstrated emotion recognition in the facial expressions task was not related to musical sophistication. Altogether, our findings point to the necessity of employing different methods to examine the variables of interest and relationships between them in a more extensive fashion, as well as to the importance of ensuring reasonable statistical power with adequate sample sizes. We argue the present study can supplement to the musical sophistication literature, though the data has to be faced with critical judgement.

5. References

Abbruzzese, L., Magnani, N., Robertson, I. H., & Mancuso, M. (2019). Age and gender differences in emotion recognition. *Frontiers in psychology, 10,* 2371.

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., ... & Frieler, K. (2019). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition and Emotion*, *33*(6), 1099-1118.

Aldunate, N., & González-Ibáñez, R. (2017). An integrated review of emoticons in computer-mediated communication. *Frontiers in psychology*, *7*, 2061.

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of experimental child psychology, 83*(2), 111-130.

Barrett, K. C., Ashley, R., Strait, D. L., & Kraus, N. (2013). Art and science: how musical training shapes the brain. *Frontiers in Psychology*, *4*, 713.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental review*, *29*(3), 180-200.

Bigand, E., & Tillmann, B. (2021). Near and far transfer: Is music special?. *Memory & Cognition.* doi: 10.3758/s13421-021-01226-6

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5:9/10, 341-345.

Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*, *42*(1), 74-81.

Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2020). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*.

Corrigall, K. A., & Schellenberg, E. G. (2015). Predicting who takes music lessons: Parent and child characteristics. *Frontiers in Psychology*, *6*, 282.

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in psychology*, *4*, 222.

Criscuolo, A., Bonetti, L., Särkämö, T., Kliuchko, M., & Brattico, E. (2019). On the association between musical training, intelligence and executive functions in adulthood. *Frontiers in psychology*, *10*, 1704.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29.

Dawson, C., Aalto, D., Šimko, J., Vainio, M., & Tervaniemi, M. (2017). Musical sophistication and the effect of complexity on auditory discrimination in Finnish speakers. *Frontiers in neuroscience*, *11*, 213.

Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology*, *2*, 124. doi: 10.3389/fpsyg.2011.00124.

Egermann, H., & McAdams, S. (2012). Empathy and emotional contagion as a link between recognized and felt emotions in music listening. *Music Perception: An Interdisciplinary Journal*, *31*(2), 139-156.

Farmer, E., Jicol, C., & Petrini, K. (2020). Musicianship Enhances Perception But Not Feeling of Emotion From Others' Social Interaction Through Speech Prosody. *Music Perception*, *37*(4), 323-338.

Fodor, J. A. (1983). *The modularity of the mind*. Cambridge, MA: MIT Press.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*(9), 2038-2043.

Franklin, M. S., Sledge Moore, K., Yip, C. Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*(3), 353-365.

Frischen, U., Schwarzer, G., & Degé, F. (2019). Comparing the effects of rhythm-based music training and pitch-based music training on executive functions in preschoolers. *Frontiers in Integrative Neuroscience*, *13*, 41.

Gelding, R. W., Harrison, P. M., Silas, S., Johnson, B. W., Thompson, W. F., & Müllensiefen, D. (2021). An efficient and adaptive test of auditory mental imagery. *Psychological research*, *85*(3), 1201-1220.

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska directed emotional faces: a validation study. *Cognition and emotion*, *22*(6), 1094-1118.

Gosvvarni, U. (2011). Language, music, and children's brains: a rhythmic timing perspective on language and music as cognitive systems. *Language and music as cognitive systems*, *292*.

Grandjean, D., Bänziger, T., & Scherer, K. R. (2006). Intonation as an interface between language and affect. *Progress in brain research*, *156*, 235-247.

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental science*, *18*(5), 686-702.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. routledge.

Hawk, S. T., Fischer, A. H., & Van Kleef, G. A. (2012). Face the noise: embodied responses to nonverbal vocalizations of discrete emotions. *Journal of personality and social psychology*, *102*(4), 796.

Hawk, S. T., Van Kleef, G. A., Fischer, A. H., & Van Der Schalk, J. (2009). " Worth a thousand words": absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, *9*(3), 293.

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*, *76*(3), 486-502.

Hetland, L. (2000). Learning to make music enhances spatial reasoning. *Journal of aesthetic education*, *34*(3/4), 179-238.

Hietanen, J. K., Surakka, V., & Linnankoski, I. (1998). Facial electromyographic responses to vocal affect expressions. *Psychophysiology*, *35*(5), 530-536.

Hirnstein, M., Leask, S., Rose, J., & Hausmann, M. (2010). Disentangling the relationship between hemispheric asymmetry and cognitive performance. *Brain and cognition*, *73*(2), 119-127.

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage*, *47*(2), 735-744.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research*, *33*(3), 217-238.

Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, *42*(4), 599-623.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature reviews neuroscience*, *11*(8), 599-605.

Kraus, N., & White-Schwoch, T. (2017). Neurobiology of everyday communication: What have we learned from music?. *The Neuroscientist*, *23*(3), 287-298.

Larrouy-Maestri, P., Harrison, P. M., & Müllensiefen, D. (2019). The mistuning perception test: A new measurement instrument. *Behavior Research Methods*, *51*(2), 663-675.

Lausen, A., & Hammerschmidt, K. (2020). Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Humanities and Social Sciences Communications*, *7*(1), 1-17.

Lee, H., & Müllensiefen, D. (2020). The Timbre Perception Test (TPT): A new interactive musical assessment tool to measure timbre perception ability. *Attention, Perception, & Psychophysics*, *82*(7), 3658-3675.

Levitin, D. J. (2012). What does it mean to be musical?. *Neuron*, *73*(4), 633-637.

Liebenthal, E., Silbersweig, D. A., & Stern, E. (2016). The language, tone and prosody of emotions: neural substrates and dynamics of spoken-word emotion perception. *Frontiers in neuroscience*, *10*, 506.

Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., & Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports*, *6*(1), 1-13.

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021.

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior research methods*, *45*(4), 1234-1245.

Lima, C. F., Correia, A. I., Müllensiefen, D., & Castro, S. L. (2020). Goldsmiths Musical Sophistication Index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality and music preferences. *Psychology of Music*, *48*(3), 376-388.

Lopes, M. R. R. N. M. (2015). Emotional contagion in voice-to-voice service encounters: a dynamic approach to the influence of customers on employees' behavior and welfare.

MacGregor, C., & Müllensiefen, D. (2019). The Musical Emotion Discrimination Task: A new measure for assessing the ability to discriminate emotions in music. *Frontiers in psychology*, *10*, 1955.

Manera, V., Grandi, E., & Colle, L. (2013). Susceptibility to emotional contagion for negative emotions improves detection of smile authenticity. *Frontiers in Human Neuroscience*, *7*, 6.

Mankel, K., Barber, J., & Bidelman, G. M. (2020). Auditory categorical processing for speech is modulated by inherent musical listening skills. *NeuroReport*, *31*(2), 162.

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences*, *115*(51), 13129-13134.

Martins, M., Pinheiro, A. P., & Lima, C. F. (2021). Does Music Training Improve Emotion Recognition Abilities? A Critical Review. *Emotion Review*, *13*(3), 1-12.

McKenzie, C., Ostevik, A., Hodgetts, B., Cummine, J., & Aalto, D. (2019). Associations Between Musical Experience and Auditory Discrimination. *Canadian Acoustics*, *47*(3), 30-31.

Miendlarzewska, E. A., & Trost, W. J. (2014). How musical training affects cognitive development: rhythm, reward and other modulating variables. *Frontiers in neuroscience*, *7*, 279.

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

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: no causal effect of music practice on music ability. *Psychological science, 25*(9), 1795-1803.

Mullensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of nonmusicians: An index for assessing musical sophistication in the general population. *PLOS ONE, 9*, e89642. doi: 10.1371/journal.pone.0089642

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Park, M., Gutyrchik, E., Welker, L., Carl, P., Pöppel, E., Zaytseva, Y., ... & Bao, Y. (2015). Sadness is unique: neural processing of emotions in speech prosody in musicians and non-musicians. *Frontiers in Human Neuroscience, 8,* 1049.

Parsons, C. E., Young, K. S., Jegindø, E. M. E., Vuust, P., Stein, A., & Kringelbach, M. L. (2014). Music training and empathy positively impact adults' sensitivity to infant distress. *Frontiers in Psychology, 5*, 1440.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in psychology*, *2*, 142.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107-120.

Porflitt, F., & Rosas, R. (2020). Musical sophistication explains a good deal of cognitive performance. A cross-sectional study of musicians and non-musicians. *Revista de investigación musical*, 147.

Radocy, R. E., & Boyle, J. D. (2012). *Psychological foundations of musical behavior*. Charles C Thomas Publisher.

Roden, I., Könen, T., Bongard, S., Frankenberg, E., Friedrich, E. K., & Kreutz, G. (2014). Effects of music training on attention, processing speed and cognitive music abilities—findings from a longitudinal study. *Applied Cognitive Psychology*, *28*(4), 545-557.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews*, *32*(4), 863-881.

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current directions in psychological science*, *26*(6), 515-520.

Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & cognition*, *48*(8), 1429-1441.

Sauter, D. (2006). An investigation into vocal expressions of emotions: the roles of valence, culture, and acoustic factors. *Ann Arbor, 1050*, 48106-1346.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408-2412.

Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, *31*(1), 97-113.

Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. *Frontiers in psychology*, *4*, 511.

Schellenberg, E. G., & Peretz, I. (2008). Music, Language and Cognition: Unresolved Issues. *Trends in cognitive sciences*, *12*(2), 45-46.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*(3), 341-349.

Scott, S. K., Sauter, D., & McGettigan, C. (2010). Brain mechanisms for processing perceived emotional vocalizations in humans. In *Handbook of behavioral neuroscience* (Vol. 19, pp. 187-197). Elsevier.

Strait, D. L., O'Connell, S., Parbery-Clark, A., & Kraus, N. (2014). Musicians' enhanced neural differentiation of speech sounds arises early in life: Developmental evidence from ages 3 to 30. *Cerebral Cortex, 24*, 2512–2521.

Swaminathan, S., & Schellenberg, E. G. (2016). Music Training. In *Cognitive Training* (pp. 137-144). Springer, Cham.

Swaminathan, S., & Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychonomic Bulletin & Review*, *24*(6), 1929-1934.

Swaminathan, S., & Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. *Scientific reports*, *8*(1), 1-7.

Swaminathan, S., & Schellenberg, E. G. (2020). Musical ability, music training, and language ability in childhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(12), 2340.

Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude?. *Intelligence*, *62*, 119-124.

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in neurosciences*, *29*(7), 382-390.

Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in psychology*, *5*, 658.

Thompson, W. (2010, December). Emotional communication in the human voice. In *Proceedings of the Music Cognition II: Music and Emotions*

Conferences,(Macau: University of Saint Joseph), doi (Vol. 10, No. 2.1, pp. 2661-5683).

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica*, *158*, 407-424.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help?. *Emotion*, *4*(1), 46.

Twaite, J. T. (2016). *Examining Relationships between Basic Emotion Perception and Musical Training in the Prosodic, Facial, and Lexical Channels of Communication and in Music*. City University of New York.

Vaughn, K. (2000). Music and mathematics: Modest support for the oft-claimed relationship. *Journal of Aesthetic Education*, *34*(3/4), 149–66.

Wagner, H. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal Of Nonverbal Behavior*, *17*, 3-28. doi: 10.1007/bf00987006

Weijkamp, J., & Sadakata, M. (2017). Attention to affective audio-visual information: Comparison between musicians and non-musicians. *Psychology of Music*, *45*(2), 204-215.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature reviews neuroscience*, *8*(7), 547-558.

Appendix A

Correlations Between Emotion Recognition (Hu scores by Task) and Musical Sophistication (Gold-MSI General Factor and Subscales)

	Gold-MSI					
Hu Scores	AE	PA	MT	SA	E	GF
(Stimulus x Emotion)	*r*	*r*	*r*	*r*	*r*	*r*
Faces						
Anger	-.13	-.06	-.21	-.17	-.36[a]	-.18
Disgust	-.01	-.23	-.21	-.16	-.37*	-.20
Fear	-.24	-.03	-.08	.01	-.04	-.05
Happiness	-.31	-.31	-.27	-.23	-.36[b]	-.31
Neutral	-.18	-.28	-.34	-.18	-.29	-.27
Sadness	-.18	-.14	-.24	-.12	-.26	-.24
Prosody						
Anger	-.22	.07	-.01	.03	.00	-.00
Disgust	-.02	.06	.11	.22	.06	.18
Fear	-.35*	.17	-.06	-.01	.11	-.08
Happiness	-.01	-.28	-.00	-.04	-.09	-.01
Neutral	.12	-.00	.16	-.11	-.02	.06
Sadness	.10	.08	.12	-.22	.02	-.02
Vocalizations						
Anger	.10	-.04	.04	.09	-.06	.11
Disgust	.05	.09	.03	.17	.24	.11
Fear	-.07	-.00	.01	.06	.14	.01
Happiness	-.18	-.09	-.17	.01	-.12	-.13
Neutral	.31*	.03	.23	.02	-.21	.14
Sadness	-.04	.13	.03	.27	-.09	.15

Note. AE – Active Engagement; PA – Perceptual Abilities; MT – Musical Training; SA – Singing Abilities; E – Emotions; GF – General Factor; *r* – Pearson's coefficients

*p < .05, [a]p = 0.52, [b]p = 0.57

Appendix B

Demographic and Background Characteristics of the Participants (*n* = 31)

Characteristics	Musicians (*n* =21)	Controls (*n* = 10)
Age in years	32 (2.3)	23 (1.5)
Music training in years	11 (0.9)	----------
Age of training onset (years)	11 (1.1)	----------
Average practice hours per week	7 (1.2)	----------
Gold-MSI (general factor)	98 (2.4)	64 (5.5)
Socioeconomic Status (SES)[1]	2 (0.4)	5 (1.0)
Raven's APM-12 (problems solved)[2]	8 (0.6)	9 (0.8)

Note. Standard deviations in parentheses.

[1] SES is scored between 1 and 9. [2] Raven's Advanced Progressive Matrices (12-item short form) raw scores (range: 0–12)