



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Música e Emoções Vocais: Associações Entre Treino Musical, Aptidão Musical, e o Processamento Temporal de Emoções Vocais

Aíssa Mariama Nascimento Baldé

Mestrado em Ciências em Emoções

Orientador:

Doutor César Lima, Professor Auxiliar,
Iscte – Instituto Universitário de Lisboa

Novembro, 2021



CIÊNCIAS SOCIAIS
E HUMANAS

Música e Emoções Vocais: Associações Entre Treino Musical, Aptidão Musical, e o Processamento Temporal de Emoções Vocais

Aíssa Mariama Nascimento Baldé

Mestrado em Ciências em Emoções

Orientador:
Doutor César Lima, Professor Auxiliar,
Iscte – Instituto Universitário de Lisboa

Novembro, 2021

Dedicatória

Aos meus amigos e família

Agradecimentos

Dou por terminada uma das etapas mais importantes e desafiantes da minha vida. Foi com muito esforço que cheguei aqui, mas nada disto teria sido possível sem o apoio incondicional de várias pessoas, que passo a citar.

Primeiro, gostaria de agradecer aos meus pais. Obrigada, mãe, por todo o amor. Nunca teria chegado tão longe sem a tua ajuda e apoio incondicional. Se estou onde estou hoje, é por tua causa. Obrigada, pai. Espero que onde quer que estejas, sintas orgulho da pessoa que me tornei.

Ao professor César Lima, por todo o apoio e orientação ao longo desta jornada. Por me inspirar a ser melhor e por acreditar em mim. Obrigada por toda a disponibilidade, paciência e gentileza. Adicionalmente, gostaria ainda de agradecer ao professor Oleksander Horchak por toda a sua ajuda nas análises estatísticas.

Às minhas melhores amigas, hoje e sempre, Marta Costa, Sofia Relvas, Beatriz Gonçalves, Beatriz Martins, Beatriz Silvestre, e Margarida Nunes. Nunca teria chegado tão longe sem vocês. Obrigada por acreditarem sempre em mim.

À Ana Isabel, à Matilde, e à Sofia Menezes, as minhas eternas companheiras durante as fases mais duras deste processo. Obrigada pela vossa ajuda, foi indispensável.

E finalmente, à Marta, Teresa, Inês, e Tatiana, por me terem aturado durante este mestrado, e por viverem comigo esta aventura.

Obrigada a todos!

Resumo

Existe um interesse crescente na possibilidade de as capacidades musicais se transferirem positivamente para as competências emocionais. A presente dissertação pretende contribuir para a investigação neste tópico, explorando associações entre duas dimensões da sensibilidade à música — treino musical formal e variação natural em competências musicais — e o reconhecimento de emoções vocais. Foi usado um paradigma de gating auditivo, no qual frases expressas em quatro emoções (alegria, raiva, tristeza, medo) e neutralidade, foram apresentadas de forma parcial em segmentos de duração crescente, até à apresentação completa do estímulo. Participantes com e sem treino musical foram comparados ($N = 107$), e o nível de competência musical dos participantes sem treino foi medido, de forma a identificar indivíduos com capacidades musicais “naturalmente” boas, mesmo na ausência de treino musical. Os resultados revelaram que o treino musical esteve positivamente associado ao reconhecimento de emoções vocais quando apresentadas na totalidade. Os músicos também foram mais exatos a reconhecer emoções em prosódia quando tinham acesso a informação acústica apenas parcial. Contudo, não houve diferenças significativas entre grupos no ponto de reconhecimento das emoções. Quanto aos participantes sem treino musical, aqueles com competências musicais naturalmente elevadas também apresentaram algumas vantagens no reconhecimento de emoções, mas as associações entre variação nas competências de percepção musical e o reconhecimento de emoções foram genericamente fracas ou não significativas. Na sua globalidade, estes resultados sugerem que o treino musical está associado a melhor desempenho no reconhecimento de emoções vocais, e realçam a importância de considerar a natureza multifacetada da musicalidade.

Palavras-chave: Treino musical, sofisticação musical, reconhecimento de emoções, gating

Classificação APA PsycINFO:

2300 Psicologia Experimental Humana

2326 Percepção Auditiva e Vocal

2360 Motivação e Emoção

Abstract

There is a growing interest in the possibility that musical abilities transfer positively to emotional abilities. This dissertation aims to contribute to the research in this topic, exploring associations between two dimensions of music sensitivity — formal musical training and natural variation in musical skills — and the recognition of emotions in prosody. An auditory gating paradigm was used, in which sentences expressing four basic emotions (joy, anger, sadness, fear) and neutrality, were presented to participants in segments of partial acoustic information, in gates of increasing duration, until the full utterance was presented. Musically trained and untrained individuals were compared ($N = 107$), and, additionally, the level of musical ability of the non-musician participants was measured, in order to identify individuals with “naturally” good musical abilities, even in the absence of musical training. The results revealed that musical training was positively associated with the recognition of vocal emotions when presented in full. Musicians were also more accurate at recognizing emotions in prosody when they had access to partial acoustic information only. However, there were no significant differences between groups in emotion identification points. As for participants without musical training, those with naturally good musical skills also showed some advantages in recognizing emotions, but the associations between variation in musical perception skills and vocal emotion recognition were generally weak or non-significant. Overall, these results suggest that music training is associated with better performance in the recognition of vocal emotions and highlight the importance of considering the multifaceted nature of musicality.

Keywords: Music training, musical sophistication, emotion recognition, gating

APA PsycINFO Classification:

2300 Human Experimental Psychology

2326 Auditory & Speech Perception

2360 Motivation & Emotion

Índice Geral

Introdução	1
Capítulo I – Enquadramento Teórico.....	5
Prosódia Emocional.....	5
Associações entre Música e Prosódia Emocional	6
<i>Treino Musical</i>	6
<i>Capacidades Musicais “Naturais” e Reconhecimento de Prosódia Emocional</i>	9
Tarefas de Reconhecimento de Prosódia Emocional	11
<i>Paradigma de Gating</i>	12
Objetivos e Hipóteses.....	17
Capítulo II - Método	21
Participantes	21
Instrumentos	23
<i>Capacidades e experiência musical (autorrelato)</i>	23
<i>Capacidades objetivas de percepção musical</i>	24
<i>Reconhecimento de prosódia emocional</i>	26
<i>Seleção de estímulos</i>	26
<i>Construção dos gates</i>	29
<i>Tarefa experimental</i>	29
<i>Funcionamento cognitivo geral</i>	30
Procedimento.....	31
Análise de Dados.....	32
<i>Medidas cognitivas e musicais</i>	32
<i>Musicalidade</i>	32
<i>Gating</i>	32
<i>Análises estatísticas</i>	33
Capítulo III. Resultados	35
Dados cognitivos e de musicalidade	35
Exatidão ao longo dos gates	37
Reconhecimento Acima do Nível do Acaso	40
Pontos de Identificação da Emoção (EIPs)	41
Treino Musical Formal.....	42
<i>Exatidão do reconhecimento</i>	42
<i>Desempenho em condições de reconhecimento mais exigentes</i>	43
<i>EIPs</i>	45

<i>Duração do treino musical</i>	46
<i>Capacidades cognitivas</i>	48
Competências de Percepção Musical: Autorrelato.....	49
Competências de Percepção Musical: Medidas Objetivas	50
Diferentes Níveis de Competência Musical em Não-Músicos.....	52
Capítulo IV. Discussão	57
Referências.....	69
Anexos	77
Anexo A - Confusões mais comuns entre emoções em cada intervalo de gate	77
Anexo B - Informação adicional acerca do modelo global que estimou exatidão ao longo dos gates em função da emoção e grupo	78
Anexo C - Distribuição de pontos de identificação corretos e incorretos para cada emoção em função do intervalo de gate	86

Índice de Quadros

Quadro 2.1. Características acústicas e perceptivas dos estímulos selecionados para cada emoção.....	28
Quadro 3.1. Estatísticas descritivas das medidas musicais e cognitivas da amostra completa, e separadamente para músicos e não-músicos.....	35
Quadro 3.2. Correlações entre a duração do treino musical e as variáveis musicais e cognitivas.....	36
Quadro 3.3. Média de categorizações corretas para cada emoção, gate e grupo.....	39
Quadro 3.4. Média de valores <i>Hu</i> em função da emoção, gate e grupo.....	39

Índice de Figuras

Figura 2.1. Distribuição do treino musical formal na amostra ($N = 107$).....	23
Figura 2.2. Exemplo de ensaio no subteste de Melodia.....	26
Figura 2.3. Exemplo de ensaio na tarefa de gating.....	30
Figura 3.1. Exatidão do reconhecimento de diferentes emoções prosódicas, em função da duração de gate e do grupo (músicos e não-músicos).....	37
Figura 3.2. Pontos de identificação da emoção (EIP) para cada emoção (ms).....	41
Figura 3.3. Diferenças entre músicos e não-músicos na exatidão do reconhecimento de emoções.....	43
Figura 3.4. Diferenças entre músicos e não-músicos na exatidão do reconhecimento de emoções quando a informação acústica era incompleta (desempenho entre G50 e G1000).....	44
Figura 3.5. Diferenças entre músicos e não-músicos na exatidão do reconhecimento de emoções quando a informação acústica era incompleta (desempenho no G100).....	45
Figura 3.6. EIPs para cada emoção em função do grupo.....	46
Figura 3.7. Diferenças na exatidão do reconhecimento de emoções em função de diferentes níveis de musicalidade.....	52
Figura 3.8. Diferenças entre músicos, não-músicos de alta-musicalidade e não-músicos de baixa musicalidade na exatidão do reconhecimento de emoções quando a informação acústica era incompleta (desempenho entre G50 e G1000).....	54
Figura 3.9. EIPs para cada emoção em função do nível de musicalidade dos participantes.....	55

Introdução

A comunicação vocal é uma função biológica essencial para o estabelecimento e manutenção de relações sociais. Durante as interações sociais apoiamo-nos em pistas acústicas verbais e não-verbais para decifrarmos informação linguística e emocional relevante, o que nos permite responder de forma adequada às necessidades da outra pessoa e regular o nosso comportamento.

De forma semelhante, a música funciona como um sinal acústico extremamente eficiente na comunicação de emoções (Swaminathan & Schellenberg, 2015). A música é um canal importante de comunicação durante a infância, promovendo a atenção, a regulação social, e a sincronia emocional entre bebés e cuidadores (Nakata & Trehub, 2004; Trehub, 2003). Para além disso, está também no centro de comportamentos afiliativos, promovendo comportamento pró-social (Tarr et al., 2014) e a sincronização entre indivíduos (Overy, 2012; Trost et al., 2017). À semelhança de expressões faciais ou vocais, as emoções musicais são rapidamente identificadas pelos ouvintes (Thompson, 2009), e estudos apontam para a existência de universalidade no reconhecimento de emoções básicas expressas através da música (Fritz et al., 2009).

Neste sentido, a música e a voz são dois domínios auditivos que se cruzam na capacidade de expressar emoções, o que tem levado a discussões científicas sobre a possibilidade de ser nessa componente que jaz a possível origem evolutiva comum entre a música e a linguagem (e.g., Bhatara et al., 2014). Ao contrário dos aspetos semânticos ou sintáticos presentes na linguagem e na música, que são distintamente humanos, os aspetos acústicos relacionados com a comunicação emocional a partir da voz estão presentes noutras espécies (e.g., Fitch, 2006). Evidência empírica indica também que a música e a fala usam um código acústico comum para a expressão de emoções. Juslin e Laukka (2003) realizaram uma meta-análise na qual compararam os perfis acústicos das diferentes emoções na música e na prosódia da fala, e descobriram que o conjunto de pistas acústicas utilizadas para expressar emoções básicas é bastante semelhante nos dois domínios. A alegria, por exemplo, tende a ser expressa e percecionada, quer na música quer na fala, quando a velocidade/tempo são mais rápidos, a intensidade é média-alta, e a frequência fundamental é média-alta e variável. Já a tristeza segue, de forma semelhante na música e na fala, um padrão acústico oposto: ou seja, é expressa e percecionada nos dois domínios por velocidade/tempo mais lentos, intensidade baixa, e frequência fundamental baixa. Os autores explicam que somos capazes de percecionar emoções na voz e em música com base em pistas acústicas semelhantes porque a música é bastante eficaz em mimetizar os atributos emocionais da fala, funcionando como uma ‘voz super-expressiva’, capaz de recrutar

redes cerebrais envolvidas no processamento de emoções vocais. Estudos mais recentes complementam esta proposta, sugerindo que características musicais de baixo-nível apresentadas de forma isolada (e.g., timbre de um instrumento musical), são tão eficazes quanto expressões vocais na comunicação de significado emocional (Liu et al., 2018). Estes resultados apoiam a hipótese heurística, segundo a qual as emoções musicais seriam capazes de ‘invadir’ as redes neurais dedicadas ao processamento de emoções vocais (e.g. Juslin & Västfjäll, 2008; Peretz, 2010). Segundo esta hipótese, o envolvimento continuado com a música poderia promover alterações nos circuitos cerebrais dedicados ao processamento de emoções vocais, via neuroplasticidade (Peretz, 2010).

Se de facto existe uma partilha de mecanismos acústicos e neurocognitivos entre as emoções musicais e vocais, é também esperado que exista uma associação entre sensibilidade à música e sensibilidade às emoções vocais. Pessoas com maior sensibilidade à música, devido a exposição, treino e/ou fatores genéticos, deverão ter também maior sensibilidade às emoções em fala. Uma forma de testar esta possibilidade é explorar diferenças entre músicos e pessoas sem treino musical formal, chamados “não-músicos”. O treino musical envolve a mestria de várias competências musicais, e os músicos dedicam grande parte do seu tempo à prática do instrumento. Para além disso, a proficiência musical tem uma forte componente expressiva, que passa pela capacidade do músico de comunicar emoções (Sloboda & Juslin, 2001), com investigação que aponta para associações positivas entre o número de anos de treino musical e o reconhecimento de emoções em excertos musicais (e.g., Livingstone et al., 2010). Existem ainda estudos empíricos que observaram associações positivas entre a expertise musical e o reconhecimento de emoções expressas na voz (e.g., Correia et al., 2020; Lima & Castro, 2011; Thompson et al., 2004).

Ao considerar o nível de competência musical da amostra de não-músicos, a partir da aplicação de medidas que permitam identificar indivíduos com capacidades de percepção musical “naturalmente” boas, tal como indivíduos que se envolvem ativamente com música (envolvimento informal com a música, i.e., consumo de música no dia-a-dia, ida a concertos, importância dada à música), é possível perceber o papel que outros tipo de “sensibilidade” musical, para além do treino musical formal, poderão ter nas associações entre música e competências emocionais reportadas na literatura (Martins et al., 2021). Torna-se relevante considerar esta questão tendo em conta investigação recente que demonstrou associações entre boas capacidades musicais e o reconhecimento de prosódia emocional mais apurado, mesmo na ausência de treino musical (Correia et al., 2020).

A possível relação entre competências musicais e o reconhecimento de emoções em voz tem sido um tópico de crescente interesse, tendo sido explorado quer em artigos empíricos (e.g., Correia

et al., 2020), quer em revisões da literatura (Martins et al., 2021; Nussbaum & Schweinberger, 2021). Nesse sentido, a presente dissertação pretende contribuir para este debate, explorando associações entre duas dimensões da sensibilidade à música — treino musical formal e variação natural em competências musicais — e o reconhecimento de emoções em prosódia, usando um paradigma de *gating* auditivo. Este paradigma tem sido usado em estudos focados no curso temporal do reconhecimento de emoções vocais, e tem como principal característica o facto de ser sensível à natureza dinâmica da informação auditiva, decompondo os estímulos em diferentes intervalos de tempo (*ms*) ou *gates* (e.g., Castiago & Pinheiro, 2019; Pell & Kotz, 2011). Os gates são depois apresentados aos participantes de forma sucessiva, começando com gates de durações mais curtas (em que os participantes têm acesso apenas à informação acústica inicial do estímulo), e progredindo para gates de duração mais longa até ser apresentado o estímulo completo (Grosjean, 1985). Recorrendo a esta tarefa será assim possível perceber se os músicos e indivíduos com boas competências musicais apresentam vantagens no reconhecimento de emoções em estádios de processamento acústico mais precoces (i.e., quando são expostos a menos informação acústica), tornando-os desta forma mais “rápidos” a reconhecer com exatidão emoções na voz. Tanto quanto sabemos, esta possibilidade ainda não foi testada na literatura.

A presente dissertação segue a seguinte estrutura: no primeiro capítulo será apresentado o enquadramento teórico, no qual será primeiramente discutida literatura que concerne à prosódia emocional, e o reconhecimento temporal da mesma. De seguida serão discutidas associações entre o treino musical formal, capacidades musicais naturais, e o reconhecimento de prosódia emocional. Finalmente, o tipo de metodologia geralmente utilizada para aceder ao reconhecimento explícito de prosódia emocional no contexto de associações entre música e emoções é debatido, e o *gating* auditivo é introduzido como um paradigma ideal para explorar possíveis diferenças no curso temporal de reconhecimento de emoções vocais entre músicos e não-músicos.

No capítulo que se segue (Capítulo II), o método utilizado na presente dissertação é apresentado. Neste capítulo será feita a descrição da amostra, dos instrumentos utilizados, e do procedimento. Finalmente, será apresentado o plano de análises estatísticas para os dados obtidos.

No terceiro capítulo, são apresentados os resultados, tendo em conta as hipóteses teóricas delineadas no Capítulo I. Recorrendo a modelos estatísticos de efeitos mistos, são exploradas associações entre a exatidão no reconhecimento de emoções e competências musicais, tal como possíveis diferenças a nível da informação acústica necessária para reconhecer explicitamente prosódia emocional em função do treino musical e competência musical. Finalmente, no último capítulo (Capítulo IV) os resultados são discutidos criticamente, considerando a literatura existente,

sendo ainda apresentadas as principais limitações, contribuições da presente investigação, e sugestões para estudos futuros.

Capítulo I – Enquadramento Teórico

Prosódia Emocional

A capacidade de descodificar informação emocional na voz é uma ferramenta social valiosa (Thönnissen et al., 2010). As emoções vocais permitem a comunicação entre indivíduos, servindo um papel crítico na direção da atenção visual para eventos emocionais relevantes no ambiente, como faces (Brosch et al., 2009; Rigoulot & Pell, 2012). À semelhança do que ocorre com o reconhecimento de expressões emocionais na face (e.g., Ekman et al., 1969), resultados empíricos apontam para a universalidade no reconhecimento de emoções básicas expressas na voz (e.g., Chronaki et al., 2018; Pell & Skorup, 2008; Sauter et al., 2010), o que reitera a eficácia da fala na comunicação de estados emocionais.

A prosódia emocional, em particular, diz respeito às modificações suprasegmentais na voz quando falamos, e são produzidas durante um episódio emocional (Grandjean, 2021). Estas modificações concernem modulações na intensidade (mais intensa ou menos intensa), altura tonal (variação na frequência fundamental; F0), ritmo da voz (duração das sílabas e pausas), e qualidade da voz ou timbre (distribuição de energia espectral) entre expressões (e.g., Scherer et al., 2003; Wildgruber et al., 2006). Perante um tom de voz emocional, estes elementos acústicos são percecionados e integrados com o conteúdo verbal da voz (estrutura segmental), o que permite percepcionar teor emocional na fala durante as interações.

A capacidade de percecionar e distinguir explicitamente emoções vocais exige a capacidade de integrar vários parâmetros acústicos, numa janela temporal de milissegundos (Schirmer & Kotz, 2006), com evidência que mostra que somos capazes de extrair informação emocional de expressões vocais com níveis de desempenho quatro ou cinco vezes superiores ao nível do acaso (Scherer et al., 2003). Relativamente à percepção e processamento de emoções vocais, Schirmer e Kotz (2006) propõem um modelo que considera a compreensão de prosódia emocional um processo composto por múltiplas fases e subprocessos que são representados de forma diferenciada em regiões cerebrais relacionadas, globalmente, com a sensação, emoção e cognição. Segundo a proposta destas autoras, a percepção de prosódia emocional ocorre em três estádios: 1) extração de pistas acústicas relevantes para a percepção de emoções, 2) captação de saliência emocional a partir de um conjunto de pistas acústicas, e 3) avaliação cognitiva do significado emocional.

O primeiro estádio engloba o processamento sensorial do sinal acústico, sendo que este processo é mediado por áreas sensoriais, como o córtex auditivo primário, secundário e terciário, e áreas subcorticais, como o tálamo auditivo e a amígdala (Grandjean, 2021). A nível de resolução temporal, a análise de informação sensorial relacionada com frequência e intensidade parece ocorrer nos

primeiros 100 ms a seguir ao início do estímulo (Kotz & Paulmann, 2011; Paulmann & Kotz, 2008; Pinheiro et al., 2013). Depois deste primeiro processamento mais global, as expressões vocais emocionais recrutam áreas relacionadas com vias ventrais, que codificam o significado emocional das vocalizações (Estádio 2). Durante este estádio, diferentes pistas acústicas relevantes para percecionar prosódia emocional são integradas à medida que o processamento progride de áreas sensoriais posteriores para regiões temporais mais anteriores no giro temporal superior e sulco temporal anterior (Grandjean, 2021), com uma latência de 200 ms (Schirmer & Kotz, 2006). Após a extração de significado emocional nestas regiões cerebrais, o sinal acústico fica disponível para elaboração cognitiva (Estádio 3), como por exemplo, julgamentos emocionais explícitos. Este processo é mediado pelo córtex inferior direito e córtex orbito frontal (Grandjean, 2021).

Existe variabilidade na nossa capacidade de percecionar prosódia emocional, sendo que o processamento de emoções vocais pode ser modulado por fatores como o foco atencional e variáveis interindividuais (e.g., sexo; ver Schirmer & Kotz, 2006). Considerando a natureza dinâmica do processamento auditivo (Kraus & Chandrasekaran, 2010), a presente dissertação explora se diferenças individuais nas competências musicais e expertise musical poderiam estar associadas a diferenças no processamento de prosódia emocional.

Associações entre Música e Prosódia Emocional

Treino Musical

Aprender a tocar um instrumento musical é uma tarefa complexa que envolve a interação de diversas modalidades cognitivas, resultando em mudanças comportamentais, estruturais e funcionais no sistema auditivo (Herholz & Zatorre, 2012; Kraus & Chandrasekaran, 2010). No que toca a diferenças funcionais relacionadas com a expertise musical, participantes com treino musical apresentam maior atividade neuronal quando ouvem o instrumento que tocam, comparativamente a não-músicos, sendo que a magnitude da ativação está positivamente correlacionada com a duração do treino musical (Pantev et al., 1998). Os músicos apresentam também diferenças na estrutura cerebral relativamente a não-músicos, nomeadamente, diferenças no volume de matéria cinzenta em áreas motoras, auditivas, e visuo-espaciais relevantes para tocar instrumentos musicais (e.g., Gaser & Schlaug, 2003).

O treino musical também está positivamente associado às capacidades de percepção de fala, com estudos que reportam associações positivas entre o treino musical e o desempenho em tarefas de percepção de fala em condições de escuta adversas (e.g. Coffey et al., 2017; Strait et al., 2012; mas ver Boebinger et al., 2015). Outros estudos apontam ainda para diferenças entre músicos e não-músicos na percepção de fonemas em estádios precoces de processamento (e.g., Kühnis et al., 2013).

A proficiência musical também está associada a respostas cerebrais mais eficientes a violações de altura tonal na fala, o que sugere vantagens na percepção de prosódia (Moreno et al., 2009). Propostas para os possíveis mecanismos subjacentes aos efeitos de transferência do treino musical para a percepção de fala apontam para a existência de uma sobreposição entre as redes neurais que processam os aspectos acústicos da música e da fala. A hipótese OPERA, proposta por Patel (2011, 2014), sugere que o treino musical coloca exigências acrescidas nestes circuitos partilhados, o que por sua vez, promove estas redes a processarem informação percetiva relevante para o processamento da fala de forma mais apurada, o que se traduziria numa vantagem para os músicos.

Considerando os benefícios reportados a nível da fala, seria esperado que o envolvimento continuado com a música estivesse também associado a uma melhor capacidade de percecionar prosódia emocional na voz. A investigação que explora associações entre o treino musical e o reconhecimento de prosódia emocional surge maioritariamente de designs transversais, nos quais grupos de músicos e não-músicos são comparados relativamente ao desempenho em tarefas de reconhecimento de emoções. Thompson et al. (2004) observaram vantagens de pessoas com treino musical na identificação de emoções em sequências de tons que mimetizam prosódia emocional (experiência 1). No entanto, é de notar que nesta primeira experiência, os participantes sem treino musical responderam ao acaso em todas as emoções, o que poderá ser indicativo de uma compreensão pobre das instruções, ou da ineficácia dos estímulos em expressar a emoção pretendida. Numa segunda experiência, músicos e não-músicos ouviram expressões emocionais na sua língua materna (inglês) e numa língua estrangeira (tagalog). Os músicos apresentaram vantagens específicas à emoção, nomeadamente no reconhecimento da tristeza e medo, e não tendo sido reportado um efeito principal significativo do treino musical. A terceira experiência revelou que crianças que tiveram aulas de música durante um ano, apresentaram resultados superiores numa tarefa de discriminação de emoções, comparativamente a crianças sem aulas de música. No entanto, devido à pouca variedade nos estímulos (todas as frases faladas por uma mulher), a validade ecológica dos resultados é baixa. Já Lima e Castro (2011) testaram músicos e não-músicos numa tarefa de reconhecimento de prosódia emocional, apresentando frases com conteúdo semântico neutro, expressas em seis emoções (raiva, nojo, medo, alegria, tristeza, e surpresa). Os participantes músicos foram ainda divididos em dois grupos, mediante a faixa etária (jovens e meia-idade), de forma a compreender possíveis efeitos do treino musical a longo prazo. Foi encontrada uma associação positiva entre o treino musical e maior exatidão no reconhecimento de todas as emoções, sendo que esta vantagem se manteve independentemente da idade.

Na tentativa de compreender se a vantagem do treino musical no reconhecimento de emoções se estende para o reconhecimento de vocalizações não verbais, como gargalhadas e choro (para além

da prosódia emocional), Correia et al. (2020) testaram músicos e não-músicos no reconhecimento de emoções, comparando o desempenho dos grupos em três tarefas de reconhecimento de emoções: reconhecimento de prosódia emocional, reconhecimento de vocalizações não verbais, e reconhecimento de emoções em expressões faciais. Os músicos tiveram um melhor desempenho que os não-músicos nas duas tarefas de reconhecimento de emoções vocais, no entanto esta vantagem não emergiu para o reconhecimento de expressões faciais. Os autores sugerem que uma possível associação entre o treino musical e o reconhecimento de emoções poderá ser exclusiva à modalidade auditiva. Para além disso, também verificaram que a duração do treino musical esteve positivamente correlacionada com o reconhecimento de emoções, quer em prosódia emocional, quer em vocalizações não verbais.

Na mesma linha, Farmer et al. (2020) exploraram se os benefícios do treino musical se generalizam para diferentes canais de comunicação emocional. Para esse fim, realizaram um estudo no qual participantes com e sem treino musical foram expostos a clips de vídeo nos quais dois agentes interagiam em três condições: visual, auditiva e audiovisual. De seguida deveriam julgar, o mais rapidamente possível, qual a emoção expressa durante a interação (alegria ou raiva). Deveriam indicar também em que medida eles próprios experienciaram a emoção expressa pelos agentes. À semelhança dos resultados reportados em Correia et al. (2020), os músicos tiveram um melhor desempenho que os não-músicos, mas essa vantagem só se verificou na condição de áudio, não tendo sido identificadas diferenças na condição visual ou audiovisual. Para além disso, também não foram reportadas diferenças na emoção sentida, apenas na reconhecida. Apesar de ser necessária mais investigação para determinar associações entre o treino musical, o processamento na modalidade visual e a integração multimodal, estes resultados sugerem uma relação seletiva com emoções expressas na modalidade auditiva (para uma revisão, ver Martins et al., 2021).

A investigação até agora apresentada aponta para um quadro geralmente positivo, mas nem toda a investigação mostra associações positivas entre o treino musical e o reconhecimento de emoções em voz. Trimmer e Cuddy (2008), por exemplo, testaram músicos e não-músicos numa tarefa de categorização emocional, na qual os participantes deveriam avaliar frases e expressões melódicas análogas às mesmas. Foram apresentadas quatro emoções (alegria, tristeza, medo, raiva) e uma condição neutra. Neste estudo não foram encontradas associações entre o treino musical e a exatidão do reconhecimento de prosódia emocional. Os resultados revelaram que foi a inteligência emocional, e não o treino musical, que esteve associada a um melhor desempenho na tarefa. Também não foram reportadas associações entre o treino musical e a inteligência emocional. Os autores argumentam que a capacidade de reconhecer emoções na voz não se deve a uma maior sensibilidade acústica associada ao treino musical, mas sim a um sistema de processamento

emocional de alto nível, supramodal, i.e., independente dos domínios musical e vocal. No entanto, é importante referir que o número exato de participantes com treino musical não foi especificado. Para além disso, os participantes tinham em média 6.5 anos de treino formal, o que constitui um valor relativamente baixo, considerando que em alguns estudos anteriores que reportaram associações positivas entre o treino musical e o reconhecimento de prosódia emocional, os músicos tinham, pelo menos, oito anos de treino formal (Lima & Castro, 2011). Considerando que os efeitos do treino musical podem ser dependentes da experiência, e que a percepção de emoções na voz é um domínio biológico básico, é plausível que sejam necessários períodos de treino mais extenso para se observarem diferenças entre grupos, especialmente no contexto de tarefas de reconhecimento comportamentais (Lima & Castro, 2011). Já Park et al. (2015) exploraram diferenças entre músicos e não-músicos apresentando expressões de alegria, tristeza, e medo, enquanto captavam a atividade cerebral através de ressonância magnética funcional (fMRI). Os músicos processaram a tristeza de forma diferenciada dos participantes não-músicos, apresentando uma maior ativação no giro frontal medial, no córtex pré-frontal anterior medial, córtex cingulado posterior, e córtex retrosplenial. Os autores sugerem que estas diferenças se devem a processamento neural mais apurado para prosódia triste, tendo em conta que as ativações reportadas englobam áreas relacionadas não só com o processamento auditivo e reconhecimento de emoções, mas também áreas gerais de processamento sócio emocional. No entanto, as respostas neurais diferenciadas não se refletiram num melhor reconhecimento explícito de prosódia emocional.

Capacidades Musicais “Naturais” e Reconhecimento de Prosódia Emocional

A literatura acima revista parece apontar para uma associação positiva entre o treino musical e o reconhecimento de prosódia emocional (e.g., Correia et al., 2020; Lima & Castro, 2011; Thompson et al., 2004), apesar deste quadro não ser geral para toda a investigação debruçada sobre este tópico (e.g., Park et al., 2015; Trimmer & Cuddy, 2008). No entanto, estes resultados devem ser interpretados com cuidado. Um aspeto que tem sido debatido em literatura recente (e.g., Mankel & Bidelman, 2018; Martins et al., 2021; Nussbaum & Schweinberger, 2021) prende-se com o papel que diferenças individuais (nomeadamente predisposições) poderão ter nas associações entre o expertise musical e o desempenho em tarefas de percepção auditiva, particularmente em comparações transversais entre músicos e não-músicos. Considerando que a propensão para o treino musical tem uma componente genética (e.g., Mosing et al., 2014; Mosing & Ullén, 2016), e que fatores ambientais, como capacidades cognitivas, personalidade, e variáveis socioeconómicas motivam certos indivíduos a procurarem aulas de música, não fica claro se os grupos de músicos e não-músicos diferem apenas na experiência musical, tal como é muitas vezes presumido nos estudos

transversais (e.g., Schellenberg, 2020). A contribuição de investigação longitudinal para informar esta questão ainda é fraca, visto que a literatura é escassa e os resultados são controversos devido às limitações metodológicas e tamanho das amostras (e.g. Martins et al., 2021; McKay, 2021). Dois estudos longitudinais reportam melhorias no reconhecimento de prosódia emocional decorrentes do treino musical em populações típicas (Thompson et al., 2004; Experiência 3) e atípicas (e.g., Good et al., 2017), e dois estudos reportam efeitos nulos (Chari et al., 2019; Fuller et al., 2018).

Uma forma de explorar a contribuição de diferenças individuais em desenhos transversais é investigar se os benefícios que se observam em músicos se verificam também em indivíduos que, não tendo treino musical formal, têm ainda assim boas capacidades musicais naturais. Indivíduos que são considerados não-músicos podem ter competências de percepção musical naturalmente elevadas, apesar de não terem tido treino musical formal por variadas razões (*musical sleepers*; Law & Zentner, 2012). No que toca a capacidades auditivas, estudos indicam que, mesmo na ausência de treino musical, competências de percepção musical “naturalmente” boas estão associadas a um processamento de fala mais apurado. Especificamente, Swaminathan e Schellenberg (2017) verificaram que indivíduos com boas capacidades musicais percepcionam fonemas numa língua estrangeira de forma mais apurada, independentemente do treino musical formal. Estes resultados foram mais tarde corroborados pelos mesmos autores, que observaram associações positivas entre capacidades musicais naturais e o desempenho em testes de percepção de fala e gramática em crianças entre os seis e nove anos, mesmo quando o treino musical era controlado (Swaminathan & Schellenberg, 2019). Mankel e Bidelman (2018) reportaram que participantes sem treino musical, mas com boas competências musicais apresentaram uma codificação neural mais eficiente de fala degradada por ruído.

No entanto, os estudos acima referidos focam-se na percepção de fala, continuando a ser necessária mais investigação para perceber associações entre musicalidade natural e a capacidade de reconhecer emoções na fala. A literatura debruçada sobre associações entre sensibilidade à música e reconhecimento de emoções vocais apresenta um foco primário no treino musical formal, o que reflete uma visão incompleta daquilo que é a riqueza e variedade da musicalidade em sentido mais lato (Correia et al., 2020). Visões atuais de musicalidade consideram a experiência musical como um fenómeno multifacetado, composta por diferentes componentes que incluem, para além do treino musical, aspetos como a importância que é atribuída à música no dia-a-dia, a forma como o indivíduo responde emocionalmente à música, e o envolvimento informal com a mesma (Müllensiefen et al., 2014), além do contributo de fatores genéticos (e.g., Tan et al., 2014). Müllensiefen et al. (2014) operacionalizam um conjunto de facetas musicais num constructo psicométrico definido como *sofisticação musical*, que engloba competências musicais, proficiência

musical, conquistas musicais, e diversos comportamentos musicais que podem ser medidos em diferentes subescalas, e que podem assumir valores elevados mesmo na ausência de aulas formais de música. No contexto de investigação que explora associações entre o treino musical e o reconhecimento de emoções, raramente são implementadas medidas subjetivas e/ou objetivas que permitam identificar indivíduos com boas capacidades musicais, nomeadamente entre os participantes sem treino musical. Na prática, tipicamente, os participantes sem treino musical são apenas categorizados como não-músicos e tratados como sendo um grupo homogéneo, sendo a variação nas suas competências musicais negligenciada. Até à data, apenas um estudo considerou esta questão. Correia et al. (2020) aplicaram testes de desempenho musical e questionários de autorrelato (*Goldsmith Musical Sophistication Index*; Müllensiefen et al., 2014), e observaram que, mesmo na ausência de treino musical, participantes não-músicos que possuem boas capacidades musicais tiveram um desempenho tão bom quanto os músicos no reconhecimento de prosódia emocional e vocalizações não verbais. Estes dados são consistentes com investigação que demonstrou vantagens a nível de capacidades de percepção de fala em pessoas com elevadas competências musicais (e.g., Mankel & Bidelman, 2018), estendendo estes resultados à percepção de emoções em pistas não verbais na voz. Estes resultados salientam, mais uma vez, a importância de considerar a musicalidade como uma experiência multifacetada, que engloba contribuições não só da experiência formal (i.e., treino musical formal), mas também de fatores genéticos e ambientais, e que estes poderão contribuir para associações entre música e domínios não musicais.

Desse modo, o estudo apresentado nesta dissertação contempla não só diferenças entre músicos e não-músicos no reconhecimento de prosódia emocional, mas avalia também o nível de competência musical dos não-músicos. Pretende-se não só explorar aspectos associados ao treino formal, mas também contribuir para uma melhor compreensão do possível papel de variação natural nas competências musicais.

Tarefas de Reconhecimento de Prosódia Emocional

A literatura debruçada sobre associações entre o reconhecimento de prosódia emocional e o treino musical tipicamente apresenta as expressões emocionais na sua totalidade aos participantes, no contexto de tarefas de escolha forçada (e.g., Correia et al., 2020; Lima & Castro, 2011; Park et al., 2015; Pinheiro et al., 2015; Thompson et al., 2004; Trimmer & Cuddy, 2008). Dando um exemplo prático deste tipo de tarefa, o participante ouve uma frase completa expressa numa determinada entoação emocional (e.g., raiva) e seleciona a categoria emocional correspondente, a partir de uma lista de opções possíveis (e.g., raiva, tristeza, alegria, medo, ou neutro). Apesar desta metodologia ser informativa relativamente à exatidão geral do reconhecimento de emoções vocais, não permite

compreender *quando* é que o processo de reconhecimento efetivamente ocorre, uma vez que apenas é obtida uma medida ‘offline’ depois da exposição a toda a informação acústica. Por outras palavras, não permite determinar a quantidade de informação acústica necessária para que o participante já seja capaz de identificar a emoção veiculada. O aspetto temporal da voz é relevante, pois ao contrário do que ocorre com expressões faciais, em que toda a informação pode estar disponível em simultâneo, a estrutura prosódica da voz é intrinsecamente dinâmica (Grandjean, 2021; Schaeerlaeken & Grandjean, 2018). Algumas pistas acústicas centrais para descodificar prosódia emocional, como o nível de intensidade ou energia da voz, podem ser rapidamente percecionadas em fases iniciais do processamento, enquanto outras, como a variabilidade da frequência fundamental (F0) e a velocidade da voz podem necessitar de informação acústica que vai sendo processada ao longo do tempo (Nordström, 2019).

A dimensão temporal do processamento de emoções vocais tem sido primariamente abordada em estudos de eletroencefalografia (EEG), devido à elevada resolução temporal desta técnica. Vários estudos de EEG indicam que o componente P200 é modulado por frases prosódicas com teor emocional, comparativamente a expressões neutras (e.g., Paulmann & Kotz, 2008; Paulmann et al., 2013). Paulmann et al. (2013) sugerem que os processos envolvidos na extração de saliência emocional (ou seja, a capacidade de perceber se o estímulo é expresso com entoação emocional ou não), poderão ocorrer numa janela temporal de 200 ms após a exposição a emoções vocais. Estudos recentes sugerem que poderão existir modulações relacionadas com emoções antes dos 200 ms, nomeadamente, a nível dos 50 ms (modulações no componente P50; Liu et al., 2012). A capacidade de realizar análises mais complexas do sinal acústico (e.g., capacidade de diferenciar prosódias emocionais ou fazer elaborações cognitivas acerca da relevância social do estímulo) ocorrem em estádios mais tardios de processamento (Schirmer & Kotz, 2006). Estes processos cognitivos estão associados a modulações nos componentes N300 (Bostanov & Kotchoubey, 2004) e N400 (e.g., Liu et al., 2015). Embora o EEG permita identificar com elevada exatidão o momento em que a resposta cerebral ocorre, esta técnica é menos útil se o foco for no momento em que o participante tem acesso consciente/explícito a uma determinada representação ou resposta. Neste contexto é necessário um indicador comportamental.

Paradigma de Gating

No contexto de tarefas comportamentais, um paradigma sensível às variações temporais da fala é o paradigma de gating (Grosjean, 1985), uma técnica originalmente aplicada no contexto de processos de reconhecimento lexical e identificação de fonemas (e.g., Grosjean, 1985; Salasoo & Pisoni, 1985). Contudo, este paradigma tem sido cada vez mais utilizado no contexto do reconhecimento

temporal de prosódia emocional (e.g., Jiang et al., 2015; Nordström & Laukka, 2019; Pell & Kotz, 2011) e vocalizações emocionais não verbais (e.g., Castiajo & Pinheiro, 2019). Em estudos que utilizam este paradigma, os estímulos auditivos são divididos em segmentos (gates) de diferentes durações, que são construídos em função de incrementos de tempo específicos (e.g., ms), ou unidades linguísticas (e.g., sílabas). Os gates são depois apresentados de forma sucessiva, começando com os gates mais curtos (com menos informação acústica), e terminando com a apresentação do estímulo completo. Após ouvir cada estímulo, o participante tipicamente completa uma tarefa de categorização emocional, onde deve indicar a emoção que pensa estar presente no excerto que ouviu, e pode também indicar o nível de confiança que tem na sua resposta. Estas tarefas permitem avaliar o nível de exatidão do reconhecimento em cada gate, assim como determinar o ponto de identificação da emoção (*emotion identification point - EIP*; Pell & Kotz, 2011). O EIP da emoção corresponde ao gate específico a partir do qual o participante não altera mais a sua resposta, permitindo assim estimar a quantidade de informação percetiva mínima necessária para reconhecer explicitamente a emoção veiculada pelo estímulo (Grosjean, 1985; Pell & Kotz, 2011).

O paradigma de gating permite assim captar informação qualitativa acerca da exatidão do reconhecimento e do nível de confiança que os ouvintes têm em relação à sua resposta, assim como informação quantitativa, pois permite quantificar a informação acústica necessária para reconhecer diferentes emoções vocais (e.g., Pell & Kotz, 2011; Schaeerlaeken & Grandjean, 2018). Um dos estudos pioneiros que utilizaram o paradigma de gating para explorar o curso temporal do reconhecimento de emoções vocais comparou o curso de reconhecimento de emoções básicas (raiva, tristeza e alegria), com o reconhecimento de “atitudes” (ironia, dúvida; Grichkovstova et al., 2007). Os estímulos consistiram em duas expressões vocais gravadas por dois falantes (ator e atriz), e expressas em cada uma das emoções e atitudes exploradas no estudo. Estes autores reportaram EIPs mais curtos para emoções negativas (raiva, tristeza; entre 200–1000 ms), comparativamente a alegria (entre 800–1000 ms) e atitudes (dúvida: entre 600–1400 ms; ironia: entre 1600–1800 ms). No entanto, é de notar que foram exploradas poucas emoções (duas negativas, e uma positiva), e que o conteúdo semântico das frases utilizadas não era neutro, o que poderá ter influenciado a rapidez do reconhecimento. Estudos posteriores observaram um padrão de reconhecimento temporal diferente. Cornew et al. (2009) apresentaram pseudo-frases expressas em prosódia alegre, zangada ou neutra. As pseudo-frases diferem de frases com conteúdo semântico pois mimetizam as propriedades fonéticas e morfossintáticas da língua alvo, mas são compostas por palavras não existentes, de forma a eliminar possíveis pistas léxico-semânticas potencialmente relevantes para a descodificação da emoção (Pell et al., 2009; Scherer et al., 1991). A utilização de pseudo-frases em estudos de gating é frequente, pois permite aos investigadores estudar com maior exatidão os efeitos

isolados de prosódia emocional na fala (e.g., Cornew et al., 2009; Jiang et al., 2015; Pell & Kotz, 2011; Rigoulot et al., 2013). Os estímulos foram segmentados em incrementos de 250 ms e identificados pelos participantes numa tarefa de escolha forçada, na qual os participantes deveriam selecionar uma das três categorias emocionais em estudo (experiência 1), e numa tarefa de discriminação (experiência 2), na qual na qual os respondentes deveriam indicar se o estímulo que ouviram era emocional (se o estímulo era de raiva ou alegria) ou neutro. Os resultados da experiência 1 revelaram EIPs mais curtos para expressões neutras (444 ms), comparativamente a expressões de raiva (723 ms) e alegria (802 ms), sugerindo que existe um viés para o reconhecimento de prosódia neutra. Os autores reportaram ainda que os participantes foram mais rápidos e precisos a discriminar prosódia neutra do que prosódia emocional (estudo 2). Este dado elimina a possibilidade dos resultados no estudo 1 serem explicados por um viés na seleção da categoria neutra, em durações mais curtas do estímulo, devido a incerteza relativamente à categoria correta. É de notar que este estudo explorou apenas duas emoções, para além da condição neutra, e que a raiva continuou a ser reconhecida mais rapidamente que prosódia alegre, o que é consistente com a ideia de que estímulos negativos são processados mais rapidamente do que estímulos positivos (Paulmann & Kotz, 2018). Mais recentemente, Pell e Kotz (2011) expandiram estes resultados, incluindo cinco emoções (alegria, raiva, nojo, medo, e tristeza) e uma condição neutra. Apresentaram aos participantes pseudo-frases, divididas em sete gates correspondentes ao número de sílabas presentes no estímulo. Os resultados apontaram para EIPs mais curtos para expressões neutras (510 ms), expressões de medo (517 ms), tristeza (576 ms), e raiva (710 ms), comparativamente a expressões de alegria (977 ms) e nojo (1486 ms). Apesar de expressões neutras serem reconhecidas mais rapidamente, esta diferença não foi significativa quando comparada com o ponto de identificação para o medo ou tristeza. Logo, ao introduzir mais categorias emocionais o viés para o reconhecimento de expressões neutras foi mais fraco. Já Rigoulot et al. (2013) exploraram a hipótese de que o reconhecimento temporal mais tardio de emoções como alegria ou nojo se deve à posição das pistas acústicas relevantes para decodificar estas emoções específicas na frase. Para esse fim, desenharam um estudo no qual os participantes avaliaram o significado emocional de pseudo-frases que expressavam raiva, nojo, medo, alegria, tristeza e neutralidade, num paradigma de gating, na qual a ordem das frases foi invertida. Mais concretamente, as expressões foram segmentadas em função do número de sílabas e apresentadas ao participante a partir do offset (fim) da frase, em vez do início. Nesta modalidade de apresentação, o tempo necessário para reconhecer medo, raiva, tristeza, e neutralidade em prosódia não foi influenciado pela inversão da ordem de apresentação da frase. No entanto, a alegria e nojo foram reconhecidas significativamente mais rápido quando os participantes ouviam o fim da expressão primeiro. Estes resultados apoiam a hipótese de que a

posição das pistas emocionalmente relevantes na expressão vocal pode estar relacionada com a rapidez do reconhecimento (Paulmann & Kotz, 2018). Já em Jiang et al. (2015), os autores tentaram perceber se o reconhecimento da prosódia emocional era modulado pela cultura. Para esse fim, compararam as janelas temporais de reconhecimento de expressões de alegria, medo, raiva, tristeza, e neutralidade entre falantes de diferentes línguas (inglês e hindi). Os participantes ouviam pseudo-frases segmentadas em cinco intervalos de tempo (200 ms, 400 ms, 500 ms, 600 ms, 700 ms), seguidos da expressão completa, e os falantes nativos de inglês e hindi deveriam identificar de seguida as emoções expressas pelos falantes do seu próprio grupo, ou do grupo de língua estrangeira. O reconhecimento de emoções foi mais rápido quando os participantes ouviam expressões nativas, comparativamente a expressões ditas por falantes do outro grupo, o que aponta para um viés cultural no reconhecimento de emoções vocais. Em geral (i.e., independentemente do grupo), os EIPs foram mais curtos para expressões neutras (entre 450-542 ms) e raiva (entre 447-621 ms), seguidos de medo (entre 591 e 738 ms), tristeza (entre 622 e 755 ms), e finalmente, alegria (654-1634 ms).

Dois estudos recentes debruçaram-se ainda sobre o reconhecimento temporal de vocalizações puramente não verbais, como gargalhadas e choro. Schaeblaeken e Grandjean (2018) apresentaram vocalizações emocionais de nojo, raiva, medo, alegria, tristeza, e neutralidade, segmentadas em incrementos de 50 ms. Os resultados mostraram que emoções negativas (nojo, raiva e medo) foram reconhecidas com maior exatidão e rapidez comparativamente a expressões de alegria, tristeza, ou expressões neutras. Já Castiajo e Pinheiro (2019) apresentaram vocalizações não verbais de raiva, nojo, medo, alegria, tristeza e neutralidade. As expressões foram decompostas em sete gates, segmentados em incrementos de 100 ms. Os resultados indicaram que as expressões neutras foram reconhecidas com menos informação acústica comparativamente a outros tipos de vocalizações, seguidas de expressões de alegria (266 ms), nojo (296 ms), tristeza (305 ms), raiva (342 ms), e medo (367 ms).

No contexto de emoções vocais e musicais, Nordström e Laukka (2019) aplicaram o paradigma de gating para estudar o curso temporal do reconhecimento de emoções em prosódia emocional (Estudo 1) e em performances musicais (Estudo 2). Expressões de raiva, medo, alegria, interesse, prazer, tristeza, ternura, alívio, e neutralidade foram segmentadas em 11 gates com incrementos de 50 ms e 250 ms, de forma a obter uma resolução temporal mais fina do sinal acústico. Os autores observaram que o reconhecimento de emoções vocais e musicais segue um curso semelhante, com expressões de raiva (EIP para expressão vocal - 138 ms; EIP para expressão musical - 1481 ms), alegria (EIP para expressão vocal - 562 ms; EIP para expressão musical - 1281 ms), tristeza (EIP para expressão vocal - 575 ms; EIP para expressão musical - 1656 ms), e medo (EIP para expressão vocal - 1375 ms– EIP para expressão musical - 1500 ms) com EIPs menores em ambas as

modalidades, comparativamente a outras emoções, como interesse (EIP para expressão vocal - 1875 ms – EIP para expressão musical - 2250 ms) ou serenidade (EIP para expressão vocal - 2000 ms; EIP para expressão musical - 2875 ms). Curiosamente, tanto na música como na fala, expressões de raiva, alegria, tristeza, e neutralidade foram reconhecidas acima do nível do acaso mesmo a partir dos gates mais curtos (menos de 100 ms). Os autores sugerem que características acústicas de baixo nível (e.g., frequência e intensidade do som) poderão já ser perceptíveis para os participantes em a partir dos 50 ms, permitindo o reconhecimento explícito destas emoções em ambas as modalidades. Estes resultados complementam também a ideia de que a comunicação de emoções na fala e na música é baseada em padrões acústicos semelhantes (Juslin & Laukka, 2003; Scherer, 1995), apoiando a hipótese de um código partilhado entre os domínios.

Compilando os resultados da investigação acima apresentada é possível concluir que: (a) existe um curso de reconhecimento temporal distinto para as diferentes emoções; (b) algumas emoções negativas tendem a ser mais rapidamente identificadas do que emoções positivas (Grichkovstova et al., 2007; Jiang et al., 2015; Pell & Kotz, 2011; Rigoulot et al., 2013); (c) a comunicação de emoções vocais e musicais segue um padrão de reconhecimento temporal semelhante, que parece ocorrer com base na análise de parâmetros acústicos de baixo-nível (Nordström & Laukka, 2019).

O paradigma de gating permite então obter informação detalhada acerca do reconhecimento temporal de emoções vocais, constituindo também uma tarefa comportamental mais complexa, visto que os participantes são expostos a informação acústica de forma cumulativa. No contexto de diferenças entre músicos e não-músicos, tarefas de reconhecimento de emoções baseadas na apresentação do estímulo na sua totalidade poderão ser menos sensíveis a diferenças entre grupos, o que poderia explicar alguns dos efeitos nulos reportados na literatura (e.g., Trimmer & Cuddy, 2008). Para além do efeito do treino musical ser pequeno (Correia et al., 2020), o reconhecimento de emoções é um processo biológico básico, que tanto músicos quanto não-músicos dominam com facilidade. Neste sentido, o paradigma de gating constitui uma tarefa potencialmente mais sensível a possíveis nuances no reconhecimento de prosódia emocional, podendo ser uma tarefa capaz de isolar os efeitos do treino musical de forma mais clara.

Até à data, não foram encontrados estudos que tenham usado o paradigma de gating auditivo para perceber diferenças na quantidade de informação acústica necessária para reconhecer prosódia emocional entre músicos e não-músicos, nem em indivíduos não-músicos com boas capacidades musicais. No entanto, possíveis associações entre o treino musical e o reconhecimento temporal de emoções vocais já foi estudado a partir de medidas indiretas, nomeadamente, a partir de EEG. No contexto de vocalizações emocionais, Strait et al. (2009) observaram que, comparativamente a não-músicos, indivíduos com treino musical apresentaram uma resposta subcortical mais apurada a

choro de bebés, tal como uma melhor representação de frequências implicadas na percepção de timbre e altura tonal. Já Pinheiro et al. (2015) debruçaram-se sobre o processamento de prosódia emocional, reportando que os músicos apresentam respostas neurais diferenciadas de não-músicos a prosódia emocional em dois componentes (P50 e N100) que indexam o processamento sensorial precoce do estímulo vocal. Os resultados destes estudos apontam para uma possível vantagem dos músicos a nível da extração de propriedades acústicas básicas do sinal acústico. É possível que as vias neurocognitivas para o reconhecimento de emoções se cruzem a nível do processamento precoce, o que poderia proporcionar aos músicos vantagens no reconhecimento de prosódia emocional (Martins et al., 2021).

No entanto, estes resultados refletem diferenças no processamento implícito, o que poderá não corresponder ao tempo necessário para reconhecer emoções explicitamente (Pöppel & Bao, 2011). O paradigma de gating será útil para aprofundar conhecimento nesta questão, visto que é uma tarefa sensível ao processamento online da voz (Tyler & Wessels, 1985), e a resolução temporal mais fina permite aceder a diferentes estádios de processamento (Pell & Kotz, 2011). Logo, a tarefa de gating permitirá compreender em que estádio de processamento é que a vantagem dos músicos poderá emergir de forma explícita. A evidência reportada em estudos eletrofisiológicos (Pinheiro et al., 2015; Strait et al., 2009) e comportamentais (Nordström & Laukka, 2019), leva à predição de que os músicos sejam capazes de extrair mais informação sensorial a partir de excertos mais curtos, comparativamente a participantes não-músicos, o que refletiria o facto de necessitarem de menos informação acústica para identificar corretamente diferentes emoções vocais.

Objetivos e Hipóteses

A presente dissertação explora então possíveis diferenças entre músicos e não-músicos no reconhecimento temporal de emoções vocais, focando-se no reconhecimento de prosódia emocional. Adicionalmente, pretende-se perceber se as possíveis vantagens no processamento de emoções vocais associadas ao treino musical também se observam em pessoas que — não tendo treino musical formal — têm ainda assim competências musicais naturais elevadas. Nesse sentido, o presente estudo pretende responder às seguintes questões de investigação: Será que existem diferenças entre músicos e não-músicos na quantidade de informação acústica necessária para reconhecer emoções vocais? Qual é o potencial papel que diferenças individuais “naturais” nas capacidades musicais têm para o reconhecimento de emoções vocais?

Para responder a estas questões, foi desenhado um estudo transversal, no qual o desempenho de músicos e não-músicos foi comparado numa tarefa de reconhecimento de prosódia emocional. Considerando que o aspeto dinâmico da prosódia emocional é a variável chave para esta

investigação, o paradigma de gating auditivo é ideal, uma vez que permite controlar com maior exatidão a quantidade de informação percetiva a que os participantes têm acesso. Isto permite também compreender se as associações entre o treino musical e o reconhecimento de emoções vocais se deve à extração de pistas acústicas relevantes em estádios mais precoces do processamento, que estão associados a análises básicas de pistas auditivas e à alocação de significado emocional (e.g., Pinheiro et al., 2015; Schirmer & Kotz, 2006).

Os participantes avaliaram expressões prosódicas de alegria, raiva, tristeza, medo, e neutralidade segmentadas em 11 gates. Para avaliar capacidades de percepção musical, comportamentos e experiências musicais da amostra foi aplicada a versão portuguesa do questionário Gold-MSI (Lima et al., 2020), tal como uma adaptação online validada do *Musical Ear Test* (MET; Correia et al., 2021; Wallentin et al., 2010), que constitui uma medida objetiva de capacidades musicais. Considerando evidência indicativa de associações entre capacidades cognitivas gerais, treino musical, e capacidades de percepção musical (Swaminathan & Schellenberg, 2018), e literatura anterior que reportou associações entre capacidades cognitivas e o reconhecimento de prosódia emocional (Trimmer & Cuddy, 2008), foram aplicadas matrizes de raciocínio não verbal (MarRs-IB; Chierchia et al., 2019), de forma a compreender se possíveis associações entre o treino musical e reconhecimento de emoções são explicadas por diferenças em fatores gerais.

Com base em resultados anteriores (e.g., Correia et al., 2020; Lima & Castro, 2011; Thompson et al., 2004), é esperado que existam associações positivas entre o treino musical e o reconhecimento de prosódia emocional. Prevê-se ainda que: (1) os músicos necessitem de menos informação percetiva para reconhecer emoções vocais, comparativamente a não-músicos; e (2) que exista uma associação positiva entre competências musicais e desempenho na tarefa de reconhecimento emocional, mesmo quando o treino musical é controlado. Associado a esta segunda predição, espera-se ainda que a vantagem dos músicos no reconhecimento de emoções vocais seja maior quando comparados com não-músicos de baixa competência musical vs. não-músicos de elevada competência musical. Por outras palavras, espera-se que os não-músicos com competências musicais elevadas tenham um desempenho no reconhecimento de emoções vocais próximo ou mesmo semelhante ao dos músicos.

Capítulo II - Método

Participantes

Um total de 110 participantes foram recrutados a partir das bases de dados do Sistema de Participação em Investigação em Psicologia (SPI) do Iscte - Instituto Universitário de Lisboa e do Painel de Estudos Online (PEO) da Universidade Católica de Lisboa, assim como em resposta a anúncios divulgados nas redes sociais. Três participantes foram excluídos por não completarem todas as tarefas experimentais, o que resultou numa amostra final de 107 participantes (76 mulheres), sendo que 35 foram classificados como “músicos”, e 56 como “não-músicos”. Os restantes 16 participantes tinham níveis intermédios de treino musical, pelo que não foram categorizados, nem como músicos, nem como não-músicos, e foram apenas incluídos em parte das análises, tal como detalhado abaixo. A distribuição da amostra pelos diferentes níveis de treino musical é ilustrada na Figura 2.1. Devido a não existir um procedimento padrão para estimar o tamanho da amostra no contexto de modelos estatísticos de efeitos mistos baseados em dados não agregados (que foi a abordagem estatística utilizada na presente dissertação), foi planeado recrutar o máximo de participantes possível dentro da janela temporal definida para a recolha de dados (4 a 6 meses). O tamanho da amostra teve ainda como referência geral estudos anteriores debruçados sobre a mesma temática (e.g., $N = 80$ em Lima & Castro, 2011; $N = 169$ em Correia et al., 2020), sendo que o tamanho final da amostra se encontra entre estes valores ($N = 107$).

A amostra tinha em média 25.95 anos de idade ($DP = 9.27$; amplitude = 18–55 anos), sendo que 96.26% tinha nacionalidade portuguesa¹ e toda a amostra completou pelo menos o Ensino Secundário (12 anos de educação)². A maior parte da amostra tinha como principal ocupação ser estudante do ensino superior (59%), com 23.36% indicando trabalhar a tempo inteiro. Todos os participantes reportaram não ter dificuldades auditivas nem historial de perturbações neurológicas ou psiquiátricas. Relativamente à categorização dos participantes em função do treino musical, os participantes classificados como “músicos” deveriam cumprir os seguintes requisitos: ter pelo menos 6 anos de treino musical/instrumental formal, tal como em estudos anteriores (e.g., Correia et al., 2020; Lima & Castro, 2011) e praticar diariamente o instrumento. A amostra de músicos consistiu em 35 participantes, dos quais 28 eram instrumentistas e 7 cantores. Adicionalmente, 20 dos 35 participantes músicos reportaram tocar/cantar profissionalmente. A maior parte dos músicos reportou ter entre 6 a 9 anos de treino musical

¹ Quatro participantes reportaram não ter nacionalidade portuguesa. Entre as nacionalidades referidas encontram-se: brasileira ($n = 1$), cabo verdiana ($n = 1$), moçambicana ($n = 1$), e suíça ($n = 1$).

² Dos 107 participantes, 56.08% completou o ensino secundário, 26.17% completou licenciatura, e 17.76% reportaram ter completado o mestrado.

formal (54.29%), com 42.86% a reportar ter 10 ou mais anos de formação musical. Os instrumentos tocados pelos músicos, além da voz pelos sete que eram cantores, variaram entre piano ($n = 7$), guitarra ($n = 4$), violino ($n = 6$), saxofone ($n = 4$), clarinete ($n = 2$), baixo ($n = 1$), bateria ($n = 1$), flauta transversal ($n = 1$), órgão ($n = 1$), e viola d'arco ($n = 1$). Nenhum dos músicos reportou ter ouvido absoluto. Não foram encontradas associações significativas entre a duração do treino musical e a idade ($r = -.11, p = .25$), sexo ($r = -.11, p = .27$), e nível de escolaridade ($r = .12, p = .22$).

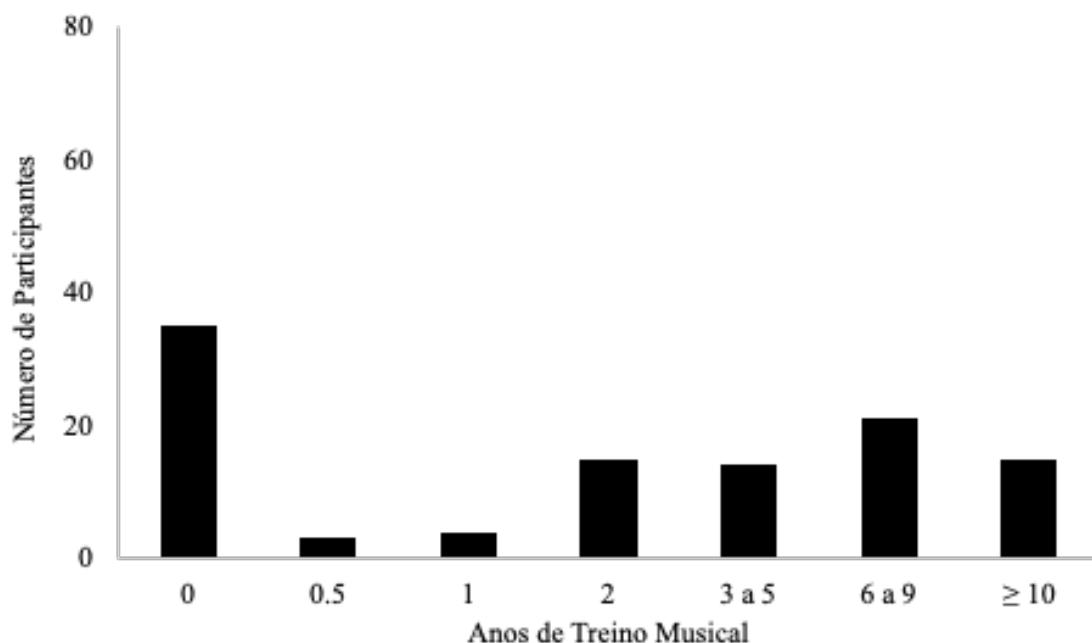
Relativamente aos restantes participantes, 87.88% tinham 0 ou até 2 anos de treino musical e foram categorizados como “não-músicos” ($n = 56$). Adicionalmente, 12.12% dos participantes apresentaram níveis intermédios de treino musical (3 a 5 anos de treino musical), e foram apenas incluídos em modelos estatísticos que trataram o treino musical como uma variável contínua. Assim, as comparações principais entre grupos foram entre os 35 músicos e os 56 não-músicos.

Entre os participantes não-músicos, parte reportou não tocar qualquer instrumento ($n = 25$), sendo que os restantes participantes indicaram ter tocado flauta ($n = 12$), guitarra ($n = 6$), baixo ($n = 1$), piano ($n = 2$), tambor ($n = 2$), kalimba ($n = 1$), e voz ($n = 7$).

O protocolo do estudo foi aprovado pela Comissão de Ética do Iscte - Instituto Universitário de Lisboa a 18/12/2020 (parecer 111/2020). O consentimento informado foi aplicado de forma digital a todos os participantes. Os participantes foram compensados, ou através de um sistema de créditos parciais (os recrutados através do SPI), ou através de um voucher (os recrutados através do PEO).

Figura 2.1.

Distribuição do treino musical formal na amostra ($N = 107$).



Instrumentos

Devido ao quadro pandémico de COVID-19, que restringiu severamente as atividades presenciais durante o período em que este trabalho foi levado a cabo, o estudo foi realizado à distância. Todas as tarefas foram adaptadas para formato *online* utilizando plataforma *Gorilla Experiment Builder* (www.gorilla.sc; Anwyl-Irvine et al., 2018).

Capacidades e experiência musical (autorrelato)

Os participantes preencheram a versão portuguesa validada do Gold-MSI (Lima et al., 2020), um questionário de autorrelato que mede a sofisticação musical de forma abrangente, considerando um conjunto alargado de facetas e comportamentos envolvidos na experiência musical (Müllensiefen et al., 2014). O instrumento é constituído por 38 itens, sendo que para 31 deles, o participante indica o nível de concordância com várias afirmações (e.g. “Eu consigo cantar ou tocar música de memória”) numa escala de *Likert* que varia entre 1 (*discordo totalmente*) a 7 (*concordo totalmente*). Nos restantes sete itens, o participante preenche escalas ordinais com sete alternativas de resposta que variam dependendo do comportamento musical que está a ser testado (e.g., “Eu consigo tocar (0-6) ou mais instrumentos”; Lima et al., 2020). Os itens podem ser agregados em cinco subescalas, que correspondem a diferentes facetas da experiência musical: Envolvimento Ativo (nove itens, e.g., “Passo muito do meu tempo livre

em atividades relacionadas com música”); Capacidades Percetivas (nove itens, e.g., “Consigo perceber quando alguém canta ou toca fora de tom”); Treino Musical (sete itens, e.g., “Tive instrução formal em teoria musical durante ____ anos”); Capacidades de Canto (sete itens, e.g., “Consigo acertar nas notas quando canto com uma gravação.”); e Emoções (seis itens, e.g., “Sou capaz de falar sobre as emoções que uma peça musical suscita em mim”). Para além destas dimensões, o Gold-MSI também permite calcular uma medida global de sofisticação musical, com base em 18 itens retirados das cinco subescalas (Índice de Sofisticação Geral), medida esta que foi utilizada nas análises estatísticas focadas neste questionário.

O Gold-MSI é um instrumento útil pois permite identificar diferenças individuais entre diferentes populações, desde músicos ativos com diferentes níveis de educação formal ou indivíduos que utilizam música de formas funcionais (e.g., DJs, educadores de música, produtores), até à população geral.

Capacidades objetivas de percepção musical

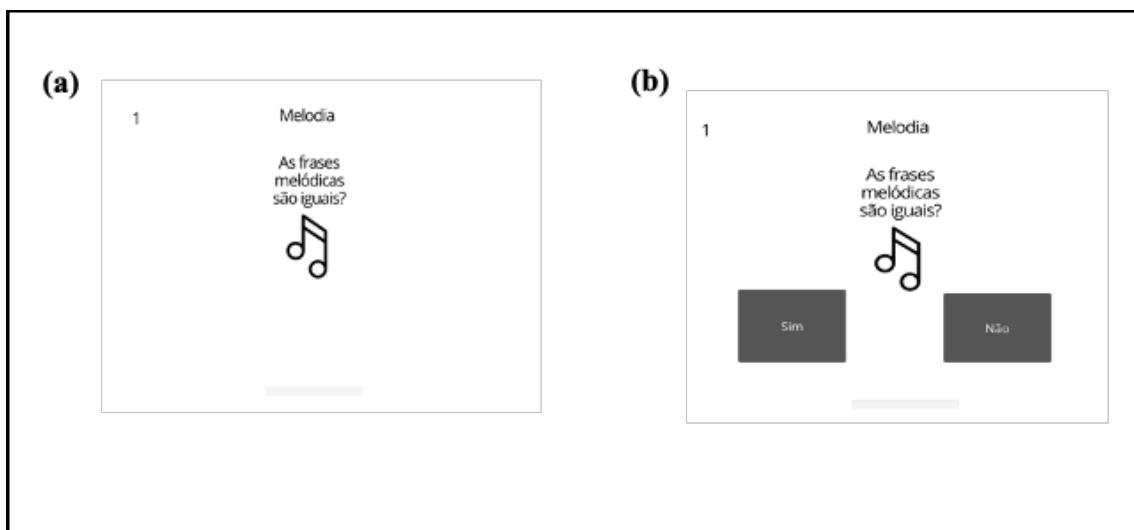
Para avaliar a capacidade de percepção musical, os participantes realizaram também o MET, cujo objetivo é medir a aptidão musical de forma objetiva e quantitativa em músicos e não-músicos (Wallentin et al., 2010). O teste consiste em 104 ensaios nos quais os participantes ouvem dois excertos musicais, e subsequentemente avaliam se são iguais ou diferentes. O teste é composto por dois subtestes, sendo a primeira metade do MET constituída pelo subteste de Melodia (52 pares de excertos melódicos tocadas em piano), e a segunda metade constituída pelo subteste de Ritmo (52 pares de excertos rítmicos tocados num bloco de madeira). Antes de cada subteste, os participantes realizam dois ensaios de treino com *feedback*. Metade dos ensaios em cada sessão são “iguais” e metade são “diferentes”. A média de respostas corretas nos 104 ensaios, e a média de respostas corretas em cada subteste foi recolhida como medida da aptidão musical dos participantes. O teste foi adaptado para a plataforma Gorilla, para ser aplicado online. Na tarefa original, as instruções e ensaios são apresentadas a partir de um ficheiro de áudio de 18 minutos, no qual as instruções para as tarefas e o número de cada ensaio são comunicados por um falante do sexo masculino. Durante cada subteste, os participantes têm uma janela temporal definida (1500 ms para os ensaios no subteste de melodia, 1659 a 3230 ms para os ensaios no subteste de ritmo) para indicarem na folha de resposta *Sim* ou *Não*.

Para a adaptação online do MET, as instruções e número de ensaios foram transformados em texto. O estímulo de cada ensaio foi digitalmente copiado a partir do ficheiro de áudio original e a duração dos intervalos entre estímulos foi preservada, de forma que a adaptação online do teste tivesse a mesma duração da versão original (18 minutos). O número do ensaio

e questão (e.g., *As frases melódicas são iguais?*) estavam visíveis no ecrã desde o início do ensaio até ao participante responder. Imediatamente após o estímulo auditivo ser apresentado, dois botões, com as opções de resposta *Sim* e *Não* ficavam visíveis e os participantes deveriam selecionar uma opção de resposta. Uma representação esquemática de um ensaio é ilustrada na Figura 2.2. A versão online do MET foi validada em Correia et al. (2021). Os resultados revelaram que a adaptação online do teste tem boas capacidades psicométricas, com níveis de consistência interna, tal como níveis de performance, semelhantes aos reportados para o teste original (Wallentin et al., 2010).

Figura 2.2.

Exemplo de ensaio no subteste de Melodia.



Notas. (a) O número do ensaio e a questão são apresentados enquanto o participante ouve o estímulo. (b) Após ouvir o estímulo, dois botões de resposta ficam visíveis, e os participantes devem responder dentro do tempo (1500 ms para o subteste de melodia). Caso o participante não responda dentro deste tempo, o próximo ensaio é automaticamente apresentado.

Reconhecimento de prosódia emocional

Para avaliar o reconhecimento de prosódia emocional, os participantes realizaram uma tarefa de *gating* auditivo, na qual os ouvintes ouvem uma quantidade cumulativa de informação vocal (Grosjean, 1985). A utilização desta tarefa permite investigar não só a evolução do reconhecimento da emoção em diferentes janelas temporais (*gates*), mas também a comparação destes resultados com os obtidos em estudos prévios que apresentaram os estímulos emocionais na sua totalidade.

Seleção de estímulos

Os estímulos consistiram em frases curtas, de conteúdo linguístico-semântico neutro (e.g., O quadro está na parede; Ela viajou de comboio), que expressavam quatro emoções (alegria, raiva, medo, tristeza) ou neutralidade através de pistas prosódicas. Para cada emoção, foram selecionados quatro estímulos de vozes femininas e quatro estímulos de vozes masculinas, ou seja, o estudo incluiu um total de 40 estímulos (oito estímulos x cinco categorias). Os estímulos produzidos por vozes femininas foram selecionados de uma base de dados previamente validada (Castro & Lima, 2010). A base de dados é constituída por um total 190 estímulos, gravados por duas falantes de português europeu, e validados por 80 participantes, sendo a base final constituída pelas expressões com níveis de reconhecimento mais elevados, reconhecidas mais rápido, e julgadas como transmitindo as emoções alvo de forma mais clara (ver Castro & Lima, 2010, para mais detalhes). Os estímulos produzidos por vozes masculinas foram selecionados de uma base de dados constituída por 224 estímulos, gravados por dois falantes de português europeu, seguindo o mesmo processo descrito em Castro e Lima (2010). Estes estímulos foram validados por 15 participantes, obtendo níveis de exatidão de reconhecimento comparáveis aos dos estímulos femininos (75% em Lima & Castro; 73% para estímulos masculinos).

Para o presente estudo foram utilizados estímulos com conteúdo semântico neutro, em vez de pseudo-frases. As pseudo-frases são frequentemente usadas em estudos que aplicam paradigmas de *gating* (Cornew et al., 2009; Jiang et al., 2015; Pell & Kotz, 2011; Rigoulot et al., 2013), de forma a evitar o efeito de pistas emocionais presentes no conteúdo linguístico que possam ser relevantes para decodificar a prosódia emocional (Pell & Kotz, 2011). Esta decisão foi tomada devido à existência de uma maior variedade de frases de conteúdo neutro do que pseudo-frases nas bases iniciais de estímulos. Para além disso, ambos os tipos estímulos, frases e pseudo-frases, foram confirmados como sendo eficazes na comunicação de emoções prosódicas (Castro & Lima, 2010). À semelhança do estudo desenvolvido na presente

dissertação, um estudo recente utilizou o paradigma de gating recorrendo a frases emocionais com conteúdo semântico neutro, e obteve resultados semelhantes aos descritos na literatura (Nordström & Laukka, 2019).

As emoções testadas foram a raiva, medo, tristeza, e alegria, tendo em conta o elevado nível de consenso em relação ao seu estatuto enquanto categorias discretas de expressão na face e voz (Pell & Kotz, 2011). Adicionalmente, estudos anteriores também já mostraram que participantes com treino musical reconhecem emoções básicas com maior exatidão comparativamente a não-músicos (e.g., Lima & Castro, 2011). Para o set final de estímulos usados aqui, foram selecionados das bases disponíveis aqueles reconhecidos com maior exatidão ($M_{\text{exatidão}} = 0.92$; $DP = 0.08$; amplitude = 0.67–1; ver Quadro 2.1 para valores de exatidão por emoção), de acordo com o mesmo procedimento seguido em estudos anteriores (e.g., Pell & Kotz, 2011; Rigoulot et al., 2013). Para assegurar variabilidade nos estímulos selecionados, as frases selecionadas para cada emoção foram selecionadas de forma em que o mesmo conteúdo semântico não se repetisse dentro da mesma categoria emocional. Os 40 estímulos finais tinham em média 1453.07 ms ($DP = 252$ ms; amplitude = 1039–2150 ms), sendo compostos por cinco sílabas em média ($DP = 1.2$; amplitude = 4–8 sílabas). O Quadro 2.1 sumariza as principais características acústicas e perceptivas dos estímulos selecionados, em função da emoção. Os parâmetros apresentados foram selecionados tendo por base a sua relevância teórica para a percepção de prosódia emocional (Anikin & Lima, 2018; Banse & Scherer, 1996; Eyben et al., 2016). O Quadro 2.1 apresenta informação referente à média e desvio padrão da frequência fundamental (em hertz), amplitude (em dB), e relação sinal-ruído (*Harmonics-to-noise ratio*; em dB), tal como parâmetros perceptivos referentes à exatidão do reconhecimento e duração dos estímulos.

Quadro 2.1.

Características acústicas e percepções dos estímulos selecionados para cada emoção.

Características percepções e acústicas	Emoção				
	Raiva	Alegria	Medo	Tristeza	Neutro
Exatidão	0.93 (3.12)	0.91 (8.20)	0.84 (10.92)	0.93 (4.93)	0.97 (4.85)
Duração (ms)	1 412.90 (222.74)	1 377.32 (181.73)	1 459.35 (279.23)	1 674.52 (306.66)	1 341.23 (144.29)
F0 (Hz)	249.54 (51.24)	289.85 (66.12)	257.33 (29.72)	164.84 (24.01)	180.74 (26.36)
HNR (dB)	11.64 (7.39)	14.00 (7.82)	12.94 (8.11)	15.30 (7.87)	12.35 (7.54)
Amplitude (Hz)	70.84 (6.49)	71.36 (6.97)	72.99 (6.85)	70.66 (6.22)	71.50 (7.11)

Notas. Os valores apresentados entre parenteses correspondem ao desvio padrão. F0 = Frequência fundamental; HNR = Harmonics-to-noise ratio.

Construção dos gates

As expressões foram editadas em Praat (Boersma & Weenik, 2021), para criar novas versões “gated” / segmentadas dos estímulos originais. Para controlar por diferenças no volume entre estímulos, e uniformizar a experiência de audição, cada expressão foi normalizada para uma amplitude máxima (*peak amplitude*) de 75dB (Pell & Kotz, 2011; Pell et al., 2009). A duração dos gates foi definida em função do tempo (ms), em vez de sílabas, para permitir um controlo mais rigoroso da quantidade de informação acústica apresentada ao participante (e.g., Castiago & Pinheiro, 2019; Nordström & Laukka, 2019). Considerando o argumento de isolar os efeitos do treino musical recorrendo a uma tarefa mais complexa, os estímulos começaram a ser segmentados aos 50 ms, de modo a obter uma resolução temporal mais fina em fases iniciais do processamento do estímulo (Nordström & Laukka, 2019). Os primeiros cinco gates foram segmentados em incrementos de 50 ms (50, 100, 150, 200, 250 ms). Visto que a duração média das frases ($M = 1472$ ms; $SD = 247$ ms) no presente estudo é inferior à de estudos anteriores (e.g., $M = 1.79$ s em Jiang et al., 2015; $M = 4.8$ s em Nordström & Laukka, 2019; $M = 1.70$ s em Pell & Kotz, 2011), a partir dos 250 ms foram aplicados incrementos de 150 ms, até aos 1000 ms da frase (400, 550, 700, 850, 1000 ms), atendendo a evidência de que todas estas emoções, incluindo a alegria, são reconhecidas em menos de um segundo (e.g., Pell & Kotz, 2011). Esta divisão criou um total de 11 gates, sendo que o último gate corresponde à expressão total (lista de gates: G50, G100, G150, G200, G250, G400, G550, G700, G850, G100, Gfull). Isto resultou num total de 440 estímulos (8 frases x 11 gates x 5 condições).

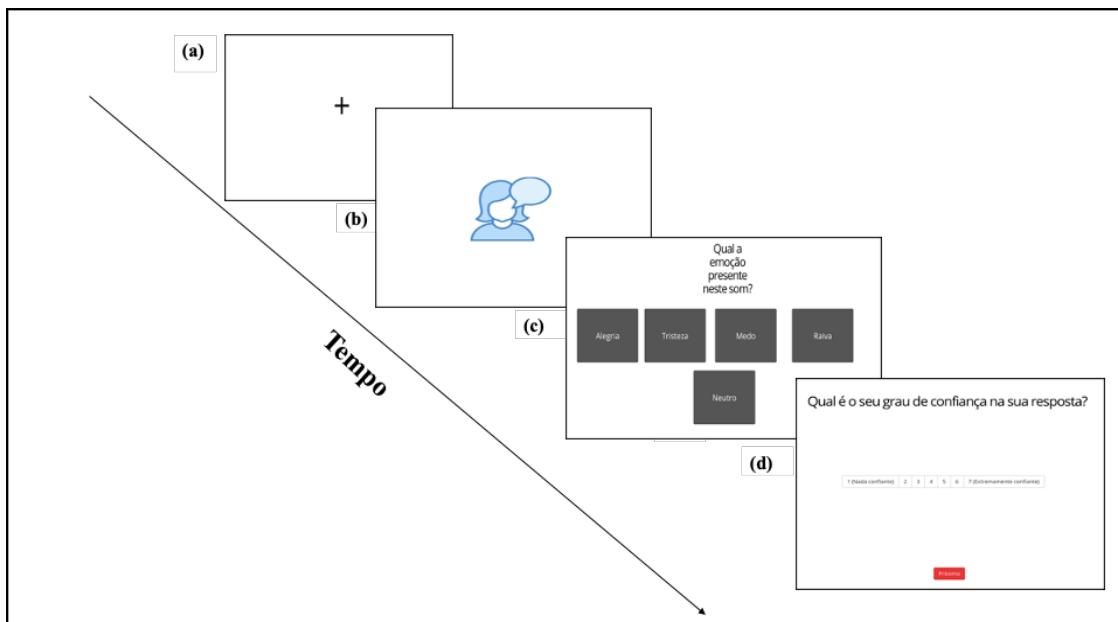
Tarefa experimental

Na tarefa de gating, os participantes receberam instruções acerca do procedimento, e realizaram seis ensaios de treino. Os estímulos foram apresentados por blocos de gate, nos quais a ordem de apresentação correspondia à duração dos gates, começando com todos os estímulos correspondentes ao gate mais curto (G50), depois todos os correspondentes ao gate seguinte (G100), e assim sucessivamente, aumentando a duração dos gates até à expressão completa (Gfull). Este modo de apresentação consecutiva por blocos é necessário para evitar possíveis enviesamentos decorrentes de ouvir primeiro o estímulo completo (Grosjean, 1985). Dentro de cada bloco, a ordem de apresentação dos estímulos foi aleatorizada para cada participante. Cada bloco foi composto por 40 estímulos, sendo que cada ensaio começou com um ponto de fixação apresentado no ecrã (1000 ms), e o estímulo auditivo. Após cada estímulo, os participantes responderam a duas questões de escolha forçada. Primeiro, realizaram uma tarefa de categorização emocional selecionando uma das cinco categorias possíveis (“raiva”, “medo”,

“alegria”, “tristeza”, “neutro”). A ordem das categorias emocionais no ecrã foi aleatorizada entre participantes. Imediatamente após a decisão da emoção expressa, os participantes avaliaram o nível de confiança face à escolha anterior numa escala de 1 (Nada confiante) a 7 (Extremamente confiante). Foram realizadas pausas entre cada bloco (e.g., Jiang et al., 2015; Nordström & Laukka, 2019). Uma representação esquemática de um ensaio é ilustrada na Figura 2.3.

Figura 2.3.

Exemplo de ensaio na tarefa de gating.



Notas. (a) Cada ensaio começa com uma cruz de fixação (1000 ms). (b) O estímulo auditivo é depois automaticamente apresentado. (c) O participante deve indicar a categoria emocional correspondente ao estímulo que ouviu. (d) De seguida, deve avaliar a confiança que tem na sua resposta.

Funcionamento cognitivo geral

Para avaliar o funcionamento cognitivo, os participantes realizaram o *Matrix Reasoning Item Bank* (MarRs-IB), que consiste em 80 matrizes de raciocínio não verbal (Chierchia et al., 2019). Esta tarefa tem um desenho semelhante a outros testes de raciocínio não verbal como as Matrizes de Raven (Raven, 2009). Cada item do *Mars-IB* contém uma matriz de nove elementos divididos em três linhas com três células cada (3x3). Oito das nove células contêm uma forma abstrata enquanto a nona célula se encontra vazia. A tarefa é encontrar a forma que falta para completar a matriz de entre um conjunto de opções. As formas e características das matrizes variam de acordo com quatro dimensões: forma, cor, tamanho e posição na matriz. O resultado

dos participantes foi calculado como a proporção do total de respostas dadas pelo participante que estavam corretas.

As instruções da tarefa foram adaptadas para português. Cada ensaio seguia a seguinte ordem: antes da matriz ser apresentada, aparecia uma cruz de fixação de 500 ms no meio do ecrã, seguida de um ecrã branco durante 100 ms. Os participantes tinham até 30 s para ver a matriz e selecionarem a opção correta. O ensaio terminava mais cedo caso o participante respondesse antes do tempo. Caso os participantes não respondessem em 25 s, aparecia um relógio do lado direito do ecrã a indicar o tempo restante (5 s).

A ordem dos ensaios foi igual para todos os participantes. Os primeiros cinco itens foram relativamente fáceis, de forma a familiarizar os participantes com a tarefa. Os participantes preenchiam itens durante oito minutos, no entanto, não receberam esta informação, nem informação relativa ao número de ensaios – a única indicação apresentada dizia respeito ao tempo que tinham para completar cada ensaio (30 s). Caso terminassem os 80 ensaios em menos de oito minutos, os mesmos ensaios seriam apresentados de novo na mesma ordem, mas as respostas da segunda ronda não seriam contabilizadas no cálculo da exatidão.

Procedimento

Cada participante foi testado numa sessão online. Para assegurar uma maior qualidade dos dados, e apoiar os participantes em dúvidas que pudessem ter, foi agendada uma sessão de Zoom com cada participante, na qual o investigador estava presente – ou seja, as experiências foram sempre acompanhadas de forma síncrona por um experimentador (que garantia que o participante estava a usar headphones, que estava num ambiente silencioso, e que se mantinha focado nas tarefas de acordo com o esperado). Após a inscrição no estudo, os participantes receberam um e-mail com instruções relativas às condições para a realização do mesmo (e.g., auscultadores funcionais, realizar a experiência num local sossegado e individualmente, manter a câmara ligada, desligar os aparelhos eletrónicos). Na sessão, os participantes receberam uma explicação oral sobre o estudo, tal como o link para a página web. Após consentirem participar no estudo, realizaram todas as tarefas de forma consecutiva (questionários sociodemográficos e Gold-MSI, Mars-Ib, MET, tarefa de gating). Poderiam interromper a sessão para fazer questões em caso de dúvida. Apesar de cada participante realizar a experiência individualmente no seu computador, as sessões de Zoom com o experimentador poderiam ter até 4 participantes presentes. A sessão teve uma duração total de 1 hora e 30 minutos.

Análise de Dados

Medidas cognitivas e musicais. Para o tratamento dos dados obtidos nas medidas cognitivas (MarsIb) e musicais (MET) foram aplicados os seguintes critérios de exclusão: para o cálculo dos scores no Mars-IB, itens em que o tempo de resposta foi inferior a 500 ms foram excluídos das análises (Chierchia et al., 2019; Correia et al., 2021). Para o cálculo dos scores no MET, foram excluídos das análises participantes que tiveram: 10 ou mais *missings* em algum dos subtestes e menos de 22 respostas corretas nos 52 ensaios de cada subteste (Correia et al., 2021). Foram excluídas das análises 14 respostas no subteste de melodia ($n = 93$), e 13 respostas no subteste de ritmo ($n = 94$).

Musicalidade. Considerando que foram aplicadas medidas de autorrelato (Gold-MSI), tal como um teste objetivo de percepção musical (MET) para identificar participantes com boas competências musicais entre não-músicos, foi calculada uma medida de *musicalidade*, que agregou os resultados obtidos em ambas as medidas. Para calcular uma medida de musicalidade geral, foi realizada uma Análise de Componentes Principais para extrair uma variável latente a partir do Índice de Sofisticação Geral (18 itens) obtido no Gold-MSI e dos scores obtidos nos subtestes de Melodia e Ritmo do MET. A análise de Componentes Principais revelou que a solução de um fator explicou 55% da variância total nos dados originais. Os subtestes de Melodia e Ritmo tiveram uma contribuição maior para a variável latente ($r = .85$ e $.86$, respectivamente), e o Fator Geral de Sofisticação Musical teve uma contribuição menor ($r = .41$). Valores mais elevados nesta medida indicam mais musicalidade. Com base nos valores obtidos nesta medida, os participantes sem treino musical (0–2 anos) foram agrupados em 2 categorias (não-músicos de *alta musicalidade*; não-músicos de *baixa musicalidade*). A divisão dos participantes por categorias foi feita com referência à média obtida nesta medida. Participantes com valores abaixo da média foram alocados no grupo de *baixa musicalidade* ($n = 24$; $M = -0.82$; $DP = 0.49$), e participantes com valores acima da média foram alocados no grupo de *alta musicalidade* ($n = 21$; $M = 0.85$; $DP = 0.58$)³.

Gating. Para as análises dos dados obtidos na tarefa de *gating*, usaram-se duas medidas principais, de acordo com os procedimentos tipicamente aplicados neste tipo de estudo (e.g., Jiang et al., 2015; Nordström & Laukka, 2019; Pell & Kotz, 2011). A primeira medida considerou a exatidão do reconhecimento em cada gate (*hit rates*), de forma a revelar como o reconhecimento muda em função da duração dos gates. Para esta análise, a medida de interesse foi a resposta dada a cada estímulo de cada gate, que pode corresponder a uma categorização

³ Excluímos desta análise 11 respostas por não cumprirem os critérios de inclusão no MET, o que impossibilitou o cálculo do score de musicalidade para estes participantes.

correta ou incorreta da emoção (hit rates). A segunda medida teve por objetivo analisar o *tempo* mínimo necessário para reconhecer emoções vocais, determinando-se para tal o EIP. Para cada um dos 40 estímulos incluídos, e para cada participante, o EIP foi definido como o gate em que o participante identificou corretamente a emoção alvo, sem mudanças subsequentes em gates mais longos para esse estímulo (Grosjean, 1985). O EIP foi calculado separadamente para cada um dos 40 estímulos, tal como avaliados por cada participante (40 estímulos x 107 participantes = 4280 EIPs no total, ou 856 por emoção). O sistema de cotação permitia que o EIP fosse atribuído ao gate mais curto se o participante desse apenas uma resposta incorreta seguida de duas identificações corretas consecutivas durante os restantes gates para esse estímulo (por exemplo, o EIP para um estímulo de raiva com o seguinte padrão de resposta em cada gate: “neutro (G50), neutro (G100), **raiva (G150)**, raiva (G200), raiva (G250), raiva (G400), raiva (G550), **neutro (G700)**, raiva (G850), raiva (G1000), raiva (GFull)”, é considerado no gate 150; Pell & Kotz, 2011; Salasoo & Pisoni, 1985). De forma a estimar o tempo necessário para reconhecer emoções, os nomes atribuídos aos gates (e.g., G50, G100, G150) foram substituídos pelos milissegundos correspondentes ao intervalo de gate (e.g., 50 ms, 100 ms, 150 ms), e o EIP médio para cada categoria emocional foi definido como a média de todos os estímulos correspondentes a essa emoção (Jiang et al., 2015; Pell & Kotz, 2011).

Adicionalmente, a proporção de participantes que escolheu cada categoria emocional para cada gate foi calculada para determinar a matriz das confusões mais comuns entre categorias emocionais (Anexo A). O hit rate corrigido (Hu) também foi calculado de forma a identificar possíveis vieses nos níveis de exatidão observados para cada categoria emocional, em cada intervalo de gate. Os valores de Hu foram calculados para cada participante e condição – os valores de Hu permitem obter uma medida de sensibilidade percetiva, sendo definidos como “a probabilidade conjunta de um estímulo ser corretamente identificado (dado que é apresentado) e de que a resposta é corretamente aplicada (dado que é utilizada)” (Wagner, 1993, p.16). Os valores de Hu podem variar entre 0 e 1, sendo que um valor de 1 indica que todos os estímulos correspondentes a determinada emoção foram corretamente classificados e que a categoria emocional nunca foi erroneamente selecionada para classificar estímulos alusivos a outras emoções. Os participantes também realizaram avaliações de confiança, mas como esta é uma medida secundária em estudos de gating, os dados não foram analisados no contexto desta dissertação, no entanto, serão analisados a posteriori para publicação científica.

Análises estatísticas. As análises estatísticas, quer dos dados de exatidão, quer dos EIP, foram baseadas numa série de modelos de efeitos mistos, baseados em dados não agregados de ensaios individuais, em que os participantes e os itens foram incluídos como random intercepts (também

foram incluídos random slopes por participante em alguns dos modelos, tal como detalhado na secção de Resultados). Estas análises contrastam com as análises tradicionais, mais típicas, de que são exemplo as ANOVAs baseadas em valores médios (agregados) de cada participante em cada condição. As suas principais vantagens são: aumentar o poder estatístico (dado que os modelos incorporam informação de todos os ensaios e participantes), assim como permitir uma generalização mais abrangente, não só entre participantes (como nas análises tradicionais), mas também entre itens.

As análises foram feitas no programa *R* Versão 4.0.5 (R Core Team, 2021), e *RStudio* Versão 1.4.1717 (RStudio Team, 2021), recorrendo a vários pacotes de R. O pacote “*tidyverse*” (Wickham et al., 2019) foi utilizado para o processamento dos dados; os pacotes “*lme4*” (Bates et al., 2015) e “*lmerTest*” (Kuznetsova et al., 2017) foram utilizados para as análises estatísticas de exatidão e tempo (EIP); o pacote “*sjPlot*” (Lüdecke, 2021) foi utilizado para reportar os principais resultados estatísticos.

Efeitos fixos foram codificados como (1, -1), sendo que -1 corresponde à categoria de referência (e.g., codificação do fator fixo “grupo”: músicos = 1; não-músicos = -1). No fator emoção, a categoria “neutro” foi definida como categoria de referência, e no fator grupo, a categoria “não-músico” foi definida como categoria de referência. Foram estimados modelos de efeitos mistos com regressão logística para analisar os dados de exatidão (hit rate como variável dependente), e modelos de efeitos mistos com regressão linear para analisar os EIPs e o desempenho quando a informação acústica era incompleta (i.e., média de desempenho em intervalos de gate mais curtos como variável dependente). Para os modelos de efeitos mistos com regressão logística e linear, utilizamos a funcionalidade de modelo misto do pacote “*afex*” no R (Singmann et al., 2021) para obter os valores de *p* via testes de *likelihood ratio* (LRT) para os efeitos principais e de interação. Os LRTs comparam modelos estatísticos hierarquicamente organizados por complexidade (e.g., um modelo simples contemplaria apenas a contribuição de um parâmetro, um modelo mais complexo que esse contemplaria a interação entre dois parâmetros, um modelo ainda mais complexo contemplaria a interação entre 3 parâmetros, etc.), e determina se adicionar complexidade ao modelo o torna mais preciso na explicação da variável dependente. Detalhes acerca destes modelos são apresentados no capítulo seguinte.

Capítulo III. Resultados

Dados cognitivos e de musicalidade

O Quadro 3.1. sumariza os resultados das medidas musicais, MET (proporção de respostas corretas) e Gold-MSI (pontuação média), tal como cognitivas, Mars-IB (proporção de respostas corretas e número de itens corretos). São apresentados resultados para a amostra total ($N = 107$), para os participantes músicos ($n = 35$), e para participantes não-músicos ($n = 56$). No Quadro 3.2. são apresentadas as correlações entre a duração do treino musical e as principais medidas cognitivas e musicais na amostra completa.

Quadro 3.1.

Estatísticas descritivas das medidas musicais e cognitivas da amostra completa, e separadamente para músicos e não-músicos

Tarefa	Amostra total ($N = 107$)	Não-Músicos ($n = 56$)	Músicos ($n = 35$)	p
MET Total	.68 (.11)	.62 (.09)	.76 (.09)	<.001
MET Melodia	.67 (.13)	.60 (.09)	.77 (.11)	<.001
MET Ritmo	.70 (.11)	.66 (.11)	.75 (.09)	<.001
Gold-MSI (sofisticação geral)	4.24 (.57)	3.90 (.56)	4.44 (.48)	<.001
Mars-IB (proporção respostas corretas)	.63 (.13)	.60 (.14)	.68 (.12)	.003
Mars-IB (itens corretos)	22.84 (6.17)	22.59 (6.12)	23.24 (6.30)	.51

Notas. Os valores de p correspondem aos valores de significância associados a testes t para comparação de médias em amostras independentes (comparações entre músicos e não-músicos).

Comparações entre grupos mostram que os músicos apresentaram valores significativamente superiores aos não-músicos no Gold-MSI, $t(89) = -5.535$, $p < .001$, no MET Total, $t(78) = -7.092$, $p = <.001$, tal como nos subtestes de Ritmo, $t(78) = -4.06$, $p < .001$, e

Melodia, $t(77) = -7.35$, $p < .001$. Em relação à medida cognitiva, comparações entre músicos e não-músicos a nível da proporção de respostas corretas, revelaram diferenças significativas entre grupos, $t(89) = -3.89$, $p < .001$ (Quadro 3.1.). No entanto, quando a comparação foi relativa ao número médio de itens corretos que cada grupo completou, não foram encontradas diferenças significativas, $t(91) = 0.67$, $p = .51$.

Também foram observadas associações significativas entre a duração do treino musical e os resultados nas facetas do Gold-MSI e nos subtestes de Melodia e Ritmo do MET, tal como uma associação de fraca magnitude entre a duração do treino musical e as capacidades cognitivas. Não foram observadas associações entre o número de itens completados no MarsIB e a duração do treino musical (Quadro 3.2.).

Quadro 3.2.

Correlações entre a duração do treino musical e as variáveis musicais e cognitivas.

Variável	r	p
MET (Total)	.56	<.001
MET Melodia	.58	<.001
MET Ritmo	.41	<.001
Gold-MSI (Sofisticação Geral)	.58	<.001
Gold-MSI: Treino Musical	.87	<.001
Gold-MSI: Envolvimento Ativo	.30	.002
Gold-MSI: Capacidades Percetivas	.40	<.001
Gold-MSI: Capacidades de Canto	.37	<.001
Gold-MSI: Emoções	.31	.001
Mars-IB (respostas corretas)	.27	.006
Mars-IB (Itens)	.08	.42

Notas. Gold-MSI = *Goldsmiths Musical Sophistication Index*; MET = *Musical Ear Test*.

Exatidão ao longo dos gates

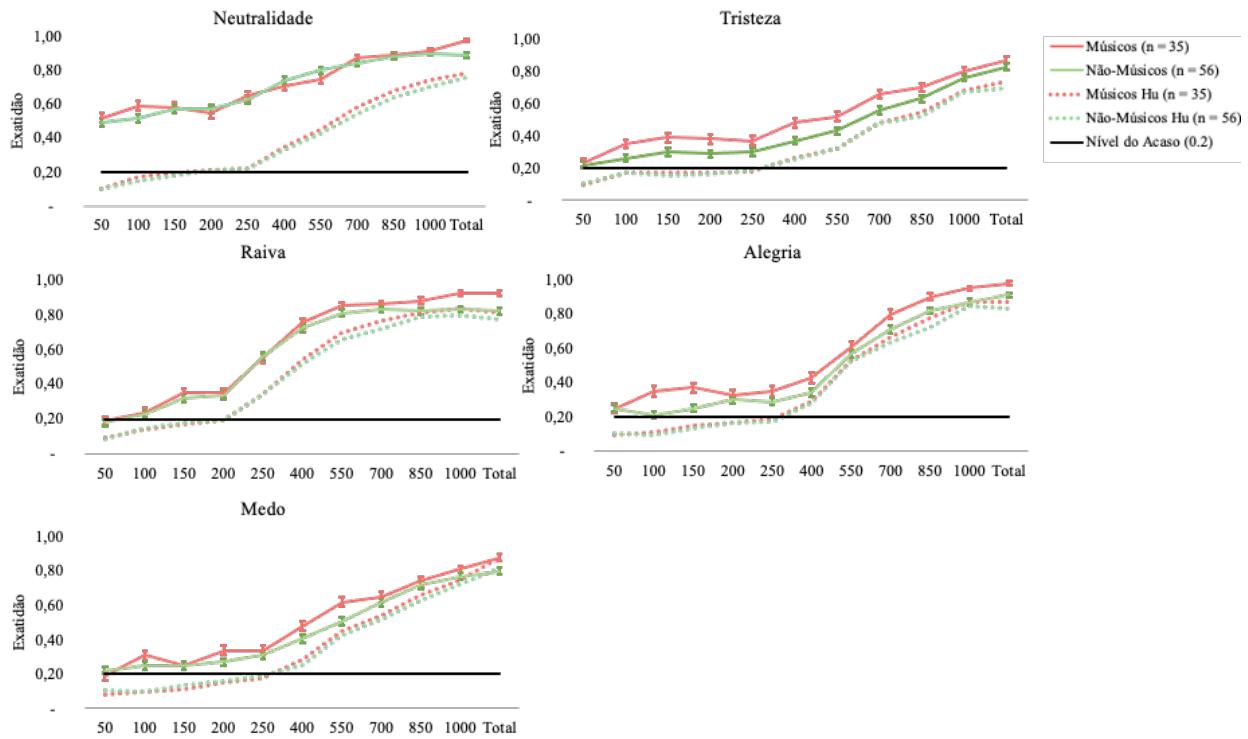
A Figura 3.1. apresenta a exatidão, em proporção bruta de respostas corretas, para cada um dos gates, emoção e grupo. Para informação adicional, as pontuações Hu também são apresentadas. Os valores médios de reconhecimento brutos e Hu são ainda apresentados nos Quadros 3.3. e 3.4., respetivamente.

Uma primeira análise dos resultados focou-se na progressão global do reconhecimento das diferentes emoções ao longo dos gates. Esta questão foi inicialmente explorada num modelo de efeitos mistos focado na exatidão do reconhecimento. Este modelo utilizou uma regressão logística, considerando as respostas dadas pelos participantes a cada estímulo como variável dependente (respostas que podiam ser corretas ou erradas; *hit rate*), e como fatores fixos o intervalo de gate (G50 até GFull), a emoção (neutralidade, raiva, tristeza, alegria, medo), o grupo (músicos vs. não-músicos), e a interação entre os três preditores. O modelo considerou o participante e o estímulo como efeitos aleatórios (*random effects*), tal como random slopes por participante em função da emoção. Considerando que o modelo máximo não convergiu, foi estimado um novo modelo simplificado, que removeu as random slopes por participantes. Tal como esperado, a exatidão melhorou com o aumento da duração dos gates, como evidenciado pelo efeito principal do gate, $\chi^2(10, N = 91) = 643.25, p < .001$. O efeito principal da emoção, $\chi^2(4, N = 91) = 186.52, p < .001$, revela também que o nível global de exatidão diferiu entre emoções. O reconhecimento de tristeza (diferença estimada = -1.33, SE = 0.10, $z = -12.86, p < .001$), medo (diferença estimada = -1.32, SE = 0.10, $z = -12.72, p < .001$), alegria (diferença estimada = -0.89, SE = 0.11, $z = -8.42, p < .001$), e raiva (diferença estimada = -0.62, SE = 0.11, $z = -5.89, p < .001$) foi significativamente menor do que o reconhecimento de neutralidade ao longo dos gates. Estes resultados sugerem que os participantes foram particularmente melhores a reconhecer neutralidade, comparativamente às restantes emoções.

O efeito principal do grupo indica os músicos foram, de forma geral, mais exatos a reconhecer as emoções do que os não-músicos, $\chi^2(1, N = 91) = 6.22, p = .01$. Finalmente, um efeito de interação significativo entre a duração do gate e emoção, $\chi^2(40, N = 91) = 119.13, p < .001$, grupo e gate, $\chi^2(10) = 56.27, p < .001$, tal como um efeito de interação significativo entre o gate, a emoção, e o grupo, $\chi^2(40, N = 91) = 72.24, p = .001$, revela que certas emoções foram reconhecidas com maior exatidão em função do intervalo de gate e do treino musical. Esta interação tripla não será decomposta aqui porque o efeito de treino musical, que é o nosso foco principal de interesse, é avaliado em detalhe nas próximas secções. Informação adicional sobre este primeiro modelo é apresentada em anexo (Anexo B).

Figura 3.1.

Exatidão do reconhecimento de diferentes emoções prosódicas, em função da duração de gate e do grupo (músicos e não-músicos).



Notas. São apresentados valores médios de exatidão. As barras de erro representam o erro padrão. Os valores de hit rate corrigidos (valores Hu) são representados pelas linhas a tracejado. A linha horizontal a preto representa o nível do acaso (0.2).

Quadro 3.3.

Média de categorizações corretas para cada emoção, gate e grupo.

Emoção	Músicos (n=35)										
	Gate (ms)										
	50	100	150	200	250	400	550	700	850	1000	Total
Neutralidade	.52 (.50)	.59 (.49)	.58 (.49)	.55 (.50)	.65 (.48)	.71 (.45)	.75 (.43)	.87 (.34)	.89 (.31)	.91 (.29)	.98 (.13)
Raiva	.19 (.39)	.24 (.43)	.35 (.48)	.35 (.48)	.55 (.50)	.75 (.44)	.85 (.35)	.86 (.34)	.88 (.32)	.92 (.27)	.92 (.27)
Medo	.19 (.39)	.31 (.46)	.25 (.43)	.34 (.47)	.34 (.47)	.48 (.50)	.62 (.48)	.65 (.48)	.74 (.44)	.81 (.39)	.88 (.32)
Tristeza	.23 (.42)	.35 (.48)	.39 (.49)	.38 (.49)	.37 (.48)	.48 (.50)	.52 (.50)	.66 (.47)	.70 (.46)	.80 (.40)	.87 (.34)
Alegria	.25 (.44)	.35 (.48)	.37 (.48)	.33 (.47)	.35 (.48)	.43 (.50)	.61 (.49)	.80 (.40)	.90 (.30)	.95 (.22)	.98 (.14)
Não Músicos (n=56)											
Emoção	Gate (ms)										
	50	100	150	200	250	400	550	700	850	1000	Total
Neutralidade	.49 (.50)	.52 (.50)	.57 (.50)	.57 (.50)	.63 (.48)	.74 (.44)	.8 (.40)	.84 (.37)	.88 (.33)	.90 (.29)	.89 (.31)
Raiva	.18 (.39)	.23 (.42)	.32 (.47)	.34 (.48)	.56 (.5)	.72 (.45)	.81 (.39)	.83 (.38)	.82 (.38)	.83 (.37)	.82 (.39)
Medo	.22 (.41)	.25 (.44)	.25 (.44)	.27 (.45)	.31 (.46)	.41 (.49)	.51 (.50)	.62 (.49)	.72 (.45)	.77 (.42)	.80 (.40)
Tristeza	.22 (.42)	.26 (.44)	.30 (.46)	.29 (.45)	.30 (.46)	.37 (.48)	.43 (.50)	.56 (.50)	.63 (.48)	.76 (.43)	.83 (.38)
Alegria	.25 (.44)	.21 (.41)	.25 (.44)	.3 (.46)	.29 (.45)	.34 (.47)	.57 (.50)	.71 (.45)	.82 (.39)	.87 (.34)	.91 (.28)

Notas. Valores de desvio padrão apresentados entre parênteses.

Quadro 3.4.

Média de valores Hu em função da emoção, gate e grupo.

Emoção	Músicos (n = 35)										
	Gate (ms)										
	50	100	150	200	250	400	550	700	850	1000	Total
Neutralidade	.10	.17	.20	.21	.22	.35	.45	.58	.68	.74	.78
Raiva	.09	.14	.17	.19	.34	.54	.69	.76	.81	.84	.81
Medo	.08	.10	.11	.15	.18	.29	.45	.55	.66	.75	.88
Tristeza	.09	.17	.17	.17	.18	.26	.32	.48	.54	.68	.74
Alegria	.10	.11	.15	.16	.19	.29	.52	.66	.77	.87	.87
Não Músicos (n = 56)											
Emoção	Gate (ms)										
	50	100	150	200	250	400	550	700	850	1000	Total
Neutralidade	.10	.15	.18	.21	.22	.33	.43	.54	.64	.70	.76
Raiva	.09	.15	.18	.19	.34	.52	.65	.72	.78	.79	.77
Medo	.10	.10	.13	.16	.19	.26	.42	.52	.63	.72	.82
Tristeza	.10	.17	.15	.16	.19	.25	.32	.48	.52	.68	.70
Alegria	.10	.09	.13	.17	.18	.28	.53	.64	.73	.85	.83

Reconhecimento Acima do Nível do Acaso

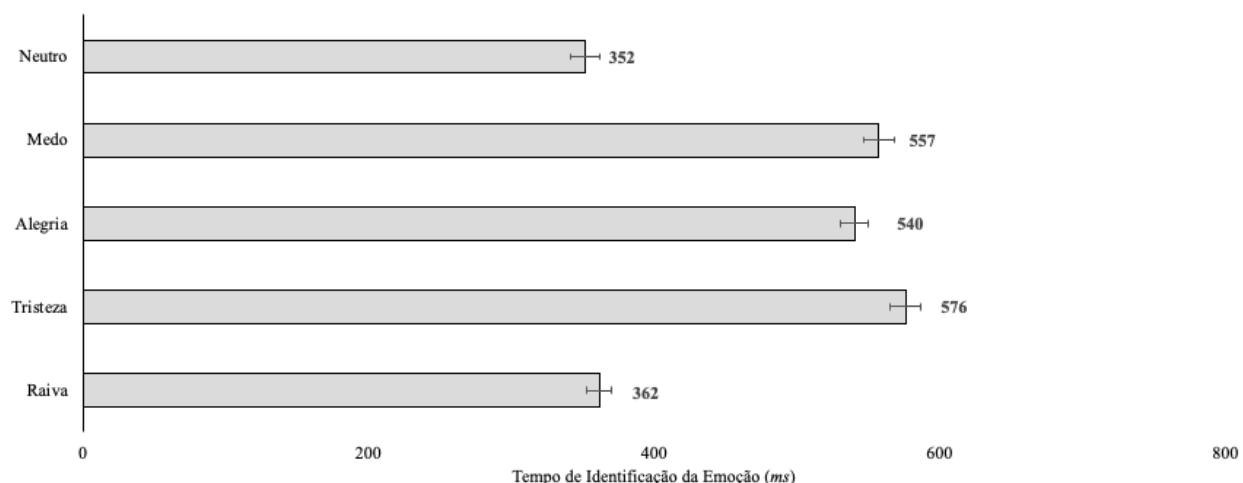
Em relação ao gate mais curto em que cada emoção foi reconhecida significativamente acima do nível do acaso, foram aplicados testes t para comparar a média de reconhecimento de cada emoção ao longo dos *gates* com o nível do acaso (.20). Os resultados mostraram que a neutralidade foi reconhecida acima do nível do acaso logo no primeiro gate, ou seja, quando os participantes eram expostos apenas aos 50 ms iniciais da expressão, $t(106) = 11.14, p < .001$. A alegria também foi reconhecida acima do nível do acaso aos 50 ms, $t(106) = 2.61, p = .01$, mas o mesmo não foi observado para as outras emoções (todos os $p > 0.1$). A partir do G100, ou seja, quando expostos a 100 ms da expressão, os participantes reconheceram adicionalmente raiva, $t(106) = 3.91, p < .001$, tristeza, $t(106) = 4.55, p < .001$, e medo, $t(106) = 3.23, p = .002$, acima do nível do acaso. Apesar destes resultados sugerirem um reconhecimento precoce das emoções prosódicas, os dados na neutralidade em particular devem ser interpretados com cautela: é possível que, face a informação acústica muito limitada, os participantes tenham sido enviesados a responder com a categoria “neutro”, independentemente do tipo de estímulo, dando origem a níveis de exatidão artificialmente elevados. Esta possibilidade é suportada pelos baixos valores de Hu (uma medida que corrige vieses) para esta categoria no primeiro gate (ver Figura 3.1. e Quadro 3.4), assim como por uma análise da matriz de confusão, em que se observa que as confusões mais comuns no primeiro gate foram justamente com neutralidade (Anexo A).

Pontos de Identificação da Emoção (EIPs)

A Figura 3.2. apresenta os EIPs médios para cada categoria emocional, uma medida que permite estimar o tempo (em ms) associado à estabilização do reconhecimento (correto) de um significado emocional discreto (ver Métodos). É possível observar pela Figura 3.2., que as emoções associadas a um EIP mais curto foram a neutralidade e raiva, seguidas de alegria, medo e tristeza. A distribuição de EIPs por intervalo de gate em função da categoria emocional é apresentado em anexo (Anexo C).

Figura 3.2.

Pontos de identificação da emoção (EIP) para cada emoção (ms).



Notas. Valores de EIP apresentados correspondem à média de EIPs para os estímulos correspondentes a cada emoção. As barras de erro representadas na figura correspondem aos valores do erro-padrão.

Treino Musical Formal

Exatidão do reconhecimento. A Figura 3.3. apresenta as diferenças entre músicos e não-músicos no reconhecimento da expressão completa, ou seja, a exatidão apenas relativa ao último gate. Para comparações com estudos anteriores sobre treino musical e o reconhecimento de emoções, que se basearam em estímulos completos (e.g., Correia et al., 2020; Lima & Castro, 2011; Thompson et al., 2004), as primeiras análises de possíveis efeitos de grupo foram assim focadas nas respostas ao último gate.

Foi estimado um modelo com regressão logística que incluiu o grupo (músicos vs. não-músicos), tipo de emoção (neutralidade, raiva, tristeza, alegria, medo) e a interação entre ambas as variáveis como fatores fixos, tal como random intercepts por participante e estímulo⁴. Os resultados mostraram um efeito principal de grupo, $\chi^2(1, N = 91) = 8.19, p = .004$, um efeito principal da emoção, $\chi^2(4, N = 91) = 32.09, p < .001$, e uma interação significativa entre ambos os termos, $\chi^2(4, N = 91) = 12.88, p = .01$.

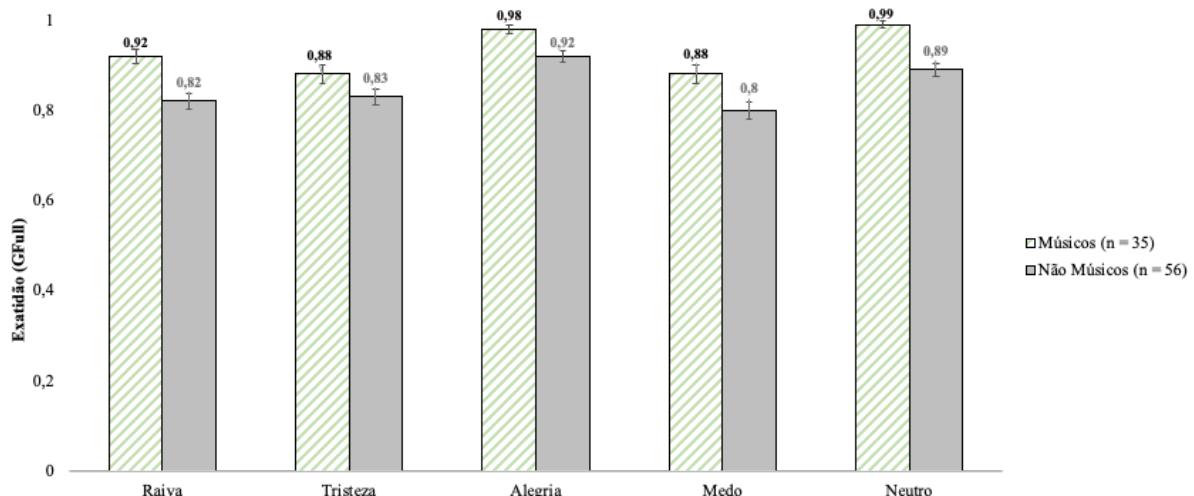
No que toca ao efeito principal da emoção, os participantes foram, de forma geral, menos precisos a reconhecer tristeza (diferença estimada = -1.49, SE = 0.34, $z = -4.45, p < .001$), medo (diferença estimada = -1.62, SE = 0.34, $z = -4.80, p < .001$), e raiva (diferença estimada = -1.19, $z = -4.45, SE = 0.34, p = .004$), comparativamente a neutralidade. No caso da alegria, a diferença face à neutralidade não foi significativa ($p = 0.9$). O efeito principal de grupo indica que houve uma vantagem global dos músicos no reconhecimento do estímulo completo, mas a interação entre grupo e emoção sugere que esta vantagem pode não ter sido semelhante para todas as emoções.

Análises posteriores do efeito de interação revelaram que os músicos foram melhores que os não-músicos a reconhecer neutralidade (diferença estimada = 1.95, SE = 0.59, $z = 3.29, p = .005$), e marginalmente mais exatos que os não-músicos no reconhecimento de raiva (diferença estimada = 0.94, SE = 0.39, $z = 2.40, p = .07$), e alegria (diferença estimada = 1.19, SE = 0.53, $z = 2.20, p = .08$). Não foram encontradas diferenças significativas para as restantes emoções (todos $ps > .1$).

⁴ No primeiro modelo estimado, foram ainda consideradas random slopes por participante para o fator emoção. Considerando que este modelo não convergiu (o que indica a necessidade de simplificar o modelo estimado), o mesmo modelo foi novamente estimado removendo as random slopes por participante.

Figura 3.3.

Diferenças entre músicos e não-músicos na exatidão do reconhecimento de emoções.



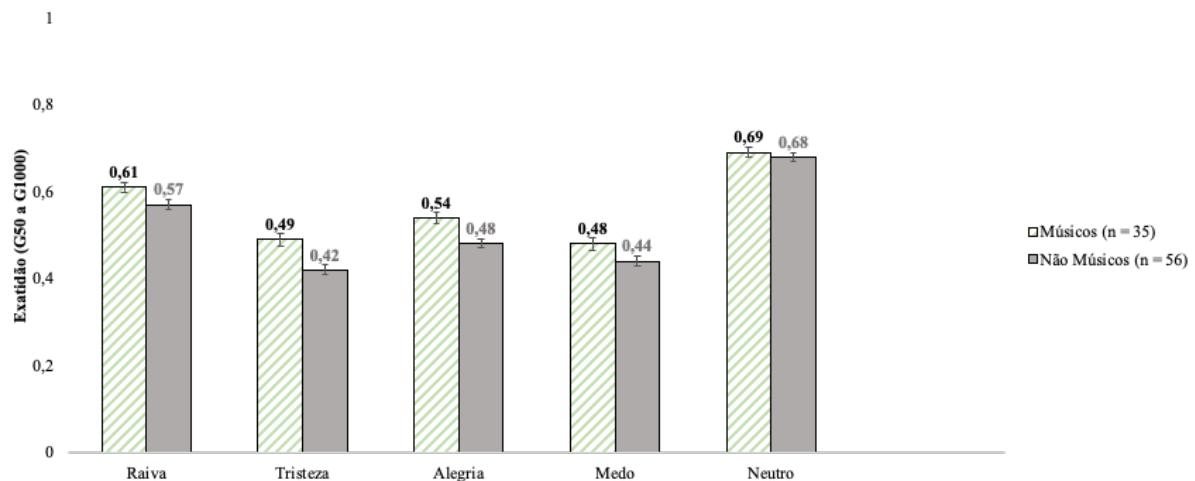
Notas. As barras de erro representadas na figura correspondem aos valores do erro-padrão.

Desempenho em condições de reconhecimento mais exigentes. Numa segunda análise de efeitos de grupo, perguntámos se os músicos foram mais exatos do que os não-músicos quando a informação acústica era incompleta – ou seja, o desempenho neste contexto foi centrado no desempenho médio entre os gates 1 (50 ms) e 10 (1000 ms). Na Figura 3.4. é apresentado o desempenho quando a informação acústica era incompleta, em função da categoria emocional e grupo. Esta questão foi explorada num modelo de efeitos mistos com regressão linear que considerou a média de desempenho entre o G50 e o G1000 como variável dependente e o grupo, emoção, e interação entre ambos como fatores fixos. Adicionalmente, foram considerados random intercepts por participante e estímulo, tal como random slopes por participante para o fator emoção. Os resultados revelaram um efeito principal de grupo, $\chi^2(1, N = 91) = 4.43, p = .04$, o que aponta para uma vantagem dos músicos no reconhecimento de emoções quando a informação era incompleta (diferença estimada = 0.04, SE = 0.02, $t = 2.13, p = .04$). A ausência de um efeito de interação significativo entre grupo e emoção, $\chi^2(4, N = 91) = 2.62, p = .62$, sugere que esta vantagem é geral para todas as emoções. O modelo revelou ainda um efeito principal de emoção, $\chi^2(4, N = 91) = 29.35, p < .001$, o que indica que existiram diferenças significativas no desempenho dos participantes mediante a emoção. Análises de contrastes entre categorias emocionais indicaram que, comparativamente à neutralidade, os participantes tiveram um pior desempenho do reconhecimento de tristeza (diferença estimada = -0.23, SE = 0.05, $t = -4.82, p < .001$), medo (diferença estimada = -0.22, SE = 0.05, $t = -4.96, p < .001$), e alegria (diferença estimada = -0.17, SE = 0.04, $t = -4.1, p < .001$), quando expostos a menos

informação acústica. Não foram encontradas diferenças significativas entre raiva e neutralidade ($p = 0.9$), sugerindo que estas emoções foram as que os participantes reconheceram com melhor exatidão com base em menos informação acústica.

Figura 3.4.

Diferenças entre músicos e não-músicos na exatidão do reconhecimento de emoções quando a informação acústica era incompleta (desempenho entre G50 e G1000).



Notas. As barras de erro representadas na figura correspondem aos valores do erro-padrão.

Para explorar a hipótese de que o efeito de treino musical pode ser mais claro quando as condições de reconhecimento são particularmente exigentes, foram estimados modelos de efeitos mistos com regressão logística focados no desempenho a nível do primeiro gate em que o reconhecimento das emoções foi acima do nível do acaso. No caso de neutralidade e alegria, que foram reconhecidas acima do nível do acaso logo aos 50 ms, foi estimado um modelo misto que considerou o hit rate no G50 como variável dependente, e o grupo, a emoção, e a interação entre ambos como preditor. Adicionalmente, foram considerados random intercepts por participante e estímulo⁵.

Os resultados não revelaram um efeito principal de grupo, $\chi^2(1, N=91) = 0.01, p = .93$, o que demonstra que a nível dos 50 ms, os músicos não apresentaram vantagens gerais sobre os não-músicos no reconhecimento de emoções. Também não foi reportada uma interação significativa entre grupo e emoção, $\chi^2(4, N=91) = 2.19, p = .70$, o que indica que a ausência

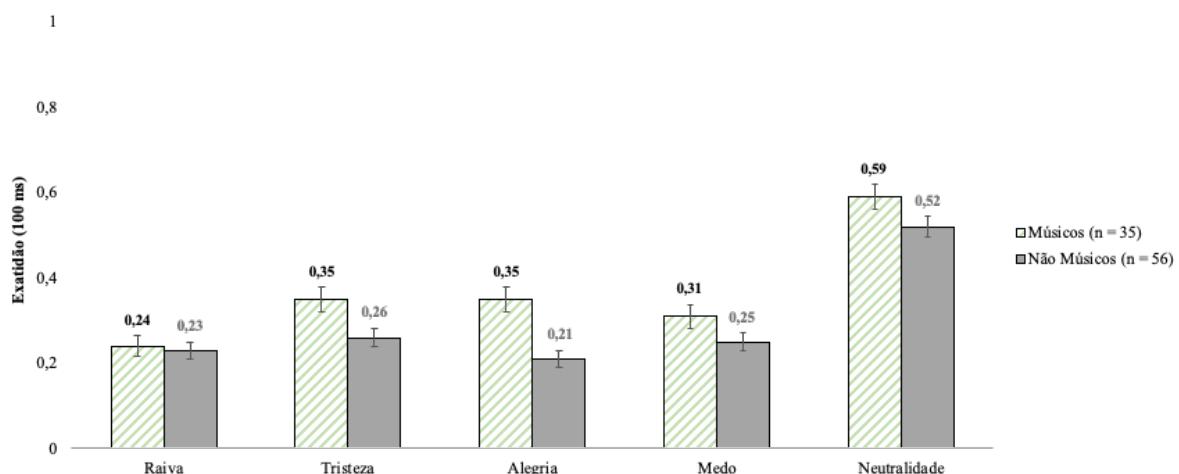
⁵ No primeiro modelo estimado, foram ainda consideradas random slopes por participante para o fator emoção. Considerando que este modelo não convergiu (o que indica a necessidade de simplificar o modelo estimado), o mesmo modelo foi novamente estimado removendo as random slopes por participante.

da vantagem do treino musical foi observada para todas as emoções, incluindo as que eram reconhecidas acima do nível do acaso nesta fase, a alegria e neutralidade.

Para as restantes emoções (raiva, medo e tristeza), que foram reconhecidas acima do nível do acaso aos 100 ms, o mesmo modelo acima descrito foi estimado considerando o hit rate no G100 (100 ms). Os resultados revelaram um efeito principal de grupo, $\chi^2(1, N = 91) = 10.09, p = .001$, revela que os músicos tiveram, de forma genérica, um desempenho superior aos não-músicos no reconhecimento de emoções vocais, a nível dos 100 ms. A interação entre grupo e emoção não foi significativa, $\chi^2(4, N = 91) = 6.49, p = 0.1$, o que sugere a vantagem associada ao treino musical foi semelhante para todas as emoções. As diferenças entre músicos e não-músicos na média de reconhecimento entre emoções está presente na Figura 3.5.

Figura 3.5.

Diferenças entre músicos e não-músicos na exatidão do reconhecimento de emoções quando a informação acústica era incompleta (desempenho no G100).



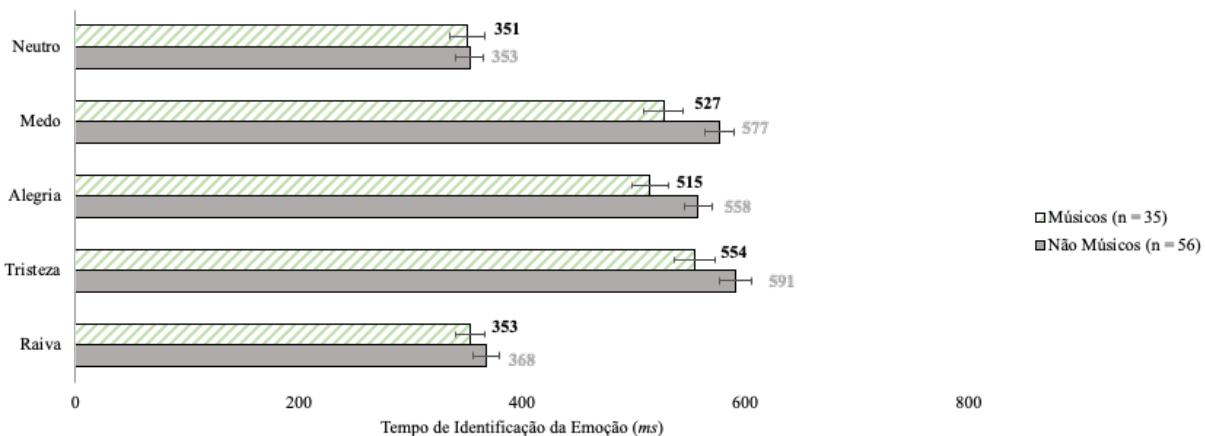
Notas. As barras de erro representadas na figura correspondem aos valores do erro-padrão.

EIPs. A Figura 3.6. apresenta os EIPs associados a cada uma das emoções, em função do grupo. Para testar a hipótese de que os músicos precisam de menos informação acústica até estabilizarem a emoção reconhecida na voz, os grupos foram comparados a nível dos EIPs, a partir de um modelo de efeitos mistos com regressão linear que considerou os EIPs para cada estímulo como variável dependente, e o grupo, emoção, e interação entre ambos como fatores fixos. O participante e estímulo foram inseridos como fatores aleatórios. Adicionalmente, foram ainda consideradas random slopes por participante em função da emoção. Os resultados não revelaram um efeito principal de grupo, $\chi^2(1, N = 91) = 1.82, p = .18$, nem um efeito de interação

entre emoção e grupo significativo, $\chi^2(4, N = 91) = 2.29, p = .68$, o que indica que não existiu uma vantagem significativa na quantidade de informação acústica associada à estabilização da categoria emocional reconhecida. O modelo revelou, no entanto, um efeito principal da emoção, o que sugere que certas emoções foram reconhecidas com base em menos informação acústica, comparativamente a outras. Comparações entre categorias emocionais revelaram que os participantes necessitaram de mais informação acústica para reconhecer tristeza (diferença estimada = 224.20, $SE = 51.90, t = 4.71, p < .001$), medo (diferença estimada = 216.8, $SE = 49.6, t = 4.37, p < .001$), e alegria (diferença estimada = 181, $SE = 47.8, t = 3.79, p = .003$), comparativamente a estímulos neutros. Também necessitaram de mais informação acústica para reconhecer raiva, no entanto, esta diferença não foi significativa ($p = 0.1$; Figura 3.2.).

Figura 3.6.

EIPs para cada emoção em função do grupo.



Notas. Valores de EIP apresentados correspondem à média de EIPs para os estímulos correspondentes a cada emoção. As barras de erro representadas na figura correspondem aos valores do erro-padrão.

Duração do treino musical. De forma a complementar os resultados obtidos para o treino musical, tentamos perceber se a duração do treino musical poderia ser uma medida mais sensível aos possíveis efeitos do treino musical, comparativamente à categorização binária entre músicos e não-músicos feita nas secções anteriores. Para esse fim, as análises focadas no treino musical foram repetidas utilizando a duração do treino musical como preditor, em vez categorização binária entre músicos e não-músicos. O número de anos de treino musical foi medido a partir do item 36 do *Goldsmiths Musical Sophistication Index* (GOLD-MSI; Lima et al., 2020), “Tive 0 / 0.5 / 1 / 2 / 3-5 / 6-9 / 10 ou mais anos de instrução formal na prática de um

instrumento musical (incluindo a voz) durante a minha vida”, medido numa escala que varia entre 1 (0 anos) e 7 (10 anos).

O primeiro modelo, explorou os efeitos da duração do treino musical no reconhecimento da expressão completa, considerando como variável dependente o hit rate face à expressão completa. Como fatores fixos foram consideradas as seguintes variáveis: emoção, a resposta do participante ao item 36 do Gold-MSI, e a interação entre ambos os fatores. O participante e o estímulo foram considerados como efeitos aleatórios⁶. Os resultados revelaram um efeito principal da duração do treino musical, $\chi^2(1, N = 107) = 9.24, p = .002$, tal como um efeito de interação entre emoção e a duração do treino musical, $\chi^2(4, N = 107) = 12.79, p = .05$, sugerindo que mais anos de treino musical estiveram positivamente associados à exatidão do reconhecimento de emoções, mas com efeitos diferentes entre emoções. Análises subsequentes dos efeitos de interação revelaram que mais anos de treino musical estiveram associados a um melhor reconhecimento de alegria (diferença estimada = 0.37, $SE = 0.12, z = 3.23, p = .001$), neutralidade (diferença estimada = 0.39, $SE = 0.16, z = 2.41, p = 0.02$), e raiva (diferença estimada = 0.21, $SE = 0.09, z = 2.28, p = .02$). No caso do medo, a duração do treino musical teve um efeito marginalmente significativo (diferença estimada = 0.12, $SE = 0.07, z = 1.86, p = .06$) No caso da tristeza, a associação entre a duração do treino musical e a exatidão do reconhecimento não foi significativa (diferença estimada = 0.07, $SE = 0.1, z = 0.75, p = .46$).

O segundo modelo foi focado no efeito da duração do treino musical no reconhecimento de emoções quando a informação acústica era incompleta (desempenho entre G50 e G1000). Foi estimado um modelo de efeitos misto com regressão linear que considerou como variável dependente a média de desempenho entre os 50 ms e os 1000 ms. Foram considerados como fatores fixos a emoção, a resposta ao item 36 do Gold-MSI, tal como a interação entre os fatores. O participante e o estímulo foram ainda considerados como efeitos aleatórios. Adicionalmente, foram ainda considerados random slopes por participante para a variável emoção. Os resultados revelaram um efeito marginal da duração do treino musical, $\chi^2(1, N = 107) = 3.61, p = .06$, o que sugere que mais anos de treino musical estiveram associados a um melhor desempenho quando a informação acústica era incompleta, apesar desta associação ser fraca. O efeito de interação entre emoção e a duração do treino musical não foi significativo, $\chi^2(4, N = 107) = 6.25, p = .18$, o que indica que esta vantagem foi geral a todas as categorias emocionais.

⁶ No primeiro modelo estimado, foram ainda consideradas random slopes por participante para o fator emoção. Considerando que este modelo não convergiu (o que indica a necessidade de simplificar o modelo estimado), o mesmo modelo foi novamente estimado removendo as random slopes por participante.

Possíveis associações entre a duração do treino musical e o curso temporal de reconhecimento de emoções vocais foi explorada num terceiro modelo focado nos EIPs. Foi estimado um modelo de efeitos mistos com regressão linear, que considerou como variável dependente os EIPs de cada participante a cada estímulo. A emoção, a resposta ao item 36 do Gold-MSI, e a interação entre os fatores foram considerados como fatores fixos. O participante e o estímulo foram considerados como fatores aleatórios, tendo sido ainda considerados random slopes por participante para a variável emoção. Os resultados não revelaram um efeito principal da duração do treino musical, $\chi^2(1, N = 107) = 1.60, p = .21$, nem um efeito de interação entre emoção e treino musical significativo, $\chi^2(4, N = 107) = 5.50, p = .24$. Estes resultados evidenciam a ausência de vantagens associadas a mais anos de treino musical na quantidade de informação acústica necessária para reconhecer emoções vocais.

Capacidades cognitivas. Considerando que comparações entre grupos revelaram diferenças significativas entre músicos e não-músicos na tarefa cognitiva (Quadro 3.1.), tal como associações significativas entre a duração do treino musical e os scores obtidos nesta medida (MarsIb; Quadro 3.2.), as análises de exatidão presentes na secção Treino Musical Formal foram repetidas considerando ainda a média obtida na tarefa cognitiva como fator adicional. O primeiro modelo, focado na exatidão face à expressão completa (hit rate), considerou o grupo, o tipo de emoção, a interação entre ambos os termos, e a média de desempenho na medida Mars-Ib como preditores. O modelo revelou um efeito marginal das capacidades cognitivas, $\chi^2(1, N = 91) = 3.34, p = .07$, no entanto, o efeito do grupo manteve-se significativo, $\chi^2(1, N = 91) = 4.20, p = .04$, tal como a interação entre emoção e grupo, $\chi^2(4, N = 91) = 12.81, p = .01$. Comparativamente ao modelo que considerou apenas a contribuição do grupo e emoção, o efeito principal do grupo ficou mais fraco ($p = .004$ vs. $p = .04$; ver sub-secção 3.5.1), mas o efeito de interação manteve-se igual de um modelo para ou outro. Tendo em conta as associações reportadas entre a duração do treino musical e os scores medida cognitiva, um segundo modelo considerou a contribuição das competências cognitivas no modelo de exatidão que considerou o treino musical enquanto variável contínua (duração do treino musical; Gold-MSI36). O modelo considerou o hit rate como variável dependente, e a emoção, duração e interação entre termos como fatores fixos. Adicionalmente, os valores médios MarsIb foram adicionados como preditor. O modelo não revelou um efeito das capacidades cognitivas, $\chi^2(1, N = 107) = 2.28, p = .13$, e o efeito principal da duração do treino musical, $\chi^2(1, N = 107) = 6.60, p = .01$, tal como de interação, $\chi^2(4, N = 107) = 12.78, p = .01$, mantiveram-se significativos. Comparativamente ao modelo que considerou apenas a duração do treino musical e emoção, o efeito principal da duração do treino musical diminui, ($p = .002$ vs. $p =$

.01), mas o efeito de interação manteve-se igual de um modelo para o outro. Em conjunto, estes resultados indicam que os efeitos do treino musical na exatidão do reconhecimento de emoções vocais se mantêm mesmo quando temos em consideração diferenças individuais nas capacidades cognitivas.

Competências de Percepção Musical: Autorrelato

Para explorar a hipótese de que capacidades de percepção musical estão associadas a um melhor reconhecimento de emoções vocais, independentemente do treino musical, as primeiras análises focaram-se nas competências de percepção musical autorrelatadas pelos participantes. Para esse fim, o score da faceta “Capacidades Percetivas” do Gold-MSI (Lima et al., 2020) foi inserido como preditor nos modelos de efeitos mistos descritos na secção Treino Musical Formal. O foco nesta faceta em particular surge de resultados prévios que reportaram que competência de percepção musical autorrelatada esteve particularmente associada ao reconhecimento de emoções vocais (Correia et al., 2020). Modelos de mediação também revelaram um efeito indireto do treino musical, a partir do papel mediador das competências de percepção musical autorrelatadas.

Um primeiro modelo foi focado na exatidão do reconhecimento, e considerou o hit rate face à expressão completa como varável dependente. A emoção, o valor médio dos participantes à faceta “Capacidades Percetivas”, e a interação entre os fatores foram considerados como fatores fixos. O participante e o estímulo foram ainda inseridos como fatores aleatórios⁷. Os resultados não mostraram um efeito principal das competências percetivas autorrelatadas na exatidão do reconhecimento, $\chi^2(1, N = 107) = 1.10, p = .29$. Também não foram reportados efeitos de interação significativos, $\chi^2(4, N = 107) = 3.92, p = .42$, o que aponta para a ausência de associação entre as capacidades percetivas autorrelatadas e o reconhecimento de emoções vocais.

O segundo modelo explorou possíveis contribuições de competências percetivas autorrelatadas no reconhecimento de emoções quando a informação acústica era incompleta (desempenho entre G50 e G1000). Foi estimado um modelo de efeitos misto com regressão linear que considerou como variável dependente a média de desempenho entre os 50 ms e o 1000 ms. Foram considerados como fatores fixos a emoção e o valor médio obtido na faceta “Capacidades Percetivas”, tal como a interação entre os dois fatores. O participante e o estímulo foram ainda considerados como fatores aleatórios. Adicionalmente, foram ainda considerados

⁷ No primeiro modelo estimado, foram ainda consideradas random slopes por participante para o fator emoção. Considerando que este modelo não convergiu (o que indica a necessidade de simplificar o modelo estimado), o mesmo modelo foi novamente estimado removendo as random slopes por participante.

random slopes por participante para a variável emoção. Os resultados revelaram um efeito principal das competências percetivas autorrelatadas, $\chi^2(1, N = 107) = 5.24, p = .02$, sugerindo que indivíduos com scores mais elevados nesta faceta tiveram um melhor desempenho no reconhecimento de emoções quando a informação acústica era incompleta. A interação entre capacidades percetivas autorrelatadas e emoção não foi significativa, $\chi^2(4, N = 107) = 3.10, p = .54$, indicando que o efeito das competências autorrelatadas foi geral e não específico a certas categorias emocionais. Para perceber se a vantagem das competências percetivas se mantinha, independentemente do treino musical, o mesmo modelo foi repetido, considerando adicionalmente a duração do treino musical como preditor. Os resultados revelaram que o efeito principal das competências percetivas se mantinha, apesar de marginalmente significativo, $\chi^2(1, N = 107) = 3.00, p = .08$, e a interação entre as competências percetivas autorrelatadas e a emoção continuou não significativa, $\chi^2(4, N = 107) = 1.98, p = .74$. Mesmo controlando pelo efeito do treino musical, as competências de percepção musical mantiveram-se associadas a uma vantagem geral no reconhecimento de emoções vocais em condições mais desafiantes (i.e., quando a informação é incompleta).

Possíveis associações entre competências de percepção musical autorrelatadas e o curso temporal de reconhecimento de emoções vocais foi ainda explorada num terceiro modelo focado nos EIPs. Foi estimado um modelo de efeitos mistos com regressão linear, que considerou como variável dependente os EIPs. A emoção, o valor médio obtido na faceta “Capacidades Percetivas”, tal como a interação ambos foram considerados como fatores fixos. O participante e o estímulo foram inseridos como efeitos aleatórios. Adicionalmente, foram incluídas random slopes por participante para a variável emoção. Os resultados revelaram um efeito marginalmente significativo de competências percetivas autorrelatadas, $\chi^2(1, N = 107) = 3.40, p = .07$, não tendo sido encontrado um efeito de interação entre as competências percetivas autorrelatadas e emoção, $\chi^2(4, N = 107) = 4.10, p = .39$. Participantes que reportaram ter melhores competências de percepção musical parecem ter necessitado de menos informação acústica para reconhecer emoções vocais (diferença estimada = -0.38, $SE = 20.61, t = -1.84, p = .07$). No entanto, quando a duração do treino musical foi inserida no modelo, o efeito das competências de percepção musical tornou-se nulo ($p = .12$).

Competências de Percepção Musical: Medidas Objetivas

Para além das capacidades percetivas autorrelatadas, os participantes completaram ainda testes objetivos de percepção musical. Associações entre o score obtido no MET e o reconhecimento de emoções vocais foram exploradas a partir dos modelos de efeitos mistos descritos na secção

Treino Musical Formal, considerando ainda a média global no MET como preditor, em vez do fator grupo.

Um primeiro modelo foi focado na exatidão do reconhecimento. Foi estimado um modelo de efeitos mistos com regressão logística, que considerou como variável dependente o hit rate face à expressão completa. A emoção, o valor médio obtido no MET, e a interação entre ambos foram considerados como fatores fixos. O participante e o estímulo foram considerados como efeitos aleatórios⁸. Os resultados revelaram um efeito do MET marginalmente significativo, $\chi^2(1, N = 107) = 2.84, p = .09$ na exatidão do reconhecimento, o que parece apontar para uma vantagem (apesar de fraca) na exatidão do reconhecimento associada a competências percetivas objetivas. O efeito de interação entre o MET e a emoção não foi significativo, $\chi^2(4, N = 107) = 5.18, p = .27$, o que indica que a vantagem das competências de percepção musical poderá ser global, e não específica à categoria emocional. No entanto, quando considerámos a duração do treino musical no modelo, este deixou de ser significativo ($p = .48$).

O segundo modelo foi focado no reconhecimento de emoções quando a informação acústica era incompleta (desempenho entre G50 e G1000). Foi estimado um modelo de efeitos mistos com regressão linear que considerou como variável dependente a média de desempenho entre o G50 e o G1000. A emoção, o valor médio obtido no MET, e a interação entre ambos, foram considerados como fatores fixos. Foram ainda incluídos random intercepts para participante e estímulo, tal como random slopes por participante para a variável emoção. Os resultados não revelaram um efeito principal do MET, $\chi^2(1, N = 107) = 1.20, p = .27$, no desempenho entre os gates 1 e 10. O efeito de interação entre o MET e a emoção também não foi significativo, $\chi^2(4, N = 107) = 1.24, p = .87$, o que demonstra que capacidades de percepção musical mais elevadas não estiveram associadas ao desempenho quando a informação acústica era incompleta.

Possíveis associações entre competências musicais objetivas e o curso temporal de reconhecimento de emoções vocais foram ainda exploradas num terceiro modelo focado nos EIPs. Foi estimado um modelo de efeitos mistos com regressão linear, que considerou como variável dependente os EIPs. A emoção, o valor médio obtido no MET, tal como a interação entre ambos foram considerados como fatores fixos. O participante e o estímulo foram considerados como fatores aleatórios. Adicionalmente, foram ainda incluídos random slopes por participante para a variável emoção. Os resultados não revelaram um efeito principal do MET, $\chi^2(1, N = 107) = 0.22, p = .64$, nos EIPs. O efeito de interação entre o MET e a emoção

⁸ No primeiro modelo estimado, foram ainda consideradas random slopes por participante para o fator emoção. Considerando que este modelo não convergiu (o que indica a necessidade de simplificar o modelo estimado), o mesmo modelo foi novamente estimado removendo as random slopes por participante.

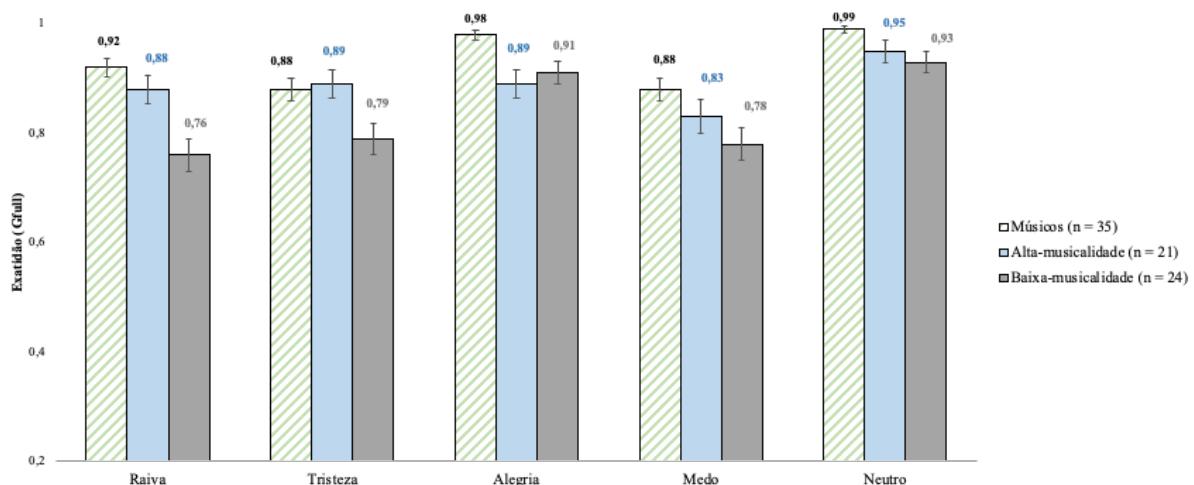
também não foi significativo, $\chi^2(4, N = 107) = 2.08, p = .72$, o que revela a ausência de associações entre capacidades musicais percetivas mais elevadas (avaliadas de forma objetiva) e o tempo necessário para reconhecer emoções vocais.

Diferentes Níveis de Competência Musical em Não-Músicos

Para explorar a hipótese de que as associações entre o treino musical e o reconhecimento de prosódia emocional seriam mais salientes quando a comparação era entre músicos e não-músicos de baixa musicalidade, comparativamente a músicos de alta musicalidade, os modelos de efeitos mistos testados na secção Treino Musical Formal foram repetidos, considerando como preditor a musicalidade dos participantes, separados em três grupos: músicos (3), não-músicos de alta musicalidade (2) e não-músicos de baixa musicalidade (1). Na Figura 3.7. são apresentadas as médias de exatidão face à expressão completa em função do nível de musicalidade dos participantes.

Figura 3.7.

Diferenças na exatidão do reconhecimento de emoções em função de diferentes níveis de musicalidade.



Nota. As barras de erro representadas na figura correspondem aos valores do erro-padrão.

O primeiro modelo considerou possíveis efeitos de diferentes níveis de competência musical no reconhecimento da expressão completa. Para esse fim, foi estimado um modelo misto com regressão logística, que considerou o hit rate face à expressão completa como variável dependente. A emoção, musicalidade, e interação entre ambos foram considerados como fatores

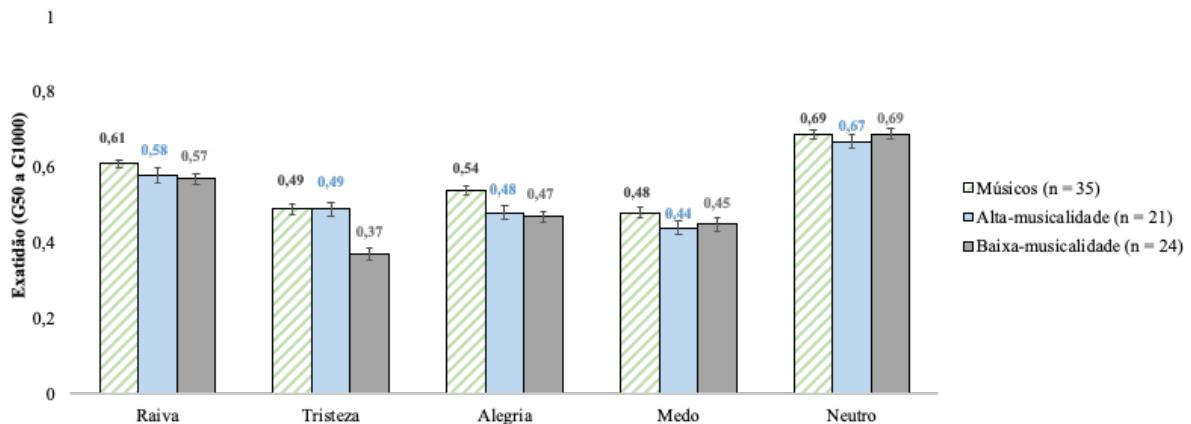
fixos. O participante e o estímulo foram considerados como efeitos aleatórios⁹. Os resultados revelaram um efeito principal da musicalidade, $\chi^2(2, N = 80) = 7.40, p = .03$, tal como um efeito de interação entre musicalidade e emoção, $\chi^2(8, N = 80) = 22.65, p = .004$, o que indica que exatidão do reconhecimento diferiu entre grupos, e que essas diferenças variaram em função da categoria emocional. Análises dos contrastes obtidos no modelo principal revelaram que os não-músicos de baixa musicalidade tiveram um desempenho significativamente pior que os músicos (diferença estimada = -1.11, $SE = 0.40, z = -2.77, p = .02$) no reconhecimento da expressão completa, enquanto o contraste entre músicos e não-músicos de alta musicalidade não foi significativo (diferença estimada = -0.69, $SE = 0.44, z = -1.58, p = .25$). Análises das interações revelaram que os não-músicos de baixa musicalidade tiveram um desempenho significativamente pior que os músicos no reconhecimento de raiva (diferença estimada = -1.42, $SE = 0.42, z = -3.07, p = .03$). Quando a comparação foi entre não-músicos de alta musicalidade e músicos, a diferença não foi significativa ($p > .1$). Para as restantes emoções não foram encontradas diferenças significativas entre grupos (todos os $ps > .1$).

Na Figura 3.8. é apresentado o desempenho quando a informação acústica era incompleta, em função da categoria emocional e do nível de musicalidade dos participantes. Relativamente a possíveis diferenças no desempenho quando a informação acústica era incompleta, foi estimado um modelo de efeitos mistos com regressão linear que considerou a média de desempenho entre o G50 e o G1000 como variável dependente, e a emoção, musicalidade, e a interação entre ambos, como preditores. Os resultados não revelaram um efeito principal da musicalidade, $\chi^2(2, N = 80) = 4.28, p = .12$, no entanto revelaram um efeito de interação significativo, $\chi^2(8, N = 80) = 42.30, p < .001$, sugerindo que as diferenças entre grupos foram específicas a categorias emocionais, e não gerais. Análises das interações apontam para diferenças no caso da tristeza, no qual não-músicos de baixa musicalidade tiveram um desempenho significativamente pior que músicos quando expostos a menos informação acústica (diferença estimada = -0.12; $SE = 0.03; t = -3.99; p = .001$). A diferença entre músicos e não-músicos de alta musicalidade não foi significativa ($p > 0.1$).

⁹ No primeiro modelo estimado, foram ainda consideradas random slopes por participante para o fator emoção. Considerando que este modelo não convergiu (o que indica a necessidade de simplificar o modelo estimado), o mesmo modelo foi novamente estimado removendo as random slopes por participante.

Figura 3.8.

Diferenças entre músicos, não-músicos de alta-musicalidade e não-músicos de baixa musicalidade na exatidão do reconhecimento de emoções quando a informação acústica era incompleta (desempenho entre G50 e G1000).



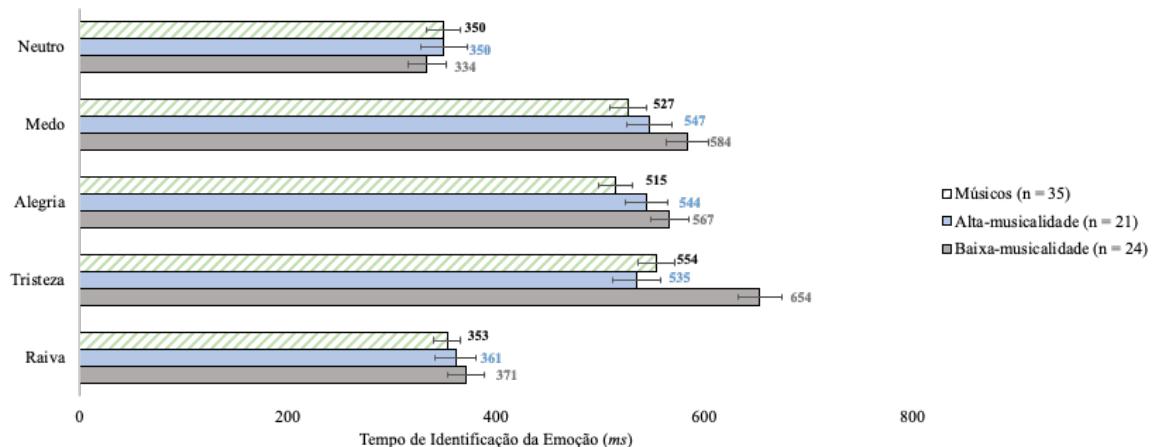
Nota. As barras de erro representadas na figura correspondem aos valores do erro-padrão.

Na Figura 3.9. são apresentados os EIPs para cada emoção em função do nível de musicalidade dos participantes. A possibilidade de que a quantidade de informação acústica necessária para reconhecer emoções vocais poderá diferir entre participantes com diferentes níveis de competência musical foi explorada a partir de um modelo de efeitos mistos com regressão linear. O modelo considerou os EIPs como variável dependente, e a emoção e musicalidade como fatores fixos. O participante e estímulo foram ainda considerados como fatores aleatórios¹⁰. Os resultados não revelaram um efeito principal da musicalidade, $\chi^2(2, N = 80) = 2.19, p = .33$, nem um efeito de interação significativo, $\chi^2(8, N = 80) = 3.50, p = .90$, o que aponta para a ausência de vantagens na quantidade de informação acústica necessária para reconhecer emoções vocais associada a diferentes níveis de musicalidade.

¹⁰ O primeiro modelo máximo (que incluiu random slopes por participante) estimado para os EIPs não convergiu, logo o modelo apresentado considerou apenas efeitos aleatórios de participante e estímulo.

Figura 3.9.

EIPs para cada emoção em função do nível de musicalidade dos participantes.



Notas. Valores de EIP apresentados correspondem à média de EIPs para os estímulos correspondentes a cada emoção. As barras de erro representadas na figura correspondem aos valores do erro-padrão.

Capítulo IV. Discussão

A presente investigação explorou associações entre treino musical, capacidades musicais, e o processamento temporal de emoções vocais. Especificamente, examinamos se músicos apresentam vantagens em comparação com não-músicos no reconhecimento temporal de expressões vocais de alegria, raiva, medo, tristeza, e neutralidade; e se indivíduos sem treino musical, mas com boas competências musicais apresentam as mesmas vantagens que os músicos a nível do reconhecimento de emoções vocais. Foi aplicado um paradigma de gating auditivo para aceder ao aspetto temporal da fala, no qual os estímulos correspondentes a cada emoção foram decompostos em 11 segmentos de diferentes durações, e apresentados aos participantes forma cumulativa.

São apresentados cinco resultados principais. Primeiro, os resultados globais obtidos na tarefa de gating replicam resultados anteriores, revelando diferenças no curso do reconhecimento temporal entre emoções vocais. Também foram encontradas associações entre o treino musical e o reconhecimento global ao longo dos gates. Em segundo, observamos uma vantagem associada ao treino musical no reconhecimento da expressão completa (GFull), tal como associações entre ter mais anos de treino musical e a exatidão no reconhecimento da expressão completa. Terceiro, os músicos tiveram um desempenho global superior aos não-músicos quando as condições de escuta eram mais desafiantes, i.e., quando a informação acústica era apresentada apenas de forma parcial (média de desempenho entre gates 50 ms e 1000 ms), e quando expostos a apenas 100 ms da expressão vocal. No entanto, não foram observadas diferenças a nível dos EIPs. Quarto, foram encontradas associações entre competências de percepção musical e o reconhecimento de prosódia emocional, mas estas foram fracas e não significativas em alguns casos. Quinto, quando comparámos os grupos em função do nível de musicalidade dos não-músicos, os não-músicos com boas competências musicais tiveram um desempenho tão bom quanto os músicos na exatidão do reconhecimento, tal como no desempenho em condições mais desafiantes.

De modo geral, os resultados na tarefa de gating são consistentes com a literatura anterior que aplicou este paradigma no contexto do curso temporal de reconhecimento de emoções vocais noutras línguas (e.g., Jiang et al., 2015; Nordström & Laukka, 2019; Pell & Kotz, 2011). Como referido por Pell e Kotz (2011), o paradigma de gating é uma medida que permite obter dois tipos de indicadores de reconhecimento, permitindo obter informação qualitativa, acerca da exatidão do reconhecimento, e informação quantitativa, referente à quantidade de variação acústica necessária para atingir diferentes níveis de exatidão, e eventualmente, estabilizar a

resposta (expresso pelos EIPs). A nível qualitativo, a exatidão do reconhecimento de emoções foi aumentando à medida que os participantes foram expostos a mais informação acústica. Os resultados revelaram ainda que reconhecimento global não foi igual para todas as emoções. Especificamente, neutralidade e raiva parecem ter sido reconhecidas com maior exatidão ao longo dos gates. Também foram observados níveis elevados de exatidão para a alegria, mas em durações da gate mais longas (a partir dos 400 ms até à expressão completa). Os participantes foram, de forma geral, capazes de reconhecer todas as emoções acima do nível do acaso a nível dos 50 ms e 100 ms. Quando expostos a 50 ms do estímulo completo, os participantes já foram capazes de identificar neutralidade e alegria acima do acaso. Aos 100 ms, reconheceram também as restantes emoções (raiva, tristeza e medo), acima do nível do acaso. Estes resultados replicam literatura anterior indicativa de que emoções básicas expressas em prosódia e música são reconhecidas com base em menos de 250 ms (Nordström & Laukka, 2019). Especificamente, estes autores observaram que os participantes foram capazes de reconhecer raiva, tristeza, e neutralidade acima do nível do acaso a partir dos 50 ms, e medo e alegria a partir dos 100 ms. Apesar do padrão de reconhecimento de emoções variar entre estudos, as mesmas emoções foram reconhecidas com base em 100 ms do estímulo. Estes resultados apoiam a ideia de que a comunicação emocional (a partir de prosódia emocional), é um processo extremamente rápido e eficiente, que permite a diferenciação entre emoções com base na análise das características sensoriais de baixo nível, que são rapidamente expressas (e.g., intensidade e frequência da voz) e que estão imediatamente disponíveis, i.e., que não necessitam de se desenvolver até fases avançadas do estímulo (Nordström & Laukka, 2019).

A nível quantitativo, os EIPs foram mais curtos para neutralidade (352 ms), raiva (362 ms), e alegria (540 ms), seguidos de medo (557 ms), e tristeza (576 ms), sugerindo que participantes necessitaram de menos informação acústica para reconhecer neutralidade e raiva. Comparações dos presentes EIPs com os reportados anteriormente (e.g., Cornew et al., 2009; Grichkovstova et al., 2007; Jiang et al., 2015; Pell & Kotz, 2011), revelam que a tristeza e medo foram reconhecidos em janelas temporais semelhantes a estudos anteriores (i.e., tristeza entre 575–622 ms; medo entre 447–621 ms), no entanto, alegria foi reconhecida mais rapidamente relativamente a literatura anterior (entre 562 ms–1634 ms). Crucialmente, os participantes necessitaram de menos informação acústica para reconhecer alegria, comparativamente às emoções negativas (à exceção de raiva), o que parece divergir da ideia de que apresentamos vantagens na descodificação de prosódia negativa (e.g., Pell & Kotz, 2011). Estes resultados poderão dever-se a diferenças metodológicas entre estudos. A literatura de gating muitas vezes divide os estímulos em gates com base no número de sílabas (Pell & Kotz,

2011; Rigoulot et al., 2013) ou recorrendo a incrementos de gate mais longos do que aqueles aplicados neste estudo (e.g., incrementos de 250 ms em Cornew et al., 2009). Visto que o controlo da informação acústica foi mais minucioso neste estudo, i.e., os gates começaram a ser segmentados a nível dos 50 ms após o início do estímulo, isto poderá ter permitido uma análise mais detalhada dos estímulos de alegria, o que resultou num EIP menor para esta emoção. Adicionalmente, o medo e a tristeza poderão ser mais difíceis de reconhecer a partir de excertos tão breves, o que poderá ter aumentado o tempo de estabilização da resposta. Estes resultados também foram verificados em Nordström e Laukka (2019), no qual foram aplicados critérios de construção de gates semelhantes (incrementos de 50 ms). Foram observados EIPs mais curtos para expressões prosódicas de alegria (562 ms) comparativamente a tristeza e medo (575 ms e 1375 ms, respetivamente). No entanto, é de notar que existe muita variabilidade a nível das metodologias aplicadas em estudos de gating, quer a nível dos estímulos utilizados (e.g., presença ou ausência de conteúdo semântico, nuance emocional), tal como nos critérios utilizados para dividir dos gates (e.g., em função do número de sílabas ou incrementos de tempo), o que poderá afetar o tempo necessário para reconhecer emoções vocais, e tornar difícil definir um EIP “geral” para as diferentes emoções.

Para explorar os efeitos de treino musical, o nosso ponto de interesse principal, começamos por tentar estabelecer um ponto de comparação com investigação anterior, pelo que as análises de exatidão focaram-se, inicialmente, na resposta dos participantes à expressão completa (GFull). Os resultados revelaram um efeito principal de grupo, o que sugere que os músicos foram, de forma geral, melhores que os não-músicos a reconhecer emoções vocais. Estes resultados corroboram investigação anterior que mostrou que os músicos apresentam vantagens no reconhecimento de emoções prosódicas (e.g., Correia et al., 2020; Farmer et al., 2020; Lima & Castro, 2011). Os músicos foram particularmente melhores que os seus pares no reconhecimento de expressões completas de neutralidade, raiva e alegria. Estes resultados poderão dever-se ao facto de a alegria e neutralidade serem distintivas e a nível de altura tonal encontrarem nos pólos extremos do contínuo (pitch), como expresso pela frequência fundamental destas emoções (Pinheiro et al., 2015). As expressões prosódicas de alegria apresentam a média de F0 mais elevada (ver secção de Métodos; Quadro 2.1.) enquanto as expressões de neutralidade se encontram entre as emoções com F0 mais baixas (ver secção de Métodos; Quadro 2.1.), o que poderá tornar estas emoções mais salientes a nível percepção, tornando-as mais fáceis de identificar, comparativamente às restantes emoções vocais. Considerando a hipótese de que os músicos são mais eficientes a captar pistas acústicas de baixo nível (e.g., Martins et al., 2021; Strait et al., 2009), como é o caso da altura tonal, estes poderão

ter sido mais eficientes a distinguir estas emoções das restantes. Adicionalmente, o treino musical parece estar especificamente associado a diferenças no processamento de altura tonal na fala (e.g., Marques et al., 2007). Curiosamente, não foram encontradas diferenças entre grupos no reconhecimento de tristeza. Tendo em conta que esta emoção apresenta a média de F0 mais baixa, faria sentido que os músicos fossem mais precisos a distinguir esta emoção das restantes. Na verdade, a ausência de diferenças entre músicos e não-músicos no reconhecimento de explícito de tristeza, já foi verificado anteriormente (Lima & Castro, 2011).

Relativamente à raiva, estes resultados são consistentes com estudos anteriores que observaram vantagens associadas à expertise musical no reconhecimento explícito de raiva (Pinheiro et al., 2015; Thompson et al., 2004). Foram também observadas associações positivas entre a duração do treino musical e a exatidão do reconhecimento, sugerindo que mais anos de treino musical contribuíram para um melhor reconhecimento de prosódia emocional, tal como observado anteriormente para emoções vocais (e.g., Correia et al., 2020), e musicais (Livingstone et al., 2010). Este resultado apoia a associação entre treino formal e melhor desempenho nesta competência.

O possível contributo de fatores de domínio geral (capacidades cognitivas) para as associações encontradas foi considerado nas análises, tendo em conta investigação indicativa de que o treino musical está correlacionado com capacidades cognitivas de modo geral (e.g., Swaminathan & Schellenberg, 2018). Logo, torna-se importante perceber se as vantagens observadas nos músicos se devem a um processo de transferência entre domínios musicais e emocionais, ou ao facto dos participantes com treino musical apresentarem um funcionamento cognitivo superior a participantes sem treino musical. Adicionalmente, o presente estudo observou associações positivas entre os valores obtidos na tarefa cognitiva e o treino musical, tal como diferenças significativas entre grupos nesta medida. Estudos anteriores debruçados sobre a mesma temática também observaram associações positivas entre o treino musical e o desempenho cognitivo (e.g., Correia et al., 2020; Lima & Castro, 2011; Thompson et al., 2004). No entanto, os resultados não foram explicados por diferenças em fatores gerais. Apesar de terem sido encontradas associações significativas entre a duração do treino musical e os resultados obtidos na medida cognitiva, tal como diferenças entre grupos (músicos tiveram, em média, um desempenho superior aos não-músicos), quando este fator foi adicionado aos modelos focados no treino musical, o efeito da expertise musical manteve-se significativo. Este resultado é consistente com investigação anterior (e.g., Lima & Castro, 2011; Thompson et al., 2004). Desse modo, os resultados obtidos não parecem ser um artefacto de vantagens cognitivas gerais (Martins et al., 2021).

Uma das questões centrais desta investigação prendeu-se com a possibilidade de existirem diferenças entre músicos e não-músicos a nível do curso temporal de emoções vocais. Para esse fim, as análises focaram-se por um lado, no desempenho dos participantes em condições de escuta mais desafiantes (i.e., quando a informação acústica era incompleta), e por outro, na quantidade de informação acústica que os participantes necessitaram até estabilizarem o reconhecimento das diferentes emoções vocais (i.e., EIPs). Um resultado novo revela que o treino musical esteve associado a um melhor reconhecimento de prosódia emocional em condições de escuta mais desafiantes. Especificamente, modelos focados no desempenho no primeiro gate em que a emoção é reconhecida acima do acaso, indicam que os músicos foram, de modo geral, melhores a reconhecer emoções quando expostos a 100 ms da expressão emocional, comparativamente a não-músicos. Com acesso a 100 ms da expressão acústica, os músicos foram capazes de extrair as principais pistas acústicas relevantes para percecionar saliência emocional, tal como diferenciar categorias emocionais, e reconhecê-las explicitamente. Estes resultados complementam investigação anterior que explorou diferenças entre músicos e não-músicos aplicando medidas implícitas de atividade cerebral (Pinheiro et al., 2015; Strait et al., 2009). Strait et al. (2009) observou que os músicos apresentam uma codificação subcortical mais eficiente de choro quando comparados a não-músicos. Já Pinheiro et al. (2015) reportaram modulações nas respostas neurais de músicos a prosódia emocional a nível dos 50 ms e 100 ms após o início do estímulo. Estas janelas temporais estão associadas a estádios precoces de processamento de emoções vocais, que englobam a extração e análise de pistas acústicas básicas (e.g., Nussbaum & Schweinberger, 2021; Schirmer & Kotz, 2006). Os músicos apresentaram ainda vantagens sobre os não-músicos quando a informação acústica é incompleta, como expresso melhor desempenho dos músicos na exatidão entre o G50 e G1000. Esta vantagem foi geral para o reconhecimento de todas as emoções.

Relativamente aos EIPs, a hipótese de que os músicos poderiam necessitar de menos informação acústica para reconhecer emoções vocais não foi verificada. Não foram encontradas diferenças significativas entre músicos e não-músicos a nível da quantidade de informação percetiva que necessitaram para reconhecer emoções vocais. Visto que os músicos foram mais precisos a reconhecer emoções com base em informação acústica parcial, seria esperado que também fossem mais “rápidos” a reconhecer as emoções vocais. O EIP é na verdade um indicador do gate em que os participantes estabilizam o significado emocional, ou seja, indicam o gate a partir do qual adicionar mais informação percetiva (gates de durações mais longas) já não contribui para um melhor reconhecimento da emoção. Por esta razão, este indicador seria indicativo de um processo mais global de integração de pistas acústicas. O facto de os músicos

terem sido mais exatos a reconhecer emoções vocais quando a informação acústica era parcial sugere que a vantagem observada nos músicos pode dever-se a uma análise mais fina do sinal acústico nas várias fases de processamento, mais do que a processos integrativos de alto nível. À medida que recebem mais informação acústica, os músicos não estabilizam a sua resposta, pois continuam a beneficiar da quantidade cumulativa de informação acústica que vão recebendo. Indivíduos com treino musical poderão ser capazes de captar mais nuances no sinal acústico, o que auxiliaria numa diferenciação cada vez melhor das emoções entre si, e culminaria num nível de exatidão superior aos seus pares quando chegam à expressão completa.

De forma a complementar a evidência de efeitos de expertise musical quando comparamos grupos com ou sem treino musical, o presente estudo mediu ainda a variação interindividual nas competências musicais da amostra global, de forma a perceber se diferenças no nível de musicalidade dos participantes (mesmo na ausência de treino musical), estão associadas a variabilidade no reconhecimento de prosódia emocional. Esta análise foi feita de duas formas: primeiro considerando o peso individual de competências de percepção musical objetivas (medidas a partir de testes de performance—MET) e autorrelatadas (Gold-MSI); e segundo, com base numa medida de “musicalidade” agregada, construída a partir dos scores obtidos no Gold-MSI e MET. Escolhemos estas medidas (Gold-MSI e MET) considerando o facto de que têm sido amplamente utilizadas para avaliar diferenças e competências musicais (e.g., Correia et al., 2020 ,2021; Müllensiefen et al., 2014; Wallentin et al., 2010). Com base nestas análises, verificámos que as associações entre competências de percepção musical e o reconhecimento de prosódia emocional foram fracas. Uma vantagem geral de competências de percepção musical autorrelatadas emergiu para o desempenho apenas quando a informação acústica era incompleta. O efeito manteve-se mesmo considerando a duração do treino musical. Um efeito marginal de competências de percepção musical autorrelatadas também se verificou a nível dos EIPs, no entanto, ao considerar a duração do treino musical, este efeito desapareceu. No caso das medidas de percepção musical objetivas (captadas a partir do MET), foi encontrado um efeito marginal no reconhecimento da expressão completa, no entanto, quando o treino musical foi incluído no modelo, o efeito deixou de ser significativo. Logo os resultados são, de modo geral, inconclusivos relativamente ao papel das competências percetivas, quando considerando as medidas de autorrelato e objetivas de forma individual. Estes resultados não replicam os relatados por Correia et al. (2020), no qual as competências de percepção musical explicaram a associação entre o treino musical e o reconhecimento de prosódia emocional e vocalizações verbais. São propostas duas explicações possíveis para as diferenças de resultados entre estudos. Por um lado, podemos atribuir esta discrepança a diferenças metodológicas.

Comparativamente a Correia et al. (2020), o tamanho amostra do presente estudo foi inferior ($N = 107$ vs. 169), o que poderá ter contribuído para efeitos menos robustos. Adicionalmente, alguns participantes foram eliminados por não cumprirem os critérios de inclusão no MET (14 respostas no subteste de melodia, e 13 respostas no subteste de ritmo), o que diminuiu ainda mais a amostra em análise. Para além disso, os testes de percepção musical aplicados em Correia et al. (2020) acederam a mais capacidades de percepção musical (quatro tarefas musicais que acederam à percepção de batida musical, memória melódica, discriminação de atura tonal, e discriminação de duração), enquanto o teste aplicado na presente investigação acedeu apenas a aspectos de percepção melódica e rítmica. Apesar do MET ser um bom indicador online de habilidade musical (Correia et al., 2021), poderá não ter sido tão eficaz na captação de diferenças individuais na competência musical. Adicionalmente, a modalidade de recolha de dados também diferiu entre estudos: laboratorial/presencial em Correia et al. (2020), e online no presente estudo. No contexto online é mais difícil exercer um controlo rigoroso das condições experimentais (e.g., qualidade do som, ruído), um aspeto particularmente relevante para o sucesso das tarefas auditivas. Por outro lado, este resultado poderá ser evidência de nuances no reconhecimento de prosódia emocional associados ao treino musical. O treino musical pareceu ser um melhor preditor para o desempenho nesta tarefa, comparativamente às competências percetivas (quando consideradas de forma isolada). Como discutido em Correia et al. (2020), a apresentação da expressão completa ao participante poderá não ser uma tarefa sensível o suficiente para captar possíveis vantagens “adicionais” causadas por experiência musical mais extensa. A tarefa de gating poderá constituir, nesse sentido, uma tarefa mais sensível por ser mais complexa, o que também explicaria porque é que a associação com competências musicais só emergiu nas condições de informação incompleta.

Finalmente, com base na medida de musicalidade agregada, os participantes não-músicos foram divididos em grupos de baixa e alta musicalidade, e o seu desempenho no reconhecimento de emoções vocais foi comparado ao desempenho dos músicos. Os resultados apontaram para diferenças significativas entre grupos na exatidão do reconhecimento face à expressão completa. Especificamente, os participantes com baixa musicalidade tiveram um desempenho significativamente pior que os músicos, no entanto, comparações entre músicos e não-músicos de alta musicalidade revelaram que o desempenho de ambos os grupos foi semelhante. Os não-músicos de baixa musicalidade foram particularmente piores a reconhecer raiva comparativamente às restantes emoções. Os participantes com baixa musicalidade tiveram mais dificuldade no reconhecimento de tristeza quando a informação acústica era incompleta, relativamente aos músicos. Não foram verificadas diferenças significativas entre músicos e não-

músicos de alta musicalidade a nível do desempenho face a menos informação acústica. No que toca aos EIPs não foram detetadas diferenças em função do nível de musicalidade dos participantes. Em conjunto, estes resultados revelam que quando consideramos o nível de competência musical da amostra, participantes com capacidades musicais “naturalmente” elevadas parecem ser tão bons quanto os músicos no reconhecimento de prosódia emocional. Estes resultados replicam resultados anteriores, nos quais músicos e não-músicos com elevadas capacidades musicais tiveram um desempenho semelhante no reconhecimento de prosódia emocional e vocalizações não verbais (Correia et al., 2020). Também complementam investigação anterior, que observou associações entre capacidades de percepção musical e percepção de fala, mesmo considerando o treino musical (Mankel & Bidelman, 2018; Swaminathan & Schellenberg, 2017, 2019). Adicionalmente, a presente investigação estende estes resultados, pois verificou que não-músicos com boas capacidades musicais apresentam não só vantagens a nível do reconhecimento da expressão completa, mas também quando são expostos a menos informação acústica, ou seja, quando as condições de escuta são mais desafiantes. Estes resultados dão suporte à hipótese de que existe uma sobreposição das vias neurocognitivas que processam emoções na música e voz e realçam o papel que fatores pré-existentes e de experiência informal com a música, para além do treino musical, poderão ter nas associações reportadas entre musicalidade e outras competências não musicais.

Considerando que a percepção de emoções vocais é um pré-requisito para uma comunicação bem-sucedida, é importante perceber que diferenças individuais poderão estar associadas a melhorias nestas competências. O presente estudo contribui para esta questão explorando se diferenças interindividuais nas capacidades musicais estavam associadas a melhores capacidades de percepção de prosódia emocional, tal como a um reconhecimento mais rápido (reconhecimento com base em menos informação acústica) da mesma. Os resultados encontrados realçam a importância que a música e a educação musical poderão ter para o desenvolvimento pessoal e qualidade de vida. Para além disso, se as vias neurocognitivas para o processamento de música e emoções vocais de facto se cruzam, é possível que o treino musical tenha de facto um papel benéfico para indivíduos com défices a nível de percepção de emoções vocais. Três estudos longitudinais exploraram o impacto do treino musical no reconhecimento de emoções vocais em indivíduos com implantes cocleares (Chari et al., 2019; Fuller et al., 2018; Good et al., 2017), mas os resultados foram nulos. Estes resultados não confirmam a ausência de relação, continuando a se necessária mais evidência longitudinal que informe esta questão. Esta investigação veio ainda contribuir para a literatura debruçada sobre associações entre competências musicais e não musicais (neste caso, emocionais), de duas

formas. Em primeiro lugar, considerou a natureza dinâmica da voz, aplicando uma tarefa comportamental mais sensível aos aspectos temporais da fala. A informação acústica que nos permite percecionar emoções na voz não está imediatamente disponível, e vai sendo integrada à medida que o indivíduo é exposto a mais informação acústica, com evidência de janelas temporais diferentes para o reconhecimento de emoções básicas (Nordström & Laukka, 2019; Pell & Kotz, 2011). Até à data, modulações no curso temporal do reconhecimento de emoções vocais associadas à expertise musical foram exploradas a partir de medidas implícitas de atividade cerebral (e.g., Pinheiro et al., 2015), sendo pela primeira vez explorado com um indicador comportamental neste projeto. Em segundo lugar, dá resposta a uma lacuna na literatura que tem vindo a ser alvo de discussão recentemente (e.g., Correia et al., 2020; Martins et al., 2021) e que diz respeito à necessidade de considerar diferentes formas a partir das quais os indivíduos se podem envolver com música, e aplicar medidas que permitam identificar indivíduos com boas competências musicais, em desenhos transversais. Desta forma, é possível aceder ao papel de fatores preexistentes e compreender a natureza da associação entre música e emoção.

É importante referir algumas limitações que poderão ter contribuído para os resultados observados. O presente estudo foi realizado online, e apesar de terem sido tomadas medidas para assegurar o rigor experimental (e.g., a experiência foi acompanhada pelo investigador; participantes deveriam ter a câmara ligada), não é possível assegurar que todos os participantes realizaram a experiência sob as mesmas condições (e.g., qualidade/volume dos auscultadores poderá ter variado; participantes poderão ter recebido notificações de apps/sites durante a realização da experiência). Para além disso, existiram ainda problemas técnicos causados por problemas de rede/internet que dificultaram a experiência de alguns participantes. Relativamente a efeitos de treino musical, a duração do treino musical foi medida com base em escalões (e.g., 6–9 anos; 10 ou mais anos) e não como uma variável contínua (número de anos concreto), um lapso que só detetamos posteriormente à recolha de dados. Apesar de terem sido observadas vantagens do treino musical considerando a duração do treino musical desta forma, uma medida contínua poderia ter sido um indicador mais sensível de vantagens associadas à expertise musical.

Para estudos futuros, seria interessante explorar esta temática aliando o gating a um método de resposta diferente (sem ser escolha forçada). A metodologia de escolha forçada tem sido consistentemente utilizada em estudos de gating (e.g., Jiang et al., 2015; Pell & Kotz; Nordström & Laukka, 2019), mas este método é alvo de crítica por criar valores de reconhecimento inflacionados (pois os ouvintes podem “adivinar” as respostas a partir das

opções dadas) e a validade ecológica desta metodologia também é baixa (não se assemelha à forma como avaliamos emoções na vida real; e.g., Frank & Stennett, 2001; Russell, 1994). Apesar destas lacunas, continua a ser a medida geralmente aplicada pois existe consenso na sua eficácia em captar diferenças no reconhecimento de informação emocional na voz (e.g., Juslin & Laukka, 2003), e do ponto de vista experimental, é uma tarefa mais fácil de aplicar. No entanto, seria informativo explorar o reconhecimento de emoções vocais recorrendo a uma medida de captação contínua. Por exemplo, Schaerlaeken e Grandjean (2018) aplicaram um paradigma de gating onde os participantes poderiam selecionar várias emoções para categorizar o estímulo, e avaliar a sua adequabilidade numa escala contínua que variava entre 1 e 100. Esta metodologia representa uma forma alternativa à escolha forçada, e permite análises mais detalhadas do reconhecimento de emoções. Adicionalmente, literatura mostra que os indivíduos podem inferir dimensões de valência e *arousal* a partir da voz e música, além de categorias emocionais específicas (e.g., Laukka et al., 2005), tal como aspectos relacionados com a novidade e urgência a partir de expressões vocais (e.g., Nordström et al., 2017). Seria interessante aliar esta metodologia a formas alternativas de conceptualizar emoções, de forma a perceber diferenças a nível do curso temporal de reconhecimento de emoções nestas modalidades, tal como possíveis modulações associadas à sensibilidade musical.

Considerando estudos anteriores que reportaram que as vantagens da musicalidade se estendem a outros tipos de vocalização emocionais (vocalizações não verbais; Correia et al., 2020), seria útil aplicar uma medida de gating para perceber se existem diferenças associadas à expertise musical e competências musicais elevadas no reconhecimento de diferentes tipos de estímulos emocionais vocais. Explorar esta questão com um indicador comportamental sensível a aspectos temporais da fala, é particularmente pertinente, tendo em conta investigação anterior que aponta para preferências na descodificação de vocalizações não verbais em relação a prosódia emocional, com base em medidas implícitas (Pell et al., 2015). Adicionalmente, no contexto da expertise musical, já foram observadas modulações a nível subcortical para o processamento de choro (Strait et al., 2009). Por último, medidas comportamentais poderão não ser tão sensíveis a diferenças mais finas no reconhecimento de prosódia emocional associados a treino musical. Logo, complementar os resultados de gating com indicadores de atividade cerebral (EEG ou fMRI), ofereceria uma visão mais clara da natureza destas associações.

Para concluir, a presente investigação apresenta evidência que apoia a noção de que a fala e a música estão interligadas a nível das emoções: os músicos são melhores a reconhecer prosódia emocional, comparativamente a não-músicos. Ao aplicar uma medida de gating, foi possível perceber que os músicos apresentam ainda vantagens nos diferentes estádios de

processamento temporal de prosódia emocional, sendo mais precisos a extrair pistas acústicas relevantes para identificar diferentes emoções vocais quando têm acesso a informação acústica incompleta, tal como quando ouvem o estímulo completo. No entanto, esta vantagem não se reflete na “rapidez” do reconhecimento, considerando a ausência de diferenças entre grupos nos EIPs. Associações entre competências de percepção musical e o reconhecimento de prosódia emocional foram menos claras, e em alguns casos, desapareceram quando o contributo do treino musical foi considerado. Estes resultados falharam em replicar evidência anterior que observou que as competências de percepção musical tiveram um papel mediador na associação entre o treino musical e a exatidão do reconhecimento de emoções vocais (Correia et al., 2020). Quando os não-músicos foram categorizados em grupos de alta musicalidade e baixa musicalidade com base numa medida de musicalidade que agregou os valores médios obtidos nas medidas de competência musical, os resultados revelaram um padrão diferente. Nomeadamente, melhores competências musicais estiveram associadas a um reconhecimento de prosódia emocional mais preciso, mesmo na ausência de treino musical.

Estes resultados sugerem que músicos e indivíduos com boas competências musicais apresentam uma vantagem a nível acústico-percetivo, tornando-se mais precisos a reconhecer emoções à medida que são expostos a mais informação acústica. Comparativamente a não-músicos, os músicos não estabilizam as suas respostas mais cedo (como evidenciado pela ausência de diferenças nos EIPs) pois beneficiam sempre de mais informação acústica para fazer julgamentos emocionais. A associação fraca entre as competências percetivas e o reconhecimento de emoções poderá dever-se a diferenças metodológicas entre estudos, ou, alternativamente, pode ser indicativo de nuances no reconhecimento associados ao treino musical, no entanto é necessária mais investigação para informar esta questão. Em conjunto, estes resultados enfatizam a necessidade de considerar a natureza multifacetada da musicalidade.

□

Referências

- Anikin, A., & Lima, C. F. (2018). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *The Quarterly Journal of Experimental Psychology*, 71(3), 622–641. <https://doi.org/10.1080/17470218.2016.1270976>
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>
- Bhatara, A., Laukka, P., & Levitin, D. J. (2014). Expression of emotion in music and vocal communication: Introduction to the research topic. *Frontiers in Psychology*, 5, 399. <https://doi.org/10.3389/fpsyg.2014.00399>
- Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America*, 137(1), 378–387 <https://doi.org/10.1121/1.4904537>
- Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer [Computer program]. <http://www.praat.org/>
- Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, 41(2), 259–268. <https://doi.org/10.1111/j.1469-8986.2003.00142.x>
- Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2009). Cross-modal emotional attention: Emotional voices modulate early stages of visual processing. *Journal of Cognitive Neuroscience*, 21(9), 1670–1679. <https://doi.org/10.1162/jocn.2009.21110>
- Castiajo, P., & Pinheiro, A. P. (2019). Decoding emotions from nonverbal vocalizations: How much voice signal is enough? *Motivation and Emotion*, 43, 803–813. <https://doi.org/10.1007/s11031-019-09783-9>
- 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>
- Chari, D. A., Barrett, K. C., Patel, A. D., Colgrove, T. R., Jiradejvong, P., Jacobs, L. Y., & Limb, C. J. (2019). Impact of auditory-motor musical training on melodic pattern recognition in cochlear implant users. *Otology & Neurotology*, 41(4), 422–431. <https://doi.org/10.1097/MAO.0000000000002525>
- Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S. J. (2019). The matrix reasoning item bank (MaRs-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*, 6(10), 190232. <https://doi.org/10.1098/rsos.190232>
- Chronaki, G., Wigelsworth, M., Pell, M. D., & Kotz, S. A. (2018). The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Scientific Reports*, 8(1), 8659. <https://doi.org/10.1038/s41598-018-26889-1>
- Coffey, E., Chepesiuk, A., Herholz, S. C., Baillet, S., & Zatorre, R. J. (2017). Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Frontiers in Neuroscience*, 11, 479. <https://doi.org/10.3389/fnins.2017.00479>
- Coffey, E., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, 352, 49–69. <https://doi.org/10.1016/j.heares.2017.02.006>

- Cornew, L., Carver, L., & Love, T. (2009). There's more to emotion than meets the eye: A processing bias for neutral content in the domain of emotional prosody. *Cognition & Emotion*, 24(7), 1133–1152. <https://doi.org/10.1080/02699930903247492>
- Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2020). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*. Advance online publication. <https://doi.org/10.1037/emo0000770>
- Correia, A. I., Vincenzi, M., Vanzella, P., Pinheiro, A. P., Lima, C. F., & Schellenberg, E. G. (2021). Can musical ability be tested online? *Behavior Research Methods*. Advance online publication. <https://doi.org/10.3758/s13428-021-01641-2>
- Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164(3875), 86–88. <https://doi.org/10.1126/science.164.3875.86>
- Eyben, F., Scherer, K., Schuller, B., Sundberg, J., André, E., Busso, C., Devillers, L., Epps, J., Laukka, P., Narayanan, S., & Truong, K. P. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2), 190–202. <https://doi.org/10.1109/TAFFC.2015.2457417>
- Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2020). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*. Advance online publication. <https://doi.org/10.1037/emo0000770>
- Correia, A. I., Vincenzi, M., Vanzella, P., Pinheiro, A. P., Lima, C. F., & Schellenberg, E. G. (2021). Can musical ability be tested online? *Behavior Research Methods*. Advance online publication. <https://doi.org/10.3758/s13428-021-01641-2>
- Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1), 173–215. <https://doi.org/10.1016/j.cognition.2005.11.009>
- Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology*, 80(1), 75–85. <https://doi.org/10.1037/0022-3514.80.1.75>
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, 19(7), 573–576. <https://doi.org/10.1016/j.cub.2009.02.058>
- Fuller, C. D., Galvin III, J. J., Maat, B., Başkent, D., & Free, R. H. (2018). Comparison of two music training approaches on music and speech perception in cochlear implant users. *Trends in Hearing*, 22, 1–22. <https://doi.org/10.1177/2331216518765379>
- Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, 23(27), 9240–9245. <https://doi.org/10.1523/JNEUROSCI.23-27-09240.2003>
- Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear and Hearing*, 38(4), 455–464. <https://doi.org/10.1097/AUD.0000000000000402>
- Grandjean, D. (2021). Brain networks of emotional prosody processing. *Emotion Review*, 13(1), 34–43. <https://doi.org/10.1177/1754073919898522>
- Grichkovstova, I., Lacheret, A., Morel, M., Beaucousin, V., & Tzourio-Mazoyer, N. (2007). Affective speech gating. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 805–808). Saarland University.
- Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*, 38, 299–310. <https://doi.org/10.3758/BF03207159>

- Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, 76(3), 486–502. <https://doi.org/10.1016/j.neuron.2012.10.011>
- Jiang, X., Paulmann, S., Robin, J., & Pell, M. D. (2015). More than accuracy: Nonverbal dialects modulate the time course of vocal emotion recognition across cultures. *Journal of Experimental Psychology: Human Perception and Performance*, 41(3), 597–612. <https://doi.org/10.1037/xhp0000043>
- 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>
- Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *The Behavioral and Brain Sciences*, 31(5), 559–621. <https://doi.org/10.1017/S0140525X08005293>
- Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistic Compass*, 5(3), 108–125. <https://doi.org/10.1111/j.1749-818X.2010.00267.x>
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599–605. <https://doi.org/10.1038/nrn2882>
- Kühnis, J., Elmer, S., Meyer, M., & Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia*, 51(8), 1608–1618.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. <http://dx.doi.org/10.18637/jss.v082.i13>
- Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, 19(5), 633–653. <https://doi.org/10.1080/02699930441000445>
- Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PLoS ONE*, e52508. <https://doi.org/10.1371/journal.pone.0052508>
- 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>
- Lima, C. F., Correia, A. I., Müllensiefen, D., & Castro, S. L. (2020). Goldsmiths Musical Sophistication Index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality and music preferences. *Psychology of Music*, 48(3), 376–388. <https://doi.org/10.1177/0305735618801997>
- Liu, P., Rigoulot, S., & Pell, M. D. (2015). Culture modulates the brain response to human expressions of emotion: Electrophysiological evidence. *Neuropsychologia*, 67, 1–13. <https://doi.org/10.1016/j.neuropsychologia.2014.11.034>
- 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>
- Liu, X., Xu, Y., Alter, K., & Tuomainen, J. (2018). Emotional connotations of musical instrument timbre in comparison with emotional speech prosody: Evidence from acoustics and event-related potentials. *Frontiers in Psychology*, 9, 737. <https://doi.org/10.3389/fpsyg.2018.00737>
- Livingstone, S. R., Muhlberger, R., Brown, A. R., & Thompson, W. F. (2010). Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal*, 34(1), 41–64. <https://doi.org/10.1162/comj.2010.34.1.41>

- Lüdecke, D. (2021). *sjPlot: Data visualization for statistics in social science* (R package version 2.8.9) [Computer software]. <https://CRAN.R-project.org/package=sjPlot>.
- Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences*, 115(51), 13129–13134. <https://doi.org/10.1073/pnas.1811793115>
- Marques, C., Moreno, S., Luís Castro, S., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, 19(9), 1453–1463. <https://doi.org/10.1162/jocn.2007.19.9.1453>
- Martins, M., Pinheiro, A. P., & Lima, C. F. (2021). Does music training improve emotion recognition abilities? A critical review. *Emotion Review*, 13(3), 199–210. <https://doi.org/10.1177/17540739211022035>
- McKay, C. M. (2021). No evidence that music training benefits speech perception in hearing-impaired listeners: A systematic review. *Trends in Hearing*, 25, 1–16. <https://doi.org/10.1177/2331216520985678>
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, 19(3), 712–723. <https://doi.org/10.1093/cercor/bhn120>
- Mosing, M. A., & Ullén, F. (2016). Genetic influences on musical giftedness, talent, and practice. In G. E. McPherson (Ed.), *Musical prodigies: Interpretations from Psychology, Education, Musicology, and Ethnomusicology* (pp. 156–167). Oxford University Press.
- Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: No causal effect of music practice on music ability. *Psychological Science*, 25(9), 1795–1803. <https://doi.org/10.1177/0956797614541990>
- 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), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior & Development*, 27(4), 455–464. <https://doi.org/10.1016/j.infbeh.2004.03.002>
- Nordström, H. (2019). *Emotional communication in the human voice* [Doctoral thesis, Stockholm University]. Stockholm University. <http://su.diva-portal.org/smash/record.jsf?dswid=-5250&pid=diva2%3A1304804>
- Nordström, H., & Laukka, P. (2019). The time course of emotion recognition in speech and music. *The Journal of the Acoustical Society of America*, 145(5), 3058–3074.
- Nordström, H., Laukka, P., Thingujam, N. S., Schubert, E., & Elfenbein, H. A. (2017). Emotion appraisal dimensions inferred from vocal expressions are consistent across cultures: A comparison between Australia and India. *Royal Society Open Science*, 4(11), 170912. <https://doi.org/10.1098/rsos.170912>
- Nussbaum, C., & Schweinberger, S. R. (2021). Links between musicality and vocal emotion perception. *Emotion Review*, 13(3), 211–224. <https://doi.org/10.1177/17540739211022803>
- Overy, K. (2012). Making music in a group: Synchronization and shared experience. *Annals of the New York Academy of Sciences*, 1252, 65–68. <https://doi.org/10.1111/j.1749-6632.2012.06530.x>
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392(6678), 811–814. <https://doi.org/10.1038/33918>
- 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, 1049. <https://doi.org/10.3389/fnhum.2014.01049>
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, 2, 142. <https://doi.org/10.3389/fpsyg.2011.00142>
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, 308, 98–108. <https://doi.org/10.1016/j.heares.2013.08.011>
- Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport*, 19, 209–213. <https://doi.org/10.1097/WNR.0b013e3282f454db>
- Paulmann, S., & Kotz, S. (2018). The electrophysiology and time course of processing vocal emotion expressions. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 459–472). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198743187.001.0001>
- Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, 4, 345. <https://doi.org/10.3389/fpsyg.2013.00345>
- Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS One*, 6(11), Article e27256. <https://doi.org/10.1371/journal.pone.0027256>
- Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication*, 50(6), 519–530. <https://doi.org/10.1016/j.specom.2008.03.006>
- Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, 33(2), 107–120. <https://doi.org/10.1007/s10919-008-0065-7>
- 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.biopsych.2015.08.008>
- Peretz, I. (2010). Towards a neurobiology of musical emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 99–126). Oxford University Press.
- Pinheiro, A. P., Del Re, E., Mezin, J., Nestor, P. G., Rauber, A., McCarley, R. W., Gonçalves, O. F., & Niznikiewicz, M. A. (2013). Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: An electrophysiological investigation. *Psychological Medicine*, 43(3), 603–618. <https://doi.org/10.1017/S003329171200133X>
- 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>
- Pöppel, E., & Bao, Y. (2011). Three modes of knowledge as basis for intercultural cognition and communication: A theoretical perspective. In S. Han & E. Pöppel (Eds.), *Culture and neural frames of cognition and communication* (pp. 215–231). Springer. https://doi.org/10.1007/978-3-642-15423-2_14
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Raven, J. (2009). The raven progressive matrices and measuring aptitude constructs. *The International Journal of Educational and Psychological Assessment*, 2, 2–38.

- Rigoulot, S., & Pell, M. D. (2012). Seeing emotion with your ears: Emotional prosody implicitly guides visual attention to faces. *PLoS One*, 7(1), e30740. <https://doi.org/10.1371/journal.pone.0030740>
- Rigoulot, S., Wassiliwizky, E., & Pell, M. D. (2013). Feeling backwards? How temporal order in speech affects the time course of vocal emotion recognition. *Frontiers in Psychology*, 4, 367. <https://doi.org/10.3389/fpsyg.2013.00367>
- RStudio Team. (2021). *RStudio: integrated development for R*. <http://www.rstudio.com>
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115(1), 102–141. <https://doi.org/10.1037/0033-2909.115.1.102>
- Salasoo, A., & Pisoni, D. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language*, 24, 210–231. [https://doi.org/10.1016/0749-596X\(85\)90025-7](https://doi.org/10.1016/0749-596X(85)90025-7)
- Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107(6), 2408–2412. <https://doi.org/10.1073/pnas.0908239106>
- Schaerlaeken, S., & Grandjean, D. (2018). Unfolding and dynamics of affect bursts decoding in humans. *PLoS one*, 13(10), e0206216. <https://doi.org/10.1371/journal.pone.0206216>
- Schellenberg, E. G. (2020). Music training, individual differences, and plasticity. In M. S. C. Thomas, D. Mareschal, & I. Dumontheil (Eds.), *Educational neuroscience: Development across the life span* (1st ed., pp. 413–438). Routledge. <https://doi.org/10.4324/9781003016830>
- Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, 9(3), 235–248. [https://doi.org/10.1016/S0892-1997\(05\)802310](https://doi.org/10.1016/S0892-1997(05)802310)
- Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15(2), 123–148. <https://doi.org/10.1007/BF00995674>
- Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 433–456). Oxford University Press.
- 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>
- Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2021). *Afex: Analysis of factorial experiments* (R package version 1.0-1) [Computer software]. <https://CRAN.R-project.org/package=afex>
- Sloboda, J. A., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 71–104). Oxford University Press.
- Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency: Effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience*, 29(3), 661–668. <https://doi.org/10.1111/j.1460-9568.2009.06617.x>
- Swaminathan, S., & Schellenberg, E. G. (2015). Current emotion research in music psychology. *Emotion Review*, 7(2), 189–197. <https://doi.org/10.1177/1754073914558282>
- Swaminathan, S., & Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychonomic Bulletin & Review*, 24(6), 1929–1934. <https://doi.org/10.3758/s13423-017-1244-5>

- Swaminathan, S., & Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. *Scientific Report*, 8, 9223. <https://doi.org/10.1038/s41598-018-27571-2>
- Swaminathan, S., & Schellenberg, E. G. (2019). Musical ability, music training, and language ability in childhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(12), 2340–2348. <https://doi.org/10.1037/xlm0000798>
- Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology*, 5, 658. <https://doi.org/10.3389/fpsyg.2014.00658>
- Tarr, B., Launay, J., & Dunbar, R. I. (2014). Music and social bonding: "Self-other" merging and neurohormonal mechanisms. *Frontiers in Psychology*, 5, 1096. <https://doi.org/10.3389/fpsyg.2014.01096>
- Thompson, W. F. (2009). *Music, thought, and feeling: Understanding the psychology of music*. Oxford University Press.
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, 4(1), 46–64. <https://doi.org/10.1037/1528-3542.4.1.46>
- Thönnissen, H., Boers, F., Dammers, J., Chen, Y. H., Norra, C., & Mathiak, K. (2010). Early sensory encoding of affective prosody: Neuromagnetic tomography of emotional category changes. *Neuroimage*, 50(1), 250–259. <https://doi.org/10.1016/j.neuroimage.2009.11.082>
- Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, 6(7), 669–673. <https://doi.org/10.1038/nn1084>
- 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>
- Trost, W. J., Labbé, C., & Grandjean, D. (2017). Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia*, 96, 96–110. <https://doi.org/10.1016/j.neuropsychologia.2017.01.004>
- Tyler, L. K., & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception & Psychophysics*, 38(3), 217–222. <https://doi.org/10.3758/BF03207148>
- 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>
- Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, 20(3), 188–196. <https://doi.org/10.1016/j.lindif.2010.02.004>
- Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino, L., Francois, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Muller, K., Ooms, J., Robinson, D., Seidel, D., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), Article 1686. <https://doi.org/10.21105/joss.01686>
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, 156, 249–268. [https://doi.org/10.1016/S0079-6123\(06\)56013-3](https://doi.org/10.1016/S0079-6123(06)56013-3)

Anexos

Anexo A

Confusões mais comuns entre emoções em cada intervalo de gate

Valores a negrito indicam a proporção de participantes que selecionaram a emoção pretendida. Valores a vermelho indicam que uma proporção maior de participantes selecionou esta categoria emocional em vez da emoção correta.

Estímulo	Gate (ms)	Emoção Selecionada				
		Neutro	Tristeza	Raiva	Medo	Alegria
Neutro	50	0.50	0.15	0.05	0.11	0.19
	100	0.55	0.15	0.04	0.09	0.17
	150	0.57	0.16	0.05	0.09	0.14
	200	0.56	0.17	0.05	0.06	0.16
	250	0.64	0.13	0.05	0.05	0.13
	400	0.66	0.16	0.06	0.06	0.06
	550	0.71	0.16	0.02	0.05	0.05
	700	0.85	0.08	0.01	0.02	0.05
	850	0.88	0.07	0.00	0.02	0.03
	1000	0.90	0.06	0.01	0.02	0.01
	Gfull	0.93	0.04	0.01	0.01	0.02
Tristeza	50	0.48	0.23	0.05	0.15	0.08
	100	0.45	0.31	0.03	0.15	0.07
	150	0.49	0.34	0.02	0.11	0.04
	200	0.51	0.32	0.02	0.10	0.05
	250	0.52	0.33	0.02	0.10	0.04
	400	0.40	0.31	0.15	0.09	0.04
	550	0.38	0.36	0.18	0.05	0.03
	700	0.30	0.60	0.02	0.06	0.02
	850	0.27	0.66	0.01	0.04	0.02
	1000	0.17	0.78	0.00	0.03	0.02
	Gfull	0.11	0.85	0.01	0.02	0.01
Raiva	50	0.32	0.12	0.18	0.18	0.14
	100	0.32	0.10	0.24	0.17	0.16
	150	0.25	0.10	0.33	0.17	0.14
	200	0.30	0.07	0.37	0.14	0.11
	250	0.17	0.04	0.55	0.12	0.12
	400	0.16	0.09	0.60	0.08	0.08
	550	0.10	0.09	0.66	0.10	0.05
	700	0.07	0.01	0.85	0.04	0.04
	850	0.06	0.01	0.85	0.04	0.04
	1000	0.06	0.02	0.87	0.02	0.02
	Gfull	0.06	0.03	0.86	0.03	0.02
Medo	50	0.44	0.13	0.15	0.20	0.08
	100	0.34	0.12	0.14	0.27	0.12
	150	0.30	0.14	0.17	0.26	0.14
	200	0.25	0.15	0.19	0.30	0.12
	250	0.23	0.20	0.13	0.32	0.12

	400	0.19	0.17	0.13	0.40	0.11
	550	0.13	0.15	0.10	0.49	0.14
	700	0.10	0.14	0.08	0.62	0.06
	850	0.06	0.11	0.06	0.72	0.05
	1000	0.05	0.08	0.05	0.78	0.04
	Gfull	0.03	0.07	0.05	0.82	0.04
Alegria	50	0.32	0.10	0.16	0.17	0.25
	100	0.28	0.07	0.19	0.19	0.26
	150	0.21	0.12	0.16	0.20	0.31
	200	0.18	0.13	0.20	0.17	0.31
	250	0.18	0.11	0.23	0.16	0.32
	400	0.24	0.12	0.17	0.13	0.34
	550	0.22	0.09	0.09	0.10	0.50
	700	0.11	0.04	0.04	0.06	0.74
	850	0.07	0.02	0.03	0.04	0.84
	1000	0.04	0.01	0.02	0.02	0.90
	Gfull	0.03	0.02	0.01	0.01	0.94

Anexo B

Informação adicional acerca do modelo global que estimou exatidão ao longo dos gates em função da emoção e grupo

São apresentados os contrastes para os efeitos de interação significativos, grupo e gate, grupo e emoção, gate e emoção, respetivamente.

Contrastes entre grupos em função da duração do gate (Interação Grupo x Gate, $\chi^2 = 56.37^{***}, p < .001$)

Contraste	Gate	Estimativa	DP	z	p
m–nm	G50	0,02	0,17	0,11	.910
m–nm	G1000	0,55	0,19	2,90	.038
m–nm	GFull	1,13	0,22	5,25	<.0001
m–nm	G100	0,43	0,17	2,50	.111
m–nm	G150	0,28	0,17	1,67	.470
m–nm	G200	0,22	0,17	1,31	.569
m–nm	G250	0,21	0,17	1,26	.569
m–nm	G400	0,31	0,17	1,82	.412
m–nm	G550	0,28	0,17	1,64	.470
m–nm	G750	0,38	0,17	2,17	.213
m–nm	G850	0,41	0,18	2,26	.192

Notas. m = músicos; nm = não-músicos

Contrastes entre grupo em função da emoção (Interação Grupo x Emoção, $\chi^2 = 20.20^{*}, p <.001$)**

Contraste	Emoção	Estimativa	DP	z	p
m - nm	Raiva	0,35	0,16	2,15	.094
m - nm	Triste	0,41	0,16	2,58	.040
m - nm	Alegria	0,63	0,17	3,76	.0008
m - nm	Medo	0,27	0,16	1,67	.189
m - nm	Neutro	0,25	0,17	1,52	.189

Notas. m = músicos; nm = não-músicos

Contrastes entre gates em função da emoção (Interação Gate x Emoção, $\chi^2 = 119.13^{*}, p <.001$)**

Contraste	Emoção	Estimativa	DP	z	p
G50 – G1000	Raiva	-4,19	0,36	-11,67	<.0001
G50 – Gfull	Raiva	-3,96	0,35	-11,20	<.0001
G50 – G100	Raiva	-0,36	0,34	-1,05	1.0000
G50 – G150	Raiva	-0,95	0,34	-2,78	0.5965
G50 – G200	Raiva	-0,99	0,34	-2,90	0.4168
G50 – G250	Raiva	-2,02	0,34	-5,98	<.0001
G50 – G400	Raiva	-3,11	0,35	-9,03	<.0001
G50 – G550	Raiva	-3,66	0,35	-10,52	<.0001
G50 – G700	Raiva	-3,71	0,35	-10,66	<.0001
G50 – G850	Raiva	-3,82	0,35	-10,89	<.0001
G1000 – G100	Raiva	3,83	0,36	10,77	<.0001
G1000 – GFull	Raiva	0,23	0,37	0,63	1.0000
G1000 – G150	Raiva	3,24	0,35	9,20	<.0001
G1000 – G200	Raiva	3,20	0,35	9,10	<.0001
G1000 – G250	Raiva	2,17	0,35	6,19	<.0001
G1000 – G400	Raiva	1,08	0,36	3,03	0.2823
G1000 – G550	Raiva	0,53	0,36	1,46	1.0000
G1000 – G700	Raiva	0,48	0,36	1,34	1.0000
G1000 – G850	Raiva	0,37	0,36	1,03	1.0000
GFull – G100	Raiva	3,60	0,35	10,28	<.0001
GFull – G150	Raiva	3,01	0,35	8,68	<.0001
GFull – G200	Raiva	2,97	0,35	8,57	<.0001
GFull – G250	Raiva	1,94	0,35	5,62	<.0001
GFull – G400	Raiva	0,85	0,35	2,41	1.0000
GFull – G550	Raiva	0,29	0,35	0,83	1.0000
GFull – G700	Raiva	0,25	0,35	0,71	1.0000
GFull – G850	Raiva	0,14	0,36	0,39	1.0000
G100 – G150	Raiva	-0,59	0,34	-1,74	1.0000
G100 – G200	Raiva	-0,63	0,34	-1,86	1.0000
G100 – G250	Raiva	-1,66	0,34	-4,96	0.0001

G100 – G400	Raiva	-2,75	0,34	-8,06	<.0001
G100 – G550	Raiva	-3,30	0,35	-9,58	<.0001
G100 – G700	Raiva	-3,35	0,35	-9,72	<.0001
G100 – G850	Raiva	-3,46	0,35	-9,96	<.0001
G150 – G200	Raiva	-0,04	0,33	-0,12	1.0000
G150 – G250	Raiva	-1,08	0,33	-3,25	0.1404
G150 – G400	Raiva	-2,17	0,34	-6,41	<.0001
G150 – G550	Raiva	-2,72	0,34	-7,96	<.0001
G150 – G700	Raiva	-2,76	0,34	-8,09	<.0001
G150 – G850	Raiva	-2,87	0,34	-8,35	<.0001
G200 – G250	Raiva	-1,04	0,33	-3,13	0.2085
G200 – G400	Raiva	-2,13	0,34	-6,30	<.0001
G200 – G550	Raiva	-2,68	0,34	-7,85	<.0001
G200 – G700	Raiva	-2,72	0,34	-7,98	<.0001
G200 – G850	Raiva	-2,83	0,34	-8,24	<.0001
G250 – G400	Raiva	-1,09	0,34	-3,25	0.1398
G250 – G550	Raiva	-1,64	0,34	-4,84	0.0002
G250 – G700	Raiva	-1,69	0,34	-4,97	0.0001
G250 – G850	Raiva	-1,80	0,34	-5,26	<.0001
G400 – G550	Raiva	-0,55	0,35	-1,60	1.0000
G400 – G700	Raiva	-0,60	0,35	-1,73	1.0000
G400 – G850	Raiva	-0,71	0,35	-2,03	1.0000
G550 – G700	Raiva	-0,04	0,35	-0,13	1.0000
G550 – G850	Raiva	-0,15	0,35	-0,44	1.0000
G700 – G850	Raiva	-0,11	0,35	-0,32	1.0000
G50 – G10	Triste	-2,73	0,34	-8,13	<.0001
G50 – GFull	Triste	-3,21	0,34	-9,44	<.0001
G50 – G100	Triste	-0,43	0,33	-1,28	1.0000
G50 – G150	Triste	-0,63	0,33	-1,89	1.0000
G50 – G200	Triste	-0,55	0,33	-1,65	1.0000
G50 – G250	Triste	-0,56	0,33	-1,69	1.0000
G50 – G400	Triste	-1,00	0,33	-3,02	0.2880
G50 – G550	Triste	-1,23	0,33	-3,71	0.0280
G50 – G700	Triste	-1,85	0,33	-5,56	<.0001
G50 – G850	Triste	-2,09	0,33	-6,28	<.0001
G1000 – GFull	Triste	-0,48	0,34	-1,41	1.0000
G1000 – G100	Triste	2,30	0,33	6,90	<.0001
G1000 – G150	Triste	2,10	0,33	6,32	<.0001
G1000 – G200	Triste	2,18	0,33	6,55	<.0001
G1000 – G250	Triste	2,17	0,33	6,51	<.0001
G1000 – G400	Triste	1,73	0,33	5,20	<.0001
G1000 – G550	Triste	1,50	0,33	4,53	0.0009
G1000 – G700	Triste	0,89	0,33	2,67	0.8239

G1000 – G850	Triste	0,64	0,33	1,92	1.0000
GFull – G100	Triste	2,79	0,34	8,23	<.0001
GFull – G150	Triste	2,59	0,34	7,67	<.0001
GFull – G200	Triste	2,67	0,34	7,88	<.0001
GFull – G250	Triste	2,65	0,34	7,85	<.0001
GFull – G400	Triste	2,21	0,34	6,56	<.0001
GFull – G550	Triste	1,99	0,34	5,90	<.0001
GFull – G700	Triste	1,37	0,34	4,06	0.0071
GFull – G850	Triste	1,12	0,34	3,32	0.1111
G100 – G150	Triste	-0,20	0,33	-0,60	1.0000
G100 – G200	Triste	-0,12	0,33	-0,36	1.0000
G100 – G250	Triste	-0,14	0,33	-0,41	1.0000
G100 – G400	Triste	-0,57	0,33	-1,74	1.0000
G100 – G550	Triste	-0,80	0,33	-2,43	1.0000
G100 – G700	Triste	-1,42	0,33	-4,30	0.0026
G100 – G850	Triste	-1,66	0,33	-5,03	0.0001
G150 – G200	Triste	0,08	0,33	0,24	1.0000
G150 – G250	Triste	0,06	0,33	0,20	1.0000
G150 – G400	Triste	-0,37	0,33	-1,14	1.0000
G150 – G550	Triste	-0,60	0,33	-1,83	1.0000
G150 – G700	Triste	-1,22	0,33	-3,71	0.0280
G150 – G850	Triste	-1,46	0,33	-4,44	0.0013
G200 – G250	Triste	-0,01	0,33	-0,04	1.0000
G200 – G400	Triste	-0,45	0,33	-1,38	1.0000
G200 – G550	Triste	-0,68	0,33	-2,07	1.0000
G200 – G700	Triste	-1,30	0,33	-3,94	0.0114
G200 – G850	Triste	-1,54	0,33	-4,67	0.0005
G250 – G400	Triste	-0,44	0,33	-1,33	1.0000
G250 – G550	Triste	-0,67	0,33	-2,02	1.0000
G250 – G700	Triste	-1,28	0,33	-3,90	0.0133
G250 – G850	Triste	-1,53	0,33	-4,63	0.0006
G400 – G550	Triste	-0,23	0,33	-0,69	1.0000
G400 – G700	Triste	-0,84	0,33	-2,57	1.0000
G400 – G850	Triste	-1,09	0,33	-3,31	0.1154
G550 – G700	Triste	-0,62	0,33	-1,88	1.0000
G550 – G850	Triste	-0,86	0,33	-2,63	0.9078
G700 – G850	Triste	-0,25	0,33	-0,75	1.0000
G50 – G1000	Alegria	-4,00	0,36	-11,05	<.0001
G50 – GFull	Alegria	-4,66	0,39	-11,82	<.0001
G50 – G100	Alegria	-0,18	0,34	-0,54	1.0000
G50 – G150	Alegria	-0,41	0,34	-1,23	1.0000
G50 – G200	Alegria	-0,44	0,34	-1,31	1.0000
G50 – G250	Alegria	-0,44	0,34	-1,31	1.0000

G50 – G400	Alegria	-0,74	0,34	-2,19	1.0000
G50 – G550	Alegria	-1,75	0,34	-5,24	<.0001
G50 – G700	Alegria	-2,67	0,34	-7,86	<.0001
G50 – G850	Alegria	-3,44	0,35	-9,86	<.0001
G1000 – GFull	Alegria	-0,66	0,41	-1,59	1.0000
G1000 – G100	Alegria	3,82	0,36	10,61	<.0001
G1000 – G150	Alegria	3,59	0,36	10,02	<.0001
G1000 – G200	Alegria	3,56	0,36	9,94	<.0001
G1000 – G250	Alegria	3,56	0,36	9,95	<.0001
G1000 – G400	Alegria	3,26	0,36	9,11	<.0001
G1000 – G550	Alegria	2,25	0,36	6,29	<.0001
G1000 – G700	Alegria	1,33	0,36	3,68	0.0304
G1000 – G850	Alegria	0,57	0,37	1,53	1.0000
GFull – G100	Alegria	4,47	0,39	11,41	<.0001
GFull – G150	Alegria	4,25	0,39	10,86	<.0001
GFull – G200	Alegria	4,22	0,39	10,79	<.0001
GFull – G250	Alegria	4,22	0,39	10,79	<.0001
GFull – G400	Alegria	3,92	0,39	10,02	<.0001
GFull – G550	Alegria	2,90	0,39	7,44	<.0001
GFull – G700	Alegria	1,99	0,39	5,05	0.0001
GFull – G850	Alegria	1,22	0,40	3,04	0.2705
G100 – G150	Alegria	-0,23	0,33	-0,69	1.0000
G100 – G200	Alegria	-0,26	0,33	-0,77	1.0000
G100 – G250	Alegria	-0,26	0,33	-0,77	1.0000
G100 – G400	Alegria	-0,55	0,33	-1,66	1.0000
G100 – G550	Alegria	-1,57	0,33	-4,72	0.0004
G100 – G700	Alegria	-2,49	0,34	-7,37	<.0001
G100 – G850	Alegria	-3,25	0,35	-9,40	<.0001
G150 – G200	Alegria	-0,03	0,33	-0,08	1.0000
G150 – G250	Alegria	-0,03	0,33	-0,08	1.0000
G150 – G400	Alegria	-0,33	0,33	-0,98	1.0000
G150 – G550	Alegria	-1,34	0,33	-4,05	0.0072
G150 – G700	Alegria	-2,26	0,34	-6,72	<.0001
G150 – G850	Alegria	-3,02	0,35	-8,78	<.0001
G200 – G250	Alegria	0,00	0,33	0,00	1.0000
G200 – G400	Alegria	-0,30	0,33	-0,90	1.0000
G200 – G550	Alegria	-1,32	0,33	-3,97	0.0101
G200 – G700	Alegria	-2,23	0,34	-6,64	<.0001
G200 – G850	Alegria	-3,00	0,35	-8,70	<.0001
G250 – G400	Alegria	-0,30	0,33	-0,90	1.0000
G250 – G550	Alegria	-1,31	0,33	-3,97	0.0101
G250 – G700	Alegria	-2,23	0,34	-6,64	<.0001
G250 – G850	Alegria	-3,00	0,34	-8,70	<.0001

G400 – G550	Alegria	-1,02	0,33	-3,07	0.2484
G400 – G700	Alegria	-1,93	0,34	-5,75	<.0001
G400 – G850	Alegria	-2,70	0,35	-7,83	<.0001
G550 – G700	Alegria	-0,92	0,34	-2,73	0.6849
G550 – G850	Alegria	-1,68	0,34	-4,89	0.0002
G700 – G850	Alegria	-0,77	0,35	-2,20	1.0000
G50 – G1000	Medo	-3,07	0,34	-9,03	<.0001
G50 – GFull	Medo	-3,48	0,35	-10,08	<.0001
G50 – G100	Medo	-0,45	0,34	-1,33	1.0000
G50 – G150	Medo	-0,35	0,34	-1,05	1.0000
G50 – G200	Medo	-0,61	0,34	-1,80	1.0000
G50 – G250	Medo	-0,70	0,34	-2,09	1.0000
G50 – G400	Medo	-1,31	0,34	-3,91	0.0130
G50 – G550	Medo	-1,86	0,34	-5,56	<.0001
G50 – G700	Medo	-2,19	0,34	-6,53	<.0001
G50 – G850	Medo	-2,71	0,34	-8,02	<.0001
G1000 – GFull	Medo	-0,41	0,35	-1,18	1.0000
G1000 – G100	Medo	2,62	0,34	7,78	<.0001
G1000 – G150	Medo	2,72	0,34	8,08	<.0001
G1000 – G200	Medo	2,47	0,34	7,34	<.0001
G1000 – G250	Medo	2,37	0,34	7,05	<.0001
G1000 – G400	Medo	1,77	0,33	5,29	<.0001
G1000 – G550	Medo	1,21	0,33	3,63	0.0366
G1000 – G700	Medo	0,88	0,34	2,64	0.8767
G1000 – G850	Medo	0,36	0,34	1,08	1.0000
GFull – G100	Medo	3,03	0,34	8,85	<.0001
GFull – G150	Medo	3,13	0,34	9,14	<.0001
GFull – G200	Medo	2,87	0,34	8,42	<.0001
GFull – G250	Medo	2,78	0,34	8,14	<.0001
GFull – G400	Medo	2,17	0,34	6,41	<.0001
GFull – G550	Medo	1,62	0,34	4,78	0.0003
GFull – G700	Medo	1,29	0,34	3,80	0.0197
GFull – G850	Medo	0,77	0,34	2,25	1.0000
G100 – G150	Medo	0,10	0,34	0,29	1.0000
G100 – G200	Medo	-0,16	0,33	-0,47	1.0000
G100 – G250	Medo	-0,25	0,33	-0,76	1.0000
G100 – G400	Medo	-0,86	0,33	-2,59	0.9946
G100 – G550	Medo	-1,41	0,33	-4,25	0.0031
G100 – G700	Medo	-1,74	0,33	-5,24	<.0001
G100 – G850	Medo	-2,26	0,34	-6,75	<.0001
G150 – G200	Medo	-0,25	0,33	-0,76	1.0000
G150 – G250	Medo	-0,35	0,33	-1,05	1.0000
G150 – G400	Medo	-0,95	0,33	-2,88	0.4434

G150 – G550	Medo	-1,51	0,33	-4,55	0.0008
G150 – G700	Medo	-1,84	0,33	-5,53	<.0001
G150 – G850	Medo	-2,36	0,34	-7,05	<.0001
G200 – G250	Medo	-0,10	0,33	-0,29	1.0000
G200 – G400	Medo	-0,70	0,33	-2,12	1.0000
G200 – G550	Medo	-1,25	0,33	-3,79	0.0201
G200 – G700	Medo	-1,58	0,33	-4,78	0.0003
G200 – G850	Medo	-2,11	0,33	-6,31	<.0001
G250 – G400	Medo	-0,60	0,33	-1,83	1.0000
G250 – G550	Medo	-1,16	0,33	-3,50	0.0599
G250 – G700	Medo	-1,49	0,33	-4,49	0.0011
G250 – G850	Medo	-2,01	0,33	-6,02	<.0001
G400 – G550	Medo	-0,55	0,33	-1,69	1.0000
G400 – G700	Medo	-0,88	0,33	-2,68	0.7848
G400 – G850	Medo	-1,40	0,33	-4,24	0.0033
G550 – G700	Medo	-0,33	0,33	-1,00	1.0000
G550 – G850	Medo	-0,85	0,33	-2,57	1.0000
G700 – G850	Medo	-0,52	0,33	-1,57	1.0000
G50 – G1000	Neutro	-2,49	0,35	-7,19	<.0001
G50 – GFull	Neutro	-3,30	0,40	-8,28	<.0001
G50 – G100	Neutro	-0,25	0,33	-0,76	1.0000
G50 – G150	Neutro	-0,32	0,33	-0,98	1.0000
G50 – G200	Neutro	-0,27	0,33	-0,81	1.0000
G50 – G250	Neutro	-0,67	0,33	-2,04	1.0000
G50 – G400	Neutro	-1,10	0,33	-3,33	0.1094
G50 – G550	Neutro	-1,36	0,33	-4,10	0.0061
G50 – G700	Neutro	-1,94	0,34	-5,75	<.0001
G50 – G850	Neutro	-2,26	0,34	-6,60	<.0001
G1000 – GFull	Neutro	-0,81	0,41	-1,95	1.0000
G1000 – G100	Neutro	2,24	0,35	6,46	<.0001
G1000 – G150	Neutro	2,17	0,35	6,25	<.0001
G1000 – G200	Neutro	2,23	0,35	6,41	<.0001
G1000 – G250	Neutro	1,82	0,35	5,23	<.0001
G1000 – G400	Neutro	1,39	0,35	3,98	0.0098
G1000 – G550	Neutro	1,13	0,35	3,23	0.1464
G1000 – G700	Neutro	0,55	0,36	1,54	1.0000
G1000 – G850	Neutro	0,23	0,36	0,63	1.0000
GFull – G100	Neutro	3,05	0,40	7,64	<.0001
GFull – G150	Neutro	2,97	0,40	7,46	<.0001
GFull – G200	Neutro	3,03	0,40	7,60	<.0001
GFull – G250	Neutro	2,62	0,40	6,57	<.0001
GFull – G400	Neutro	2,20	0,40	5,48	<.0001
GFull – G550	Neutro	1,94	0,40	4,83	0.0002

GFull – G700	Neutro	1,35	0,41	3,33	0.1094
GFull – G850	Neutro	1,03	0,41	2,52	1.0000
G100 – G150	Neutro	-0,07	0,33	-0,22	1.0000
G100 – G200	Neutro	-0,02	0,33	-0,05	1.0000
G100 – G250	Neutro	-0,42	0,33	-1,28	1.0000
G100 – G400	Neutro	-0,85	0,33	-2,57	1.0000
G100 – G550	Neutro	-1,11	0,33	-3,34	0.1068
G100 – G700	Neutro	-1,69	0,34	-5,00	0.0001
G100 – G850	Neutro	-2,01	0,34	-5,87	<.0001
G150 – G200	Neutro	0,06	0,33	0,17	1.0000
G150 – G255	Neutro	-0,35	0,33	-1,06	1.0000
G150 – G400	Neutro	-0,78	0,33	-2,35	1.0000
G150 – G550	Neutro	-1,04	0,33	-3,12	0.2105
G150 – G700	Neutro	-1,62	0,34	-4,79	0.0003
G150 – G850	Neutro	-1,94	0,34	-5,66	<.0001
G200 – G250	Neutro	-0,41	0,33	-1,23	1.0000
G200 – G400	Neutro	-0,84	0,33	-2,52	1.0000
G200 – G550	Neutro	-1,09	0,33	-3,29	0.1222
G200 – G700	Neutro	-1,68	0,34	-4,95	0.0001
G200 – G850	Neutro	-2,00	0,34	-5,82	<.0001
G250 – G400	Neutro	-0,43	0,33	-1,29	1.0000
G250 – G550	Neutro	-0,69	0,33	-2,06	1.0000
G250 – G700	Neutro	-1,27	0,34	-3,74	0.0246
G250 – G850	Neutro	-1,59	0,34	-4,62	0.0006
G400 – G550	Neutro	-0,26	0,34	-0,77	1.0000
G400 – G700	Neutro	-0,84	0,34	-2,47	1.0000
G400 – G850	Neutro	-1,16	0,35	-3,36	0.0991
G550 – G700	Neutro	-0,58	0,34	-1,71	1.0000
G550 – G850	Neutro	-0,90	0,35	-2,61	0.9492
G700 – G850	Neutro	-0,32	0,35	-0,91	1.0000

Anexo C

Distribuição de pontos de identificação corretos e incorretos para cada emoção em função do intervalo de gate

Distribuição de pontos de identificação corretos e incorretos para cada emoção em função do intervalo de gate. Total de incorretos corresponde aos estímulos dentro de cada categoria emocional nos quais os participantes não estabilizaram a sua resposta.

Emoção	Pontos e identificação da Emoção											Total Correto	Total Incorreto
	Intervalo de gate (ms)												
	50	100	150	200	250	400	550	700	850	1000			
Neutro	85	41	185	48	103	93	82	77	49	34	797	59	
Raiva	25	45	108	64	181	153	74	54	30	25	759	97	
Alegria	20	15	60	43	43	115	170	149	114	57	786	70	
Medo	18	18	49	42	49	99	112	122	93	85	687	169	
Tristeza	10	15	63	29	53	79	96	127	94	100	666	190	

Nota. Valores de EIPs baseados num total de 4280 observações (107 participantes x 40 estímulos).