DOI 10.20504/opus2026.32.34
Dossiê A pesquisa em Pedagogia Vocal na atualidade brasileira: recortes e diálogos interdisciplinares

Efeitos da imagética auditiva na afinação vocal: influência do tempo de retenção e da experiência musical na imitação de melodias

Mariana Moga de Moura e **Silva**
http://orcid.org/0000-0001-5723-5087
Universidade Federal do ABC

Patrícia **Vanzella**
http://orcid.org/0000-0002-7709-1495
Universidade Federal do ABC

João Ricardo **Sato**
http://orcid.org/0000-0002-7503-9781
Universidade Federal do ABC

Bóris **Marin**
http://orcid.org/0000-0001-5503-648X
Universidade Federal do ABC

Guilherme Alves Delmolin de **Oliveira**
http://orcid.org/0000-0002-3782-8843
Universidade Federal do ABC

Daniel Quintela **Bertuzzi**
https://orcid.org/0009-0007-9537-0061
Universidade Federal do ABC

Resumo: A imagética auditiva refere-se ao uso da mentalização de sons para aprimorar a precisão motora e a performance musical. No ensino do canto, essa estratégia é amplamente utilizada por professores, mas ainda pouco explorada sob uma perspectiva neurocientífica. O objetivo deste estudo foi investigar se a criação de imagética auditiva auxilia o desempenho vocal em termos de acurácia de frequência fundamental (F0), em contextos musicais tonais (familiares) e atonais (não-familiares). Como objetivos específicos buscou-se 1) Avaliar se a imagética auditiva contribui para a acurácia da afinação; 2) Verificar o impacto da imagética auditiva em diferentes contextos musicais (tonal e atonal) no desempenho de imitação vocal; 3) Analisar a correlação entre o contexto musical e a habilidade de criação da imagética auditiva. Foram avaliados 132 participantes (não-cantores, amadores e profissionais) em uma tarefa de imitação vocal de 30 melodias, testando duas condições: (i) vocalização imediata após a escuta e (ii) vocalização após um intervalo de quatro segundos para mentalização. Os resultados indicaram que a acurácia na afinação foi maior na condição sem tempo de mentalização ($Z = 1032$, $p < 0,001$), sugerindo que o tempo de retenção pode prejudicar a performance, possivelmente devido à sobrecarga cognitiva. Além disso, observou-se correlação moderada entre habilidades de imagética e precisão na afinação ($\rho = -0,368$), assim como uma relação entre habilidades perceptuais e controle da imagética ($\rho = 0,562$). Concluímos que a imagética auditiva pode ser uma ferramenta eficaz para o aprimoramento vocal, mas sua aplicação deve ser cuidadosamente planejada, aliada a um treinamento perceptivo eficiente para evitar interferências cognitivas.

Palavras-chave: Imagética auditiva. Afinação. Neurociência cognitiva. Pedagogia vocal.

Effects of auditory imagery on vocal tuning: Influence of retention time and musical experience on the imitation of melodies

Abstract: Auditory imagery refers to the use of sound mentalization to improve motor precision and musical performance. In the teaching of singing, this strategy is widely used by teachers, but still little explored from a neuroscientific perspective. The aim of this study was to investigate whether the creation of auditory imagery helps vocal performance in terms of fundamental frequency accuracy (F0) in tonal (familiar) and atonal (unfamiliar) musical contexts. The specific objectives were: 1) To evaluate whether auditory imagery contributes to the accuracy of tuning; 2) To verify the impact of auditory imagery in different musical contexts (tonal and atonal) on the performance of vocal imitation; 3) To analyze the correlation between the musical context and the ability to create auditory imagery. A total of 132 participants (non-singers, amateurs, and professionals) were evaluated in a 30-melody vocal imitation task, testing two conditions: (i) immediate vocalization after listening and (ii) vocalization after a four-second interval for mentalization. The results indicated that the accuracy of tuning was higher in the condition without mentalization time ($Z = 1032$, $p < 0.001$), suggesting that retention time may impair performance, possibly due to cognitive overload. In addition, a moderate correlation was observed between imagery skills and tuning accuracy ($\rho = -0.368$), as well as a relationship between perceptual skills and imagery control ($\rho = 0.562$). We conclude that auditory imagery can be an effective tool for vocal improvement, although its application must be carefully planned, combined with efficient perceptual training to avoid cognitive interference.

Keywords: Auditory imagery. Vocal tuning. Cognitive neuroscience. Vocal pedagogy.

SILVA et. al.. Efeitos da imagética auditiva na afinação vocal: influência do tempo de retenção e da experiência musical na imitação de melodias. *Opus* (Assoc. Nac. Pesqui. Pós-Grad. Música), João Pessoa, v. 32, e263234, p. 1-25, 2026. http://dx.doi.org/10.20504/opus2026.32.34
Recebido em 24/02/2025, aprovado em 23/07/2025, publicado em 24/08/2026
Editor Responsável: Dr. Vinícius Eufrásio
Editores de Seção: Dra. Joana Mariz e Dra. Viviane Kubo

Quando estudamos música fazemos uso, muitas vezes, de recursos como a imagética, também chamada de "ensaios mentais". Somos capazes de imaginar as notas da música e de também transformá-las. Recorrer ao uso de imagética é uma prática frequente entre músicos e professores de música, que geralmente a usam em diferentes modalidades (auditiva, motora *etc*.) não apenas para a internalização de canções, mas também para a leitura musical silenciosa, por exemplo (Lotze, 2013, p. 1). Pesquisas nas ciências do esporte têm apresentado associações robustas entre o uso de imagética no treino e o desempenho de atletas (Simonsmeier *et al*., 2021, p. 17). Perguntamo-nos, então: será que o ensaio mental também ajuda os cantores no seu desempenho em termos de acurácia?

Para investigar se há relação entre imagética e a acurácia na afinação do canto, utilizamos como arcabouço teórico o modelo proposto por Pfordresher, Halpern & Greenspon (2015), o *Multimodal Model of Imagery Association*, que estabelece uma associação das imagéticas auditiva e motora para o indivíduo conseguir imitar melodias com acurácia. O modelo prediz que o uso eficiente do feedback auditivo deve basear-se na habilidade de mapear uma imagem mental de planejamento motor na imagética auditiva. Ele ainda sugere que o mapeamento das associações entre imagética motora e auditiva faz o caminho oposto da imagética auditiva para a motora e sugere que tarefas de imagética podem ser úteis como uma maneira de melhorar habilidades de imitação vocal.

1. O Modelo de Associação de Imagética Multimodal (*Multimodal Imagery Association Model*)

A afinação no canto, ou falta dela, é objeto de interesse de diversas áreas da ciência, da música à psicologia (Pfordresher *et al*., 2015, p. 227). Também em 2015, os pesquisadores Peter Pfordresher, Andrea Halpern e Emma Greenspon propuseram um modelo cognitivo que teoriza as associações entre a imagética auditiva e motora com a percepção e ação, chamado Modelo de Associação de Imagética Multimodal (MAIM). Este arcabouço trata da questão da desafinação no nível da relação da percepção com a produção, especificamente para tarefas de imitação, e define falha na afinação neste tipo de tarefa como déficit na imitação de *pitch*[1] (DIP) ou *Voice Deficit Pitch Imitation*.

No MAIM, a tradução sensóriomotora é considerada representante de um exemplo de uma classe mais ampla de associações mentais que possui papel importante na percepção e planejamento motor. Neste modelo, a imagética é considerada um fator essencial para tradução sensório-motora e que tem como essência induzir uma experimentação subjetiva quasi-perceptual em diferentes modalidades sem o estímulo externo, imagética perceptiva (auditiva, por exemplo) ou ação (imagética motora). A imagética envolve associações automáticas entre diferentes modalidades (visão, audição, propriocepção, motora etc.) e, no presente contexto, refere-se às associações entre o *pitch height* e a regulação da tensão das pregas vocais (Pfordresher; Halpern; Greenspon, 2015, p. 244).

O modelo pressupõe que a nossa experiência musical e vocal atualiza constantemente nossos modelos internos, construtos de como nosso cérebro internaliza e processa associações causais entre percepção e ação (Pfordresher; Halpern; Greenspon, 2015, p. 243). Um modelo interno que representa a relação causal entre as ações e suas consequências é chamado de modelo antecipatório, pois estima as

[1] Usaremos *pitch* como sinônimo de notas musicais no contexto deste trabalho.

aferências sensoriais futuras com base nas eferências motoras, ou seja, ele prevê como o estado do sistema motor mudará em função de um comando motor. Em vista disso, uma cópia de um comando motor descendente que atua sobre o sistema sensório-motor é transferida a um modelo antecipatório que atua como um simulador neural do sistema musculoesquelético movendo-se no ambiente (Wolpert; Pearson; Ghez, 2014, p. 648; Nalborczyk *et al*., 2020, p. 3), o que aconteceria da mesma forma com as pregas vocais. Por outro lado, um modelo interno que calcula as eferências motoras a partir das aferências sensoriais é conhecido como modelo inverso. Tal modelo pode determinar quais comandos motores são necessários para produzir os movimentos específicos requeridos para alcançar uma consequência motora desejada atuando assim como um controlador (Wolpert; Ghahramani, 2000, p. 1216; Nalborczyk *et al*., 2020, p. 3). O MAIM propõe que no modelo inverso, a imagética auditiva teria um efeito de *priming*[2] sobre a imagética motora. Por outro lado, a função do modelo antecipatório viria de associações na direção oposta, conforme apresentado na figura 1, a imagética motora agindo sobre a auditiva.

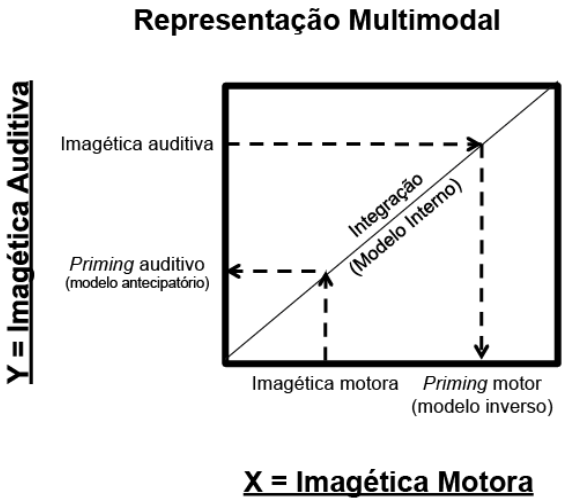

Fig. 1: Representação multimodal das associações ideais (relação linear) entre imagética e produção vocal em duas dimensões, sendo *pitch height* no eixo Y e a tensão laríngea, X. Adaptado de Pfordresher, Halpern e Greenspon (2015).

Os autores do MAIM referem-se à integração sensório-motora como o mapeamento de um determinado continuum sensório sobre um continuum motor. Em termos de acurácia de *pitch* temos as associações entre o controle motor fonatório que se dá através dos músculos intrínsecos da laringe no papel de regulação da tensão das pregas vocais e a altura percebida do som produzido.

Ao tratarmos então dos modelos antecipatório e inverso, observamos que as relações entre as modalidades seriam lineares, diretas, algo que não seria plausível na natureza. Mas, então o que seria essa falha na afinação vocal, mais especificamente o déficit de imitação de *pitch* (DIP)? Os autores

[2] Forma de memória implícita na qual o processo de apresentação de um item influencia o processamento do item seguinte, podendo facilitar o processamento deste item (*priming* positivo) ou dificultá-lo (*priming* negativo) (Baddeley; Eysenck; Anderson, 2015, p. 15)

especulam que haja alguma espécie de interferência (ou ruído) na associação entre as modalidades auditiva e motora. A figura 2 traz duas representações de como seriam essas associações deficitárias.

Especula-se que o DIP seja causado pela distorção do modelo inverso: para um indivíduo entoar uma melodia com acurácia é necessário que este seja capaz de mapear as notas escutadas de acordo com o controle muscular para gerar uma reprodução precisa dessas notas, seja a partir da memória de longo prazo ou imitando uma melodia, e se há algum tipo de falha neste caminho, o DIP acontece. Os desvios na afinação, segundo os autores, não têm a mesma chance de acontecer em todo o espaço de associações possível, mas há uma tendência maior de ocorrerem próximos a uma nota de conforto do cantor (como é mostrado no quadro B da figura 2).

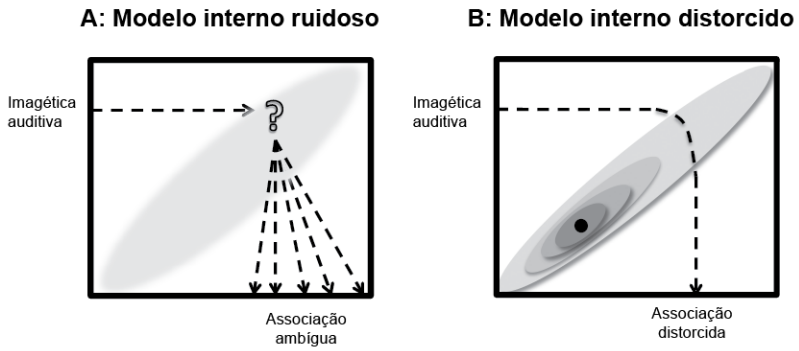

Fig 2: Representações hipotéticas das associações multimodais que podem contribuir para o déficit de imitação de *pitch*. Em A a relação não é confiável pois baseia-se num espaço equiprovável com muito ruído. Já em B, mesmo possuindo o espaço equiprovável, é incluída uma distribuição bivariada de probabilidades ao introduzir distorções sistemáticas em direção à moda da distribuição, a qual especula-se ser o *pitch* de conforto do cantor. Adaptado de Pfordresher, Halpern e Greenspon (2015).

Por conseguinte, de acordo com o MAIM, durante o desenvolvimento musical e vocal humano, dois fatores são críticos: (1) a integração de modalidades e (2) a generalização do mapeamento sensório-motor para novas vocalizações. O primeiro fator, integração, seria aquele que permite ao mapeamento interno representar relações conjuntas entre as modalidades auditivas e motoras. Sem a integração, todo o espaço é equiprovável, e nenhum mapeamento confiável é possível. O segundo fator, a generalização, é o que permite a um indivíduo ter um mapa sensório motor que não é limitado por apenas associações sensório-motoras. Diz-se que com o tempo, o mapeamento abstrato através das modalidades irá dominar a memória para mapeamentos experimentados através, possivelmente, de uma inferência bayesiana[3].

Dessa forma, foi proposto que a base da integração sensório-motora no canto, e possivelmente em outras tarefas de imitação motoras, envolve associações entre dimensões das imagens mentais baseadas na percepção, com dimensões correspondentes associadas com ações que poderiam reproduzir a percepção real.

[3] A inferência bayesiana é um método estatístico que utiliza o Teorema de Bayes para atualizar as probabilidades de hipóteses ou parâmetros à medida que novas evidências ou dados são observados. Ela se baseia na interpretação probabilística dos parâmetros, tratando-os como variáveis aleatórias com distribuições de probabilidade (distribuições a priori). A partir dos dados observados, a distribuição a priori é atualizada para uma distribuição a posteriori, que reflete o conhecimento combinado da informação prévia e dos dados (Gelman *et al*., 2013, p. 1, p. 6).

Sugere-se assim, que cantores com dificuldades de afinação não são capazes de criar imagens mentais vívidas (ou nítidas), pois há alguma interferência no processo de memória de trabalho (Greenspon; Pfordresher; Halpern, 2017, p. 598). Além disso, cantores imprecisos com relação à afinação recrutam músculos extrínsecos laríngeos ao executar subvocalizações (Pruitt; Halpern; Pfordresher, 2019, p. 8). Com o estudo do canto, e o desenvolvimento do modelo antecipatório, o indivíduo passa a não necessitar dessas manobras, como é visto que músicos mais experientes possuem maior flexibilidade em apoiar-se tanto nas imagens mentais quanto no *feedback* auditivo (modelo inverso), enquanto os novatos apoiam-se basicamente no feedback (Bishop; Bailes; Dean, 2014, p. 54).

Assim, como vimos, tanto na fala quanto no canto é necessária uma eficaz integração sensório-motora e as redes neurais que devem compreender as regiões cerebrais ativadas por sinais oriundos da laringe (fonte da produção vocal), do sistema auditivo periférico e central (*feedback* auditivo), das áreas responsáveis por um gerenciamento interno (*feedback* somestésico) e pelo processamento da linguagem formam um mapeamento complexo (Dalla Bella, 2015, p. 273). Neste sentido, no canto, a habilidade em realizar esses mapeamentos pode estar associada à criação de imagética auditiva e ser um fator adjuvante para a afinação do cantor em tarefas de imitação.

Especula-se que o treinamento auditivo-vocal pode favorecer o modelo inverso do controle motor. A vocalização seria mediada por previsões baseadas em exposições a contextos musicais e em vocalizações anteriores. Portanto, de acordo com o MAIM, a imagética auditiva poderia guiar o processo de tradução sensório-motora, isto é, a habilidade de converter as notas percebidas de uma melodia para um planejamento motor (Greenspon; Pfordresher; Halpern, 2017, p. 585). Supomos, portanto, que a prática da imagética auditiva anterior à vocalização deva contribuir para uma afinação mais precisa em tarefas de imitação vocal em contextos musicais distintos – familiar ou não-familiar – com destaque para o possível efeito adjutor no contexto não-familiar.

Assim, este trabalho teve como objetivo geral investigar se a criação de imagética auditiva auxilia o desempenho vocal em termos de acurácia de frequência fundamental (F_0), em contextos musicais tonais (familiares) e atonais (não familiares). Mais especificamente buscamos ainda avaliar se o "ensaio mental" contribui para a acurácia da afinação. Também procuramos verificar o impacto da imagética auditiva em diferentes contextos musicais (tonal e atonal) no desempenho de imitação vocal. E ainda buscamos analisar a correlação entre o tipo de contexto musical (familiar ou não familiar) e a habilidade de criação da imagética auditiva.

2. Método

Este estudo transversal possui um desenho quase-experimental. O teste foi conduzido de forma totalmente remota, por meio de plataformas online, desde o recrutamento dos voluntários até sua aplicação durante o ano de 2021. Este experimento foi aprovado pelo Comitê de Ética em Pesquisa da UFABC com parecer número 4.520.470.

O recrutamento de participantes ocorreu por meio das redes sociais e páginas institucionais do grupo de pesquisa NEUROMÚSICA UFABC e do Programa de Pós-Graduação em Neurociência e Cognição da Universidade Federal do ABC. Os voluntários se inscreveram para participar deste experimento através de formulário específico indicado no material divulgado e, na sequência, foi enviado um link por e-mail para que respondessem um questionário. Um total de 165 voluntários responderam integralmente o questionário, porém apenas 132 desses participaram da etapa experimental síncrona, com média de idade de 35,28 ± 10,57 anos (68% mulheres). Essa amostra foi composta de 47 cantores profissionais, 39 cantores amadores e 46 não-cantores.

2.1 Estímulos, instrumentos e tarefas

Tarefa de imitação vocal

A tarefa foi programada no software livre "PsychoPy2" (Peirce *et al*., 2019) e apresentada na plataforma online "Pavlovia", a qual possui integração. Uma vez que tal plataforma não permitia gravação de áudio, optou-se por realizar, paralelamente, uma sessão de gravação síncrona das vozes dos participantes através da plataforma *Cleanfeed* durante a apresentação dos estímulos no "Pavlovia".

Para este protocolo, foram compostas por um compositor integrante do grupo de pesquisa 60 melodias, utilizando-se para a escrita o software livre *MuseScore* 3. Dessas, 20 melodias apresentavam caráter atonal e as outras 40 melodias eram de caráter tonal (com 20 melodias em tonalidade maior e 20 em tonalidade menor). As melodias passaram por uma validação online utilizando o servidor *SoSci Survey*. Foram convidados 13 músicos profissionais, professores e estudantes de música de diversas cidades para avaliar se o contexto musical das melodias era tonal ou atonal. Das 60 melodias escutadas, 10 tiveram seu caráter julgado como ambíguo com mais de 50% das respostas consideradas discrepantes. Deste grupo de melodias ambíguas, apenas 7 melodias foram utilizadas no experimento. No total, 30 melodias de quatro notas foram apresentadas para todos os participantes, sendo dez tonais maiores, dez tonais menores e dez atonais. Além dessas, outras 4 melodias foram utilizadas apenas na fase de treinamento. Em todos os contextos avaliados (tonal maior, tonal menor e atonal), as melodias incluíram intervalos maiores e menores de 2ªs, 3ªs e 6ªs, 5ªs justas e 4ªs justas e aumentadas. Cada uma das melodias tinha duração de quatro segundos e quatro notas de igual duração.

A tarefa teve duas condições: I. vocalizar a melodia imediatamente após escutá-la, sem tempo de retenção para criação de imagética auditiva (SM) e II. vocalizar após tempo de retenção (4s) para criação de imagética auditiva (CM) da melodia apresentada.

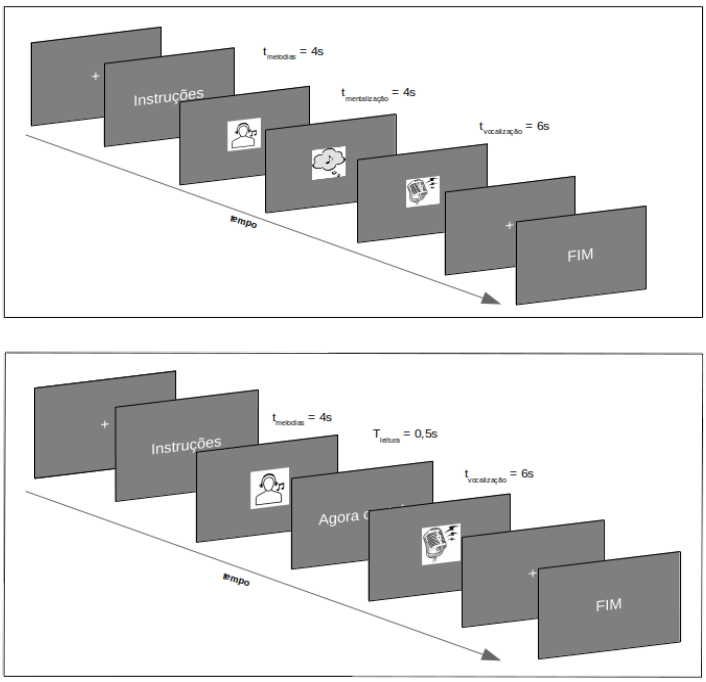

Fig. 3: Ordem das telas nas duas condições. Superior: com tempo para criação de imagética auditiva (CM); Inferior: sem tempo (SM).

Questionário

No questionário enviado aos voluntários após a manifestação de interesse em participação no experimento, utilizamos dois instrumentos: a Escala *Bucknell* de Imagética Auditiva e o Índice *Goldsmiths* de Sofisticação Musical. Os participantes também responderam questões sobre sua experiência vocal e musical e autopercepção corporal.

I) Índice *Goldsmiths* de Sofisticação Musical (*Goldsmiths Musical Sophistication Index*)

O Índice *Goldsmiths* de Sofisticação Musical foi desenvolvido como uma maneira de avaliar a musicalidade na população em geral ao observar diferenças amplas na sofisticação musical dos indivíduos (Müllensiefen *et al*., 2014). Seus criadores buscaram observar a população como um todo e não focar em populações específicas como amúsicos ou músicos profissionais. O instrumento foi adaptado para a língua portuguesa (Gold-MSI-P) em Portugal por Lima *et al*. (2018) e é esta versão que utilizamos neste estudo.

O índice contém 39 itens de autorrelato: 31 em formato de escalas *likert* e 8 questões que constituem 5 subfatores — Engajamento ativo (*Active engagement* – AE; 9 itens), Emoção (*Emotion* – E; 6 itens), Habilidades de canto (*Singing abilities* – SA; 7 itens), Habilidades perceptuais (*Perceptual abilities* – PA; 9 itens), Treinamento musical (*Musical training* – MT; 7 itens) – além de um fator comum, a Sofisticação Musical Geral (*General musicality*; 18 itens), indicado pela correlação entre os 5 subfatores. Na construção do instrumento, observou-se que indivíduos com uma maior sofisticação musical são capazes de responder de forma flexível a uma variedade de situações que envolvem música, seja na escuta ou na prática musical.

II) "Escala *Bucknell* de Imagética Auditiva" (*Bucknell Auditory Imagery Scale*)

A Escala *Bucknell* de Imagética Auditiva (EBIA) foi desenvolvida por Andrea Halpern (2015) para avaliar como indivíduos criam imagens auditivas. Três domínios da experiência auditiva são contemplados neste instrumento: musical, verbal e ambiental. O instrumento EBIA está dividido em duas subescalas, EBIA-V[4] que avalia a nitidez das imagens auditivas e EBIA-C que qualifica o controle de imagens auditivas. Halpern explica que a dimensão de controle de imagens mentais foi proposta por Gordon (1949) quando este criou o *Test of Visual Imagery Control*. Contudo, neste teste não há questões relativas ao domínio auditivo.

Na subescala de nitidez, EBIA-V, cada questão apresenta uma situação seguida da descrição de um som relevante àquela circunstância e o participante deve responder em uma escala *likert* de 1 (sem imagem mental) a 7 (imagem auditiva vívida como som real). Os itens foram desenvolvidos de forma a gerar três componentes principais sendo a Componente 1 com itens envolvendo diferentes ambientes e seus sons, como uma sala de aula, ouvir rádio dentro do carro etc. Na Componente 2, encontramos itens envolvendo música e dança. Já a Componente 3 possui itens relacionados a vozes, independentemente de ser em um contexto musical ou em diferentes situações. A outra subescala, EBIA-C, também possui três componentes principais, sendo a Componente 1 com itens verbais e ambientais, exceto por um item musical. A componente 2 possui todos os itens relacionados à música, exceto o item que contém a música *Happy Birthday*, que nas duas escalas localiza-se mais próximo ao conteúdo verbal, e um item mais ambiental. A componente 3 possui, em sua maioria, itens relacionados à dimensão verbal. Também em forma de escala *likert* de 7 pontos, os participantes devem responder de 1 (sem imagem mental) a 7 (muito fácil de mudar a imagem).

A autora pondera que até o momento da publicação do trabalho com a escala, os dados disponíveis não permitiam dizer se controle e nitidez seriam domínios separados do autorrelato de imagens mentais, mas que a alta correlação entre as duas subescalas e sua idêntica correlação com uma escala de imagética já existente[5], deve-se talvez ao fato de que cada item de controle deve iniciar-se com uma imagem nítida. Como o instrumento não havia sido traduzido para a língua portuguesa até a data deste estudo, realizamos uma tradução e algumas adaptações se mostraram necessárias. Em nossa tradução, fizemos adaptações culturais para nosso contexto, como a própria autora sugere.

Procedimento

Após inscrição por e-mail, os voluntários receberam o link para responderem o questionário na plataforma *SoSci Survey*. Uma vez finalizada essa etapa, foi agendada uma chamada de voz via *Cleanfeed* para realização da parte prática, com a tarefa de vocalização das melodias. Instruções também foram

[4] Do inglês, *vividness*.
[5] *Vividness of Visual Imagery Questionnaire* (VVIQ) publicado em Marks (1973).

enviadas ao participante para que no dia da coleta síncrona de voz estivesse em um ambiente silencioso e que usasse fones de ouvido e microfones em seu computador, além de silenciar outros programas e aparelhos que pudessem gerar algum tipo de som e distração.

Na data e horário agendados, a ligação era estabelecida via *Cleanfeed*. Caso necessário, ajustes como a regulagem do volume de microfone eram então realizados. Em seguida, era enviado o link (gerado na hora) para realização do teste no *Pavlovia* e pedia-se para o participante aguardar até que o download de todos os recursos finalizasse. Uma vez carregado o experimento, o participante era instruído sobre a tarefa e eventuais dúvidas eram esclarecidas. Foi expressamente solicitado aos participantes que mantivessem abertas apenas as duas abas/janelas, a do *Pavlovia* e a do *Cleanfeed*, durante a realização do teste. Na sequência, a gravação de voz era iniciada. A duração do teste de imitação vocal era de aproximadamente oito minutos e, no total, a sessão durava em média aproximadamente 16 minutos.

A sessão foi dividida em dois blocos, cada um com quinze melodias, e as condições foram pareadas nos dois blocos, sendo que no primeiro bloco havia oito *trials* na condição CM, enquanto no segundo havia sete. A ordem de apresentação das melodias, bem como a condição em cada *trial*, foram sorteadas previamente pela função *random* no Python3. Antes de iniciar, os participantes realizaram um breve treinamento com quatro melodias similares às apresentadas nas condições experimentais, sendo dois *trials* na condição CM e dois *trials* na condição SM.

2.2 Forma de análise dos dados

Adotamos a afinação temperada igual (Menezes, 2014, p. 264) e a variação de F_0 foi calculada em *cents*. A plataforma *Cleanfeed* gera gravações em um canal (mono) no formato *Waveform Audio File* (wav) a uma taxa de amostragem de 48kHz e *bitrate* 768 kbps. Cada participante teve uma gravação única de todas suas 34 melodias executadas, incluindo as do treinamento. Na pré-análise, foram retiradas manualmente de cada arquivo as melodias de treinamento, bem como ruídos, cliques e demais sinais que pudessem interferir na análise automatizada. Com os arquivos limpos, foi utilizada a biblioteca *pydub* para separar as 30 melodias de cada arquivo, adotando níveis abaixo de −18 dBFS por mais de 1 s como critério de corte. Nesta fase, foram excluídas as melodias que tinham sido cantadas com menos de quatro notas.

Com as melodias de cada participante separadas, utilizamos um programa desenvolvido em Python que primeiramente extrai as notas das partituras geradas no MuseScore 3 através do *toolkit Music21* (Cuthbert & Ariza, 2010) para, em seguida, ler os arquivos de gravação dos participantes utilizando o algoritmo de extração de *pitch* pYIN[6] (Mauch & Dixon, 2014). Este algoritmo seleciona pares com a nota alvo e sua probabilidade, sendo a diferença entre o *pitch* alvo e o observado definida como desvio. O pYIN divide o espaço (ou faixa) de *pitches* em M = 480 faixas (*bins*) contendo 4 oitavas de 55Hz (A1) a 880Hz (A5) em passos de 10 cents (0,1 semitom) e cada faixa de *pitch* é modelada utilizando um Modelo Oculto

[6]Sigla para *Probabilistic Yin*.

de Markov[7] de forma que as probabilidades dos candidatos a *pitch* obtidos pela autocorrelação são dadas como probabilidades de observação. Dessa maneira, temos a distribuição de probabilidades de cada candidato.

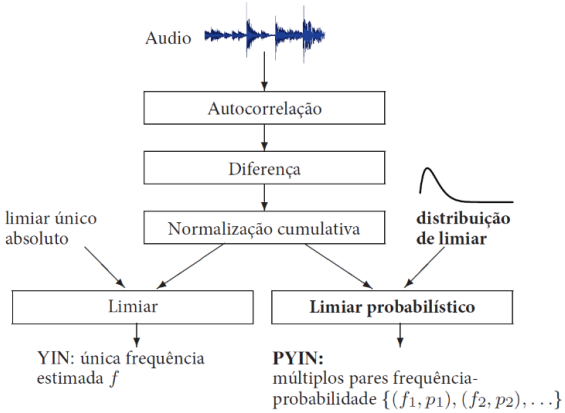

Fig 4: Comparação entre os algoritmos YIN e pYIN, cuja principal modificação se dá no resultado, com o algoritmo original, YIN, estimando uma única frequência f, e no pYIN são estimados os pares frequência-probabilidade, o que traz maior precisão. Adaptado de Mauch e Dixon (2014).

Contudo, como a amostra era muito heterogênea e incluía desde participantes sem nenhuma experiência no canto até cantores profissionais, observou-se uma distribuição dos desvios que podia variar entre -3000 e 3000 *cents*, aproximadamente, o que significa que alguns participantes vocalizaram entre duas e três oitavas abaixo ou acima do tom apresentado e, para aproveitar todos os dados, utilizamos a estatística circular. Para isso, cada desvio de cada nota foi transformado em componentes trigonométricos (x, dado pelo cosseno; y dado pelo seno) (Equação 1), dentro do período de 1200 *cents* (total de *cents* dentro de uma oitava):

$$x = cos\left(2\pi \frac{desvio}{1200}\right), y = sin\left(2\pi \frac{desvio}{1200}\right) \quad (1)$$

Em seguida, transformaram-se os valores de x e y novamente para *cents*, de forma que esses componentes trigonométricos nos fornecem agora os desvios de F_0 em valores absolutos de *cents* (em uma escala de 0 a 600), suprimindo assim a dimensão do *pitch height*. Para isso, utiliza-se uma outra função trigonométrica, o arco tangente ou tangente inverso das coordenadas x e y, dada por $\tan^{-1}2(y, x)$. O arco tangente é o ângulo entre o eixo x e uma linha que contém a origem (0, 0) e um ponto com coordenadas (x,y). O ângulo é fornecido em radianos no intervalo $]-\pi, \pi]$. Finalmente, para expressar em *cents* é multiplicado o resultado por $600/\pi$. E, por fim, o resultado é convertido em valor absoluto para evitar estatística circular (Equação 2):

[7]Modelo estatístico que descreve um sistema composto por estados ocultos e observações visíveis. Ele é definido por dois processos estocásticos: (1) um processo de Markov oculto, que modela a transição entre estados através de uma matriz de transição, e (2) um processo de emissão, que determina a probabilidade de gerar uma observação a partir de um estado. O HMM é caracterizado por um conjunto de estados, observações, uma matriz de transição, uma matriz de emissão e um vetor de probabilidades iniciais. É amplamente utilizado em áreas como reconhecimento de fala, processamento de linguagem natural e bioinformática (Rabiner, 1989, p. 257).

$$valor\ (cents) = \left|tan^{-1}2(y,x) \times \left(\frac{600}{\pi}\right)\right| \qquad (2)$$

Para verificar a existência ou não de correlação entre as variáveis, de acordo com os nossos objetivos e hipóteses, realizamos a correlação de Spearman. Para comparar a afinação dos participantes nos contextos tonal e atonal e nas condições SM e CM foi realizado o teste não-paramétrico de postos sinalizados de Wilcoxon. Para verificar a significância da diferença específica entre os contextos tonal maior, tonal menor e atonal, foi realizado o teste não-pa- ramétrico de Friedman seguido do teste post-hoc de Nemenyi. As análises foram realizadas em Python3 utilizando as bibliotecas[8] específicas de cálculo e estatística.

3. Resultados

Com relação ao objetivo geral deste estudo (avaliar se a criação da imagética, leia-se o tempo de retenção da melodia para a criação da imagética auditiva, contribui para o desempenho em termos de acurácia na afinação), observamos que o melhor desempenho em termos de afinação na tarefa de imitação vocal se deu nos trials SM (figura 4). O teste não-paramétrico de Wilcoxon indicou que a diferença entre as condições SM e CM dadas pelas médias dos DF_0 é estatisticamente significante (Z = 1032, p < 0,001), bem como da proporção de erros (Z = 897, p < 0,001).

Ao observarmos o desempenho de cada grupo vemos que, de maneira geral, como esperado, cantores profissionais tiveram um melhor desempenho nas tarefas de imitação, dada pelo desvio de F_0, com menor taxa de erro. Notamos ainda que nos três grupos o desempenho nas tarefas de imitação vocal foi melhor na condição SM do que na condição CM (figura 5). Ressaltamos aqui que a proporção de erros foi mais alta nos três grupos pois é uma medida mais conservadora do que o próprio DF_0.

Ao compararmos a afinação nos contextos tonal e atonal (Figura 6), não observamos diferença significativa entre DF_0 em cada um no teste de Wilcoxon (Z = 3907, p = 0,94), mas sim para a proporção de erros (Z = 1989,5, p < 0,00). Por outro lado, observamos que a afinação foi melhor nos trials com melodias tonais maiores ao compararmos com tonais menores e atonais nos três grupos, com significância dada pelo Teste de Friedman tanto em termos de DF_0 ($\chi 2$ = 11,58, p = 0,003) quanto na proporção de erros ($\chi 2$ = 22,73, p < 0,00).

[8] Bibliotecas *pandas, numpy, seaborn, matplotlib, scipy, pingouin, statsmodels.*

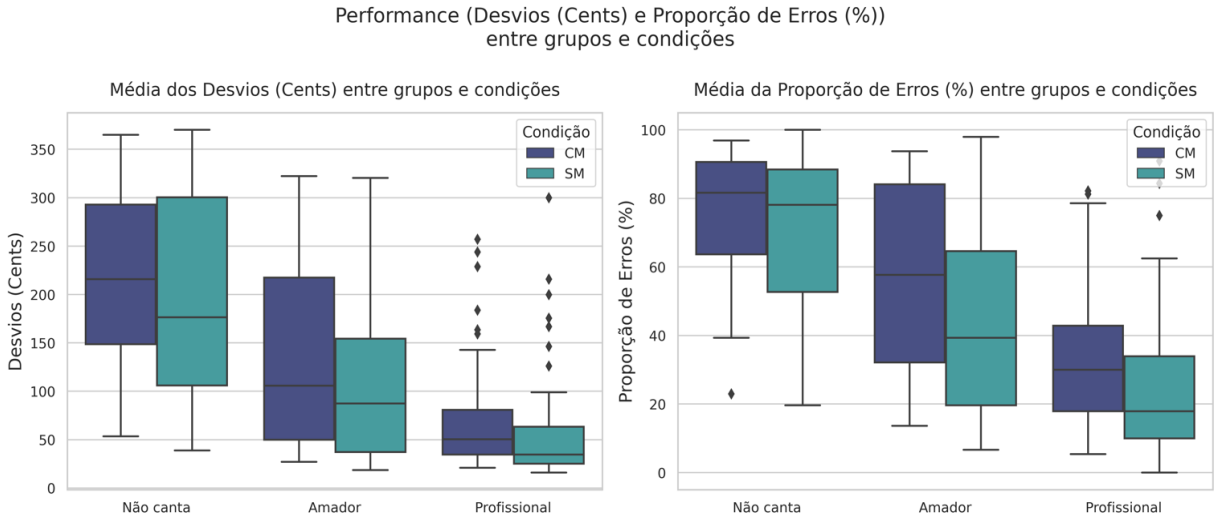

Fig. 5: Comparação do desempenho entre grupos dados pelos desvios de F_0 em *cents* (à esquerda) e pela proporção de erros (à direita). CM = com tempo de mentalização; SM = sem tempo de mentalização.

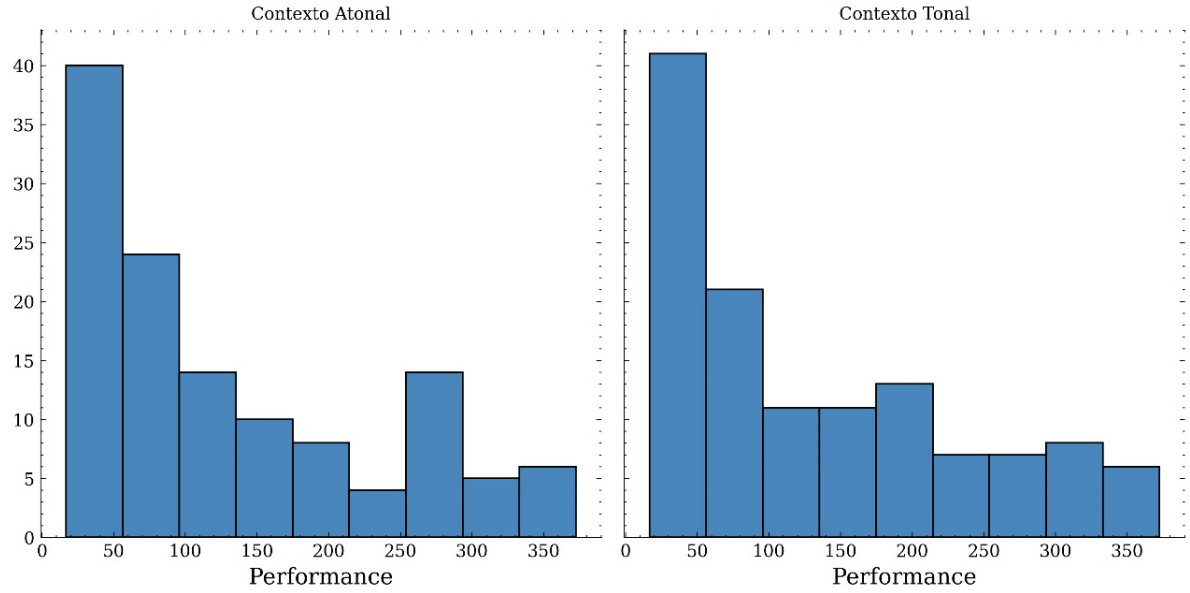

Fig. 6: Distribuição dos desvios da frequência fundamental DF_0 (dados como performance, sendo frequência absoluta) em cada contexto.

A análise post-hoc de Nemenyi indicou diferenças significativas entre o contexto maior e menor em DF_0 (p = 0,002) e na proporção de erros (p = 0,023), e entre os contextos maior e atonal em DF_0 (p = 0,042) e na proporção de erros (p = 0,001), não havendo diferenças estatisticamente significativas entre melodias menores e atonais em DF_0 (p = 0,636) e na proporção de erros (p = 0,099).

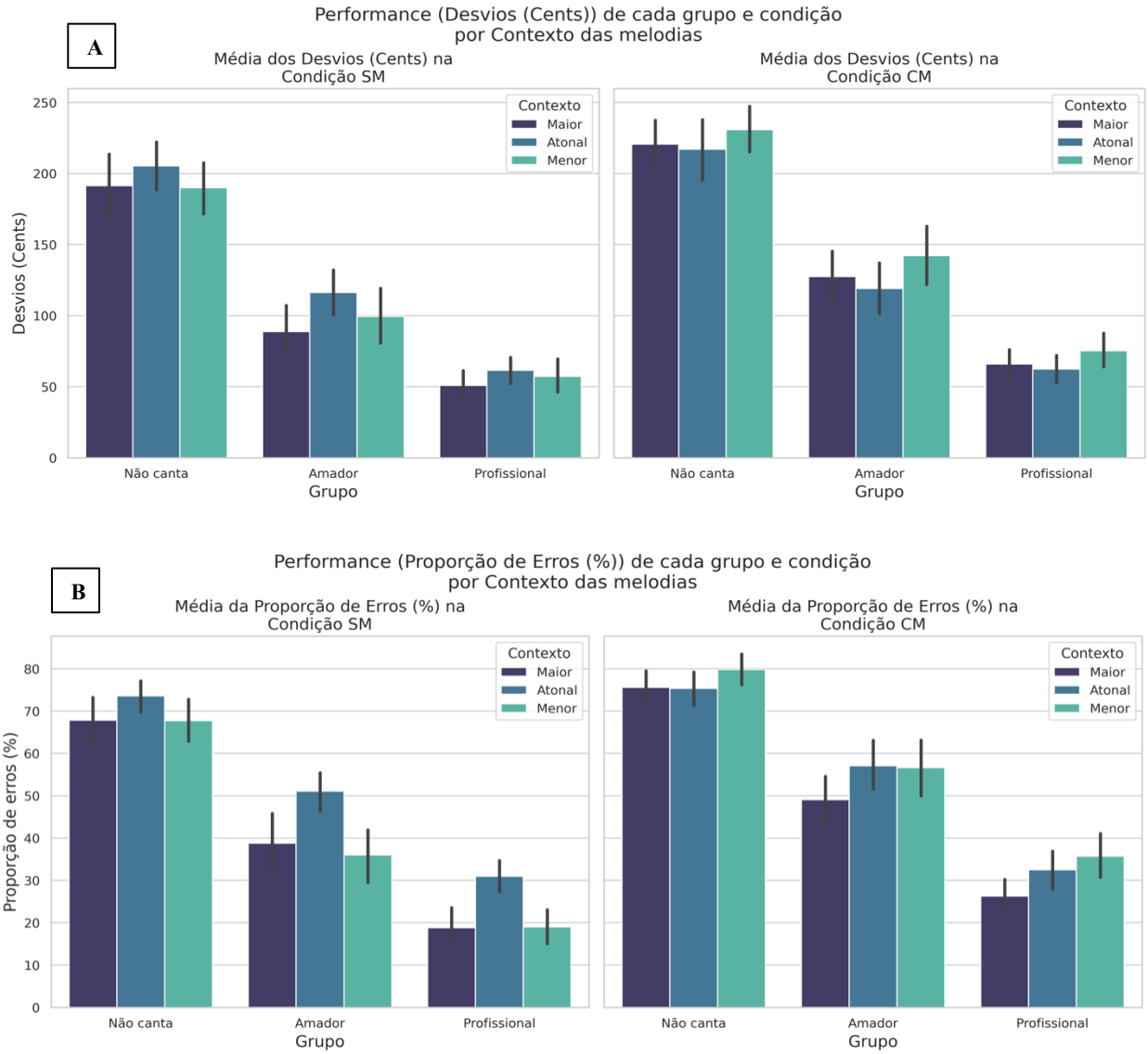

Fig. 7: (A) Na parte superior da figura, temos o desempenho de cada grupo em cada contexto musical na condição sem tempo de mentalização (SM) (à esquerda) e na condição com tempo de mentalização CM (à esquerda) dado pelos desvios em *cents*. Na parte inferior da figura, em (B), temos a proporção de erros de cada grupo em cada contexto na condição SM (à esquerda) e CM (à direita).

Como a amostra não segue uma distribuição normal, utilizamos o teste de Kruskal-Wallis para verificar as diferenças entre os grupos nas variáveis analisadas. Para os casos com efeito significativo ($p < 0,001$), foram realizadas comparações entre os pares de grupos. Observamos que o grupo de profissionais

apresentou maior habilidade de imagética, tanto de controle quanto de nitidez, e maior musicalidade geral e de suas habilidades, como vemos na tabela 1, a seguir:

Habilidades	Não-cantores	Amadores	Profissionais	Pontuação máxima possível
Engajamento Ativo*	3,73 (0,98)	4,36 (0,99)	5,05 (0,92)	7
Habilidades Perceptuais*	5,04 (0,91)	5,60 (1,01)	6,48 (0,82)	7
Treinamento Musical*	3,09 (1,42)	4,43 (1,21)	5,86 (0,55)	7
Habilidades Canto*	3,47 (0,88)	4,69 (1,23)	5,68 (0,88)	7
Emoções*	5,66 (0,76)	5,77 (0,98)	6,23 (0,79)	7
Musicalidade Geral*	3,68 (0,74)	4,85 (0,98)	5,81 (0,64)	7
Controle Imagético*	5,30 (1,46)	5,77 (1,14)	6,09 (0,99)	7
Nitidez Imagética*	5,22 (1,51)	5,49 (1,21)	5,72 (1,04)	7
Taxa de Erros	0,73 (0,29)	0,48 (0,38)	0,27 (0,31)	0**
Média desvios F_0	209,83 (127,83)	116,56 (115,116)	62,73 (62,73)	0**

**Desvio e taxa de erros mínimos

Tab. 1: Escores médios de cada grupo em cada habilidade (*) e desempenho (taxa de erros % e desvio de F_0 em *cents*) na tarefa de imitação. Desvios-padrão indicados entre parênteses (Dp). Teste de Kruskal-Wallis para as variáveis (p <0,001).

Observamos então que a subescala de controle imagético se correlacionou moderadamente com DF_0 ($\rho = -0{,}368, p \leq 0{,}0$) e com a taxa de erros ($\rho = -0{,}383, p \leq 0{,}0$). Já para os subitens de musicalidade também obtivemos correlação moderada tanto entre o engajamento ativo e DF_0 ($\rho = -0{,}308, p \leq 0{,}0$) como com a taxa de erros ($\rho = -0{,}342, p \leq 0{,}0$). As correlações entre as habilidades perceptuais e DF_0 ($\rho = -0{,}459, p \leq 0{,}0$) e proporção de erros ($\rho = -0{,}491, p \leq 0{,}0$) também se mostraram moderadas, mas um pouco mais significativas. As correlações mais significativas foram observadas entre o treinamento musical e DF_0 ($\rho = -0{,}560, p \leq 0{,}0$), e taxa de erros ($\rho = -0{,}584, p \leq 0{,}0$). Observamos correlações significativas também entre as habilidades de canto e DF_0 ($\rho = -0{,}537, p \leq 0{,}0$) e proporção de erros ($\rho = -0{,}568, p \leq 0{,}0$). Isso foi encontrado para o índice de sofisticação musical geral e DF_0 ($\rho = -0{,}559, p \leq 0{,}0$) e proporção de erros ($\rho = -0{,}592, p \leq 0{,}0$). Por outro lado, as habilidades perceptuais tiveram uma correlação positiva significativa com a subescala de controle imagético ($\rho = 0{,}562, p \leq 0{,}0$) e moderada com a subescala de nitidez imagética ($\rho = 0{,}436, p \leq 0{,}0$).

4. Discussão

Neste experimento, investigamos se a imagética auditiva, criada a partir da retenção de uma melodia na memória por um curto intervalo de tempo, é um fator que contribui para o desempenho de afinação. Avaliamos também se a criação da imagética auditiva em contextos musicais distintos – tonal maior, tonal menor e atonal (este último menos familiar à maior parte da população) – se correlaciona com o desempenho em termos de afinação nas tarefas de imitação vocal. Buscamos avaliar, igualmente, se maior afinação dada pela acurácia na imitação das melodias (medida pelo desvio de F_0 em *cents)* se correlaciona com maior índice de musicalidade e de suas habilidades e se a criação de imagética auditiva após a escuta

das melodias se correlaciona com índice de musicalidade e de suas habilidades.

Os resultados revelaram correlação negativa moderada entre habilidades perceptuais e DF_0, e uma relação mais significativa com a taxa de erros. Em outras palavras, quanto mais desenvolvidas as habilidades de percepção, melhor o desempenho de afinação na tarefa de imitação vocal. Essa associação também ocorre entre as habilidades de canto e o nível do treinamento musical. Essas associações, ainda que não tão fortes, nos dão pistas da alimentação dos modelos internos através da prática. Interessantemente, a correlação entre o desempenho de afinação na tarefa de imitação vocal com o engajamento ativo é mais fraca que os atributos anteriores (habilidades perceptuais, de canto e treinamento musical), o que nos indica que a mera exposição e o envolvimento passivo de um indivíduo em atividades musicais talvez não sejam suficientes para capacitá-lo a atualizar eficientemente seus modelos internos com relação à integração sensório-motora. Por outro lado, vimos que a sofisticação musical geral, avaliada pelo índice de sofisticação da bateria Gold-MSI-P, apresentou a maior correlação negativa com os desvios e taxa de erros como era de se esperar. Em outras palavras, quanto maior o nível de musicalidade geral, melhor o desempenho de afinação na tarefa de imitação vocal. O termo "musicalidade" envolve diversas concepções e é problematizado na literatura (Müllensiefen *et al*., 2014, p. 2). Tem-se discutido amplamente se habilidades musicais podem ser herdadas, inatas ao indivíduo, ou se podem ser desenvolvidas pela exposição ao ambiente e às experiências de vida. Segundo Honing *et al*. (2015, p. 2), a musicalidade pode ser definida como um conjunto de traços que se desenvolvem naturalmente de forma espontânea e é baseada em e limitada por nossos sistemas cognitivos e biológicos.

Em relação à comparação do desempenho de afinação vocal nos diferentes contextos, vimos que, de maneira geral, não houve diferença significativa entre os contextos tonal e atonal, tanto no cálculo dos desvios como da taxa de erros, mas observamos que o desempenho foi melhor nos trials com melodias maiores, comparadas com menores e atonais.

As análises *post-hoc* revelaram diferenças significativas entre contextos maior vs. menor ($p < 0,01$) e maior vs. atonal ($p < 0,05$), mas não entre menor e atonal ($p = 0,12$). Essa disparidade pode ser explicada pela hierarquia tonal (Krumhansl, 2001, p. 18), já que o modo maior é estruturalmente mais estável e culturalmente mais familiar, facilitando a antecipação melódica (Huron, 2006, p. 55). No caso de cantores, conjecturamos que há maior ocorrência de exercícios de vocalise com padrões tonais (tanto em modo maior quanto menor) nas aulas de canto. E, mesmo os indivíduos não-cantores, estão mais expostos a músicas tonais, mas não a músicas atonais. Este ponto aqui nos é caro pois, segundo David Huron (2006, p. 348), se um ouvinte aplicar um esquema tonal para sua escuta, mesmo se a melodia que está ouvindo for atonal, então sua predição de notas seguintes irá necessariamente falhar pois haverá uma quebra da expectativa daquela melodia. A explicação para isso se dá pelo fenômeno da hierarquia tonal que, segundo Krumhansl (2001, p. 63), enquanto escutamos uma música, o nosso cérebro analisa estatisticamente os componentes tonais da escala e calcula as chances de cada nota aparecer. Ela explica que na estrutura tonal, cada nota da escala tem um peso, no sentido de ter mais ou menos chance de aparecer em um determinado contexto (a tônica, por exemplo, é a nota que ocorre com maior frequência no decorrer de

uma obra tonal). Por outro lado, na música atonal não há uma hierarquia definida e, sobretudo na escrita atonal dodecafônica, todos os 12 tons dentro de uma oitava têm chances iguais de aparecer.

Assim, podemos pensar então na noção de expectativa e antecipação das notas de uma melodia e quanto isso pode afetar o desempenho de afinação numa tarefa de imitação de melodias, como a nossa. Segundo Huron (2006, p. 43), as expectativas que fazemos ao escutar uma música são construídas através de circuitos neurais cujas ativações dependem do padrão das entradas sensoriais. Na prática, esses circuitos facilitam certos comportamentos motores, através da integração sensório-motora, e/ou causam o direcionamento da atenção em determinadas maneiras, e podem ainda estar associados a sentimentos motivadores. Se há uma predição de eventos que evocam respostas positivas então uma recompensa límbica de valência positiva será iniciada. As redes que falham em predizer os estados através desse reconhecimento de padrões atrofiam enquanto aquelas que obtém sucesso em predizer são mantidas e reforçadas. Então, como em nosso experimento as melodias foram embaralhadas e sorteadas de forma que não há um padrão lógico para a exposição do contexto, a quebra de expectativa acontece a todo momento e o sujeito não consegue antecipar o próximo passo. Talvez este seja um aspecto de confusão em nosso desenho. Relacionado ainda à expectativa, precisamos mencionar que a orientação diatônica presente na música ocidental baseia-se nesta espécie de "Gestalt tonal" das mensagens musicais, uma tentativa de completar as frases musicais com notas que correspondem àquele contexto, e isso traria uma sensação de satisfação pela expectativa correspondida. Outro ponto é que em tarefas de imitação de padrões melódicos e de fala, cantores parecem exibir uma tendência em se manter nos padrões melódicos escutados anteriormente quando passados para um padrão de notas faladas, o que sugere uma maior saliência do contorno melódico (Wisniewski; Mantell; Pfordresher, 2013, p. 90) e marca assim sua influência nas expectativas seguintes.

Relacionado a essa questão, de fato, o MAIM não consegue prever a mudança de tom pois é mapeada uma nota cantada por vez. Esta é uma limitação deste modelo pois no canto propriamente dito temos uma sequência de sons no tempo que, como vimos, produzem um contínuo de tons, ou uma melodia. O fato de que cantores desafinados possuem desempenho superior em tarefas de imitação de melodias do que na imitação de tons sozinhos sugere que o mapeamento multimodal não é executado independentemente em cada ponto no tempo e pode ser sensível à informação relativa e não somente à altura do som (Pfordresher; Halpern; Greenspon, 2015, p. 247). Devemos ainda pontuar que a percepção de intervalos pode ser facilitada em determinadas tonalidades, e mesmo músicos não irão experimentar os intervalos melódicos da mesma maneira pois seu processo de aprendizagem pode diferir (Huron, 2006, p.120). Assim, a noção de familiaridade do contexto pode estar relacionada à presença de determinados intervalos e isso pode influenciar as predições feitas pelos modelos internos do ouvinte nas próximas notas da melodia. Apesar de sua relevância, uma análise detalhada das sequências intervalares utilizadas em nossas melodias não foi realizada neste trabalho, mas será feita em momento posterior.

Observamos que o tempo de retenção na condição CM, de mesma duração da melodia escutada (t = 4 s), pareceu atrapalhar o desempenho em vez de ajudar, como sugerem os resultados que mostram melhor acurácia nas tentativas SM. Além disso, a proporção de erros foi muito mais baixa

consistentemente nos três grupos nesta condição com diferença estatística ao compararmos com a condição CM. De acordo com o MAIM, a formação da imagética auditiva ocorre durante a escuta (Pfordresher; Halpern; Greenspon, 2015, p. 247). Assim, supomos que a CM levou à perda de atenção e à deterioração da precisão na imitação vocal. De fato, a memória de trabalho tem relação intrínseca com as funções executivas, incluindo a atenção (Baddeley, 2012, p. 23). Estudos recentes corroboram a relação entre memória de trabalho e processamento melódico. Especula-se que cantores usam mais a expectativa dentro do contexto tonal do que a própria memória de trabalho (Ammirante; Russo, 2023, p. 9) no caso de melodias tonais.

De fato, ao olharmos para cada grupo deste estudo, o desempenho na afinação de profissionais foi muito superior ao de amadores e não-cantores, como esperado, mas os achados recentes nos dão pistas da relação entre mudanças de altura e memória de trabalho. Apenas é pertinente apontarmos que a proporção de erros total de profissionais foi relativamente alta na condição CM, algo que não estávamos esperando. Especulamos que causas relacionadas a fatores como ansiedade com o próprio desempenho podem ter impactado esse grupo pois, para profissionais, pode haver uma cobrança maior em obter um resultado "satisfatório" e correções demasiadas podem ocorrer, como ajustes extras, algo que justamente impediria as predições do modelo antecipatório, aumentando o erro no modelo inverso. Como a condição CM também apresentou maior proporção de erros nos outros dois grupos comparados à outra condição, isto poderia, além do maior erro no modelo inverso, ter contribuído para a dificuldade em se compreender a tarefa mesmo após a realização dos trials de treinamento e tempo para esclarecimento de dúvidas. De fato, a proporção de erros dos não-cantores é altíssima na condição CM. Isto é, a variabilidade vista nos desvios dos não-cantores é muito maior, o que sugere que cantores com pouca ou nenhuma experiência utilizam justamente mais a memória operacional para sons, apoiando-se somente no modelo inverso e possuem pouca vivência musical de forma a realizar antecipações das notas e atualizar seu modelo antecipatório, algo que pode melhorar com a prática ainda que amadora.

Ao analisarmos os escores da Escala Bucknell de Imagética Auditiva, nota-se que o controle imagético, dado pela subescala EBIA-C, correlaciona-se com os subíndices de musicalidade e apenas observamos correlação positiva significativa entre as habilidades perceptuais com a subescala de controle imagético e correlação positiva moderada com a nitidez imagética dada pela subescala EBIA-V. Na escala EBIA, profissionais destacaram-se no controle imagético, enquanto não cantores e amadores não diferiram significativamente. Observamos, então, que a subescala de controle imagético se correlaciona fraca e negativamente com a DF_0 e com a proporção de erros. Há uma relação, ainda que não robusta, entre menor proporção de erros e DF_0, com maiores escores na habilidade em modificar as imagens auditivas. Alguns estudos mostram associação da subescala de controle com desempenho na afinação (Gelding; Thompson; Johnson, 2015, 2019; Greenspon; Pfordresher; Halpern, 2013), enquanto outros a de nitidez (Greenspon; Pfordresher; Halpern, 2013; Gelding; Thompson; Johnson, 2019). De fato, o MAIM pressupõe que a imagética seria o centro das associações sensório-motoras que levariam a habilidades como a imitação vocal e outras que também residem em associações intermodais (Pfordresher, Halpern & Greenspon, 2015),

processos mentais fortemente ligados ao corpo, às relações sociais, ao ambiente e às experiencias sensório-motoras e à ação em si (Iachini, 2011, p. 63).

Podemos ainda especular que então o mapeamento é distorcido, semelhante a uma das simulações feitas no estudo original do MAIM. Nossos resultados sugerem uma relação entre a imagética e o desempenho em termos de afinação, apesar de correlações fracas encontradas. Em outros estudos, há uma correlação positiva entre nitidez de imagética auditiva autodeclarada e acurácia de imitação de altura (Pruitt; Halpern; Pfordresher, 2019, p. 8). Especulamos que cantores mais experientes, treinados formalmente ou não, e pessoas com uma sofisticação musical mais desenvolvida, possuam habilidade de criação imagética mais refinada, com imagens mais nítidas e mais controle em transformações imagéticas.

Como foi mencionado, Bruder e Wöllner (2019, p. 3) sugerem que cantores que possuem acurácia maior do mapeamento das imagens auditivas para imagens motoras – suas imagens mentais possuem maior nitidez – subvocalizam mais. Os autores explicam que a subvocalização seria uma característica intrínseca do loop fonológico descrito no modelo de memória de trabalho de Baddeley, um dos sistemas escravos que acompanham as funções executivas. O *loop* fonológico está envolvido no processamento da informação auditiva e é composto por dois componentes: um armazenamento fonológico que segura a informação acústica ou falada e um processo articulatório que registra visualmente o material guardado no armazenamento fonológico pela subvocalização. Segundo Hubbard (2018, p. 17), a presença de fala irrelevante pode interferir na lembrança ao afetar o armazenamento fonológico, enquanto a articulação irrelevante tende a prejudicar a recodificação dessa informação. Essa abordagem levou pesquisadores a sugerirem que a voz interior e o ouvido interior seriam processos distintos e que a informação somestésica na subvocalização contribuiria para a imagética auditiva.

No entanto, Pruitt, Halpern e Pfordresher (2019, p. 10) encontraram justamente um resultado oposto, no qual cantores mais experientes subvocalizam menos, sugerindo que com a prática as ações tornam-se mais eficientes e o movimento mais refinado. Os dados deles sugerem, portanto, que as pessoas usem as duas imagens ao mesmo tempo e que imitadores não acurados devam usar a imagética motora como uma maneira de tentar encenar uma vaga imagem auditiva. Segundo os autores do MAIM, a formação da imagética auditiva leva diretamente a uma imagética motora associada, com a força da associação sendo determinada pela acurácia e precisão do mapeamento sensório-motor (Pfordresher *et al*., 2015). De fato, lembramos que não-cantores obtiveram pior desempenho em termos de afinação, mas que a proporção de erros do grupo de profissionais foi relativamente alta na condição CM no contexto atonal, algo inesperado, enquanto não-cantores obtiveram desempenho melhor nesta mesma condição comparando-se com SM. Especulamos então o fato de que profissionais identificam o contexto musical como não-tonal e imediatamente tentam decodificá-lo, causando ajustes extras, algo que justamente barraria as predições do modelo antecipatório e aumentaria o erro no modelo inverso.

Cantores tendem a cantar em seu registro de conforto (Pfordresher; Halpern; Greenspon, 2015, p. 245) e Ammirante e Russo (2023, p. 8) apontam ainda que "cantabilidade" pode ser confundida com familiaridade, sendo mais difícil cantar *pitches* que ocorrem menos frequentemente, o que reforça o papel do mapeamento como mecanismo da expectativa melódica.

5. Limitações e perspectivas futuras

Este estudo propôs verificar possíveis associações entre imagética e a integração sensóriomotora na imitação do canto, ainda que de forma indireta. Com os dados apresentados pudemos estabelecer correlações entre esses dois aspectos. Contudo, há apenas pistas de como a imagética se insere nos processos dos modelos internos. As relações intervalares, que podem ter um papel importante no processamento das melodias, e consequente atualização nas predições dos modelos antecipatório e inverso, também não puderam ser analisadas em mais detalhes no escopo deste trabalho.

Procuramos controlar a apresentação dos estímulos, duração, timbre, ordem de apresentação, andamento, articulação, mas é indispensável apontar que o estudo da afinação no canto envolve uma série de problemas pela quantidade de variáveis envolvidas, seja na complexidade do estímulo como na familiaridade do estímulo, memória de trabalho, funções executivas como a atenção (Halpern; Pfordresher, 2022, p. 2) e, principalmente, habilidades perceptuais como o feedback auditivo (Loui *et al*., 2015). De fato, como os estímulos eram todos novos, melodias inéditas aos participantes, há um componente de sobrecarga cognitiva para tentar decodificar esses padrões que pode ter impactado no desempenho dos participantes. Este fator pode ser investigado com mais cuidado numa pesquisa futura, atentando-se ainda para o intervalo entre trials e entre blocos, de forma não cansar os participantes.

É importante destacar que, em muitos estudos com cantores amadores, as condições experimentais estavam longe de simular a realidade de uma apresentação no palco. Isso pode gerar estresse nos participantes, resultando em um baixo desempenho na afinação devido à ansiedade de se apresentar em um ambiente não familiar, sendo o estresse um fator que poderia impactar na afinação.

Devemos ainda pontuar as limitações desse experimento quanto ao fato de ser um estudo totalmente online. Tan *et al*. (2021) apontam, em recente revisão, a confiabilidade dos testes e medidas coletadas de forma remota na comparação de medidas feitas presencialmente e online em tarefas de imitação de notas isoladas e de melodias nas quais observaram uma heterogeneidade substancial no desempenho da afinação nos diferentes estudos.

Mesmo com as recomendações e requisitos mínimos de equipamento explicitamente expostos na divulgação da pesquisa e no questionário que os participantes responderam, não há como haver uma homogeneidade em termos de hardware, isto pode apresentar diferenças, ainda que mínimas na captação do sinal. É preciso, ainda, destacar o fato de que o nível de atenção dos participantes pode ter variado em função da possibilidade de terem realizado o experimento em ambientes compartilhados. Mais uma vez, as recomendações e requisitos para que o ambiente fosse silencioso sem foco de distrações e que a pessoa estivesse sozinha foram apresentados na divulgação e muitos estavam dividindo o espaço com outras pessoas já que o contexto do período de coleta era justamente de alta de casos de covid-19 no país, o que pode ter interferido na capacidade de concentração na tarefa ao impactar no estado de humor dos participantes.

6. Conclusões

Este estudo demonstra que a afinação vocal é mediada por múltiplos fatores, incluindo prática ativa, familiaridade tonal e eficiência na integração sensório-motora. O uso da imagética auditiva na afinação pode ter efeitos contraditórios dependendo do tempo de retenção e do nível de experiência musical. A dificuldade em contextos atonais ressalta a necessidade de treinamento específico para expandir a flexibilidade cognitiva de cantores, enquanto a relação modesta entre imagética auditiva e afinação sugere que outros mecanismos (ex.: memória de trabalho) merecem investigação futura.

A descoberta de que o tempo para mentalização prejudicou a afinação vocal sugere que há uma interferência cognitiva que impacta o desempenho dos cantores. Esse achado pode ser explicado pela sobrecarga da memória de trabalho. A retenção prolongada da melodia pode aumentar a demanda por recursos cognitivos, resultando em dificuldades na conversão da imagética auditiva em comandos motores precisos. Além disso, é possível que a ansiedade do desempenho tenha um papel relevante nesse processo, levando a ajustes excessivos e, consequentemente, a um maior índice de erro na afinação. Estudos futuros poderiam explorar essa relação testando diferentes tempos de retenção, como 1s, 2s, a fim de identificar um ponto de equilíbrio. O experimento, apesar de breve, demonstrou exigir alta carga cognitiva pela exigência da flexibilidade na mudança entre os contextos musicais, e talvez mais pausas fossem necessárias entre blocos de *trials*.

Os resultados também indicaram que melodias em modo maior foram cantadas de forma mais acurada do que as menores e atonais, o que pode ser explicado pela hierarquia tonal. Essa teoria sugere que ouvintes internalizam estatísticas musicais ao longo da vida, tornando certas sequências melódicas mais previsíveis e acessíveis para a reprodução vocal. Como vimos, a quebra de expectativas tonais pode aumentar a dificuldade no processamento cognitivo, exigindo maior esforço para a imitação. No contexto da pedagogia vocal, a exposição frequente a escalas tonais maiores pode contribuir para um processamento mais eficiente dessas melodias, enquanto a ausência de padrões tonais familiares nas melodias atonais pode dificultar sua retenção e reprodução. Esse achado levanta a possibilidade de que treinos vocais com material atonal possam ajudar a reduzir essa disparidade, aumentando a flexibilidade cognitiva dos cantores ao lidar com estruturas melódicas menos previsíveis.

A diferença de desempenho entre cantores profissionais, amadores e não-cantores sugere que a prática musical modifica a relação entre imagética auditiva e controle motor na afinação vocal. Enquanto os profissionais apresentaram um desempenho superior, os não-cantores se saíram melhor na condição sem tempo de mentalização, indicando que a estratégia cognitiva utilizada por cada grupo pode ser diferente. Cantores experientes podem se apoiar em modelos internos já estabelecidos, permitindo ajustes automáticos na afinação, enquanto não-cantores dependem mais do feedback auditivo imediato. Esse achado reforça a ideia de que a prática musical altera a forma como o cérebro processa a informação auditiva e a transforma em comandos motores, sugerindo que a imagética pode ser mais útil para cantores experientes que já possuem uma base sólida de controle motor. Estudos futuros poderiam investigar em que momento da formação vocal a imagética auditiva se torna uma ferramenta mais eficaz para a afinação.

Referências

AMMIRANTE, P., & RUSSO, F. A. Towards a Vocal Constraints Model of Melodic Expectancy: Evidence from Two Listening Experiments. *Music & Science*, *6*. 2023. https://doi.org/10.1177/20592043231179410.

BADDELEY, Alan. Working memory: Theories, models, and controversies. *Annual review of psychology*, v. 63, n. 1, p. 1-29, 2012.

BADDELEY, A., EYSENCK, M.W., & ANDERSON, M.C. *Memory* (2nd ed.). Psychology Press. 2015. https://doi.org/10.4324/9781003453536.

BISHOP, Laura; BAILES, Freya; DEAN, Roger T. Performing musical dynamics: How crucial are musical imagery and auditory feedback for expert and novice musicians? *Music Perception*: An Interdisciplinary Journal, v. 32, n. 1, p. 51-66, 2014. https://doi.org/10.1525/mp.2014.32.1.51

BRUDER, C. & WÖLNER, C. Subvocalization in singers: Laryngoscopy and surface EMG effects when imagining and listening to song and text. *Psychology of Music*. 2019. https://doi.org/10.1177/0305735619883681.

CUTHBERT, M. S. & ARIZA, C. "music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data." in J. Stephen Downie and Remco C. Veltkamp (Eds.). *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, August 9-13, 2010, Utrecht, Netherlands. p. 637-642. https://shorturl.at/au6y6.

DALLA BELLA, S. Defining poor-pitch singing: A problem of measurement and sensitivity. Music Perception: An Interdisciplinary Journal, 32(3), 272-282. 2015

GELDING, R. W., THOMPSON, W. F. & JOHNSON, B. W. The pitch imagery arrow task: effects of musical training, vividness, and mental control. *PLoS One* 10, e0121809, 2015. https://doi.org/10.1371/journal.pone.0121809.

GELDING, R; THOMPSON, W. F.; JOHNSON, B. W. Musical imagery depends upon coordination of auditory and sensorimotor brain activity. *Scientific Reports.* 9:16823. 2019 https://doi.org/10.1038/s41598-019-53260-9.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; DUNSON, D. B.; VEHTARI, A.; RUBIN, D. B. *Bayesian Data Analysis*. 3. ed. Boca Raton: Chapman and Hall/CRC, 2013.

GREENSPON, E. B., PFORDRESHER, P. Q., & HALPERN, A. R. *pitch* Imitation Ability in Mental Transformations of Melodies. *Music Perception: An Interdisciplinary Journal,* v. 34, n. 5, p. 585-604. 2017. https://doi.org/10.1525/mp.2017.34.5.585

HALPERN, A. R. Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain,* v. 25, n. 1, p. 37-47. 2015. https://doi.org/10.1037/pmu0000081.

HALPERN, A.R., PFORDRESHER, P.Q. What do less accurate singers remember? Pitch-matching ability and long-term memory for music. Atten Percept Psychophys 84, p. 260-269. 2022. https://doi.org/10.3758/s13414-021-02391-1.

HONING H, TEN CATE C, PERETZ I, TREHUB SE. Without it no music: cognition, biology and evolution of musicality. *Phil. Trans. R. Soc.* B 370: 20140088. 2015. http://dx.doi.org/10.1098/rstb.2014.0088.

HUBBARD, T. Some Methodological and Conceptual Considerations in Studies of Auditory Imagery. *Auditory Perception & Cognition. 2018*. https://doi.org/10.1080/25742442.2018.1499001.

HURON, David. *Sweet anticipation*: Music and the psychology of expectation. MIT press, 2006.

IACHINI, Tina. Mental imagery and embodied cognition: a multimodal approach. *Journal of Mental Imagery*, v. 35, n. 3-4, p. 1-66, 2011. https://shorturl.at/q6Uyv.

KRUMHANSL, Carol L. *Cognitive foundations of musical pitch*. Oxford University Press, 2001.

LIMA, C. F., CORREIA, A. I., MÜLLENSIEFEN, D., & CASTRO, S. L. Goldsmiths Musical Sophistication Index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality, and music preferences. *Psychology of Music*. 2018. https://doi.org/10.1177/0305735618801997.

LOTZE, Martin. Kinesthetic imagery of musical performance. *Frontiers in human neuroscience*, v. 7, p. 280, 2013. https://doi.org/10.3389/fnhum.2013.00280.

LOUI, P., DEMOREST, S. M., PFORDRESHER, P. Q., & IYER, J. Neurological and developmental approaches to poor pitch perception and production. *Annals of the New York Academy of Sciences*, v. 1337, n. 1, p. 263-271, 2015. https://doi.org/10.1111/nyas.12623.

MARKS, David F. Visual imagery differences in the recall of pictures. *British Journal of Psychology*, v. 64, n. 1, p. 17-24, 1973. DOI: 10.1111/j.2044-8295.1973.tb01322.x.

MAUCH, M., & DIXON, S. pYIN: A fundamental frequency estimator using probabilistic threshold distributions. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Florence, Italy, 2014, pp. 659-663, https://doi.org/10.1109/ICASSP.2014.6853678.

MENEZES, F. *A acústica musical em palavras e sons*. 2ª edição. São Paulo. Ateliê Editorial. 2014.

MÜLLENSIEFEN D, GINGRAS B, MUSIL J, STEWART L. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE,* v. 9, n. 2: e89642. 2014. https://doi.org/10.1371/journal.pone.0089642.

NALBORCZYK, L.; GRANDCHAMP, R; KOSTER, E.; PERRONE-BERTOLOTTI, M.; LOEVENBRUCK, H. Can we decode phonetic features in inner speech using surface electromyography?. *PLoS ONE*, Public Library of Science, v. 15, n. 5, 2020. https://doi.org/10.1371/journal.pone.0233282.

PEIRCE, J., GRAY, J. R., SIMPSON, S., MACASKILL, M., HÖCHENBERGER, R., SOGO, H., ... & LINDELØV, J. K. PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, v. *51, n*. 1, 195-203. 2019. https://doi.org/10.3758/s13428-018-01193-y.

PETER Q. PFORDRESHER, STEVEN M. DEMOREST, SIMONE DALLA BELLA, SEAN HUTCHINS, PSYCHE LOUI, JOANNE RUTKOWSKI, GRAHAM F. WELCH; Theoretical Perspectives on Singing Accuracy: An Introduction to the Special Issue on Singing Accuracy (Part 1). *Music Perception*. February 2015; v. 32, n. 3, p. 227–231. doi: https://doi.org/10.1525/mp.2015.32.3.227.

PFORDRESHER, P. Q, HALPERN, A. R., & GREENSPON, E. A Mechanism for Sensorimotor Translation in Singing: The Multi-Modal Imagery Association (MMIA) Model. Music Perception, v. 32, n. 3, p. 242-253. 2015 https://doi.org/10.1525/mp.2015.32.3.242.

PRUITT, T. A., HALPERN, A. R., & PFORDRESHER, P. Q. Covert singing in anticipatory auditory imagery. *Psychophysiology*, p. 1–12. 2019 https://doi.org/10.1111/psyp.13297.

RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, v. 77, n. 2, p. 257-286, 1989. https://doi.org/10.1109/5.18626.

SIMONSMEIER, B.A., ANDRONIE, M, BUECKER, S & FRANK, C. The effects of imagery interventions in sports: a meta-analysis*, International Review of Sport and Exercise Psychology*, v. 14, n. 1, p. 186-207, 2021. https://doi.org/10.1080/1750984X.2020.1780627.

SUNDBERG, J. *Ciência da Voz*. Trad. Gláucia Salomão. São Paulo, Edusp. 2015.

TAN, Y.T, PERETZ, I. MCPHERSON, G.E., WILSON, S.J.; Establishing the Reliability and Validity of Web-based Singing Research. Music Perception 1 April 2021; v. 38, n. 4, p. 386–405. doi: https://doi.org/10.1525/mp.2021.38.4.386.

WISNIEWSKI, M G.; MANTELL, J T.; PFORDRESHER, Peter Q. Transfer effects in the vocal imitation of speech and song. *Psychomusicology*: Music, Mind, and Brain, v. 23, n. 2, p. 82, 2013. https://doi.org/10.1037/a0033299.

WOLPERT, D. M. & GHAHRAMANI, Z. Computational principles of movement neuroscience. *Nature Neuroscience* supplement. volume 3. 2000. p. 1212-1217. https://doi.org/10.1038/81497.

WOLPERT, D. M., PEARSON, K. G. & GHEZ, C. P. J. O planejamento do movimento. *In*: KANDEL, E.R., JESSEL, T. M., SIEGELBAUM, S. A., HUDSPETH, A. J. (eds.) *Princípios de Neurociências*. 5ª ed. Porto Alegre. AMGH. P. 645-665. 2014.

Mariana Moga de Moura e Silva é professora de canto e preparadora vocal. Mestra em Neurociência e Cognição pela Universidade Federal do ABC (UFABC). Pós-graduada em Formação Integrada em Voz Artística pelo Centro de Estudos da Voz (CEV) em São Paulo. Licenciada em Educação Musical pela Universidade Federal de São Carlos (UFSCar). Realiza pesquisa interdisciplinar nas áreas da Voz, Audição e Neurociência Cognitiva e Computacional. É integrante do grupo de pesquisa NeuroMúsica UFABC. Graduanda em Fonoaudiologia pela Faculdade de Medicina da Universidade de São Paulo (USP) e membro do Laboratório de Investigação Fonoaudiológica em Voz (LIFVoz) da FMUSP.
E-mail: marianamogacanto@gmail.com

Patrícia Vanzella formou-se pela Escola de Comunicações e Artes da Universidade de São Paulo e pela Academia Franz Liszt de Budapest. Realizou doutorado em Piano/MúsicNEUROMÚSICA na Catholic University of America e pós-doutorado no programa de pós-graduação em Engenharia da Informação da Universidade Federal do ABC. De 1993 a 2017, integrou o corpo docente do Departamento de Música da Universidade de Brasília. Em 2015 implementou e desde então coordena o projeto Neurociência e Música na Universidade Federal do ABC, instituição na qual é Professora Adjunta no Centro de Matemática, Computação e Cognição (http://neuro.ufabc.edu.br/music). Como solista, atuou frente aos principais conjuntos sinfônicos do país, como OSESP, OSUSP e Orquestra Sinfônica do Teatro Nacional Claudio Santoro. Como camerista apresentou-se em concertos na Europa, América e Ásia. Tem desenvolvido projetos de pesquisa em neurociência cognitiva da música em parceria com pesquisadores da University of Toronto, Harvard University, University of London, Max Planck Institute for Human Cognitive and Brain Sciences (Leipzig), Universidade Federal de São Paulo, Universidade Federal do ABC, Universidade de São Paulo e Instituto do Cérebro do Hospital Israelita Albert Einstein em São Paulo. Paralelamente às atividades de pesquisa, dedica-se a divulgar o conhecimento científico por meio da realização de ações educativas e culturais, como os Concertos Sinápticos e o ciclo de palestras Conversas Neuromusicais. É líder do Grupo de Pesquisa NEUROMÚSICA UFABC (CNPq) e presidente da Associação Brasileira de Cognição e Artes Musicais, ABCM (https://www.abcogmus.org) na gestão 2023-2026.
E-mail: pvanzella@yahoo.com

João Ricardo Sato é Professor Adjunto da Universidade Federal do ABC. Por cinco anos, foi coordenador do Núcleo de Cognição e Sistemas Complexos, unidade estratégica vinculada diretamente à reitoria da UFABC que tem como objetivo a realização de atividades em equipes interdisciplinares voltadas para a pesquisa, ensino e extensão na área de neurociências e cognição. A linha de pesquisa atual é interação entre neurociências e ciências exatas, com principal foco no neurodesenvolvimento, bases neurais de transtornos psiquiátricos, conectividade cerebral, inteligência artificial e processamento de sinais neurais. Formou-se no Bacharelado em Estatística pela USP em 2002, tendo concluído três iniciações científicas como bolsista PIBIC/CNPQ na área de séries temporais financeiras, com o intuito de atuar no mercado financeiro. No doutorado que se iniciou em 2004, decidiu focar sua pesquisa no desenvolvimento de novos métodos estatísticos para modelar a conectividade cerebral por meio de sinais de ressonância magnética funcional. Este trabalho interdisciplinar realizou-se sob a orientação de um docente do Departamento de Estatística um docente da Faculdade de Medicina da USP. Em 2006, realizou estágio de pesquisa (doutorado-sanduíche) no Instituto de Psiquiatria do King's College London, onde integrou técnicas computacionais de aprendizado de máquina à sua linha de pesquisa. Após a conclusão do doutorado em 2007, atuou como consultor estatístico e pesquisador do Hospital das Clínicas de São Paulo. Além disso, também atuou como professor no curso de pós-graduação lato sensu da Universidade Metodista em São Bernardo do Campo. Em janeiro de 2009, foi contratado pela UFABC para o cargo de professor adjunto do Centro de Matemática, Computação e Cognição. Em julho de 2009 tornou-se o coordenador do Núcleo de Cognição e Sistemas Complexos (NCSC). Os docentes associados ao NCSC foram a equipe idealizadora do Bacharelado em Neurociências (pioneiro no Brasil) e do programa de pós-graduação em Neurociências e Cognição da UFABC. Esta equipe também foi responsável por todo o planejamento da infra-estrutura para pesquisa científica em neurociências, hoje disponível no primeiro andar do Bloco Delta no Campus de

São Bernardo do Campo. Atualmente é coordenador do Núcleo Interdisciplinar de Neurociência Aplicada da Universidade Federal do ABC.
E-mail: jrsatobr@gmail.com

Bóris Marin trabalha em Neurociência Computacional como professor da Universidade Federal do ABC (UFABC). Anteriormente, foi pós-doutorando pela Universidade de São Paulo, Research Associate da UniversSource Brain (modelagem neuronal colaborativa), Open Worm (simulação de um organismo completo - C elegans) e Geppettoiversidade de São Paulo. Tem experiência na área de Biofísica, com ênfase em Neurociência Teórica e Computacional, atuando principalmente nos seguintes temas: modelagem e simulação de redes de neurônios biológicos, métodos qualitativos da teoria de bifurcações, desenvolvimento aberto e colaborativo de modelos computacionais e software científico. É editor da linguagem de especificação de modelos NeuroML, e colaborador dos projetos Open Source Brain (modelagem neuronal colaborativa), Open Worm (simulação de um organismo completo - C elegans) e Geppetto (framework para biologia computacional).
E-mail: borismarin@gmail.com

Guilherme Alves Delmolin de Oliveira é graduado em Psicologia pela Universidade Metodista de São Paulo (2014). Mestre em Neurociência e Cognição pela Universidade Federal do ABC (2018) eneuropsicológicaesmo programa. Atualmente é professor no curso de Psicologia da Universidade Municipal de São Caetano do Sul. É colaborador e pesquisador vinculado ao Projeto Neurociência e Música na UFABC (CNPq - http://neuro.ufabc.edu.br/music) desenvolvendo pesquisa interdisciplinar nas áreas de Psicologia do Desenvolvimento, Cognição Musical e Neurociências. Realizou aprimoramento profissional em Psicologia do Esporte pela Universidade Metodista de São Paulo (2017). Realizou atualização profissional no Centro Paulista de Neuropsicologia CPN/NANI no ambulatório de triagem com foco em distúrbios da aprendizagem e avaliação neuropsicológica interdisciplinar. Tem experiência na área de Psicologia, com ênfase em Psicologia Cognitiva e Psicologia Clínica, atuando com ênfase nos seguintes temas: avaliação neuropsicológica, psicoterapia de crianças, jovens e adultos e psicologia da música.
E-mail: guiharris@gmail.com

Daniel Quintela Bertuzzi é Mestre em Neurociência e Cognição pela Universidade Federal do ABC (UFABC). Possui graduação em Ciência e Tecnologia (2017) e em Neurociência (2020), ambas pela UFABC. Membro do grupo NEUROMUSICA UFABC.
E-mail: dqbertuzzi@gmail.com

Disponibilidade de dados: Os dados estarão disponíveis mediante solicitação.

Conflito de Interesses: Nenhum conflito de interesses foi declarado.

Modalidade de avaliação: Duplo-cega por pares.

A *Opus* está licenciada sob uma licença CC-BY.