

Análise Multimodal de Estados Emocionais em Performance Musical: Uma Abordagem 4E

Cristiano Américo de Oliveira
ORCID: <https://orcid.org/0009-0009-9074-6183>

Resumo

Este artigo propõe um sistema de captação multimodal, utilizando câmera e inteligência artificial (IA), para analisar os estados emocionais de intérpretes musicais em diferentes condições de performance (ex: de memória, lendo partitura). A pesquisa se fundamenta na cognição 4E (incorporada, embebida, enativa e estendida), na psicologia ecológica de Gibson, na neurociência afetiva de António Damásio, na teoria das metáforas conceituais e na arte sônica. O objetivo é demonstrar como variações faciais e oculomotoras co-variam com saliências texturais/temporais do som e com estados afetivo-atencionais reportados pelo intérprete, evitando a mera classificação de emoções e focando em padrões corporais situados. Serão coletados dados de vídeo, áudio e marcadores faciais/oculares, sincronizados e processados por IAs, para posterior triangulação interpretativa. Os resultados esperados incluem a visualização de dados e a discussão de como a expressividade corporal se acopla à performance musical, contribuindo para uma compreensão mais ecológica e multissensorial da significação musical. Este estudo representa uma etapa inicial de uma pesquisa de doutorado mais ampla.

Palavras-chave: Cognição 4E; Performance Musical; Emoção; Análise Facial; Inteligência Artificial; António Damásio.

Abstract

This article proposes a multimodal capture system, using camera and artificial intelligence (AI), to analyze the emotional states of musical performers under different performance conditions (e.g., from memory, sight-reading). The research is grounded in 4E cognition (embodied, embedded, enactive, and extended), Gibson's ecological psychology, António Damásio's affective neuroscience, conceptual metaphor theory, and sonic art. The aim is to demonstrate how facial and oculomotor variations co-vary with textural/temporal sound saliences and with reported affective-attentional states, avoiding mere emotion classification and focusing on situated bodily patterns. Video, audio, and facial/oculomotor marker data will be collected, synchronized, and processed by AIs for subsequent interpretive triangulation. Expected results include data visualization and a discussion of how bodily expressivity couples with musical performance, contributing to a more ecological and multisensory understanding of musical meaning. This study represents an initial stage of a broader doctoral research.

Keywords: 4E Cognition; Musical Performance; Emotion; Facial Analysis; Artificial Intelligence; António Damásio.

1. Introdução

A performance musical é um fenômeno complexo que transcende a mera execução técnica de uma obra, englobando a expressão de emoções e a intrincada interação do intérprete com o instrumento, o ambiente e a própria música. Tradicionalmente, a análise da expressividade musical tem se restringido a aspectos puramente sonoros ou a relatos subjetivos dos músicos. Contudo, abordagens contemporâneas da cognição, como a Cognição 4E (Embodied, Embedded, Enactive, Extended), oferecem um arcabouço teórico robusto para investigar a performance musical como um processo situado e incorporado (GALLAGHER, 2023; TOFFOLO, 2023). Esta perspectiva desafia a visão cartesiana de uma mente isolada no cérebro, propondo que a cognição se estende ao corpo, ao ambiente e às interações dinâmicas que o indivíduo estabelece com o mundo.

Este artigo delineia uma proposta de pesquisa inovadora que visa desenvolver e aplicar um sistema de captação multimodal para analisar os estados emocionais de intérpretes musicais durante a execução de peças em diferentes condições, como performance de memória e leitura de partitura. A hipótese central que guia este estudo é que variações faciais e oculomotoras do intérprete co-variam de forma significativa com saliências texturais e temporais do som produzido, bem como com os estados afetivo-atencionais reportados pelo próprio músico. Em vez de buscar uma classificação categórica e reducionista de emoções, o foco será na descrição de padrões corporais situados (rosto, olhos, cabeça) em acoplamento dinâmico com o contexto performativo, em estrita consonância com os princípios da cognição 4E e da psicologia ecológica de James J. Gibson (TOFFOLO, 2023).

O sistema proposto integrará uma câmera de alta velocidade com ferramentas avançadas de Inteligência Artificial (IA) para a detecção e análise de landmarks faciais, pose da cabeça e movimentos oculares. Os dados visuais serão meticulosamente sincronizados com a captação de áudio de alta qualidade, permitindo uma análise temporal detalhada da relação entre a expressividade corporal e as características sonoras. A triangulação interpretativa dos dados coletados — que incluirá dados multimodais, o referencial teórico da cognição 4E e da neurociência afetiva (com ênfase nas contribuições de António Damásio sobre a relação entre corpo, emoção e sentimento) (DAMÁSIO, 1994; DAMÁSIO, 1999), e os relatos subjetivos dos intérpretes — permitirá uma compreensão mais aprofundada e

holística dos estados emocionais manifestados durante a performance musical. Esta pesquisa, embora constitua uma etapa inicial de um projeto de doutorado mais abrangente, almeja contribuir significativamente para a literatura científica brasileira na área da cognição musical e neurociência da música, oferecendo novas perspectivas metodológicas e teóricas.

2. Referencial Teórico

2.1. Cognição 4E: Incorporada, Embebida, Enativa e Estendida

A Cognição 4E representa uma virada paradigmática nas ciências cognitivas, distanciando-se da concepção tradicional que via a cognição como um processo meramente cerebral e computacional. Como aponta Gallagher (2023), essa abordagem é, na verdade, um conjunto de diversas perspectivas que, apesar de suas diferenças, compartilham um núcleo comum de oposição ao cognitivismo clássico. As quatro dimensões que a compõem são:

- Incorporada (Embodied): A cognição é intrinsecamente moldada pela natureza do corpo do organismo. Nossas experiências sensorio-motoras não são apenas um subproduto, mas fundamentam e influenciam ativamente nossos processos cognitivos. O corpo, portanto, não é um mero recipiente para a mente, mas uma parte constitutiva e integrante dela (VARELA; THOMPSON; ROSCH, 2017).
- Embebida (Embedded): A cognição não ocorre em um vácuo, mas em um contexto ambiental específico, sendo profundamente influenciada por ele. O ambiente não é um pano de fundo passivo, mas um componente ativo que molda, restringe e possibilita o comportamento cognitivo. A interação contínua e dinâmica com o ambiente é crucial para a emergência da cognição (CLARK, 2008).
- Enativa (Enacted): A cognição emerge da interação dinâmica e recíproca entre o organismo e seu ambiente. Em vez de simplesmente processar informações pré-existentes do mundo, os organismos ativamente dão sentido ao seu mundo através de suas ações e percepções. O ato de conhecer é indissociável do ato de fazer, e vice-versa (VARELA; THOMPSON; ROSCH, 2017).

- Estendida (Extended): Em certas circunstâncias, os processos cognitivos podem transcender os limites do cérebro e do corpo, incorporando ferramentas, tecnologias e outros aspectos do ambiente como partes funcionais do sistema cognitivo. Um exemplo paradigmático é o uso de um caderno para auxiliar a memória, onde o caderno se torna uma extensão da capacidade cognitiva do indivíduo (CLARK; CHALMERS, 1998).

No contexto da performance musical, a abordagem 4E permite uma compreensão do intérprete não como um cérebro isolado que executa uma partitura, mas como um sistema dinâmico e integrado. Neste sistema, o corpo do músico, o instrumento, o espaço físico da performance e as interações sociais (mesmo que implícitas) são elementos constitutivos da experiência cognitiva e emocional. A expressividade do músico, sob esta ótica, não é meramente um produto interno, mas uma manifestação enativa e incorporada de seu engajamento profundo com a música.

2.2. Psicologia Ecológica e Affordances de James J. Gibson

A Psicologia Ecológica, desenvolvida por James J. Gibson (1979), complementa a visão 4E ao enfatizar a relação direta e significativa entre o organismo e seu ambiente. Gibson introduziu o conceito de affordances, que se refere às possibilidades de ação que o ambiente oferece a um organismo, dadas suas capacidades e características. As affordances não são propriedades objetivas do ambiente nem meras percepções subjetivas, mas sim relações intrínsecas entre o organismo e o ambiente, percebidas diretamente.

Como destaca Toffolo (2023), a teoria de Gibson é fundamental para a superação do dualismo cartesiano na psicologia cognitiva, propondo que "perceber não é o processo de mediação entre mundo e mente no qual os órgãos dos sentidos apenas carregam, de forma neutra, informações do mundo para que o processamento mental gere significados (ou qualquer outra explicação idealista/subjetivista), mas ao contrário, o que as coisas significam no ato de perceber são as ações que elas permitem ao indivíduo (affordances), que está no mundo" (TOFFOLO, 2023, p. 5). Isso implica que a percepção é sempre uma percepção-ação, onde o organismo está ativamente engajado com o ambiente.

Na performance musical, o instrumento oferece affordances específicas para a ação do músico (ex: a tensão das cordas, a resistência das teclas, a ergonomia do arco). O espaço de performance, a acústica do local e a presença de uma audiência também fornecem affordances que moldam intrinsecamente a execução e a expressividade. A percepção do músico é, portanto, uma percepção-ação, onde o ato de perceber já é um engajamento ativo com as possibilidades de ação que o ambiente musical oferece. A captação de movimentos faciais e oculares pode, assim, revelar como o intérprete explora e responde a essas affordances em tempo real, ajustando sua performance de forma adaptativa e contínua.

Gibson também aborda a importância da multissensorialidade, afirmando que "uma percepção nunca ocorre em um único sistema, mas sim em mais de um ao menos, ao mesmo tempo. A percepção sonora nunca é somente sonora, a percepção tátil nunca é somente tátil e assim por diante" (TOFFOLO, 2023, p. 6). Essa equivalência entre os sistemas perceptuais é uma redundância importante que fortalece as significações emergentes no ato perceptivo, o que é diretamente relevante para a análise multimodal proposta neste artigo.

2.3. Neurociência Afetiva e a Teoria dos Marcadores Somáticos de António Damásio

As contribuições de António Damásio (1994, 1999) são cruciais para a compreensão dos estados emocionais e sua intrínseca relação com o corpo e a cognição. Damásio propõe uma distinção fundamental: as emoções são reações corporais automáticas e não conscientes a estímulos relevantes, enquanto os sentimentos são a experiência consciente e subjetiva dessas emoções e dos estados corporais associados. A Hipótese dos Marcadores Somáticos postula que as emoções e os sentimentos desempenham um papel crucial na tomada de decisões e na regulação do comportamento, atuando como "sinais" que guiam o organismo. Esses "marcadores somáticos" são alterações fisiológicas (como batimentos cardíacos, tensão muscular, expressões faciais, alterações posturais) que são associadas a experiências emocionais passadas e que influenciam processos cognitivos subsequentes, muitas vezes de forma inconsciente.

No contexto da performance musical, as emoções não são meros acompanhamentos da execução, mas elementos intrínsecos que moldam a interpretação, a expressividade e a

experiência subjetiva do músico. As expressões faciais e oculares, muitas vezes involuntárias ou microexpressões, podem ser consideradas manifestações visíveis desses marcadores somáticos, refletindo estados emocionais e afetivos profundos do intérprete. A análise detalhada dessas microexpressões e movimentos pode oferecer resultados valiosos sobre os estados inconscientes ou de pleno sentimento que Damásio descreve, especialmente quando correlacionados com o relato subjetivo do músico e as características sonoras da performance. A música, por sua capacidade única de evocar e modular emoções, torna-se um campo fértil para investigar a dinâmica complexa entre corpo, emoção e cognição, conforme proposto por Damásio.

Arturo Carsetti, em sua metabiologia, explora a complexidade da vida e da cognição, ressaltando que "a vida se esconde no equilíbrio criativo em jogo, um equilíbrio que deve ser constantemente renovado sob pena de dissipar a própria vida" (CARSETTI, 2020, p. 1). Essa perspectiva se alinha com a visão de Damásio sobre a homeostase e a regulação biológica como fundamentos da emoção e da consciência, sugerindo que a performance musical, como um ato vital e expressivo, é um reflexo desse equilíbrio dinâmico e contínuo.

2.4. Metáforas Conceituais na Semântica Musical

A teoria das metáforas conceituais, desenvolvida por Lakoff e Johnson (1980), sugere que o sistema conceitual humano é fundamentalmente metafórico, estruturado por mapeamentos sistemáticos de domínios de experiência mais concretos e sensório-motores para domínios mais abstratos. Na música, esta teoria se manifesta de forma proeminente na maneira como descrevemos e compreendemos fenômenos musicais, frequentemente utilizando termos que remetem a experiências corporais, espaciais e táteis (ex: "música brilhante", "textura áspera", "melodia ascendente", "harmonia densa").

Nogueira (2011) destaca a relevância da metáfora conceitual para a semântica da música, argumentando que "a música, como linguagem, não se restringe à sua estrutura formal, mas adquire significado através de sua interação com a experiência humana, mediada por esquemas imagéticos e metáforas" (NOGUEIRA, 2011, p. 2). Essa perspectiva é crucial para a interpretação dos dados multimodais, pois permite ir além da mera descrição de movimentos e sons, buscando as relações semânticas e expressivas que o intérprete

estabelece com a música. As variações faciais e oculares podem ser interpretadas como manifestações corporais dessas metáforas, revelando como o músico encarna e expressa o sentido musical. Por exemplo, um "brilho" espectral na música pode co-variá com um leve erguer de sobrancelhas ou um alargamento dos olhos, indicando uma correspondência entre a percepção sonora e a expressão facial que reflete uma metáfora conceitual subjacente e um engajamento afetivo. A compreensão dessas metáforas é essencial para desvendar a riqueza da comunicação não-verbal na performance musical (NOGUEIRA, 2018).

2.5. Análise Espectromorfológica e Arte Sônica

A análise do som na performance musical não se restringe apenas a parâmetros tradicionais como altura, ritmo e timbre. A espectromorfologia, desenvolvida por Denis Smalley e explorada por autores como Lasse Thoresen e Manuella Blackburn, oferece um arcabouço para descrever as "formas sonoras" e suas transformações ao longo do tempo (THORESEN, 2007; BLACKBURN, 2011). A tipomorfologia de Schaeffer, adaptada por Thoresen, permite categorizar objetos sonoros com base em suas características morfológicas e energéticas, fornecendo uma linguagem precisa para a descrição de eventos sonoros complexos.

Manuella Blackburn (2011) destaca a importância das "formas visuais do som" na composição, argumentando que a espectromorfologia não apenas descreve o som, mas também pode informar estratégias composicionais, utilizando o vocabulário espectromorfológico para moldar o material sonoro. Essa perspectiva é particularmente relevante para a análise multimodal, pois sugere uma intrínseca ligação entre a percepção auditiva e a visualização das características do som, o que pode se manifestar na expressividade corporal do intérprete.

Trevor Wishart, em sua obra "On Sonic Art" (1996), expande a compreensão do som para além dos limites da música tradicional, propondo o conceito de "arte sônica". Wishart enfatiza a importância da estrutura e da organização dos sons em si, independentemente de sua categorização como "música" ou "efeitos sonoros". Ele argumenta que a escuta atenta ao mundo e a observação das paisagens sonoras naturais e eventos estruturais são fundamentais para a compreensão da arte sônica. Para Wishart (1996, p. 17), "este livro não é sobre como

fazer, embora eu vá discutir em algum momento novas técnicas... mas sim sobre o porquê". Essa busca pelo "porquê" da organização sonora ressoa com a investigação dos estados emocionais na performance, onde a expressividade do intérprete pode ser vista como uma forma de dar sentido e propósito à organização dos eventos sonoros.

A integração da análise espectromorfológica e dos princípios da arte sônica na pesquisa proposta permite uma compreensão mais rica das saliências texturais e temporais do som. Ao correlacionar essas características sonoras com os dados faciais e oculares do intérprete, é possível investigar como o músico reage e molda o som em um nível micro e macro, revelando a complexa interação entre a percepção sonora, a ação corporal e a expressão emocional. A análise multimodal, nesse contexto, busca desvendar como o intérprete "encarna" a estrutura sonora e a transforma em uma experiência emocionalmente carregada.

3. Sistema Proposto: Captação Multimodal e Análise por IA

O sistema de captação e análise proposto visa coletar dados multimodais de alta qualidade e processá-los para extrair informações relevantes sobre os estados emocionais do intérprete musical. A arquitetura do sistema é baseada em uma abordagem não invasiva e markerless, utilizando tecnologias de visão computacional e inteligência artificial de ponta.

3.1. Aquisição e Sincronização de Dados

- **Câmera:** Será empregada uma câmera de alta velocidade (1080p/4K a ≥ 120 fps) para capturar as expressões faciais e movimentos oculares do intérprete com alta fidelidade temporal. Uma lente de 35–85 mm e iluminação difusa frontal/lateral serão utilizadas para otimizar a qualidade da imagem e minimizar sombras indesejadas. O posicionamento da câmera será fixo em relação ao rosto do intérprete (meio-busto), com o eixo óptico próximo ao olhar (± 10 – 15°), visando minimizar distorções e garantir a visibilidade dos detalhes faciais e oculares.
- **Áudio:** Um microfone dedicado de alta qualidade (≥ 48 – 96 kHz) será utilizado para a captação de áudio com fidelidade sonora. Idealmente, o áudio será gravado no mesmo

arquivo de vídeo para garantir sincronismo nativo. Caso a gravação seja separada, um "clap" inicial será utilizado, e técnicas de correção por correlação cruzada áudio-audio serão aplicadas para garantir uma sincronização precisa e robusta.

- Sincronização: A sincronização entre os dados de vídeo e áudio será realizada através de timestamps do contêiner de vídeo. Em cenários onde os arquivos são separados, técnicas avançadas de correlação cruzada de sinais serão aplicadas para alinhar as séries temporais com precisão sub-milissegundo, essencial para a análise de eventos rápidos e microexpressões.

3.2. Processamento por Inteligência Artificial (IA)

O processamento dos dados visuais será conduzido por algoritmos de IA de visão computacional, focando na extração de características faciais e oculares sem a necessidade de marcadores físicos, o que aumenta a naturalidade da performance:

- Detecção de Landmarks: Algoritmos de detecção de landmarks faciais 2D (pontos chave na face, olhos, íris) serão aplicados por frame para mapear com precisão a geometria facial do intérprete.
- Pose da Cabeça: A pose da cabeça (yaw, pitch, roll) será estimada utilizando técnicas de PnP (Perspective-n-Point) com um modelo facial canônico, estabelecendo correspondências robustas entre os landmarks 2D detectados e um modelo 3D pré-definido.
- Olhar Relativo e Piscada: O vetor de olhar relativo será calculado com base na geometria (posição do centro da íris em relação aos cantos oculares). A detecção de piscadas será realizada utilizando o Eye Aspect Ratio (EAR) e probabilidades associadas, permitindo quantificar a frequência e duração desses eventos.
- Qualidade e Suavização: A qualidade dos dados extraídos será continuamente monitorada por scores de confiança do detector, com rejeição automática de outliers. Suavização temporal (ex: filtro de Savitzky–Golay) será aplicada para reduzir ruídos e artefatos, garantindo a integridade dos onsets e eventos rápidos, que são cruciais para a análise expressiva.

3.3. Sinais Produzidos (Séries Temporais)

O sistema gerará séries temporais detalhadas de diversos sinais, que constituirão a base para a análise quantitativa e qualitativa:

- Cabeça: Dados de yaw, pitch, roll (rotações nos eixos vertical, horizontal e longitudinal, respectivamente); velocidades e acelerações angulares, que indicam a dinâmica dos movimentos da cabeça.
- Olhos: Taxa e duração de piscadas; vetor de olhar (distribuição por setores, indicando a direção e foco do olhar); velocidades e acelerações angulares do olhar.
- Face: Variações nas distâncias entre landmarks faciais (ex: abertura da boca, elevação das sobrancelhas, contração dos lábios), que podem indicar a ativação de músculos faciais específicos e microexpressões.
- Áudio: Parâmetros acústicos como intensidade, frequência fundamental (pitch), variações de timbre (espectro), ataque e decaimento dos sons, e densidade textural.

4. Análise de Dados e Triangulação Interpretativa

A análise dos dados coletados será realizada de forma integrada, buscando padrões de co-variação entre os sinais multimodais e os relatos subjetivos do intérprete. A triangulação interpretativa será o pilar metodológico, permitindo uma compreensão aprofundada dos estados emocionais manifestados durante a performance musical.

4.1. Correlação Multimodal

Serão empregadas técnicas estatísticas e de aprendizado de máquina para identificar correlações significativas entre os parâmetros de áudio (intensidade, pitch, timbre, textura) e os sinais visuais (movimentos da cabeça, olhar, expressões faciais). Por exemplo, a ocorrência de microexpressões faciais específicas ou padrões de movimento ocular pode ser correlacionada com mudanças abruptas na dinâmica sonora ou em características espectromorfológicas do som (THORESEN, 2007; BLACKBURN, 2011). A análise

espectromorfológica, conforme discutido por Thoresen e Blackburn, oferece uma linguagem precisa para descrever as "formas sonoras" que podem ser diretamente relacionadas às respostas corporais do intérprete.

4.2. Relato do Intérprete e Triangulação

Após cada sessão de performance, o intérprete será convidado a fornecer um relato detalhado de sua experiência, focando em momentos de particular intensidade emocional, dificuldades técnicas, ou insights interpretativos. Este relato será guiado por um protocolo semiestruturado, que incluirá perguntas sobre suas intenções expressivas, percepções sensoriais e estados afetivos. A triangulação ocorrerá pela comparação e contextualização dos dados multimodais objetivos com os relatos subjetivos do intérprete, permitindo:

- **Validação:** Verificar se os padrões corporais e sonoros observados correspondem às experiências emocionais reportadas.
- **Aprofundamento:** Compreender as nuances e as causas subjacentes de certas manifestações expressivas que não seriam evidentes apenas pelos dados objetivos.
- **Geração de Hipóteses:** Formular novas hipóteses sobre a relação entre cognição, emoção e performance musical, especialmente no que tange aos marcadores somáticos de Damásio e à forma como o corpo expressa sentimentos inconscientes ou de pleno engajamento (DAMÁSIO, 1994).

4.3. Conexão com a Cognição 4E e a Psicologia Ecológica

A interpretação dos resultados será fortemente embasada nos princípios da Cognição 4E e da Psicologia Ecológica. A co-variação entre os sinais multimodais será analisada como evidência da natureza incorporada, embutida e enativa da cognição musical. Os movimentos faciais e oculares, por exemplo, não serão vistos como meros indicadores de emoções internas, mas como parte integrante do processo de dar sentido ao ambiente musical e de se engajar ativamente com as affordances oferecidas pelo instrumento e pelo contexto performático (GIBSON, 1979; TOFFOLO, 2023). A expressividade do intérprete, nesse

sentido, é uma manifestação de sua interação dinâmica e adaptativa com o mundo sonoro, refletindo um "equilíbrio criativo" que Carsetti (2020) descreve como essencial para a vida e a cognição.

5. Considerações Finais

Este artigo apresentou uma proposta de pesquisa para a análise multimodal de estados emocionais em performance musical, fundamentada em abordagens contemporâneas da cognição, como a Cognição 4E, a Psicologia Ecológica de Gibson, a neurociência afetiva de Damásio, a teoria das metáforas conceituais e a arte sônica. O sistema de captação e análise, baseado em tecnologias de IA e visão computacional, visa coletar dados detalhados de expressões faciais, movimentos oculares e características sonoras, permitindo uma triangulação interpretativa com os relatos subjetivos dos intérpretes.

O objetivo principal não é categorizar emoções de forma simplista, mas sim desvendar os padrões corporais situados que co-variam com as saliências sonoras e os estados afetivo-atencionais do músico. Acreditamos que essa abordagem mais ecológica e multissensorial da significação musical contribuirá para uma compreensão mais rica e holística da expressividade na performance. Ao focar na interação dinâmica entre organismo e ambiente, e na manifestação de marcadores somáticos e metáforas conceituais, esta pesquisa almeja lançar luz sobre os processos cognitivos e emocionais subjacentes à criação e interpretação musical.

Esta etapa inicial de pesquisa de doutorado, ao integrar diversas perspectivas teóricas e metodologias avançadas, busca não apenas expandir o conhecimento na área da cognição musical, mas também oferecer ferramentas e subsídios para educadores, terapeutas e próprios músicos, promovendo uma apreciação mais profunda da complexidade e riqueza da experiência musical humana. Acreditamos que a análise detalhada dos movimentos involuntários e das microexpressões, conforme sugerido pela teoria de Damásio, em conjunto com a compreensão da organização sonora proposta por Wishart e a análise espectromorfológica, pode revelar aspectos inconscientes e profundamente enraizados da experiência emocional do intérprete, contribuindo para uma visão mais completa da performance como um ato de engajamento total do ser.

Referências

- ALONSO, D.; RAMOS, A. A marca do cognitivo e cognição 4E. *Revista Portuguesa de Filosofia*, v. 78, n. 2, p. 1001-1020, 2022.
- BEZERRA, A. P.; FIALHO, J. Cognição 4E e música: revisão integrativa da literatura. *Opus*, v. 29, n. 1, p. 1-23, 2023.
- BLACKBURN, M. The visual sound-shapes of spectromorphology: an illustrative guide to composition. *Organised Sound*, v. 16, n. 1, p. 5-13, 2011.
- CARSETTI, A. *Metabiology: Non-standard Models, General Semantics and Natural Evolution*. Cham: Springer, 2020.
- CLARK, A. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press, 2008.
- CLARK, A.; CHALMERS, D. The extended mind. *Analysis*, v. 58, n. 1, p. 7-19, 1998.
- DAMÁSIO, A. R. *O Erro de Descartes: Emoção, Razão e o Cérebro Humano*. São Paulo: Companhia das Letras, 1994.
- DAMÁSIO, A. R. *O Sentimento de Si: O Corpo, a Emoção e a Neurobiologia da Consciência*. São Paulo: Companhia das Letras, 1999.
- GALLAGHER, S. *Embodied and Enactive Approaches to Cognition*. Cambridge: Cambridge University Press, 2023.
- GIBSON, J. J. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.
- LAKOFF, G.; JOHNSON, M. *Metaphors We Live By*. Chicago: University of Chicago Press, 1980.
- NOGUEIRA, M. A contribuição da metáfora conceitual para a semântica da música. In: *Anais do XXI Congresso da ANPPOM*. Uberlândia, 2011. p. 1-8.
- NOGUEIRA, M. Resposta de orientação musical. *Música Hodie*, v. 16, n. 1, p. 100-115, 2018.
- THORESEN, L. *Spectromorphological analysis of sound objects: an adaptation of Pierre Schaeffer's typomorphology*. Oslo: Norwegian Academy of Music, 2007.
- TOFFOLO, R. B. G. A cognição musical deve ser mais que musical. In: *Anais do XXXIII Congresso da ANPPOM*. João Pessoa, 2023. p. 1-11.
- VARELA, F. J.; THOMPSON, E.; ROSCH, E. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press, 2017.
- WISHART, T. *On Sonic Art*. New York: Harwood Academic Publishers, 1996.