UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

PIANOS AND MICROPHONES:

DOES THE TYPE OF MUSICAL TRAINING AFFECT EMOTION RECOGNITION?

Ricardo Miguel Correia Francisco

MESTRADO INTERUNIVERSITÁRIO EM
NEUROPSICOLOGIA CLÍNICA E EXPERIMENTAL

2021

UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

PIANOS AND MICROPHONES:

DOES THE TYPE OF MUSICAL TRAINING AFFECT EMOTION RECOGNITION?

Ricardo Miguel Correia Francisco

Dissertação orientada pela Professora Doutora Ana Pinheiro
Coorientada pelo Professor Doutor César Lima

MESTRADO INTERUNIVERSITÁRIO EM NEUROPSICOLOGIA CLÍNICA E EXPERIMENTAL

2021

Abstract

Music, emotion, and language have been subjects of interest in neurosciences research due to their relationship as means of social communication. It has been widely acknowledged that the musicians' brain may help explain this relationship for it is an adequate example of cross-domain neuroplasticity. Indeed, musical performance presupposes the activation of different sensory and motor systems associated with a facilitated response to emotional auditory information. Nonetheless, literature is scarce when defining the concept of "musical expertise". A few studies have accounted for factors other than general musical training in auditory emotional processing, however, no study has tackled the implications of vocal musical training.

Vocal musical training is considered to have different neural implications from instrumental musical training, since the instrument of a singer is contained within the body. Singers have shown to have an enhanced activation of the auditory feedback system in comparison to non-musicians and instrumentalists, hence enabling a facilitated response to the production and recognition of vocal emotional information. The following study sets to explore the underlying differences in emotional auditory processing taking into consideration the type of musical training (vocal *vs* instrumental).

Nine singers, thirteen instrumentalists, and nine non-musicians were recruited for an emotional recognition task. Participants listened to nonverbal vocalizations and prosodic speech and had to categorize those stimuli in terms of their emotional quality (anger, disgust, fear, happiness, neutral, sadness).

We found no significant differences in accuracy measures and response times between the three groups. A main effect of stimulus type (speech prosody vs. vocalizations) was found, in which emotional vocalizations were faster and more accurately recognized in comparison to speech prosody stimuli. Furthermore, an interaction effect between emotion and type of stimulus was observed.

We propose that the emotional recognition's task results were affected by the reduced number of participants recruited. It might also reflect the need to assess other possible cross-domain influencing factors.

Happiness, and disgust were the most accurately recognized emotions in the nonverbal vocal emotions condition. In the prosody condition, participants exhibited higher rates of accuracy in fear, but not in vocalizations. We propose that the acoustic ambiguity of fearful vocalizations might be reduced by the inherently longer duration of prosodic stimuli.

Additionally, a correlation analysis for musical ability, engagement, and emotional recognition was performed, foregrounding the importance of individual differences in cross-domain effects of music.

Keywords: Emotional Recognition; Musical training; Transfer effects.

Resumo

A relação entre música, linguagem e emoções tem sido tema de debate na comunidade científica, nomeadamente devido às similaridades que estes conceitos apresentam como meios de comunicação social. De facto, muitos descrevem a música como “a língua das emoções” e, como tal, a investigação em neurociências tem estudado esta proposta recorrendo à população mais fluente neste “idioma”, os músicos.

Devido à sua complexidade inerente, o treino musical tem sido associado à ativação de redes neuronais relativas a funções motoras, cognitivas e sensoriais. Por exemplo, estudos realizados com o recurso a ressonância magnética (fMRI) revelam que músicos apresentam uma ativação mais acentuada em regiões de processamento sensorial, assim como uma maior densidade de substância cinzenta nestas estruturas. Do mesmo modo, estudos eletrofisiológicos reportam que a prática musical está associada a benefícios na acuidade visual, controlo motor, processamento de informação auditiva, e processamento acústico de estímulos emocionais complexos (e.g., prosódia emocional e vocalizações). De facto, a relação entre o treino musical e a eficiência da resposta subcortical a estímulos emocionais tem sido frequentemente replicada. Por este motivo, o cérebro musical é considerado um dos principais exemplos de neuroplasticidade interdomínios.

Recentemente, alguns investigadores têm salientado a relevância de outros fatores influentes no reconhecimento emocional auditivo para além do treino musical. Por exemplo, há evidências que indicam que capacidades preceptivas acústicas individuais (e.g., deteção de frequência) e relação do indivíduo com a música no quotidiano são fatores que influenciam a capacidade de discriminação emocional. Assim, verificou-se que não-músicos com boas capacidades percetivas apresentaram resultados semelhantes aos músicos em tarefas de discriminação emocional auditiva. De acordo com esta ideia, propomos que fatores como o tipo de treino musical deveriam ser abordados como potenciais fontes de diferenciação no processamento emocional auditivo, com especial foco no treino musical vocal.

O treino musical da voz tem diferentes implicações neuronais do treino musical instrumental, dado que o “instrumento” do cantor está contido no seu corpo. Assim, cantores distinguem-se de outros músicos devido à ativação de sistemas motores específicos durante a performance musical. De facto, ativação motora no aparelho vocal evoca recetores somatosensoriais que participam nos mecanismos de feedback envolvidos na manutenção de notas, produção de vocalizações e reprodução de tons emocionais. Consequentemente, há evidências de que cantores demonstram uma resposta mais controlada deste mecanismo quando comparados com controlos e outro tipo de músicos. Assim, o presente estudo visa explorar as diferenças no processamento auditivo emocional considerando as especificidades de diferentes tipos de treino musical (vocal *vs*. instrumental).

Nove cantores, treze instrumentalistas e nove participantes sem treino musical foram recrutados para a realização de duas tarefas de escolha forçada. Na primeira, foi pedido aos participantes que escutassem com atenção expressões de emoção vocalizadas e que selecionassem a resposta correta de seis possibilidades (raiva, medo, repugna, felicidade, neutro e tristeza). A segunda tarefa seguiu uma estrutura semelhante, no entanto, ao invés de vocalizações, os participantes foram expostos a frases com conteúdo semântico neutro transmitido com propriedades prosódicas emocionais (discurso prosódico). Adicionalmente, os participantes preencheram um questionário relativo ao seu relacionamento com a música (Gold-MSI) e realizaram testes de habilidade percetiva de sons, nomeadamente tarefas de discriminação de frequências sonoras, e perceção de tempo e duração.

Ao contrário do que era esperado, não foram obtidas diferenças significativas para as medidas de acuidade e tempo de respostas, entre os grupos. Propomos que a reduzida amostra e consequente reduzido poder estatístico possam ter influenciado os nossos resultados. Do mesmo modo, concluímos que estes resultados podem ser reflexivos da necessidade de explorar diferenças individuais na discriminação emocional auditiva (e.g., capacidades cognitivas, traços de personalidade, idade, entre outros).

Um efeito principal de tipo de estímulo (prosódia *vs.* vocalizações) foi observado, demonstrando que os participantes foram mais rápidos e mais precisos na discriminação de emoções para a condição das vocalizações em relação à condição da prosódia. Conforme estudos prévios, este efeito era esperado dada a complexidade neuronal do processamento semântico de frases. Ademais, um efeito de interação foi observado entre categoria emocional e tipo de estímulo, onde emoções foram diferentemente reconhecidas mediante o tipo de transmissão. Salientamos quatro emoções: felicidade, tristeza, medo e repugna.

Participantes apresentaram melhores resultados de acuidade na felicidade, tristeza e repugna quando estas eram expressas através de vocalizações. Como era expectável, na condição prosódica, a repugna foi a categoria emocional com menos acuidade de reconhecimento, assim replicando estudos anteriores. É importante salientar que, ao contrário do esperado o medo foi a emoção com menor reconhecimento na condição das vocalizações. No entanto, o mesmo não se verificou na condição do discurso prosódico, onde demonstrou ser a emoção com maior acuidade discriminatória. Propomos que a ambiguidade acústica evidenciada na expressão do medo vocalizado possa ser reduzida devido à inerente duração das frases prosódicas. Assim, exposição a demonstrações mais longas de medo na voz auxiliam a sua discriminação em relação a outras emoções com perfil acústico semelhante.

Adicionalmente, foi realizada uma análise correlacional entre as medidas descritivas individuais da amostra (sub-escalas do Gold-MSI e tarefas psicoacústicas) e acuidade no reconhecimento emocional. Associações entre as sub-escalas de habilidades de canto e envolvimento emocional com a música foram observadas, assim como com as tarefas acústicas percetuais (nomeadamente a deteção de frequência e discriminação de duração). Tendo em consideração estes resultados, salienta-se a importância de explorar fatores individuais, para além do treino musical, no processamento auditivo de emoções.

Apesar dos resultados, propomos que o cérebro de um cantor é um ótimo exemplo de neuroplasticidade induzida através da música, assim evidenciando especificidades neuronais em relação a outros tipos de treino musical. Por este motivo, encorajamos estudos futuros a explorar estas características neuronais devido ao seu potencial no auxílio do entendimento do processamento emocional auditivo.

Acknowledgments

Music icon John Lennon has stated that music should be used as a tool to spread the love. Ironically, in my particular case, music was used to test the patience of my loved ones. As such, the following section of the thesis is dedicated to them. Indeed, I would not have successfully executed this project without the support and guidance of various people.

Firstly, I would like to thank Professor Ana Pinheiro for her supervision of my master's thesis. It was truly an honour to work with one of the best researchers in auditory emotional processing in the country. Thank you for your enthusiasm for the field of music and for the opportunity of sharing that enthusiasm with me. I appreciate all your advice, ideas, support, and patience. I acknowledge the opportunity you provided me to grow as a researcher and neuropsychologist and, for that, I will be forever grateful. I would also wish to thank Professor César Lima for the support and advice. I extend my gratitude for helping in the creation and understanding of the experimental paradigm, as well as for the availability and knowledge you have provided me with. It was a pleasure learning and taking inspiration from you.

I grant my thanks to the Voice, Emotion, and Speech Lab, where I had the opportunity to develop my work and share ideas with the lab members. Thank you all for helping me create and run the experiments, as well as in data collection and writing support. I would like to give a special thanks to Inês Martins for introducing me to the field of music and emotion research, as well as all the availability and kindness.

I wish, also, to thank my family and all my friends, for they have an infinite amount of patience for my antics. My deepest gratitude to mom, dad and sister for your unqualified love, humour, and, above all, for believing in me. Furthermore, thank you for keeping my belly full and brain functioning in times of writing.

Lastly, I would like to give a special thanks to my "offscreen" supervisors. I am greatly in debt to Andreia Santiago, Margarida Marques, and Lucas Naumman for their guidance in the statistical analysis of the study, as well as for proofreading this thesis and constantly motivating me to successfully finish the project.

Once again, my deepest gratitude to all of you. I am, truly, a lucky person.

Index

List of Tables

1. Introduction

I would like to start this thesis by proposing a challenge to the reader. There is a globally known musical composition from 1988 named “Don’t worry, be happy” by Bobby McFerrin. The challenge is to carefully listen to this track without feeling a little bit happier. It may prove to be quite a troublesome task, I assure you. The same principle could be applied with an emotionally sad piece of music, or a fearful composition because music can easily induce an emotional response in the listener (Siedlecka & Denson, 2019). This would explain the reason why so many authors, rightfully, call music the “language of emotion” (Corrigall & Schellenberg, 2013; Richards, 2021).

Indeed, music and language seem to be quite identical in some respects, namely in the neural processing and transmission of emotional content (Besson et al., 2002; Brown, 2017; Castro & Lima, 2010; Kunert et al., 2015; Pinheiro et al., 2015). Hence, one must wonder: If music is the language of emotion, are musicians its “native” speakers? To further explore this question, one must firstly understand how emotional communication works.

The origin of Language

One of the most prized assets of communication is language. This pivotal skill enables the maintenance of social relationships and adaptative human behaviour through the effective communication of emotions (Juslin & Laukka, 2003). However, there has been some debate in the literature concerning the neural origin and structure of this concept (Besson & Schön, 2001; Indefrey & Levelt, 2000).

Historically, neuroscience theorists have proposed that language is an autonomous system organized into different processing submodules tasked with various aspects of language, such as: phonology, syntax, semantics, or pragmatic aspects of word production (Indefrey & Levelt, 2000; Shain et al., 2020; Besson & Schön, 2001; Chomsky, 1978, 1992). This proposal defines language as a complex and specialized neuronal skill, which is innate and effortless to the individual (Pinker, 1994). However, more recently a considerable amount of research has challenged this assumption by arguing that language is a product of the dynamic functioning of several different perceptual and sensorimotor systems (Arbib, 2013; Brown, 2017; Fedorenko, 2014; Patel et al., 1998).

For example, one of the most recent and popular theories for the emergence of oral communication comes from Arbib's (2013) “mirror neurons theory”. In the authors view, language is an innate neural capability that developed over time through processes of adaptative pressures. The author argues that the neuronal networks of language stemmed from mirror systems of gestural language and vocal pantomimes. Thus, the observation and execution of basic vocal sounds and gestures extended to the development of sensory processes and motor systems dedicated to the production and comprehension of language. Brown (2017) has argued that the imitation of vocal pantomimes may explain not only the origin of language, but also of music. Indeed, pantomimes of vocal expressions have very basic acoustic properties. The author proposes that these utterances would neither be speech, nor music-like in its acoustic features. Instead, this would be the precursor to a second stage of shared sensorimotor integration between language and music ("Musilanguage"). This stage comprises overlapping processes between music and language, enabling the production and processing of the affective melody present in speech and musical performance. Nonetheless, this would presuppose that language and music rely on similar neural mechanisms, hence begging the question: "In what way are music and language related?".

Music and Language: different domains, similar features

It is widely accepted in the scientific community that music and language are both effective means of communication, namely in signalling emotional content (Juslin & Laukka, 2003; Kraus & Slater, 2015). These domains rely on acoustic cues such as pitch, timbre and rhythm for their transmission of messages (Asaridou & McQueen, 2013; Besson & Schön, 2001; Brown, 2017; Juslin & Laukka, 2003; Lima & Castro, 2011). Patel (2003) argues that the structure of the syntactic properties of music and language are similar, thus sharing overlapping neuronal mechanisms. The author coined this as the "Shared Syntactic Integration Hypothesis" (SSIH), and it helps one to explain why similar acoustic organization of musical compositions and of vocal sounds evoke similar emotional responses (Juslin & Laukka, 2003).

Patel et al. (1998) employed an electroencephalogram (EEG) experiment to study the P600 component associated with syntactic incongruity. In this paradigm, participants listened to sentences and musical chord sequences with varying levels of syntactic incongruity. The main objective was to test if there were similarities in the neuronal processing of these stimuli. In fact, the P600 showed indistinguishable amplitude and scalp distribution for sentences and music sequences independent of congruency, hence highlighting a possible overlap between these domains.

Steinbeis and Koelsch (2008) found that music and language share common features in the processing of semantically meaningful information, as well. In two experiments using electrocenphalography EEG and functional magnetic resonance image (fMRI) it was shown that single chords (congruent and incongruent with the target word) primed the processing of subsequently presented congruent affective target words, as indicated by an increased N400 amplitude and activation of the right middle temporal gyrus in both conditions. More importantly, this study highlights that when primed by affective words, single chord incongruous to the preceding affect, also elicited an N400 response, and activated areas associated with the processing of meaning in acoustic signals, such as prosody, voices, and motion. These data indicates that musical features can influence subsequent word processing at a semantic level, which suggests that the expression of emotions in music is processed in a very similar fashion to meaning in language. In line with these findings, more recent studies have used similar brain imaging techniques along with electrophysiological measures to further solidify the hypothesis that music and language share common neuronal resources (e.g., Chiang et al., 2018; Kunert et al., 2015; Sun et al., 2018). Most studies concerning music and language have used expert musicians to ascertain whether language and music overlap with each other in sound processing (e.g., Bialystok & DePape, 2009; Kunert et al., 2015; Lima & Castro, 2011). In fact, research has highlighted that a musician's brain exhibits different processing patterns of sound in comparison to non-musicians (Münte et al., 2002; Park et al., 2014; Soncini & Costa, 2006), thus becoming an important tool in auditory research.

The musician's "super-powers"

Given the previous statement concerning a musician's brain, if the reader has experience with musical training, there might be a slight curiosity to know more about the difference between musicians and non-musicians in sound processing.

Musical performance presupposes the activation of multiple sensory, motor, and cognitive skills such as kinaesthetic control, memory, visual perception, pattern recognition, amongst other skill sets of the same importance (Barrett et al., 2013). Also, increased musical expertise was linked to the automatization of motor skills and higher motor efficiency, as well as a greater activation of sensorimotor cortices in musical practice (James et al., 2013; Pujol et al., 2000). This comes as no surprise since musical performance presupposes motor abilities to maneuverer an instrument, auditory sensory ability to manipulate the notes being played, visual acuity for musical sheet reading, and memory for practicing motor sequences and musical information, such as chord progressions (Martins, 2020). It is due to this level of cognitive demand that a musician's brain has been studied as a fruitful example of neuroplasticity in the auditory system (Münte et al., 2002).

Research has found that the activation of these different systems is associated with an enhanced processing of specific features of sound along with a faster and more robust response to patterns that are more complex (Kraus & Chandrasekaran, 2010). A study by Soncini and Costa (2006) showed that musically trained participants were better at recognizing complex acoustic cues, such as in speech-in-noise tasks. In this study, participants with music training and non-musicians were compared in the Portuguese Sentence Lists (LSP) test, where sentence recognition thresholds were investigated in quiet *vs.* noisy conditions. They concluded that in quiet situations musicians and non-musicians had similar performances. However, in the noise situation, musicians exhibited better performances, thus indicating that musical practice improves speech recognition in noisy situations. Wong et al. (2007) highlighted the facilitated processing of sound in discriminating pitch and tonal cues. Additionally, imaging techniques have been used to study this matter, and it is not surprising that several researchers found anatomical differences in the brain structure of expert musicians when compared to non-musicians (Barrett et al., 2013). For example, Acer et al. (2018) have found that musicians' brains have higher volumes of white and grey matter in somatosensory and premotor areas. Hutchinson et al. (2003) argue that cerebellar volume of musicians is correlated with lifelong intensity of practice.

Thus, musical expertise does seem to influence the processing of auditory stimuli functionally and structurally. In fact, this differentiation in the processing and recognition of auditory stimuli has been found in emotional content as well (Barrett et al., 2013).

Decoding the Language of emotion

Most researchers would argue that emotional communication is conveyed primarily through non-verbal vocalizations and speech prosody (Besson et al., 2002; Brown, 2017; Juslin & Laukka, 2003; Lima & Castro, 2011; Paulmann & Kotz, 2008). Basically, emotional prosody is the modulation of the acoustical output of vocal sounds. It comprises elements of speech regarding its acoustic properties, including pitch, loudness, tempo, intonation, rhythm, and timbre, which are not encoded by grammar or vocabulary (Brown, 2017; Lima & Castro, 2011; Pinheiro et al., 2015; Schirmer & Kotz, 2006; Zora et al., 2020). In line with this idea, when it comes to the auditory aspect of recognizing emotions, one is referring to the capability of an individual to infer emotional content in sounds through these cues (Lhavis, Allena & Scattoni, 2011; Rodrigues et al., 2017). An interesting behavioural study by Lima and colleagues (2011) compared a group of highly trained musicians with their non-musical counterparts in a forced choice emotional recognition task using short sentences expressing one of six emotions, by prosody only. A robust effect of musical expertise was found in the accuracy of emotional recognition. The authors proposed that musical expertise was associated with cross-domain benefits in emotional prosody, thus solidifying the hypothesis that music and language share common neural resources in the emotional processing of sound.

At a neural level, there is strong evidence that the perception of emotional prosody is a multi-stage process, in which different sub-stages are represented in the brain (Paulmann & Kotz, 2008; Pell & Kotz, 2011; Schirmer & Kotz, 2006). Emotional processing starts with the sensory assessment of the acoustic signal, followed by the detection of emotional salience in the vocal signal, and ending with the cognitive evaluation of its emotional significance (Pinheiro et al., 2015; Rigoulot et al., 2015). Electrocenphalography research has found that musicians seem to exhibit a faster and more robust response across these stages of emotional sound processing (Patel et al., 1998; Pinheiro et al., 2015). It is worth mentioning that different emotions have different acoustic profiles, thus making it difficult to ascertain if these putative advantages are general or specific to certain emotional categories (Banse & Scherer, 1996; Park et al., 2014, 2015). Nonetheless musical training has been widely associated with enhanced responses to emotional sounds (Martins et al., 2021).

According to the reviewed literature so far, one has established that expert musicians may, arguably, be the most fluent individuals in the processing of “emotional language”. However, research is scarce when it comes to differentiating the several factors that might define “musical expertise”. Indeed, Correia et al. (2020) proposed that the primary focus of researchers on music training reflects a narrow view of musicality, mainly because it does not account for the diverse factors other than formal lessons in years of training. This team of researchers has found that non-musicians with good auditory perceptual abilities and frequent musical engagement had a similar performance to expert musicians in an emotional recognition task. This finding establishes that years of musical training might not be the only factor that enhances the emotional processing of sound.

Pianos and Microphones

Do Chopin and Michael Jackson share the same type of musical expertise? Somehow it might seem like quite an arbitrary question and I would not argue against it. Let's address the "elephant in the room". Why Chopin and Michael Jackson, the reader might be asking?

On the one hand, Frederic Chopin studied the piano and was introduced to what we now call "classical music" since a very early age. Most of his compositional works display complex instrumental orchestration which presupposes an advanced knowledge of instrumental performance (Johnson, 2002). On the other hand, Michael Jackson began his singing career at six years of age and, since then, the performer was submitted to years of vocal training, namely the "Speech-Level-Singing" technique for control of muscles, such as the larynx, pharynx, and diaphragm (Riggs, 1992).

Indeed, both are world famous musicians with several years of musical expertise; nonetheless, one would argue that their musical training is not the same. Assuming that musical training is not the only accountable factor for facilitated emotional recognition, could the type of expertise also influence the auditory processing of emotions? Even though there has been a growth in research regarding the topic of musical training in emotional processing, to the best of our knowledge, there are no studies that probe how different types of musical training modulate the processing of emotions expressed by non-verbal vocalizations and speech prosody. Furthermore, singers have been widely overlooked (Barret, 2013).

Singing requires the production of musical sounds with one's own voice, which gathers competences of language, speech, and musical expertise (Mavridis & Pyrgelis, 2016). In fact, singing and speaking indulge in a similar activation of motor musculature, such as the larynx, the pharynx and diaphragm (Jürgens, 2009). It would, thus, come as no surprise that vocally trained individuals tend to exhibit a greater activation of sensory and motor brain regions (Zarate, 2013). When asked to sing, individuals with vocal training showed an activation in regions such as the interior insula, temporal sulcus and anterior cingulate cortex, revealing a specificity of those areas in vocal training (Zarate & Zatorre, 2008). Not many studies have been developed with singers, however some research has tackled the differences in neural activation of different types of musicians.

An interesting study by Krishnan et al. (2018) used fMRI to scan a sample of non-musicians, trained guitarists and non-classical musicians - expert beatboxers who, similarly to singers, predominantly use their vocal apparatus to produce sound. The main aim of this investigation was to ascertain whether expertise dependent plasticity is modulated by the type of instrument the musician plays. Short clips of guitar progressions and beatbox recordings were played to the participants in order to assess the activity in the scan. Results showed that all musicians exhibited a strong activation in sensorimotor regions, such as the inferior parietal cortex and the frontal cortex. More importantly, the authors highlight an instrument-expertise effect, in which musical training is associated with recruitment of auditory and sensorimotor networks depending on the expertise. I.e., guitar players responded selectively to guitar sequences, whilst beatboxers responded selectively to beatboxing sequences. Therefore, these results demonstrate that auditory perception is not simply driven by properties of an auditory stimuli, but it is influenced by the auditory-motor knowledge and experience of the listener.

Considering this, one would argue that the singer's neuronal specificities are of great value to auditory processing research, since the singer's instrument is contained within its voice. It is hypothesised that activation in the sensorimotor cortex areas for the production of sound is associated with an auditory feedback response (Mavridis & Pyrgelis, 2016). Musically trained singers are masters at regulating this feedback response, hence allowing for greater pitch and timbre control when singing (Kleber et al., 2017). A study by Banissy and colleagues (2010) found that the activation of the sensorimotor cortex and consequent auditory feedback response was associated with better performance in the discrimination of emotional auditory cues. Thus, one might infer that vocal musical training might influence emotional discrimination.

Indeed, previous research has reported that participants with musical vocal intensive training performed better at pitch, timbre and voice discrimination tasks than non-musicians (Chartrand & Belin, 2006; Kleber et al., 2017; Mavridis & Pyrgelis, 2016; Zarate, 2013; Zarate & Zatorre, 2008). Furthermore, there is evidence that expert singers are able to accurately produce different emotional acoustic profiles in terms of tempo, loudness, pitch and spectral balance (Scherer et al., 2015, 2017).

Accounting for these findings, a singer's brain might prove to be valuable to further understand the different neural specificities in vocal emotional recognition.

Let's answer our questions

Until this point, quite a few questions were presented regarding the topic of vocal emotional processing. Thus, it is worth reviewing what we have presented so far. Firstly, the existing evidence suggests that language processing functions through the dynamic operations of several brain systems, and not as an autonomous one. Secondly, language shares some features of these mechanisms with musical processing, namely in emotional content. Thirdly, musicians tend to exhibit a greater activation of sensorimotor systems which improve their recognition of the acoustic cues of emotional sound. Finally, we asked ourselves if this facilitation is modulated by the type of musical training. Considering these findings, the hypothesis that different types of musical training may exhibit differences in emotional sound processing is yet to be tested. As such, we set to further examine this question in the present study.

We tested musically trained individuals divided by musical expertise, i.e., the instrumentalists' group and the singer's group were compared with non-musically trained participants in the recognition of emotions in speech prosody and in nonverbal vocalizations (e.g., crying, or laughing).

It is worth mentioning that, although speech prosody and nonverbal vocalizations are two sources of vocal transmission, their underlying production and perception mechanisms seem to differ (Correia et al., 2020; Juslin & Laukka, 2003; Paulmann & Kotz, 2008; Pinheiro et al., 2015). As in previous research, the combination of both vocal signals allows us to determine if the effect of musical expertise is specific to emotional prosody or if it extends to the recognition of nonverbal vocal emotions (e.g., Correia et al., 2020).

We have covered a wide range of emotional categories by including six basic emotions: anger, disgust, fear, happiness, neutral, sadness. Additionally, the stimuli underwent previous perceptual and acoustic validation or were previously utilized (Castro & Lima., 2010; Castro, Lima & Scott 2013) to ensure that the expression conveyed the intended meaning. Two measures were taken for the purpose of this investigation: emotional recognition accuracy rates and response times (RT's). These measures were intended to enable the analysis of the processing speed and accuracy of participants since, as we have previously seen, musicians might show increased responsiveness to emotional auditory salience (Chartrand & Belin, 2006; Lima & Castro, 2011; Paulmann & Kotz, 2008; Pinheiro et al., 2015). The Gold-MSI and a set of auditory perceptual ability tasks (Mullensiefen et al., 2014) were applied for control purposes, since musical engagement and good musical abilities may be associated with better emotional recognition. Furthermore, these tasks may help us understand individual differences in categorical emotional recognition which have been overlooked in literature (Correia et al., 2020; Martins et al., 2021).

Based on what we reviewed so far if, emotional auditory processing is modulated by musical expertise, then it is expected that instrumentalists and singers have an advantage over non musically trained participants. Hence, this would favour the hypothesis of the shared mechanisms of music and language (Lima & Castro, 2011). Regarding the correlation analyses we expect an association between higher scores in psychoacoustic abilities higher performance in the emotional recognition task, namely in the pitch discrimination task, since this association has been previously reported (e.g., Globerson et al., 2013). We also expect an association between the self-reported musical sophistication (Gold-MSI) and emotional recognition, mainly in the singing abilities sub-scale as previously observed by Correia and colleagues (2020).

Nonetheless, what the present investigation adds is the idea that the enhanced auditory feedback mechanism in singers might influence the discrimination of vocally transmitted emotions. Like Krishnan's (2018) findings with guitarists and beatboxers, we expect to observe differences based on the specificities of musical expertise. As such, we expect the singers' group to exhibit a better performance over the instrumentalists' group and the non-musicians' group in vocal emotional recognition, in both vocalizations and speech prosody, along with a faster response time to the emotional stimuli.

2. Method

Participants

A G-power analysis was conducted with a η_p^2 value of 0.06 which allows one to detect a moderate effect size of 0.25 (Cohen, 1988). Accounting for the five measurements being collected (the Gold-MSI + three psychoacoustic tasks + emotional recognition task), the ideal sample size should consist of 96 participants (i.e., 32 per group of musical expertise). A total of 90 participants were initially scheduled to part take in the study. However, due to the unfolding pandemic restrictions during the data collection phase, we were not able to recruit the expected sample. Thus, we tested a total of 31 participants distributed into the three groups according to musical expertise: 9 non-musicians (3 men; 6 women); 13 instrumentalists (8 men; 5 women), and 9 singers (4 men; 5 women). Table 1 displays their demographic and musical background characteristics. The recruitment sites were local music schools and orchestras including *Conservatório de Música de Lisboa*, *Escola Superior de Música de Lisboa*, and local choirs such as *Coro* Gulbenkian and *Coro da Universidade de Lisboa*.

Inclusion criteria were as follows: (a) European Portuguese as the first language; (b) right handedness; (c) no history of psychiatric or neurologic illness; (d) no history of substance abuse; and (e) no auditory or visual impairments. The sample was matched for age, socioeconomic status (European Socio-economic Classification or ESeC – Rose & Harrison, 2007), and educational level. *Table 1* displays their demographic and musical background characteristics.

Participants in the musician's groups (instrumentalists and singers) were included if they had a minimum of five years of formal training and regular practice at the moment of testing (*M instrumentalists* = 10.46, *SD* = 5.09*; M singers* = 9.00, *SD* = 3.35*)*. Instrumentalists and singers did not differ in years of musical training *(p* = 0.651).

Table 1– Demographic and Musical Background Characteristics of the Participants

Characteristics	Controls	Instrumentalists	Singers
Age in Years	21.2 (2.8)	29.5 (10.7)	34.00 (9.12)
Education in years	14 (2.45)	14.54 (1.98)	16.11 (1.05)
Music training in years	___	10.46 (5.09)	9.00 (3.35)
Age of training onset	___	10.54 (3.64)	12.44 (5.19)
Average practice hours per week	___	8.38 (6.76)	4.78 (5.01)

Note: Standard Deviation in parenthesis

Materials

Self-reported musical engagement

As previously mentioned, we used the Gold-MSI Portuguese version (Lima et al., 2020) to assess musical engagement and self-reported abilities, as this instrument is suited to evaluate individual differences in samples who vary in musical skill and global involvement with music (Müllensiefen et al., 2014).

This instrument includes 38 items grouped into five subscales corresponding to different facets of musicality: Active Engagement (e.g., *I spend a lot of free time doing music related activities),* Perceptual Abilities (e.g., *I can tell when people sing or play out of time with the beat),* Music Training (e.g., *I would not consider myself a musician),* Singing Abilities (e.g., *I can sing or play music from memory),* and Emotional Involvement (e.g., *Pieces of music rarely evoke emotions in me)*. For each item the participant is asked to indicate their level of agreement with the statement using a seven-point Likert scale (i.e., 1 = *completely agree* to 7 = *completely disagree*).

Auditory performance tasks and musical abilities

Participants were assessed in their musical ability through the completion of three tasks: musical beat perception, pitch discrimination, and duration discrimination.

Musical beat perception was evaluated with the Beat Alignment Test (Mullensiefen et al., 2014). The task consists of the presentation of 17 short instrumental clips (10-16 s), overlaid with a metronome beep track. In four of these 17 excerpts the metronome track has been altered to mismatch the tempo of the instrument. Participants had to indicate whether the beep track was on or off the beat of the music clip. The order of trials was pseudo-randomized across participants and the task was applied in PsychoPy Experiment Builder version 3.2.4 (http://www.psychopy.org/).

During the pitch discrimination and duration task, participants were presented with three consecutive 250 ms pure tones – two with the same frequency (1000 Hz) and one with a higher frequency (ranging from 1 to 256 Hz). In each trial, participants had to identify which of the three tones was the highest. Similarly, in the duration discrimination task, participants were presented with three consecutive tones (two with the same length – 250 ms; and one with a longer length – 256 ms). In each trial, the participant had to indicate which of the three tones was the longest. Both tasks were implemented in the Psychoacoustics toolbox (Soranzo & Grassi, 2014), on Matlab (2016a).

Stimuli

The selected stimuli for the main task are categorized into speech prosody and non-verbal vocalizations. For each stimulus category we selected 12 different exemplars of the six basic emotional categories (happiness, sadness, disgust, fear, angry and neutral). These were taken from previously validated sets of stimuli (speech prosody – Castro & Lima, 2010; non-verbal vocalizations – Lima, Castro & Scott, 2013). Stimuli selection was based on the recognition rates reported on the validation studies from the different corpora to obtain a similar average recognition accuracy across stimuli (85.15 % for vocalizations and 75.28 % for speech prosody). Additionally, whenever possible, we avoided selecting stimuli with the highest accuracy to avoid ceiling-effects in our task. Appendix 1 displays the average accuracy ratings for the present selection.

Speech prosody stimuli consist of short sentences (*M* = 1386 ms, *SD* = 234ms) recorded by two female voices with neutral semantic content (e.g., *"O Quadro está na* *parede",* "The painting is on the wall") and variations in prosody to convey different emotions.

Non-verbal vocalizations were short vocal bursts (M = 971 ms, SD= 285 ms) with no verbal content (e.g., crying), recorded by male and female actors. The auditory stimuli were normalized to 70 dB of intensity with a Praat script.

Procedure - Emotional Recognition Task

After filling in the informed consent, participants were presented with two similar emotional recognition tasks that differed in the type of stimuli presented (i.e., non-verbal vocalizations and speech prosody). E-Prime 2.0 software was used to programme these tasks (www.psnet.com).

Using headphones, 12 stimuli of each emotional category were presented once, in a total of 72 trials. In each task, the order of presentation was pseudo-randomized ensuring that no more than two stimuli from the same emotional category appeared consecutively.

In the speech prosody task, participants were told they would listen to short sentences that were neutral in terms of semantic content and were asked to pay attention to the tone of voice. Consecutively, the labels of the seven emoticons were introduced and explained to ensure their adequate understanding (i.e., happiness, sadness, disgust, fear, angry, neutral and none of the above). Participants were then instructed to correctly identify the emotion conveyed by the auditory stimuli by pressing the corresponding emoticon on the keyboard.

In the non-verbal vocalizations task, participants were told they would listen to short non-verbal bursts and to pay attention to the emotional content of the stimuli. Similarly, to the previous task, they were then instructed to correctly identify the emotional content by pressing the correspondent key board emoticon. Both tasks were self-paced, but participants were instructed to respond as fast as possible.

Participants were tested in a quiet room after completing the background demographic questionnaires and the Gold-MSI. The order of the tasks was randomized across participants and the testing session had an approximate time of 2 hours. The auditory stimuli were presented via high quality headphones with the appropriate volume for each participant.

Data Analysis

Data analysis for the control measures and emotional recognition task was conducted using the IBM SPSS software version 25.

We chose three music perception tasks associated with prosodic recognition (beat perception, duration discriminations and pitch discrimination – Müllensiefen., 2014). It is important to notice that some data were missing, namely in the beat perception task (n = 3) due to participants failing to complete the test. Hence, the sample sized varied slightly in data analysis of this task. For comparisons between participants, we tested for the possibility of emotion-specific effects, to ensure that associations with musical expertise were not driven by a small subset of emotions via correlation.

Concerning the emotional recognition task, a response was considered accurate when it matched the intended expression of the utterance stimuli. The proportion of correct identifications for each emotion were computed individually by means of the unbiased hit rate accuracy measure, or Hu_Scores (Wagner, 1993). Hu scores represent the probability that a given emotion will be correctly utilized, thus varying between 0 and 1. These scores are calculated considering three values: the frequency of correct answers for the correct category, for example happiness (hits); the total number of stimuli presented in that category (*n* of happiness stimuli); and the total number of times in which the participant chose happiness as a response (*n* of happiness responses). According to Wagner (1993), the formula for transforming the raw scores is as follows: $(hits)^2 / (\text{total of happy stimuli} \times \text{number happiness responses})$. We proceeded to convert the raw data using this method for each participant in the different emotional categories. Hu = 0 when no stimulus from an emotional category is correctly recognized, whilst Hu = 1 when all stimuli from a given emotion have been correctly recognized and the corresponding category is always correctly used. Hence, using this measure allows us to control for possible ceiling effects and biased categorical responses.

Response times (RTs) were measured from stimulus onset till the first button press (response selection). We considered the handedness of the participants and made sure that responses were given with the index finger of the dominant hand, which was kept in the same position for the entirety of the task.

Additionally, a complementary analysis was executed to assess the possibility of emotion-specific effects, to ensure that associations with expertise were not driven by small sub-sets of emotions (Correia et al., 2020; Pinheiro et al., 2015).

3. Results

Self-Reported Music Abilities

As expected, musical expertise seems to have a significant effect on self-reported musical abilities on all sub-scales ($p < 0.05$). Tukey post-hoc comparisons on the three groups indicate that the singers and instrumentalists' groups had significantly higher ratings than the non-musical group (respectively, $p = 0.003$; and $p < 0.001$). Comparisons between the singers and instrumentalists were not statistically significant on all sub-scales, ($p > 0.001$).

Furthermore, we used Pearson's correlation to test if there was any association between self-reported abilities and emotional recognition. Globally, we found no evidence of a correlation between self-reported musical engagement and emotional recognition in both conditions ($p = .182$).

Psychoacoustic Tasks Performance

We initially questioned if we could find similar results to the self-reported measures using the psychoacoustic performance-based tasks.

A one-way ANOVA was computed to test auditory perceptual abilities with type of expertise as the independent factor and the psychoacoustic task scores as the dependent variables (respectively, duration perception, pitch perception, beat perception). Mean results are presented in Table 2. Interestingly, differences between groups were only found in the pitch recognition task [F (2, 28) = 4.76, $p = .017$]. Post-hoc comparisons were made using the Tukey test, showing that singers and instrumentalists had statistically significant higher scores than non-musicians in the pitch recognition task ($p < 0.017$). However, there was no difference between singers and instrumentalists in this task ($p = 1.000$).

We used Pearson's correlation to test if there was any association between performance on psychoacoustic tasks and emotional recognition. Globally, we found no evidence of a correlation between psychoacoustic performance and emotional recognition in nonverbal vocalizations ($r = - 0.231$, $p > 0.211$; $r = - 0.260$, $p > 0.158$; $r = - 0.123$, $p > 0.132$). However, in speech prosody, pitch discrimination exhibited a moderate association with better recognition of emotions ($r = - 0.050$, $p > 0.789$; $r = - 0.375$*, $p > 0.38$; $r = - 0.210$, $p > 0.284$).

Recognition Accuracy

The raw scores were transformed into Hu Scores and were submitted into a mixed repeated measures ANOVA with emotion (anger, disgust, fear, happiness, neutral, sadness) and stimulus type (nonverbal vocalizations and speech prosody) as the repeated-measures factors, and type of expertise as the between subjects' factor (controls, instrumentalists, and singers). Tables 2 displays the summary statistics for the observed results.

The global recognition accuracy was .76 ($SD = 0.17$) meaning that, overall, participants were able to accurately categorize the emotions presented.

An effect of stimulus type was observed, in which nonverbal vocalizations ($M =.87$; $SD = 0.17$) were associated with higher recognition rates than speech prosody stimuli ($M =.64$ ($SD = 0.25$).

A main effect of emotion was observed, in which some emotions were better recognized than others [F (5, 27) = 6.069, $p > .001$, $\eta_p^2 = .184$]. In the speech prosody condition, fear ($M = .88$, $SD = 0.34$) exhibited the highest recognition scores in comparison to all other emotions ($p = 0.001$; $p < 0.001$; $p < 0.001$; $p = 0.026$; $p = 0.003$), whilst disgust ($M =.35$, $SD = 0.53$) exhibited the statistically significant lower score in comparison to all other emotions ($p < 0.001$; $p < 0.001$; $p < 0.001$; $p < 0.001$; $p = 0.002$). Concerning nonverbal vocalizations, happiness ($M = .95$, $SD = 0.01$) displayed the highest recognition accuracy in comparison to the other emotions ($p = 0.031$; $p < 0.001$; $p = 0.014$; $p = 0.007$), except for disgust ($p = 0.473$), whilst fear exhibited the lowest score ($M =.73$, $SD = 0.39$) in comparison to all other emotions ($p = 0.01$; $p < 0.001$; $p < 0.001$; $p = 0.015$; $p < 0.001$). There were no significant differences between the three groups in emotional recognition [F (2, 27) = .408, $p > .05$, $\eta_p^2 = .029$].

Response Times

The average time to identify the target expression was computed for each emotion. Hence, we ran a mixed repeated-measures ANOVA using the type of musical expertise as the between subjects' factor, and the emotional category (anger, disgust, fear, happiness, neutral and sad) and stimulus type (nonverbal vocalization and speech prosody) as the repeated-measures factors. Table 3 displays the response times for each emotion. The results revealed a main effect of type of stimuli (nonverbal vocalizations *vs.* speech prosody): participants were faster at responding to nonverbal vocalizations (M =

1.697 ms, SD = 0.065 ms) in comparison to prosodic stimuli (M = 2532 ms, SD = 0.111 ms) [F (1, 28) = 80.518, p < 0.001, η_p^2 = .742].

An interaction effect of type of stimuli and emotion was observed [F (1, 28) = 38.766, p = 0.008, η_p^2 = .581]. Pairwise comparisons between emotional category and stimulus type show that in the case of nonverbal vocalizations, the slowest response time was fear in comparison to other emotions (p < 0.001; p < 0.001; p < 0.001 p < 0.001; p < 0.001), whilst the fastest was disgust, but only significantly in comparison to fear and sadness (p =0.069; p < 0.001; p = 0.346; p = 0.181; p = 0.021). In the case of speech prosody, participants exhibited a faster response time when categorizing neutral stimuli in comparison to all other categories, except for happy, and sad emotions (p < 0.001; p < 0.001; p < 0.001; p = 0.222; p = 0.076). The slowest response time was observed in disgust in comparison to all other categories, except for anger (p = 0.675; p = 0.035; p < 0.001; p < 0.001; p < 0.001). Table 3 illustrates the response times per emotional category in each stimulus type.

Contrary to what we expected, there were no differences in response times between the three groups, thus we observed no main effect of musical expertise in response times [F (2, 28) = 2.346, p = 0.114, η_p^2= .144].

Table 2 – Results of the tasks for the full sample as a function of musical expertise

Task	Instrumentalists	Singers	Non-musicians	*p*-value
	M (SD)	*M (SD)*	*M (SD)*	
Gold-MSI				
Active Engagement	4.87 (0.59)	4.54 (0.67)	3.33 (0.86)	p < 0.001
Perceptual Abilities	5.90 (0.63)	6.26 (0.55)	4.90 (0.99)	p = 0.001
Musical Training	5.54 (0.97)	5.62 (0.66)	1.56 (0.37)	p < 0.001
Singing Abilities	4.89 (1.09)	5.81 (0.72)	3.75 (1.23)	p < 0.001
Emotions	5.76 (0.57)	5.98 (0.60)	5.13 (0.78)	p = 0.023
General Index	5.16 (0.61)	5.52 (0.52)	3.17 (0.59)	p < 0.001
Psychoacoustic Tasks				
Beat Perception	0.42 (0.18)	0.51 (0.12)	0.54 (0.15)	p = 0.216
Pitch Discrimination	0.59 (0.24)	0.65 (0.30)	6.12 (8,41)	p = 0.017
Duration Discrimination	4.66 (2.44)	4.22 (2.07)	3.92 (1.66)	p = 0.723
Emotion Recognition				
Prosody	0.66 (0.04)	0.65 (0.05)	0.61 (0.05)	p = 0.706
Vocalizations	0.89 (0.02)	0.87 (0.03)	0.86 (0.03)	p = 0.746

Note. For the Gold-MSI and Psychoacoustic tasks the *p*-values correspond to the main effect of group (musical expertise) in a one-way analysis of variance (ANOVA). For the emotion recognition tasks, the *p*-values correspond to the main effect of group (musical expertise) in a mixed-design analyses of variances (ANOVAs), including type of expertise as the between subject factor and emotion as the repeated-measures factor. Pitch and Duration discrimination tasks are computed in frequency (Hz) and milliseconds (ms), which means that lower scores correspond to better performance.

Table 3– Performance per emotional category in accuracy and response times (RT's)

	Anger	**Fear**	**Disgust**	**Happiness**	**Neutral**	**Sadness**
Condition	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Vocalizations						
Accuracy	.88 (.026)	.73 (.039)	.92 (.035)	.95 (.013)	.85 (.038)	90 (.016)
RT's (ms)	1660 (.079)	2129 (.124)	1518 (.063)	1575 (.075)	1649 (.100)	1647 (.056)
Prosody						
Accuracy	.68 (.043)	.81 (.034)	.35 (.053)	.65 (.0.25)	.73 (.030)	.61 (.059)
RT's (ms)	2941 (.0168)	2770 (.136)	2994 (.170)	2176 (.114)	2065 (.092)	2248 (.098)

Complementary Analyses

To further understand the impact of individual differences in emotional recognition for each emotion type, we ran Pearson's correlation in each group to test the association between recognition for each emotional category and the control measurements (Gold-MSI and psychoacoustic tasks). Annexes 2 to 7 display the summary results.

According to our results, non-musicians exhibit a strong negative correlation between the self-reported singing abilities (SA) and the recognition of prosodic sadness, *r* (9) = -.870, *p* = .002; as well as a moderate correlation between self-reported emotional engagement with music (EM) and the recognition of prosodic happiness, *r* (9) = .685, *p* = .042.

Concerning psychoacoustic performance, non-musicians exhibited a moderate negative correlation between duration perception and the recognition of prosodic sadness, *r* (9) = -.688, *p* = .040, as well as a strong negative correlation between pitch perception and happy prosodic content *r* (9) = -.809, *p* = .008. Additionally, non-musicians exhibit a strong negative correlation between pitch perception and vocalized anger *r* (9) = -.928, *p* < .001.

Singers displayed a moderate association between the recognition of vocalized fear and the singing abilities sub-scale r (9) = .671, p = .048, as well as a moderate correlation between the recognition of prosodic fear and the perceptual abilities sub-scale r (9) = .713, p = .031. No significant correlations were found regarding psychoacoustic tasks (p > .05).

A moderate positive correlation between instrumentalists' self-reported singing abilities and recognition of vocalized anger was observed, r (13) = .649, p = .016.

Concerning the psychoacoustic tasks, instrumentalists' results exhibited a negative correlation between the duration discrimination task and vocalized anger r (13) = -.649, p = .016.

4. Discussion

Do different types of musical expertise influence emotional processing of the voice in different ways? We tried to answer this question by testing if vocal musical experts, instrumental musical experts, and non-musicians differed in the processing of emotional vocalizations and speech prosody. Our main hypothesis was that singer's vocal expertise facilitated the processing and discrimination of emotional prosody in comparison to instrumentalists and non-musicians (Lundy et al., 2000; Zarate, 2013).

Firstly, we asked ourselves if our participants differed in musical engagement and overall musical ability since these features may have an association with acoustic emotional recognition. As expected, musicians, regardless of type of musical training, seem to have higher rates of musical engagement and sophistication in comparison to the non-trained participants (Larrouy-Maestri et al., 2019; Müllensiefen et al., 2014). However, when it came to the psychoacoustic tasks, only the pitch perception task showed that singers and instrumentalists had significantly higher scores in comparison to non-musicians. The same was not true for the beat perception task and the duration discrimination task, where there was no difference between the three groups. In fact, there is evidence that exposure to musical training greatly influences the perception of the acoustic properties of sounds, namely frequency detection (McDermott & Oxenham, 2008; Powner, 2013). However, the same influence is yet to be replicated when it comes to the rhythmic properties of sound. For example, a study by Honing and Ladinig (2009) proposed that the processing of tempo and rhythmic properties are facilitated by actively listening to music, as opposed to having formal musical training. This might explain the similar scores in the beat and duration perception tasks in our sample. No correlations between the self-reported musical engagement scores and psychoacoustic tasks were found in relation to emotional recognition. Thus, these findings confirm that any differences in emotional recognition scores were due to musical expertise.

As expected, we observed a difference between stimulus types, in which nonverbal vocalizations were more accurately and quickly recognized than speech prosody. This is in line with previous literature showing that vocal bursts are decoded more accurately than speech prosody (Lausen & Hammerschmidt, 2020). These behavioural findings have been extended by neurophysiological evidence showing that nonverbal vocalizations facilitates early stages of perceptual processing since they require less cognitive effort due to the lack of semantic content (e.g., Lausen & Hammerschmidt, 2020; Liu et al., 2012; Martins et al., 2021; Pell et al., 2015; Pinheiro et al., 2015).

An interaction effect between stimulus type (speech prosody *vs.* nonverbal vocalizations) and emotional category was observed, in which emotional categories were differently recognized amongst themselves. Three emotions stood out: happiness, disgust, and fear.

In line with previous research, happiness was easily recognized in the nonverbal vocalizations condition (e.g., Correia et al., 2020; Kamiloğlu et al., 2020; Martins et al., 2021). Indeed, it seems that laughter has a very distinct acoustic pattern. It is usually characterized by high intensity and frequency, which may enable a facilitated processing of this emotion (Kamiloğlu et al., 2020; Liu et al., 2012). Interestingly, disgust was a close second in recognition accuracy for this condition. In accordance to Banse and Scherer (1996), naturally occurring vocal expressions of disgust mostly consist of affective bursts or vocal emblems such as “yuck!” rather than long sentences spoken with “disgust-specific” voice quality, hence explaining these results.

Fear was the least accurately recognized emotion in nonverbal vocalizations, also displaying the slowest response time. Interestingly enough, one would expect that fearful vocalizations would have a fast response time and accuracy due its evolutionary advantage associated with danger avoidance (Larson et al., 2006; Méndez-Bértolo et al., 2016; Sauter & Eimer, 2010). Nonetheless, fearful vocalizations seem to have an ambiguous acoustic profile. Indeed, research has shown that vocalized fear may share features with several other different emotions in terms of intensity, quality and frequency, such as despair and anger, making it difficult to distinguish from these emotions (Banse & Scherer, 1996; Correia et al., 2020; Lausen & Hammerschmidt, 2020).

Contrarily to what was observed in the previous condition, fear was the emotional category with highest rates of accuracy in speech prosody. We would argue that the inherent longer duration of the prosodic stimuli would help reduce the ambiguity of the fearful emotion, since research has highlighted that longer vocal stimulus are associated with higher ratings of confidence in response selection (Castiajo & Pinheiro, 2021; Lausen & Hammerschmidt, 2020; Pell & Kotz, 2011). Furthermore, in line with previous literature, disgust exhibited the lowest scores of accuracy in the speech prosody condition (e.g., Scherer, 2003). It has been established that disgust is an emotion with emphasis on the pantomime expression of repulse, as opposed to a specific set of acoustic properties in speech, thus explaining the difference of scores between the prosodic and vocalized conditions for this emotion (Banse and Scherer, 1996).

A complementary analysis of individual differences on emotional recognition was carried out between the control variables (musical engagement and auditory perceptual abilities) and the performance in the recognition of each emotion type. Concerning the Gold-MSI, we found that the singing abilities subscale was associated with a facilitated discrimination of some of the emotions in each group. For example, in non-musicians, reported singing engagement was strongly associated with prosodic sadness recognition; whilst in instrumentalists it was positively associated with anger recognition in nonverbal vocalizations; and in singers it was positively associated with the recognition of fear in nonverbal vocalizations. Correia et al. (2020) observed similar results in speech prosody conditions; however, our results show that it happens in nonverbal vocalizations as well. Furthermore, similar to previous research (e.g., Correia et al., 2020; Macgregor & Müllensiefen, 2019), the emotions subscale and perceptual abilities subscale were positively correlated, respectively, with non-musicians performance in recognizing happy prosodic stimuli, and singers ability to recognize prosodic fearful stimuli. These findings are relevant as they support the proposal that factors other than musical training (e.g., musical engagement, or singing abilities) might account for the facilitated auditory emotional discrimination response. Future studies should approach this proposal by assessing self-reported and performance based singing abilities and its relation to specific emotional discrimination.

Regarding the psychoacoustic tasks, pitch discrimination exhibited a strong negative association with prosodic recognition of happiness, as well as strong negative correlation with anger recognition in nonverbal vocalizations in non-musicians. Indeed, associations with pitch discrimination and "loud" emotions are to be expected, since these emotions have distinct higher frequencies in their transmission (Banse & Scherer, 1996; Macgregor & Müllensiefen, 2019). Interestingly, duration discrimination exhibited a moderate negative correlation with prosodic sadness in non-musicians. Indeed, the acoustic profile of sadness is characterized by slow rhythmic utterances, hence a better capacity of tempo detection might account for this association (Siedlecka & Denson, 2019). Nonetheless, the association was only present non-musicians. This might signal a further need to explore individual differences in the sample. Previous research has highlighted that emotional processing is associated with temperament dimensions and attentional control, suggesting that general psychological patterns might be challenged by individual specificities (Zagórska & Fajkowska, 2015).

In previous studies, an advantage for musicians in emotional recognition of vocalizations and speech prosody has been reported (e.g., Correia et al., 2020; Lima & Castro, 2011; Pinheiro et al., 2015; Rigoulot et al., 2015). However, contrary to our prediction, this was not the case. Even though non-musicians exhibited the lowest accuracy in emotional recognition of vocalizations and speech prosody, no significant differences were observed between the groups. We propose that this contradiction can be further explained by the reduced number of participants in the study. Due to the restrictions imposed by COVID-19 pandemic, we were not able to recruit the necessary sample, which reduced the statistical power of our findings. Nonetheless, these results may also account for the unexplored variability of impactful factors in emotional recognition of musicians. Prosodic emotional recognition in musicians has produced mixed results with several studies failing to find any significant differences between musically trained individuals and non-musicians (e.g., Chari, 2019; Fuller., 2018 Mualem & Lavidor, 2015; Thompson et al., 2004). One would argue that these inconsistent findings are due to the lack of accountability of other influencing factors besides musical training. Indeed, cross-sectional studies involving highly trained musicians do not allow one to tease apart factors such as genetic contributions, cognitive abilities, personality, or socioeconomic status that might, also, impact auditory emotional processing (Martins et al., 2021).

Our main hypothesis rested on idea that the facilitated auditory feedback response observed in musically trained singers would enable one to better recognize complex psychoacoustic properties of emotion, such as in speech prosody. Contrary to our hypothesis, there were no differences between groups in both conditions (vocalizations and speech prosody), thus supporting the idea that the type of musical expertise does not influence the processing of emotional recognition in vocalizations and prosodic verbal stimuli.

Indeed, whilst singing and speaking share common means of sound production, differences in their acoustic properties such as loudness, spectral properties, noise to harmonic ratio and frequency have been reported (e.g., Hansen et al., 2020; Juslin & Laukka, 2003; Livingstone et al., 2013; Lundy et al., 2000; Scherer, 2003). In line with this idea, research has highlighted the possibility that the auditory feedback system differs in modulation responses in singing *vs.* speaking (Blumberg, Freeman & Robinson., 2009; Natke et al., 2003). This might explain why even though singers are exceptional in the production and recognition of sang emotional vocal expressions, results seem to be scarce regarding emotional recognition in speech (Scherer, 2017; Scherer, Trznadel, Fantini & Sundberg, 2017). Further research should assess the impact of vocal music training in the recognition of emotional acoustic variations of speaking and singing.

Limitations

Shortcomings of the current study should be discussed. The first concerns our sample. As previously mentioned, our sample consists of only 31 participants divided into three groups. Ideally, for this study, at least 96 participants (32 per group) were required for solid statistical inferences. Another concern regarding our sample is that a few of the instrumentally trained participants had some form of vocal training. Even though not enough to be considered experts (more than 5 years) it may still influence the data. Another shortcoming arises from the experimental paradigm. Our design is based on a quasi-experimental method, not allowing for the random distribution of participants. Hence, we cannot ascertain that the results were not influenced by other variables that might influence emotional recognition, such as emotional intelligence, cognitive abilities, age of perceiver, or personality traits (Brück et al., 2011; Correia et al., 2020; Martins et al., 2021; Pinheiro et al., 2015; Seung et al., 2009; Trimmer & Cuddy, 2008). Future studies on the topic should use a larger and more representative sample of participants with these factors in mind, while still focusing on specificities of musical expertise on emotional recognition. It would also be interesting to assess the neuronal correlates of emotional processing in different musical experts by using EEG. Indeed, the usage of event related potentials (ERP) methodology is particularly advantageous in emotional recognition since it affords tracking neurocognitive processes as they happen in real time from stimulus onset until a response is made (Liu et al., 2012). Thus, it would prove useful to track possible differences between neuronal processing of sound in different types of musical expertise. One would, additionally, argue that a larger range of emotions should be researched. In the present study we have only used basic emotions, thus it would be of relevance to further explore more complex emotions, such as shame or guilt.

5. Concluding remarks

Are musicians the native speakers of the emotional language? So far, this question has no easy answer.

It has been well accepted that musical expertise facilitates emotional recognition, namely in auditory processing. However, only recently have researchers proposed the possibility that other factors besides musical training might help explain this facilitated response. In line with this idea, our main hypothesis focused on the specificities of musical singing training in emotional recognition of speech prosody and vocalizations. Contrary to what we expected, our results did not foreground a possible association between singing expertise and a facilitated response towards vocally transmitted emotions. Nonetheless, we found that measures such as self-reported musical engagement and auditory perceptual abilities were associated with specific emotions in each group of musical expertise.

Despite the results, previous studies highlight that a singer's brain is a good example of musically induced neuroplasticity with different characteristics than other types of expertise. Specificities regarding sensorimotor activation, feedback response modulation, and emotional processing are highlighted in vocal musical training, which may prove to be of relevance for research in auditory processing. Hence, we encourage future research to study the neural characteristics of this population by using EEG and fMRI tools. Indeed, these methods may prove to be valuable in ascertaining the neuronal differences of expertise in higher order processing mechanisms, such as auditory emotional processing. Furthermore, studies are needed to address the issue of factor variability in emotional recognition. Well-powered and designed studies should include variables such as musical engagement, personality traits, cognitive ability, and expertise to further understand the associations between music and emotion.

6. References

Acer, N., Bastepe-Gray, S., Sagiroglu, A., Gumus, K. Z., Degirmencioglu, L., Zararsiz, G., & Ozic, M. U. (2018). Diffusion tensor and volumetric magnetic resonance imaging findings in the brains of professional musicians. *Journal of Chemical Neuroanatomy*, *88*, 33–40. https://doi.org/10.1016/j.jchemneu.2017.11.003

Arbib, M. A. (2013). Précis of How the brain got language: The Mirror System Hypothesis . *Language and Cognition*, *5*(2–3), 107–131. https://doi.org/10.1515/langcog-2013-0007

Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: Evidence for shared domain-general mechanisms. *Frontiers in Psychology*, *4*(JUN), 1–14. https://doi.org/10.3389/fpsyg.2013.00321

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expressions. Journal of Personality and Social Psychology. *70*, *70*(3), 614–636.

Barrett, K. C., Ashley, R., Strait, D. L., & Kraus, N. (2013). Art and science: how musical training shapes the brain. *Frontiers in Psychology*, *4*(October), 1–13. https://doi.org/10.3389/fpsyg.2013.00713

Besson, M., Magne, C., & Schön, D. (2002). Emotional prosody: Sex differences in sensitivity to speech melody. *Trends in Cognitive Sciences*, *6*(10), 405–407. https://doi.org/10.1016/S1364-6613(02)01975-7

Besson, M., & Schön, D. (n.d.). *Comparison between Language and Music*. 232–258.

Bialystok, E., & DePape, A. M. (2009). Musical Expertise, Bilingualism, and Executive Functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 565–574. https://doi.org/10.1037/a0012735

Brown, S. (2017). A joint prosodic origin of language and music. *Frontiers in Psychology*, *8*(OCT), 1–20. https://doi.org/10.3389/fpsyg.2017.01894

Brück, C., Kreifelts, B., Kaza, E., Lotze, M., & Wildgruber, D. (2011). Impact of personality on the cerebral processing of emotional prosody. *NeuroImage*, *58*(1), 259–268. https://doi.org/10.1016/j.neuroimage.2011.06.005

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. https://doi.org/10.3758/BRM.42.1.74

Chartrand, J. P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*(3), 164–167. https://doi.org/10.1016/j.neulet.2006.06.053

Chiang, J. N., Rosenberg, M. H., Bufford, C. A., Stephens, D., Lysy, A., & Monti, M. M. (2018). The language of music: Common neural codes for structured sequences in music and natural language. *Brain and Language*, *185*(July), 30–37. https://doi.org/10.1016/j.bandl.2018.07.003

Correia, A. I., Castro, S. L., Macgregor, C., Müllensiefen, D., Schellenberg, E. G., Correia, A. I., Macgregor, C., Müllensiefen, D., & Schellenberg, E. G. (2020). Emotion With Naturally Good Musical Abilities Good Musical Abilities. *Emotion*.

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology*, *5*(APR), 1–17. https://doi.org/10.3389/fpsyg.2014.00335

Globerson, E., Amir, N., Golan, O., Kishon-Rabin, L., & Lavidor, M. (2013). Psychoacoustic abilities as predictors of vocal emotion recognition. *Attention, Perception, and Psychophysics*, *75*(8), 1799–1810. https://doi.org/10.3758/s13414-013-0518-x

Hansen, J. H. L., Bokshi, M., & Khorram, S. (2020). Speech variability: A cross-language study on acoustic variations of speaking versus untrained singing. *The Journal of the Acoustical Society of America*, *148*(2), 829–844. https://doi.org/10.1121/10.0001526

Honing, H., & Ladinig, O. (2009). Exposure Influences Expressive Timing Judgments in Music. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 281–288. https://doi.org/10.1037/a0012732

Hutchinson, S., Lee, L. H., & Gaab, N. (2003). *cercor%2F13.9.943*. 943–949.

Indefrey, P., & Levelt, W. J. M. (2000). The Neural Correlates of Language Production. *The New Cognitive Neurosciences*, 845–865.

Jürgens, U. (2009). The Neural Control of Vocalization in Mammals: A Review. *Journal of Voice*, *23*(1), 1–10. https://doi.org/10.1016/j.jvoice.2007.07.005

Juslin, P. N., & Laukka, P. (2003). Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kamiloğlu, R. G., Fischer, A. H., & Sauter, D. A. (2020). Good vibrations: A review of vocal expressions of positive emotions. *Psychonomic Bulletin and Review*, *27*(2), 237–265. https://doi.org/10.3758/s13423-019-01701-x

Kleber, B., Friberg, A., Zeitouni, A., & Zatorre, R. (2017). Experience-dependent modulation of right anterior insula and sensorimotor regions as a function of noise-masked auditory feedback in singers and nonsingers. *NeuroImage*, *147*, 97–110.

https://doi.org/10.1016/j.neuroimage.2016.11.059

Kraus, N., & Slater, J. (2015). Music and language: Relations and disconnections. In *Handbook of Clinical Neurology* (1st ed., Vol. 129). Elsevier B.V. https://doi.org/10.1016/B978-0-444-62630-1.00012-3

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., Manly, T., & Scott, S. K. (2018). Beatboxers and Guitarists Engage Sensorimotor Regions Selectively When Listening to the Instruments They can Play. *Cerebral Cortex*, *28*(11), 4063–4079. https://doi.org/10.1093/cercor/bhy208

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS ONE*, *10*(11). https://doi.org/10.1371/journal.pone.0141069

Larrouy-Maestri, P., Harrison, P. M. C., & Müllensiefen, D. (2019). The mistuning perception test: A new measurement instrument. *Behavior Research Methods*, *51*(2), 663–675. https://doi.org/10.3758/s13428-019-01225-1

Larson, C. L., Schaefer, H. S., Siegle, G. J., Jackson, C. A. B., Anderle, M. J., & Davidson, R. J. (2006). Fear Is Fast in Phobic Individuals: Amygdala Activation in Response to Fear-Relevant Stimuli. *Biological Psychiatry*, *60*(4), 410–417. https://doi.org/10.1016/j.biopsych.2006.03.079

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. https://doi.org/10.1057/s41599-020-0499-z

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), 1021–1031. https://doi.org/10.1037/a0024521

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport*, *23*(2), 108–112. https://doi.org/10.1097/WNR.0b013e32834ea757

Livingstone, S. R., Peck, K., & Russo, F. A. (2013). Acoustic differences in the speaking and singing voice. *Proceedings of Meetings on Acoustics*, *19*. https://doi.org/10.1121/1.4799460

Lundy, D. S., Roy, S., Casiano, R. R., Xue, J. W., & Evans, J. (2000). Acoustic analysis of the singing and speaking voice in singing students. *Journal of Voice*, *14*(4), 490–493. https://doi.org/10.1016/S0892-1997(00)80006-5

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*(AUG). https://doi.org/10.3389/fpsyg.2019.01955

Martins, M., Pinheiro, A. P., & Lima, C. F. (2021). Does Music Training Improve Emotion Recognition Abilities? A Critical Review. *Emotion Review*, *December*, 175407392110220. https://doi.org/10.1177/17540739211022035

Mavridis, I. N., & Pyrgelis, E. S. (2016). Brain activation during singing: "clef de Sol Activation" is the "concert" of the human brain. *Medical Problems of Performing Artists*, *31*(1), 45–50. https://doi.org/10.21091/mppa.2016.1008

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452–463. https://doi.org/10.1016/j.conb.2008.09.005

Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y. H., Vuilleumier, P., Gil-Nagel, A., & Strange, B. A. (2016). A fast pathway for fear in human amygdala. *Nature Neuroscience*, *19*(8), 1041–1049. https://doi.org/10.1038/nn.4324

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0089642

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*(6), 473–478. https://doi.org/10.1038/nrn843

Natke, U., Donath, T. M., & Kalveram, K. T. (2003). Control of voice fundamental frequency in speaking versus singing. *The Journal of the Acoustical Society of America*, *113*(3), 1587–1593. https://doi.org/10.1121/1.1543928

Park, M., Gutyrchik, E., Bao, Y., Zaytseva, Y., Carl, P., Welker, L., Pöppel, E., Reiser, M., Blautzik, J., & Meindl, T. (2014). Differences between musicians and non-musicians in neuro-affective processing of sadness and fear expressed in music. *Neuroscience Letters*, *566*, 120–124. https://doi.org/10.1016/j.neulet.2014.02.041

Park, M., Gutyrchik, E., Welker, L., Carl, P., Pöppel, E., Zaytseva, Y., Meindl, T., Blautzik, J., Reiser, M., & Bao, Y. (2015). Sadness is unique: Neural processing of emotions in speech prosody in musicians and non-musicians. *Frontiers in Human Neuroscience*, *8*(JAN), 1–8. https://doi.org/10.3389/fnhum.2014.01049

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. https://doi.org/10.1038/nn1082

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717–733. https://doi.org/10.1162/089892998563121

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport*, *19*(2), 209–213. https://doi.org/10.1097/WNR.0b013e3282f454db

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). Functional contributions of the basal ganglia to emotional prosody: Evidence from ERPs. *Brain Research*, *1217*, 171–178. https://doi.org/10.1016/j.brainres.2008.04.032

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Pell, Marc D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0027256

Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., & Gonçalves, Ó. F. (2015). The music of language: An ERP investigation of the effects of musical training on emotional prosody processing. *Brain and Language*, *140*, 24–34. https://doi.org/10.1016/j.bandl.2014.10.009

Powner, A. M. (2013). Pitch perception of musicians and non-musicians: a comparison of psychophysical tuning curves and frequency difference limens. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience*, *290*, 175–184. https://doi.org/10.1016/j.neuroscience.2015.01.033

Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience*, *22*(3), 474–481. https://doi.org/10.1162/jocn.2009.21215

Scherer, K. R. (2003). *<Scherer.SpeechComm2003.pdf>*. *40*, 227–256.

Scherer, K. R., Sundberg, J., Fantini, B., Trznadel, S., & Eyben, F. (2017). The expression of emotion in the singing voice: Acoustic patterns in vocal performance. *The Journal of the Acoustical Society of America*, *142*(4), 1805–1815. https://doi.org/10.1121/1.5002886

Scherer, K. R., Sundberg, J., Tamarit, L., & Salomão, G. L. (2015). Comparing the acoustic expression of emotion in the speakingand the singing voice SCHERER , Klaus R ., et al . Comparing the acoustic expression of emotion in the speakingand Comparing the acoustic expression of emotion in the speaking and the sin. *Computer Speech and Language*, *29*, 218–235.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Scott, S. K., Sauter, D., & McGettigan, C. (2010). Brain mechanisms for processing perceived emotional vocalizations in humans. In *Handbook of Behavioral Neuroscience* (Vol. 19, Issue C). Elsevier B.V. https://doi.org/10.1016/B978-0-12-374593-4.00019-X

Seung, J. L., Lee, H. K., Kweon, Y. S., Chung, T. L., & Lee, K. U. (2009). The impact of executive function on emotion recognition and emotion experience in patients with schizophrenia. *Psychiatry Investigation*, *6*(3), 156–162. https://doi.org/10.4306/pi.2009.6.3.156

Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, *138*(December 2019), 107307. https://doi.org/10.1016/j.neuropsychologia.2019.107307

Siedlecka, E., & Denson, T. F. (2019). Experimental Methods for Inducing Basic Emotions: A Qualitative Review. *Emotion Review*, *11*(1), 87–97. https://doi.org/10.1177/1754073917749016

Soncini, F., & Costa, M. J. (2006). Efeito da prática musical no reconhecimento da fala no silêncio e no ruído. *Pró-Fono Revista de Atualização Científica*, *18*(2), 161–170. https://doi.org/10.1590/s0104-56872006000200005

Steinbeis, N., & Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS ONE*, *3*(5), 1–7. https://doi.org/10.1371/journal.pone.0002226

Sun, Y., Lu, X., Ho, H. T., Johnson, B. W., Sammler, D., & Thompson, W. F. (2018). Syntactic processing in music and language: Parallel abnormalities observed in congenital amusia. *NeuroImage: Clinical*, *19*(May), 640–651. https://doi.org/10.1016/j.nicl.2018.05.032

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional Intelligence, Not Music Training, Predicts Recognition of Emotional Speech Prosody. *Emotion*, *8*(6), 838–849.

https://doi.org/10.1037/a0014080

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28. https://doi.org/10.1007/BF00987006

Zagórska, A., & Fajkowska, M. (2015). Individual differences in visual and auditory processing of emotional material. *Polish Psychological Bulletin*, *46*(2), 174–180. https://doi.org/10.1515/ppb-2015-0023

Zarate, J. M. (2013). The neural control of singing. *Frontiers in Human Neuroscience*, *JUN*. https://doi.org/10.3389/fnhum.2013.00237

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*, *40*(4), 1871–1887. https://doi.org/10.1016/j.neuroimage.2008.01.026

Zora, H., Rudner, M., & Montell Magnusson, A. K. (2020). Concurrent affective and linguistic prosody with the same emotional valence elicits a late positive ERP response. *European Journal of Neuroscience*, *51*(11), 2236–2249. https://doi.org/10.1111/ejn.14658

7. Annexes

Annexe 1 – Average recognition rates per emotional category and type of stimuli of selection

	Vocalizations	**Prosody**
Anger	78,33	80
Disgust	95,42	50,42
Fear	65,83	70,42
Sadness	86,25	83,33
Happiness	94,58	81,25
Neutral	90,50	86,25
Total	**85,15**	**75,21**

Note: Accuracy ratings for vocalizations were taken from Lima, Castro and Scott (2013). Prosody ratings were taken from Castro and Lima (2010).

Annexe 2 – Correlation values between emotions and control measures in non-musicians for prosody stimuli

	Anger	Disgust	Fear	Happiness	Neutral	Sadness
Task	*r value*	*r value*	*r value*	*r value*	*r value*	*r value*
Gold-MSI						
Active Engagement	,033	-,413	.122	,537	-,021	-,121
Perceptual Abilities	,538	,456	,394	,594	,375	-,175
Musical Training	,220	,286	,412	,439	-,117	-,322
Singing Abilities	,194	,302	,073	,333	-,558	**-,870****
Emotions	,487	,341	,072	**,685***	,164	-,237
Psychoacoustic Tasks						
Beat Perception	-,356	-,191	-,214	-,235	,222	,566
Pitch Discrimination	-,456	-,459	-,320	**-,809****	-,662	-,092
Duration Discrimination	,491	-,088	,201	,090	-,444	**-,688***

Note: Moderate correlations are signalled with (*). Strong correlations are signalled with (**). Pitch and duration discrimination results are in "staircase format". Thus, lower results correspond to higher performance.

Annexe 3 – Correlation values between emotions and control measures in non-musicians for vocalization stimuli

	Anger	Disgust	Fear	Happiness	Neutral	Sadness
Task	*r value*	*r value*	*r value*	*r value*	*r value*	*r value*
Gold-MSI						
Active Engagement	-,407	-,216	-,137	,439	-,028	-,555
Perceptual Abilities	,320	,053	-,129	,090	-,096	,256
Musical Training	-,061	-,417	-,546	,580	-,422	-,362
Singing Abilities	,102	-,370	-,254	,310	-,506	,093
Emotions	,476	-,103	-,318	,291	-,301	-,118
Psychoacoustic Tasks						
Beat Perception	,046	-,118	,000	,179	-,022	,275
Pitch Discrimination	**-,928****	-,299	-,469	,311	,094	-,341
Duration Discrimination	,059	,113	-,504	,004	-,664	,473

Note: Moderate correlations are signalled with (*). Strong correlations are signalled with (**). Pitch and duration discrimination results are in "staircase format". Thus, lower results correspond to higher performance.

Annexe 4 – Correlation values between emotions and control measures in singers for prosody

	Anger	Disgust	Fear	Happiness	Neutral	Sadness
Task	*r value*	*r value*	*r value*	*r value*	*r value*	*r value*
Gold-MSI						
Active Engagement	-,322	-,134	,178	-,347	-,512	-,320
Perceptual Abilities	-,258	,372	**,713***	-,080	-,264	,095
Musical Training	-,301	,372	,424	,154	,415	,138
Singing Abilities	-,201	,206	,318	,484	,042	-,065
Emotions	-,512	,110	,467	,171	-,132	-,135
Psychoacoustic Tasks						
Beat Perception	-,356	-,191	-,214	-,235	,222	,566
Pitch Discrimination	-,456	-,459	-,320	**-,809****	-,662	-,092
Duration Discrimination	,491	-,088	,201	,090	-,444	**-,688***

Note: Moderate correlations are signalled with (*). Strong correlations are signalled with (**). Pitch and duration discrimination results are in "staircase format". Thus, lower results correspond to higher performance.

Annexe 5 – Correlation values between emotions and control measures in singers for vocalizations

	Anger	Disgust	Fear	Happiness	Neutral	Sadness
Task	*r value*	*r value*	*r value*	*r value*	*r value*	*r value*
Gold-MSI						
Active Engagement	-,40	-,101	-,088	-,005	,102	,231
Perceptual Abilities	,302	-,180	,265	-,475	-,445	-,207
Musical Training	,135	,323	,227	-,064	-,236	-,314
Singing Abilities	-,234	,397	**,671***	-,034	-,011	-,293
Emotions	,323	,185	,519	-,365	-,298	-,490
Psychoacoustic Tasks						
Beat Perception	-,116	,010	-,076	,117	,454	-,304
Pitch Discrimination	,656	,336	,662	,450	,057	,433
Duration Discrimination	-,271	-,523	-,266	,217	,329	,225

Note: Moderate correlations are signalled with (*). Strong correlations are signalled with (**). Pitch and duration discrimination results are in “staircase format”. Thus, lower results correspond to higher performance.

Annexe 6 – Correlation values between emotions and control measures in instrumentalists for prosody

	Anger	Disgust	Fear	Happiness	Neutral	Sadness
Task	*r value*	*r value*	*r value*	*r value*	*r value*	*r value*
Gold-MSI						
Active Engagement	-,441	-,457	-,253	-,376	-,192	-,042
Perceptual Abilities	**,555***	,235	-,152	-,357	-,086	-,009
Musical Training	-,141	-,074	-,066	-,438	-,284	-,259
Singing Abilities	-,092	-,180	,311	-,181	-,278	-,132
Emotions	,522	,140	-,154	-,224	,113	,169
Psychoacoustic Tasks						
Beat Perception	,219	,256	,537	,371	,279	-,034
Pitch Discrimination	,327	**-,609***	,107	-,494	,160	,291
Duration Discrimination	-,248	,092	,156	,170	,225	,094

Note: Moderate correlations are signalled with (*). Strong correlations are signalled with (**). Pitch and duration discrimination results are in “staircase format”. Thus, lower results correspond to higher performance.

Annexe 7 – Correlation values between emotions and control measures in instrumentalists for vocalizations

	Anger	Disgust	Fear	Happiness	Neutral	Sadness
Task	*r value*	*r value*	*r value*	*r value*	*r value*	*r value*
Gold-MSI						
Active Engagement	,088	,050	,056	-,167	-,224	-,245
Perceptual Abilities	,219	,484	,270	,177	-,200	,147
Musical Training	,259	,296	,016	-,019	-,083	-,266
Singing Abilities	**,649***	,383	,227	,130	,470	,403
Emotions	-,121	,183	,404	,064	-,457	,011
Psychoacoustic Tasks						
Beat Perception	,054	-,313	-,060	-,098	,168	-,159
Pitch Discrimination	-,170	,022	,091	-,064	-,273	-,192
Duration Discrimination	**-,649***	,248	-,419	-,128	-,323	-,427

Note: Moderate correlations are signalled with (*). Strong correlations are signalled with (**). Pitch and duration discrimination results are in “staircase format”. Thus, lower results correspond to higher performance.