

# Corpo, Som e Emoção: fundamentos teóricos e sistema multimodal para a análise da performance musical

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

## Resumo

Este artigo apresenta uma proposta de pesquisa que integra fundamentos teóricos da cognição incorporada e da psicologia ecológica a um sistema de captação multimodal para análise de estados emocionais na performance musical. A partir das contribuições de Maturana, Varela, Gibson, Damásio e Huron, compreende-se a cognição como um processo biológico, perceptivo e afetivo, emergente da interação entre corpo, som e ambiente. O sistema desenvolvido utiliza visão computacional e inteligência artificial para extrair dados faciais e oculares de intérpretes em diferentes condições performáticas (leitura, memória e improvisação), correlacionando-os com variações sonoras e relatos subjetivos. A metodologia propõe uma triangulação interpretativa entre medidas fisiológicas, parâmetros acústicos e descrições fenomenológicas, a fim de revelar padrões expressivos associados à emoção musical. Os resultados esperados visam demonstrar que o gesto, a escuta e a emoção formam uma unidade funcional no ato performativo, contribuindo para a compreensão da expressividade musical como cognição encarnada e ecológica.

**Palavras-chave:** cognição incorporada; emoção musical; performance; inteligência artificial; análise multimodal.

## Abstract

This paper presents a research proposal that integrates theoretical foundations of embodied cognition and ecological psychology with a multimodal data capture system designed to analyze emotional states in musical performance. Drawing on the works of Maturana, Varela, Gibson, Damasio, and Huron, cognition is understood as a biological, perceptual, and affective process emerging from the dynamic coupling between body, sound, and environment. The proposed system employs computer vision and artificial intelligence to extract facial and ocular data from performers under different performance conditions (score reading, memory, and improvisation), correlating them with acoustic variations and subjective self-reports. The methodology proposes an interpretative triangulation between physiological measures, acoustic parameters, and phenomenological descriptions to uncover expressive patterns associated with musical emotion. The expected results aim to demonstrate that gesture, listening, and emotion form a functional unity in performance, contributing to a deeper understanding of musical expressivity as embodied and ecological cognition.

**Keywords:** embodied cognition; musical emotion; performance; artificial intelligence; multimodal analysis.

# 1. Introdução

A performance musical constitui um campo privilegiado para investigar a relação entre corpo, emoção e cognição. Longe de se restringir à execução técnica de uma obra, ela se configura como um processo dinâmico de interação entre o intérprete, o som e o ambiente, no qual percepção, gesto e afeto se articulam em tempo real. Compreender a expressividade musical, portanto, exige ultrapassar abordagens centradas no produto sonoro e considerar o corpo como parte integrante da construção de sentido.

Nas últimas décadas, avanços nas ciências cognitivas e na neurociência afetiva têm mostrado que a mente não é um sistema isolado, mas um fenômeno incorporado e situado. Essa perspectiva compreende a cognição como um processo emergente das interações entre organismo e ambiente, integrando dimensões biológicas, perceptivas e emocionais. No contexto performático, isso implica reconhecer que emoção, gesto e percepção sonora não são instâncias separadas, mas expressões complementares de um mesmo sistema adaptativo.

A presente pesquisa propõe o desenvolvimento de um sistema multimodal de captação e análise — combinando vídeo de alta velocidade, áudio de alta fidelidade e algoritmos de inteligência artificial — para investigar como variações faciais, oculares e sonoras de intérpretes musicais se correlacionam com estados emocionais e condições performativas distintas. A hipótese central é que essas variações revelam padrões corporais situados, evidenciando o acoplamento entre percepção, emoção e ação durante a execução musical.

Mais do que classificar emoções, o objetivo é compreender como o corpo sente, pensa e age sobre o som, configurando a performance como um processo cognitivo-afetivo incorporado. Para tanto, a pesquisa se apoia em um conjunto de referenciais teóricos complementares — biológicos, fenomenológicos e perceptivos — que fundamentam uma concepção integrada de cognição musical.

## 2. Quadro conceitual

### 2.1. Maturana: vida, autopoiese e acoplamento estrutural

A concepção biológica de Humberto Maturana constitui o ponto de partida para compreender a cognição como um fenômeno inerente à própria vida. O autor, em colaboração com Varela, define os sistemas vivos como autopoieticos, ou seja, capazes

de produzir e manter a própria organização por meio de interações recursivas com o ambiente. Essa formulação estabelece que todo ser vivo é, essencialmente, um sistema cognitivo: conhecer é viver, e viver é um processo de cognição.

Um sistema cognitivo é aquele cuja organização define um domínio de interações no qual pode agir de modo relevante para a própria manutenção, e o processo de cognição é o próprio agir ou comportar-se nesse domínio. Sistemas vivos são sistemas cognitivos, e viver, enquanto processo, é um processo de cognição. Esta afirmação é válida para todos os organismos, com ou sem sistema nervoso. (Maturana; Varela, 1980, p. 13, tradução nossa)<sup>1</sup>

Essa concepção desloca o conhecimento da esfera representacional para o domínio das interações entre organismo e meio. O conhecimento não é um conteúdo armazenado no cérebro, mas uma prática relacional de adaptação contínua. Em contextos de performance musical, isso significa que o intérprete não processa informações abstratas, mas opera como um sistema vivo que se reorganiza em função das condições sonoras, motoras e emocionais do ato performativo.

## **2.2. Varela: cognição incorporada e enativa**

Partindo da autopoiese, Francisco Varela propõe uma ampliação epistemológica que introduz a dimensão fenomenológica da experiência. Sua formulação da cognição enativa descreve o conhecimento como um ato de engajamento com o mundo, em que o sujeito não representa a realidade, mas a faz emergir por meio de sua ação.

Conhecer não é representar um mundo preexistente, mas participar na sua contínua formação por meio da ação. O conhecimento não é um espelho do mundo, mas uma prática de incorporação que o faz emergir. (Varela; Thompson; Rosch, 1991, p. 9, tradução nossa)<sup>2</sup>

Essa perspectiva torna a mente um fenômeno relacional e dinâmico, resultante do acoplamento entre corpo e ambiente. No domínio musical, essa relação manifesta-se quando o intérprete faz o som existir com o corpo e, simultaneamente, é transformado

---

<sup>1</sup> Do original: *A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain. Living systems are cognitive systems, and living as a process is a process of cognition. This statement is valid for all organisms, with and without a nervous system.* (MATURANA; VARELA, 1980, p. 13)

<sup>2</sup> Do original: *Cognition is not the representation of a pre-given world by a pre-given mind but rather the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs.* (VARELA; THOMPSON; ROSCH, 1991, p. 9)

por ele. A performance, nesse sentido, configura-se como um ciclo autopoietico e enativo no qual percepção e ação se retroalimentam continuamente.

### 2.3. Gibson: percepção ecológica e affordances

A psicologia ecológica de James J. Gibson acrescenta uma dimensão perceptiva essencial à compreensão da cognição situada. Para o autor, perceber é explorar ativamente o ambiente e identificar as informações que este oferece para a ação. As *affordances* — possibilidades de ação — definem a relação de complementaridade entre o organismo e o mundo.

As affordances do ambiente são aquilo que ele oferece ao animal — o que ele provê ou proporciona, seja para o bem ou para o mal. Refiro-me a algo que diz respeito tanto ao ambiente quanto ao animal, de um modo que nenhum termo existente abarca. Implica a complementaridade entre o animal e o ambiente. (Gibson, 1979, p. 127, tradução nossa)<sup>3</sup>

Essa formulação desloca a percepção da representação passiva para a ação exploratória. Em performance musical, o instrumento, o espaço acústico e o público constituem affordances que orientam o gesto e a expressividade. O músico age como um sistema perceptivo-ativo, ajustando-se às condições ecológicas que o som e o ambiente oferecem, num processo contínuo de descoberta e adaptação.

### 2.4. Damásio: emoção, corpo e consciência

António Damásio introduz a dimensão neurofisiológica e afetiva da cognição, mostrando que emoção e sentimento são componentes estruturais da razão e da consciência. Para o autor, os estados corporais e emocionais não são epifenômenos, mas mecanismos regulatórios fundamentais que participam da tomada de decisão e da experiência consciente.

Uma emoção é definida como um conjunto de mudanças nos estados corporais e cerebrais desencadeadas por um sistema neural especializado, que responde a conteúdos específicos das percepções — reais ou recordadas — relativos a um objeto ou evento particular. Essas modificações variam desde alterações no meio interno e nas vísceras [...] até mudanças no sistema musculoesquelético, evidentes ao observador (por exemplo, postura, expressão facial,

---

<sup>3</sup> Do original: *The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. [...] It implies the complementarity of the animal and the environment.* (GIBSON, 1979, p. 127)

comportamentos como congelamento, fuga e luta). (Bechara; Damásio, 2005, p. 339–340, tradução nossa)<sup>4</sup>

A Hipótese dos Marcadores Somáticos sustenta que o corpo armazena sinais afetivos que orientam decisões e percepções antes da deliberação racional. No campo da performance musical, esses marcadores são perceptíveis nas microexpressões, posturas e variações gestuais do intérprete, que traduzem a presença da emoção como princípio organizador da expressividade sonora.

## **2.5. Huron: expectativa musical e emoção antecipatória**

David Huron expande essa discussão para o domínio da escuta e da execução musicais ao propor a teoria ITPRA, que descreve a emoção como produto de processos de antecipação, reação e avaliação. Inspirado na neurociência e na etologia, o autor mostra que o prazer, a surpresa e a tensão musicais resultam de mecanismos biológicos de previsão e resposta.

Como a música induz ou evoca estados emocionais nos ouvintes? Diversos mecanismos foram propostos para explicar como os sons provocam emoções, incluindo respostas auditivas inatas, associações aprendidas e processos mediados por neurônios-espelho. Propõe-se que conceitos etológicos ajudam a explicar por que a música é capaz de induzir apenas certas emoções. (Huron, 2015, p. 370, tradução nossa)<sup>5</sup>

A música, assim, é entendida como um ambiente temporal de interação entre expectativa e emoção. As dinâmicas de tensão e resolução ativam respostas corporais e afetivas que integram a experiência perceptiva e performática, revelando o papel do corpo como mediador entre cognição e sentimento musical.

## **2.6. Integração teórica: corpo, emoção e ambiente na performance**

As concepções de Maturana, Varela, Gibson, Damásio e Huron convergem para uma compreensão unificada da cognição como processo incorporado, situado e afetivo. Maturana define a vida como cognição; Varela expande essa visão ao mostrar que o conhecimento é ação enativa; Gibson demonstra que a percepção é inseparável das

---

<sup>4</sup> Do original: *An emotion is defined as a collection of changes in body and brain states triggered by a dedicated brain system [...] These modifications range from changes in internal milieu and viscera [...] to changes in the musculoskeletal system that may be obvious to an external observer.* (BECHARA; DAMASIO, 2005, p. 339)

<sup>5</sup> Do original: *How does music induce or evoke feeling states in listeners? [...] It is proposed that ethological concepts help explain why music is able to induce only certain emotions.* (HURON, 2015, p. 370)

possibilidades ecológicas do ambiente; Damásio evidencia que a emoção constitui a base fisiológica e consciente da razão; e Huron descreve como esses mecanismos se expressam no campo musical, mediando expectativa, movimento e sentimento.

A partir dessa integração, a performance musical pode ser concebida como um sistema de acoplamento entre corpo, som e ambiente, no qual percepção, emoção e ação formam um circuito contínuo de significação. O corpo do intérprete atua simultaneamente como sistema vivo (Maturana), agente enativo (Varela), explorador ecológico (Gibson), organismo emocional consciente (Damásio) e agente expectante e avaliativo (Huron). Essa unidade conceitual oferece o alicerce para compreender a performance como experiência cognitiva plena — onde gesto, som e consciência se constituem mutuamente.

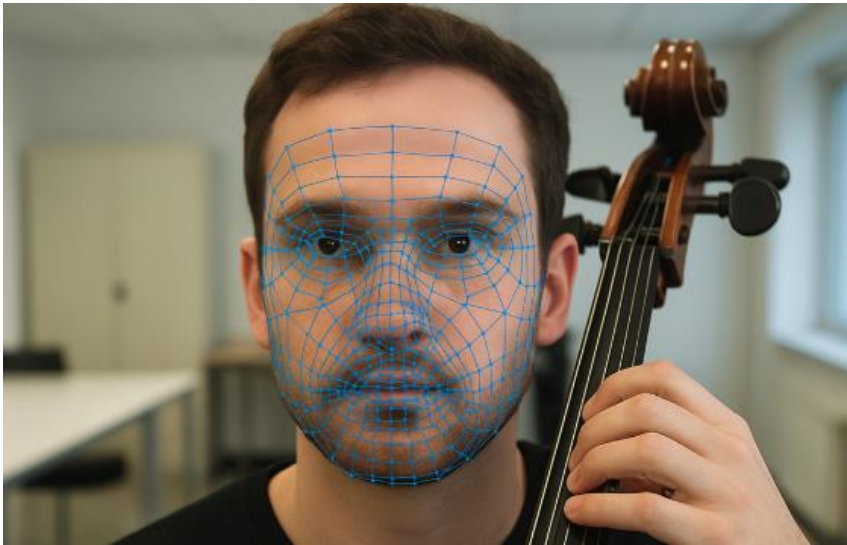
### 3. Sistema Proposto: Captação Multimodal e Análise por IA

O sistema proposto visa registrar e processar dados multimodais – visuais e sonoros – de intérpretes musicais, a fim de investigar empiricamente as relações entre emoção, percepção e ação durante a performance. A arquitetura adota uma abordagem não invasiva e *markerless*, empregando tecnologias de visão computacional e inteligência artificial (IA) para a análise facial e ocular, integradas à captação de áudio de alta fidelidade. Esse sistema permitirá observar, em tempo real, como padrões gestuais e microexpressões se articulam às variações sonoras e contextuais do ato performativo.

Tabela 1 – Estrutura geral do sistema de captação e análise

Módulo	Síntese técnica e funcional	Finalidade
Aquisição de dados	Câmera de alta velocidade ( $\geq 120$ fps, 1080p/4K) e microfone de estúdio ( $\geq 48$ –96 kHz), sincronizados por <i>timestamps</i> e correlação cruzada.	Captar com precisão microexpressões, movimentos oculares e variações sonoras.
Processamento por IA	Algoritmos de visão computacional para detecção facial, pose da cabeça e análise ocular, com filtragem e suavização temporal.	Extrair parâmetros expressivos de alta fidelidade gestual.
Geração de séries temporais	Integração dos dados de cabeça, olhos, face e áudio em bases quantitativas sincronizadas.	Criar matrizes multimodais para análise expressiva.
Análise multimodal	Aplicação de estatística e <i>machine learning</i> para correlação entre parâmetros acústicos e gestuais.	Identificar padrões de acoplamento corpo–som–emoção.
Triangulação interpretativa	Confronto entre dados objetivos e relatos subjetivos dos intérpretes.	Contextualizar e validar os resultados sob uma perspectiva fenomenológica.

Imagem 1 – Captura multimodal de performer musical com rastreamento facial baseado em IA.



Malha facial densa e pontos de detecção representando a análise de expressões e microvariações utilizadas na modelagem multimodal da performance musical. *Fonte: elaboração própria (2025).*

#### 4. Análise de Dados e Triangulação Interpretativa

A análise adota uma abordagem multimodal e interpretativa, articulando medidas objetivas (dados visuais e sonoros) e relatos subjetivos (percepções e emoções autorreferidas). O objetivo é compreender como a expressividade corporal e a dinâmica sonora se inter-relacionam na constituição do sentido musical.

A triangulação metodológica busca, portanto, evidenciar o acoplamento corpo–som–emoção, integrando três níveis de observação: fisiológico, perceptivo e fenomenológico.

Tabela 2 – Etapas de análise e relação teórica

Etapa	Procedimento	Relação teórica
<b>Correlação multimodal</b>	Identificação de co-variações entre parâmetros acústicos (intensidade, timbre, textura) e sinais faciais ou oculares (pose, piscadas, microexpressões).	Evidencia o acoplamento corpo–música descrito por <b>Maturana, Varela e Gibson</b> .
<b>Relato e triangulação</b>	Confronto dos dados objetivos com os relatos afetivos e perceptivos dos intérpretes, obtidos após cada sessão de performance.	Permite observar os <b>marcadores somáticos de Damásio</b> e os mecanismos de <b>expectativa emocional de Huron</b> .
<b>Interpretação ecológica e incorporada</b>	Análise integrada das expressões faciais e oculares como parte de um sistema perceptivo e afetivo situado, não como sinais isolados.	Fundamenta-se na <b>cognição incorporada e ecológica</b> , onde percepção, gesto e emoção são processos unitários.

## Síntese interpretativa

O sistema multimodal e o modelo de análise aqui propostos operacionalizam o arcabouço teórico do estudo, integrando:

- **Maturana e Varela** → cognição como autopoiese e acoplamento dinâmico;
- **Gibson** → percepção direta e affordances como mediação ação–ambiente;
- **Damásio** → emoção e marcadores somáticos como base da consciência;
- **Huron** → expectativa musical e regulação afetiva.

A combinação entre análise automatizada (IA), parâmetros acústicos e relatos fenomenológicos cria uma base empírica para examinar a unidade *corpo – emoção – música*, oferecendo subsídios teóricos e metodológicos tanto para a pesquisa cognitiva quanto para a prática composicional e performática contemporânea.

## 5. Considerações Finais

A pesquisa proposta articula uma convergência entre biologia, fenomenologia, psicologia ecológica e neurociência afetiva para compreender a performance musical como um fenômeno incorporado, situado e emocionalmente regulado. Ao integrar os aportes teóricos de Maturana, Varela, Gibson, Damásio e Huron, delineia-se um quadro conceitual que concebe o intérprete como sistema vivo, sensório-motor e afetivo, cuja cognição emerge do acoplamento dinâmico entre corpo, som e ambiente.

O desenvolvimento do sistema multimodal de captação — combinando visão computacional, análise acústica e relatos subjetivos — busca ampliar as ferramentas disponíveis para o estudo da expressividade musical, aproximando as ciências cognitivas das práticas performáticas. A precisão técnica das medições permitirá identificar padrões de co-variação entre movimento, som e emoção, fornecendo evidências empíricas para teorias que tratam a cognição como processo incorporado e interativo.

Mais do que classificar emoções, o objetivo é compreender como o corpo sente e pensa o som: como gestos, microexpressões e variações acústicas compõem uma rede contínua de sentido. A triangulação entre dados objetivos e experiência vivida poderá revelar como marcadores somáticos, affordances perceptivas e expectativas musicais se entrelaçam na construção do significado expressivo.

Em termos mais amplos, a proposta reforça a pertinência de uma abordagem interdisciplinar para a cognição musical, na qual métodos tecnológicos e referenciais teóricos convergem para uma compreensão mais profunda da performance como ato de conhecimento. Ao considerar o músico não apenas como executor, mas como organismo cognitivo e afetivo em constante acoplamento com o mundo, esta pesquisa contribui para repensar a relação entre arte, ciência e experiência, abrindo caminhos para novas metodologias de análise e criação musical.

## Referências

BECHARA, Antoine; DAMÁSIO, Antonio R. *The Somatic Marker Hypothesis: A Neural Theory of Economic Decision*. *Games and Economic Behavior*, v. 52, n. 2, p. 336–372, 2005. DOI: 10.1016/j.geb.2004.06.010.

DAMÁSIO, Antonio R. *O erro de Descartes: emoção, razão e o cérebro humano*. São Paulo: Companhia das Letras, 1994.

DAMÁSIO, Antonio R. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace, 1999.

GIBSON, James J. *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1979.

HURON, David. *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press, 2006.

HURON, David. *Virtuous and Sinful Emotions in Music*. *Music Perception*, v. 33, n. 3, p. 320–331, 2015. DOI: 10.1525/mp.2015.33.3.320.

MATURANA, Humberto R.; VARELA, Francisco J. *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: D. Reidel Publishing Company, 1980.

MATURANA, Humberto R.; VARELA, Francisco J. *El árbol del conocimiento: las bases biológicas del conocimiento humano*. Buenos Aires: Lumen, 2003.

VARELA, Francisco J.; THOMPSON, Evan; ROSCH, Eleanor. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press, 1991.