

Prompt Multimodal Acessível com IA Generativa para Feedback Verbal em Regência: Um Estudo Exploratório

Erickinson Lima¹

¹*Centro de Investigação em Didática e Tecnologia na Formação de Formadores, Aveiro, Portugal (erickinson.bezerra@ua.pt)*

Resumo: O presente estudo exploratório avalia um protótipo assistivo para prática de regência por estudantes cegos, baseado em prompts multimodais com IA generativa para classificação de compassos e feedback verbal corretivo. Com três participantes, houve melhora de acurácia (80,0%→94,7%) e redução de latência (P50: 6898→4999 ms). Gráficos gestuais sustentam leitura pedagógica dos erros e refinamento do *prompting*.

Palavras-chave: Projeto Maestro; Acessibilidade Musical; Performance Musical e Inclusão; PERFORMA.I.

INTRODUÇÃO

A regência é uma arte do gesto: aprende-se, em grande parte, pela observação de trajetórias gestuais, pela identificação visual do ictus e pela comparação com modelos técnicos do professor (Farberman, 1997; Rudolf, 1997). Quando esse padrão pedagógico é tratado como norma, estudantes cegos ficam em desvantagem estrutural — não por falta de potencial artístico, mas por falta de acesso equivalente ao diagnóstico técnico, à autoavaliação e à prática deliberada com feedback imediato (Lima, 2024). Se a validação do gesto depende da presença constante do docente (ou de intermediários), reduz-se a repetição com correção, aumenta-se a assimetria de oportunidades e comprometem-se autonomia, confiança e desempenho.

Essa dependência do olhar revela um problema pedagógico de base: como transformar um conhecimento predominantemente visual em um processo de aprendizagem acessível, sem empobrecer a dimensão técnica e musical da regência?

Em cursos de música, a prática fora da sala de aula é decisiva para consolidar padrões gestuais e integrar gesto e intenção sonora. Entretanto, para o estudante cego, a ausência de um mecanismo confiável de verificação e correção durante o estudo individual desloca a aprendizagem para um regime de “tentativa sem retorno”, em que o erro se acumula, a técnica se cristaliza de modo inadequado e a correção ocorre tardiamente, apenas quando há disponibilidade do professor. Nesse cenário, a barreira não está no gesto em si, mas na falta de instrumentos de mediação que permitam ao estudante monitorar a própria execução com autonomia.

Tecnologias assistivas baseadas em rastreamento gestual e feedback verbal emergem como alternativa promissora para reduzir essa lacuna, desde que atendam a requisitos específicos: operar em tempo real (Garbunova, 2018), fornecer orientações acionáveis (não apenas rótulos), lidar com variações naturais do gesto e produzir retornos consistentes que sustentem um ciclo de prática deliberada. A recente disponibilidade de IA generativa multimodal amplia esse horizonte ao permitir a integração entre evidências visuais/temporais do movimento e a produção de feedback verbal estruturado, aproximando o estudante de um tipo de diagnóstico que, historicamente, dependeu do professor presente.

Ainda assim, persistem questões críticas: qual o papel do *prompt* na confiabilidade desse processo? Como projetar instruções e saídas que sejam acessíveis, tecnicamente precisas e pedagogicamente úteis, evitando tanto explicações vagas quanto a introdução de “alucinações” e inconsistências? E, sobretudo, como avaliar um *prompt* não apenas pelo acerto classificatório, mas pela qualidade do feedback verbal como recurso de aprendizagem?

Este trabalho investiga essas questões por meio de um estudo exploratório centrado no design de um *prompt* multimodal acessível com IA generativa para prática técnica de regência. Propomos um fluxo em que trajetórias gestuais capturadas por webcam são convertidas em representações visuais, verbais e combinadas a um *prior* quantitativo de similaridade, orientando a IA generativa a produzir duas saídas complementares:

1. a estimativa do compasso executado;



2. feedback verbal corretivo, curto e acionável, voltado à repetição e ao refinamento técnico.

Como contribuições:

1. descrevemos princípios de projeto de *prompt* voltados à acessibilidade na regência;
2. apresentamos um protótipo que operacionaliza esses princípios em tempo real;
3. discutimos implicações pedagógicas para ciclos de estudo autônomo, enfatizando que a IA atua como complemento — e não substituto — da mediação docente.

PERSPECTIVA TEÓRICA

A literatura recente tem reposicionado a IA generativa menos como “oráculo” e mais como parceira de diálogo capaz de sustentar processos de aprendizagem, diagnóstico e tomada de decisão — desde que o desenho do sistema e a formulação dos *prompts* sejam tratados como parte do método. Robertson et al., (2024) argumentam, sob uma lente construtivista, que a interação com ferramentas como o ChatGPT tende a funcionar como um ciclo iterativo de co-construção de conhecimento, no qual a IA pode atuar como *more knowledgeable other*, oferecendo suporte cognitivo para que o usuário avance em tarefas complexas com apoio gradual. Essa perspectiva teórica dialoga diretamente com o problema apontado na introdução: quando o aprendizado depende de validação externa contínua, o custo pedagógico recai sobre o tempo, a repetição e a autonomia; por isso, tecnologias que favoreçam um ciclo “ação–retorno–ajuste” podem reduzir assimetrias — desde que o retorno/feedback seja confiável e orientado à prática.

Nesse cenário, a engenharia de prompts deixa de ser um detalhe operacional e passa a ser mediação pedagógica. Robertson et al., (2024) propõem que bons prompts combinam contexto, estrutura e avaliação: considerar conhecimento prévio, audiência e tom; explicitar formato e expectativas; e validar/ajustar respostas para conter desinformação e vieses.

Em um plano mais prescritivo, Hiranchan e Kiattisin (2024) reforçam a importância de frameworks de prompting com propriedades como modularidade, abstração, extensibilidade e padronização, além do uso de *constraints* (papéis, limites e requisitos de saída) para reduzir ambiguidade e elevar utilidade. Embora esse tipo de abordagem seja útil para disciplinar a interação, ela também expõe uma tensão central: quanto mais “bem escrito” o *prompt*, maior a chance de respostas convincentes — o que aumenta a responsabilidade sobre verificação e governança,

sobretudo quando o público-alvo depende do feedback como critério de correção.

É justamente em contextos de alto risco que a literatura explicita o problema da “confiança injustificada”. Endler et al., (2026) ao avaliarem um sistema de prompts para acesso a informação crítica, usam critérios de relevância, clareza e completude com procedimento de revisão e confiabilidade entre avaliadores. Os resultados mostram uma distribuição com acertos plenos e falhas significativas e, sobretudo, um achado inquietante: respostas podem ser muito claras e, ainda assim, fundamentadas de modo incorreto — o que aumenta o risco por induzir segurança indevida. Por isso, os autores defendem arquitetura híbrida, transparência técnica, mecanismos de *fallback* e supervisão humana. Essa discussão conversa com o software apresentado: quando a pessoa está treinando/praticando sozinha, o feedback verbal precisa ser, ao mesmo tempo, imediato e confiável; logo, não basta “gerar texto bom”, é necessário administrar incerteza, oferecer critérios de decisão conservadores e explicitar limites.

No campo educacional, Qi et al., (2026) investigam sistemas de feedback baseados em LLMs que simulam dinâmicas tutor–estudante e reconhecem, como obstáculos práticos, a inconsistência e a superconfiança dos modelos, sugerindo mitigação de vieses, verificação e supervisão humana. Um ponto particularmente útil para a acessibilidade é que o trabalho explicita como o formato do feedback pode favorecer certos perfis e excluir outros: quando o retorno é essencialmente textual, aprendizes “visuais” tendem a ser menos atendidos, e os autores apontam multimodalidade como direção (diagramas, recursos interativos etc.). No caso do presente trabalho, o “visual” precisa ser traduzido em “verbal”: a trajetória (gesto) vira estrutura linguística acionável. Isso recoloca a acessibilidade como problema de *design de representação* — não apenas de interface.

A multimodalidade aparece como resposta técnica para alinhar linguagem a sinais corporais/espaciais. Casas e Mitchell (2024) mostram que modelos guiados apenas por texto têm limitações para manter coerência com o contexto físico; por isso, incorporam *ControlNet* e condicionamento por mapas de profundidade, articulando *prompts* estruturados (inclusive por fala) para controlar a geração e manter aderência ao corpo e ao espaço. Ainda que o objetivo deles seja visualização imersiva, o princípio é transferível: quando há um “mundo” (corpo em movimento) por trás do texto, *entradas estruturadas* reduzem deriva e aumentam controle. Ao mesmo tempo, os autores chamam atenção para ética e riscos de vieses/representações, defendendo diretrizes e desenho intencional de *prompts* para evitar

estereótipos e distorções. Em uma tecnologia assistiva, essa cautela se traduz em garantir que o feedback não produza “normas” rígidas e inadequadas ao percurso do estudante, nem substitua o julgamento pedagógico contextual (Li, 2023).

Uma contribuição relevante para operacionalizar “qualidade de *prompt*” vem de Dahò e Caci (2026), que trabalham em contexto clínico (neuroreabilitação) e propõem um percurso transparente: elaboração de *prompts* com apoio de especialistas, avaliação do *prompt* por critérios (clareza, completude e relevância) e análise sistemática das saídas com um framework que contempla coerência interna, clareza estrutural, acurácia, adaptabilidade e abrangência. O achado é duplo: há otimismo cauteloso sobre o uso da IA como apoio (não substituto), mas persistem limites de personalização e viabilidade no “mundo real”. Para a regência acessível, esse resultado sugere um critério importante: não basta medir apenas desempenho do classificador; é preciso avaliar qualidade pedagógica do feedback (clareza, acionabilidade, adequação ao nível do estudante) e condições de uso (tempo de resposta, robustez, consistência).

Os textos apresentados convergem em quatro pontos que sustentam o trabalho apresentado.

1. A IA generativa pode apoiar autonomia quando opera como suporte em ciclos iterativos de prática e reflexão;
2. Depende de prompting estruturado (contexto–estrutura–avaliação) e de restrições explícitas que reduzam ambiguidades e orientem formato e finalidade;
3. A confiança no feedback exige governança: lidar com superconfiança, inconsistência e “clareza enganosa” demanda verificação, regras conservadoras e, quando aplicável, desenho híbrido que preserve o papel do professor;
4. Multimodalidade e representação são centrais: quando o objeto é corporal/espacial, enriquecer a entrada (sinais, resumos, imagens/diagramas) melhora alinhamento e permite transformar gesto em orientação verbal tecnicamente controlada.

A lacuna que se abre — e onde presente trabalho se insere — é que grande parte das evidências vem de domínios como segurança, programação, clínica e visualização educacional; há menos sistematização específica para práticas músico-performativas em que o “dado” é movimento e o “erro” é simultaneamente técnico e pedagógico. Ao trazer a engenharia de *prompts* para a perspectiva da acessibilidade musical (com feedback verbal corretivo e critérios de decisão

conservadores), o presente estudo pode contribuir não só com um protótipo, mas com um argumento metodológico: em contextos de inclusão, *prompt* é também *currículo*, *rubrica* e *mediação*.

O SOFTWARE

O software é um protótipo assistivo para prática técnica de regência que opera em tempo real a partir de uma webcam. A aplicação captura o movimento das mãos por visão computacional e transforma esse movimento em uma representação de trajetória bidimensional, aplicada a rotinas de limpeza e padronização para reduzir ruídos e tornar comparáveis execuções realizadas em diferentes durações. O resultado é um registro gestual consistente o suficiente para suportar análise automática e geração de retornos ao usuário, sem exigir interação visual complexa durante a prática.

A partir da trajetória, o sistema estima o compasso executado entre classes-alvo (2/4, 3/4 e 4/4, por exemplo) por meio de uma arquitetura de decisão em camadas que integra componentes determinísticos e componentes baseados em IA generativa. Em termos funcionais, a IA generativa é utilizada como um módulo de interpretação que auxilia a etapa de reconhecimento e contribui para a produção de saídas estruturadas através da API da OpenAI. Por razões de propriedade intelectual, não são apresentados detalhes operacionais, parâmetros internos, heurísticas de decisão, nem a especificação completa das representações intermediárias utilizadas no processo de classificação e validação.



Figura 1. Síntese da perspectiva do software.

Na interação com o software, o estudante executa o gesto diante da webcam e recebe, ao final de cada tentativa, feedback verbal corretivo, orientado a apoiar o ciclo execução–análise–correção durante o estudo individual. Para fins de avaliação, o sistema registra logs por tentativa (por exemplo, rótulo de referência quando disponível, classe estimada, tempo de resposta e características básicas da captura), permitindo calcular métricas de desempenho do classificador e indicadores de responsividade em contexto de uso. Dessa forma, o software busca reduzir a dependência de validação externa contínua, funcionando como complemento à mediação docente

e favorecendo maior autonomia na prática entre aulas.

SÍNTESE DA FUNÇÃO DOS PROMPTS NO SOFTWARE

O software utiliza um módulo de interpretação automatizada responsável por analisar a trajetória gestual capturada e compará-la com padrões previamente estabelecidos no banco de referências. Essa etapa tem como objetivo reconhecer a organização métrica do movimento realizado pelo usuário, considerando aspectos como forma geral da trajetória, distribuição espacial dos pontos de apoio gestual, coerência direcional e proximidade com modelos técnico-musicais de referência.

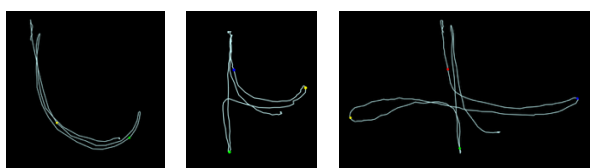


Figura 2. Gestual que compõe o banco de dados do software. Da esquerda para direita: 2/4 – 3/4 – 4/4.

A entrada desse módulo é composta por representações visuais e estruturais do gesto executado, associadas a referências previamente classificadas. O sistema organiza essas informações de modo a permitir uma avaliação comparativa entre o gesto do usuário e os padrões correspondentes aos compassos trabalhados. A análise não se limita a uma única característica isolada do movimento; ela considera o comportamento global da trajetória, sua estabilidade geométrica e sua aderência ao padrão gestual esperado.

Para aumentar a confiabilidade da avaliação, o processo combina evidências visuais e critérios internos de similaridade. Essas informações são utilizadas como suporte para a tomada de decisão, permitindo que o sistema identifique a categoria gestual mais provável entre as opções disponíveis. A resposta gerada nessa etapa é estruturada de forma objetiva, contendo o padrão reconhecido e uma justificativa sintética, suficiente para alimentar as etapas posteriores do sistema sem expor ao usuário os parâmetros internos do processo analítico.

Essa organização permite que o software opere de maneira estável e reprodutível, mantendo uma separação clara entre os dados capturados, os modelos de referência, os critérios de comparação e a resposta final apresentada ao usuário. Assim, a interpretação gestual preserva tanto a precisão técnica quanto a possibilidade de integração com recursos pedagógicos posteriores.

Após a etapa de reconhecimento, o sistema aciona um segundo módulo, voltado à geração de feedback pedagógico. Diferentemente da etapa anterior, esse

módulo não tem como finalidade classificar o gesto, mas transformar o resultado da análise em uma orientação didática adequada ao contexto de aprendizagem da regência.

A formulação do feedback considera o compasso selecionado para a prática, o grau de aproximação entre a execução do usuário e o modelo de referência, bem como critérios pedagógicos previamente definidos. A partir desses elementos, o software organiza uma resposta em linguagem acessível, priorizando clareza, objetividade e utilidade prática para o estudante.

O conteúdo gerado é orientado por princípios técnicos da regência, mas apresentado de modo formativo. Quando a execução apresenta boa correspondência com o padrão esperado, o sistema oferece reforço positivo e pequenas sugestões de refinamento. Quando a execução exige ajustes, o feedback torna-se mais diretivo, destacando aspectos como definição dos pontos gestuais, estabilidade da trajetória, clareza dos tempos, controle da amplitude e preparação do movimento seguinte.

A linguagem do feedback é controlada para evitar termos excessivamente técnicos ou informações internas do sistema. Em vez de apresentar dados computacionais, índices numéricos ou detalhes do processo de classificação, a resposta é convertida em orientação pedagógica compreensível ao usuário. Dessa forma, o software preserva a experiência educacional e mantém protegidos os elementos técnicos responsáveis pela análise.

As instruções internas de mediação utilizadas pelo software funcionam como camadas de articulação entre a análise computacional e a resposta pedagógica. A primeira camada orienta a interpretação do gesto capturado, auxiliando na identificação do padrão métrico mais compatível com a execução realizada. A segunda camada converte esse reconhecimento em uma devolutiva didática, adequada ao ensino da técnica gestual de regência.

Tabela 1. Sumário do *prompting* multimodal no protótipo (visão funcional)

Elemento	Descrição (alto nível)
Entradas	Representações multimodais da execução (trajetória renderizada/visualizações do gesto) e referências técnicas; metadados da tentativa (compasso-alvo quando selecionado, contexto de prática) e um resumo quantitativo de similaridade usado como evidência auxiliar.
Objetivos	(1) Reconhecer o compasso executado (2/4, 3/4, 4/4) de modo robusto; (2) Gerar feedback verbal, corretivo, curto e acionável, orientado à repetição e ao refinamento técnico.
Restrições	Linguagem acessível e direta; evitar termos excessivamente técnicos e qualquer exposição de parâmetros internos; manter foco no objetivo técnico do compasso; modular o nível de intervenção conforme a qualidade estimada da execução; evitar explicações longas e ambiguidades.
Formato de saída	Saída estruturada e previsível em JSON com campos predefinidos (ex.: abertura/avaliação breve, dicas técnicas e convite à repetição), facilitando consistência entre tentativas e reprodução por texto-para-fala.

Essa estrutura permite que o sistema uma reconhecimento de padrões, análise comparativa e comunicação pedagógica em um fluxo contínuo. Ao mesmo tempo, a descrição preserva os aspectos essenciais para compreensão científica do método, sem revelar detalhes específicos da implementação,

das regras internas de decisão, da organização exata dos dados ou da formulação operacional dos comandos utilizados no processamento.

PROMPT ORIENTADOR E O FEEDBACK

O *prompt* possui uma estrutura simples e pedagógica. Primeiro, define o papel do software, indicando que a resposta deve assumir a perspectiva de um professor de regência. Em seguida, apresenta o contexto da prática, isto é, o compasso trabalhado, o nível geral de desempenho e os aspectos técnicos que podem orientar a devolutiva.

Depois, o *prompt* estabelece regras de linguagem, evitando que o feedback revele informações computacionais ou detalhes internos da análise. Por fim, determina uma forma de resposta, composta por abertura, orientações técnicas e incentivo final. Essa organização garante que o feedback verbal seja claro, útil e coerente com o objetivo pedagógico do sistema.

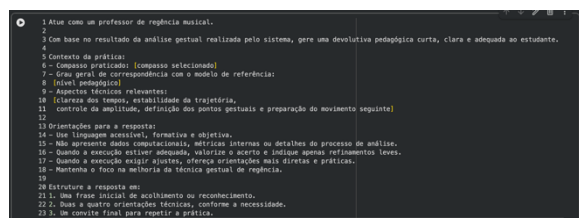


Figura 3. Exemplo de *prompt* utilizado na estrutura do software.

O feedback parte de uma execução considerada satisfatória (ver exemplo 01). Por isso, a fala começa com um reforço positivo: “Excelente. Seu compasso ternário simples está claro e bem-organizado”. Essa abertura tem função motivacional, pois reconhece o acerto do estudante antes de apresentar qualquer ajuste técnico.

Em seguida, o feedback identifica o que funcionou bem: “Os três tempos aparecem de forma perceptível, e a trajetória mantém boa continuidade técnica entre os tempos”. Pedagogicamente, isso é importante porque o aluno não recebe apenas um elogio genérico; ele entende quais aspectos da sua execução estão corretos. Nesse caso, a clareza dos tempos e a continuidade da trajetória são apontadas como qualidades do gesto.

Depois, o sistema apresenta uma sugestão de refinamento: “deixar o segundo tempo com um ponto de chegada um pouco mais definido”. Essa orientação é específica, mas leve. Ela não invalida a execução do aluno; apenas indica um ponto técnico que pode tornar o gesto mais preciso. No compasso ternário simples, o segundo tempo precisa ser perceptível para que a métrica não fique reduzida apenas ao movimento de “queda e retorno”.

A frase “prepare o retorno ao tempo um com leveza” também tem valor pedagógico. Ela orienta o

estudante a compreender que o terceiro tempo não deve ser tratado como um gesto isolado, mas como uma preparação para o início do próximo compasso. Isso reforça a ideia de continuidade musical e evita uma condução rígida ou fragmentada.

Finaliza-se com um convite “Vamos repetir mais uma vez para consolidar?” transforma o feedback em ação prática. O objetivo não é apenas corrigir, mas estimular a repetição consciente, consolidando o gesto que já está bem encaminhado.

Exemplo 01:

Excelente. Seu compasso ternário simples está claro e organizado. Os três tempos aparecem de forma perceptível, e a trajetória mantém boa continuidade técnica entre os tempos. Para refinar ainda mais, procure deixar o segundo tempo com um ponto de chegada um pouco mais definido e prepare o retorno ao tempo um com leveza. Vamos repetir mais uma vez para consolidar?

No segundo exemplo, o feedback é mais direto porque parte de uma execução que exige ajustes/correções. Ainda assim, a fala começa de maneira acolhedora: “Vamos ajustar com calma”. Essa escolha evita um tom punitivo e cria uma relação pedagógica mais positiva com o erro. O erro é tratado como parte natural do processo de aprendizagem.

Em seguida, como exemplo, o feedback organiza a correção a partir da estrutura do compasso quaternário simples. A orientação “procure definir melhor os quatro pontos do gesto” mostra ao aluno que o problema não está apenas na forma geral do movimento, mas na falta de diferenciação entre os tempos. Em regência, cada tempo precisa ter identidade gestual suficiente para ser compreendido por quem acompanha o/a regente, no caso o instrumentista ou cantor(a).

A explicação continua detalhando a função de cada ponto: o primeiro tempo deve ter chegada mais firme; o segundo precisa de direção clara; o terceiro não deve se confundir com o primeiro; e o quarto deve preparar o próximo compasso. Essa sequência transforma o feedback em um pequeno roteiro técnico. O aluno passa a saber exatamente o que observar na repetição seguinte.

A orientação sobre o primeiro tempo destaca a importância do ictus, ou seja, o ponto de chegada do gesto que comunica o pulso. Quando o primeiro tempo não é firme, a entrada musical pode ficar pouco clara. Já a indicação sobre o segundo tempo reforça a necessidade de direção, evitando que ele seja apenas uma passagem sem função expressiva.



A observação sobre o terceiro tempo é especialmente importante no compasso quaternário simples, pois ele pode ser confundido com uma nova queda ou com um gesto sem identidade. Pedagogicamente, o terceiro tempo precisa aparecer como ponto próprio dentro da organização do compasso. O quarto tempo, por sua vez, tem função preparatória: ele conduz o gesto para o próximo tempo um.

Finalmente, o feedback menciona o controle da amplitude e do retorno. Isso orienta o aluno a evitar movimentos excessivamente grandes ou acelerados, que podem comprometer a estabilidade da pulsação. A frase final “Vamos repetir juntos novamente?” mantém o caráter formativo da resposta, convidando o estudante a aplicar imediatamente as correções sugeridas.

Exemplo 02:

Vamos ajustar com calma. No compasso quaternário simples, procure definir melhor os quatro pontos do gesto: o primeiro tempo deve ter uma chegada mais firme, o segundo precisa ser conduzido com direção clara, o terceiro deve aparecer sem se confundir com o primeiro, e o quarto deve preparar o próximo compasso. Mantenha a amplitude controlada e evite acelerar o retorno. Vamos repetir juntos novamente?

Entre a Fase 1 e a Fase 2, o regime de *prompting* multimodal foi refinado para aumentar o controle, previsibilidade e robustez sem alterar o objetivo pedagógico do sistema. Na Fase 2, o prompt passou a operar com maior restrição de saída, exigindo resposta estruturada em JSON com campos fixos, o que reduziu variações na forma do retorno e facilitou validação automática e síntese por texto-para-fala. Nesse sentido, houve redução de redundâncias no contexto fornecido ao modelo e maior delimitação do escopo da resposta (foco técnico no compasso selecionado e em correções acionáveis), diminuindo verbosidade e ambiguidades.

O processo adotou critérios mais conservadores para lidar com casos ambíguos, privilegiando consistência e estabilidade do reconhecimento e do feedback verbal. Esses ajustes caracterizam uma evolução do “contrato” do *prompt* — isto é, de como evidências são apresentadas e como a resposta é restringida — e oferecem uma explicação metodologicamente plausível para a melhora observada em desempenho e previsibilidade temporal entre as fases, sem expor detalhes internos de implementação.

PROCEDIMENTO METODOLÓGICO

A metodologia foi delineada como um estudo exploratório com três participantes cegos, com foco em observar a viabilidade do software em condições de prática individual e em registrar evidências de

desempenho e usabilidade sem exigir navegação visual durante a prática.

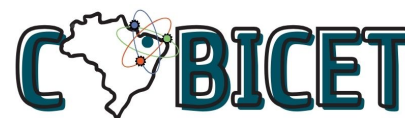
Os participantes foram convidados a executar padrões gestuais de regência diante de uma *webcam*, em sessões curtas e padronizadas por orientação rítmica (metrônomo), de modo a reduzir variações externas (andamento, pausa, posição). A cada tentativa, o sistema capturava o gesto, processava o registro e entregava uma saída estruturada (classe estimada) acompanhada de feedback verbal corretivo, além de gerar automaticamente um *log* em CSV com metadados de execução (por exemplo, carimbo temporal, classe de referência quando informada, classe estimada, latência e indicadores básicos da captura). Esse arranjo permitiu analisar o comportamento do sistema e do usuário dentro de um ciclo de prática “executar–ouvir–corrigir–repetir”, preservando o papel do professor como mediação contextual, mas reduzindo a dependência de validação contínua entre aulas.

No recorte analítico apresentado aqui, adotamos dois eixos de avaliação. O primeiro foi a latência, entendida como o tempo “end-to-end” entre o acionamento da análise pelo usuário e a disponibilização do resultado com feedback verbal. Operacionalmente, a latência foi registrada automaticamente em milissegundos no *log* de cada tentativa, permitindo calcular medidas descritivas (mediana, média e percentis) e comparar consistência temporal (por exemplo, se o sistema mantém resposta estável ao longo da sessão). Esse indicador foi escolhido por ser diretamente ligado à experiência pedagógica: quanto menor e mais estável a latência, maior a chance de o feedback funcionar como correção “no fluxo” de estudo, sem quebrar o encadeamento entre ação e ajuste.

O segundo eixo foi a análise visual dos gráficos gestuais gerados pelo software a partir das trajetórias registradas. Em vez de tratar o gesto apenas como “certo/errado”, os gráficos foram usados como evidência qualitativa para interpretar padrões de execução e possíveis fontes de erro residual. O procedimento consistiu em inspecionar, por tentativa e por sequências de tentativas, características como:

1. dispersão do traço (variabilidade espacial);
2. contorno global e organização do percurso (coerência do padrão);
3. regularidade entre repetições (consistência intra-sujeito);
4. sinais de hesitação (trechos iniciais/finais irregulares).

Essa leitura foi usada para problematizar se determinadas confusões do sistema poderiam estar associadas a variações técnicas do próprio gesto (por exemplo, instabilidade do binário) e para orientar o desenho do feedback verbal corretivo, mantendo a



análise como exploratória e voltada a subsidiar refinamentos do protótipo e hipóteses para estudos futuros.

As execuções dos participantes foram organizadas para simular prática individual realista, porém com padronização mínima para tornar as tentativas comparáveis. Em cada sessão, o participante posicionava-se diante da *webcam* de modo que a mão dominante permanecesse visível e com amplitude suficiente de movimento. A execução era guiada por um metrônomo (andamentos definidos no protocolo) e cada tentativa era realizada como um “bloco” contínuo de gesto, buscando manter regularidade do pulso e consistência do padrão espacial (ictus e retornos). Ao final de cada tentativa, o participante recebia o feedback verbal do software e, quando necessário, repetia o gesto com ajuste técnico sugerido, reforçando a dinâmica execução–correção.

Nas fases de coleta estruturada, os participantes executaram séries de tentativas para os três padrões-alvo (2/4, 3/4 e 4/4), alternando classes de forma controlada para evitar viés por repetição prolongada do mesmo compasso.

A cada tentativa, o participante iniciava e encerrava a captura gestual pelo software, realizava o gesto conforme o compasso solicitado/selecionado e aguardava o feedback verbal. Cada execução gerava um registro independente no log (com classe de referência quando informada, classe estimada e latência), permitindo relacionar “o que foi feito” com “o que o sistema devolveu”.

Na terceira coleta (30 minutos), a execução teve caráter mais próximo de treino: o participante manteve a prática concentrada em um padrão específico (especialmente o binário simples), com repetições sucessivas ao longo do tempo. O objetivo foi observar mudanças na **estabilidade gestual** durante a prática orientada: comparou-se visualmente a trajetória produzida em janelas de tempo (por exemplo, início, meio e fim da sessão), verificando se o traço tendia a ficar mais consistente (menor dispersão e contorno mais regular) à medida que o participante incorporava as correções verbais. Essa etapa complementa as métricas de desempenho ao oferecer evidência qualitativa sobre como o gesto se organiza ao longo do treino, e não apenas se foi classificado corretamente.

A participação foi voluntária e precedida de consentimento; os dados foram coletados e armazenados de forma anonimizada, em conformidade com princípios de privacidade e proteção de dados.

Participaram do estudo três estudantes cegos voluntários, com pelo menos um ano de experiência prévia em regência. A coleta foi organizada em duas

fases: na Fase 1, foram registradas 90 tentativas válidas no total (30 por compasso: 2/4, 3/4 e 4/4), executadas com metrônomo a 80 bpm; na Fase 2, após ajustes no pipeline e no regime de *prompting*, foram registradas 300 tentativas válidas no total (100 por compasso), ampliando a diversidade de execuções sob condições equivalentes.

Em todas as tentativas, o ground truth (compasso esperado) foi definido pelo compasso selecionado no próprio software antes da execução, seguindo o protocolo de blocos do pesquisador, e posteriormente comparado à predição do sistema registrada no log (expected vs predicted), o que permitiu calcular acurácia e métricas por classe, além de relacionar desempenho com latência e com os gráficos gestuais gerados.

RESULTADOS E DISCUSSÃO

Embora o foco do trabalho seja o desenho de *prompts* multimodais e a mediação pedagógica por feedback verbal, reportamos o desempenho de reconhecimento como indicador do fluxo operacional do software. Observa-se aumento de acurácia global da Fase 1 para a Fase 2, com estabilização das classes 3/4 e 4/4 — padrões que dialoga com a análise qualitativa dos gráficos gestuais e com a necessidade de *prompting* ancorado em evidência.

Tabela 2. Desempenho do *fluxo de prompting* multimodal (Fase 1 vs Fase 2).

Fase	N total	N por compasso	Acurácia global
1	90	30	80,0%
2	300	100	94,7%

Tabela 3. Métricas por classe do *fluxo de prompting* multimodal (precisão, revocação e F1).

Compasso	Fase 1 Precisão	Fase 1 Revocação	Fase 1 F1	Fase 2 Precisão	Fase 2 Revocação	Fase 2 F1
2/4	0,76	0,93	0,84	1,00	0,84	0,91
3/4	0,88	0,47	0,61	0,89	1,00	0,94
4/4	0,81	1,00	0,90	0,96	1,00	0,98

As Figuras 4 e 5 apresentam a distribuição de latência end-to-end do sistema, medida em milissegundos e composta por processamento local, custos de rede e tempo de inferência do modelo. Por se tratar de uma métrica que reflete diretamente o tempo percebido pelo usuário entre acionar a análise e receber o resultado com feedback verbal, sua interpretação é central para avaliar a viabilidade do protótipo em um ciclo de prática execução–correção. Os histogramas evidenciam não apenas diferenças no “tempo típico” de resposta, mas principalmente no comportamento da cauda (eventos raros e muito lentos), que tende a ser a principal fonte de frustração e quebra de continuidade em atividades repetitivas.

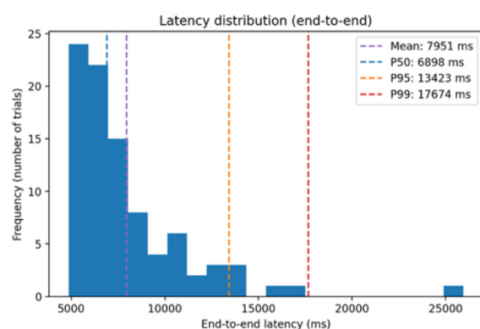


Figura 4. Gráfico de latência referente à fase 1.

Na Fase 1, observa-se uma distribuição fortemente assimétrica à direita, com a maior concentração de tentativas em torno de valores médios-altos e uma cauda extensa. Essa assimetria é explicitada pela diferença entre mediana e média ($P50=6898\text{ms}$; $\text{média}=7951\text{ms}$), indicando que poucas execuções muito lentas elevam a média. Os percentis superiores reforçam esse padrão: $P95=13423\text{ms}$ e $P99=17674\text{ms}$, o que significa que 5% das tentativas ultrapassaram $\sim 13,4\text{ s}$ e 1% ultrapassou $\sim 17,7\text{ s}$. O histograma ainda sugere a presença de outliers extremos, com ocorrências acima de 20 s (aproximando-se de $\sim 26\text{ s}$), compatíveis com episódios de instabilidade (por exemplo, jitter de rede, fila/concorrência de inferência ou reprocessamentos), que degradam a previsibilidade do retorno.

Na Fase 2, a distribuição desloca-se de forma clara para a esquerda e se torna mais concentrada, indicando melhora simultânea em tempo típico e estabilidade. A mediana cai para $P50=4999\text{ms}$ e a média para 5305ms , reduzindo a distância entre ambas e, portanto, a influência de eventos muito lentos. Mais importante, há uma contração acentuada da cauda: $P95=7060\text{ms}$ e $P99=8719\text{ms}$, o que implica que 95% das tentativas ficaram abaixo de $\sim 7,1\text{ s}$ e 99% abaixo de $\sim 8,7\text{ s}$. Embora ainda existam outliers, eles se tornam mais raros e menos extremos ($\sim 10\text{--}12\text{ s}$), sugerindo mitigação dos episódios que anteriormente prolongavam substancialmente o tempo de resposta.

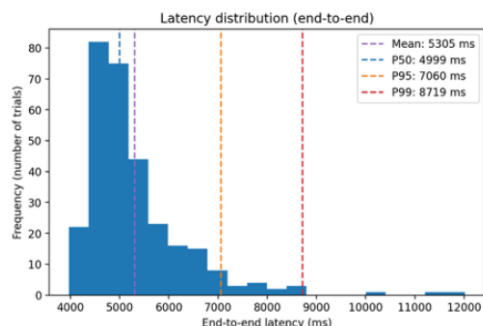


Figura 5. Gráfico de latência referente à fase 2.

Comparativamente, a melhora é consistente e particularmente expressiva nos percentis superiores. A mediana reduziu 27,5% ($6898 \rightarrow 4999\text{ms}$) e a média 33,3% ($7951 \rightarrow 5305\text{ms}$), evidenciando ganho no desempenho global. Entretanto, o efeito mais relevante para a experiência do usuário aparece na redução da variabilidade: $P95$ diminuiu 47,4% ($13423 \rightarrow 7060\text{ms}$) e $P99$ 50,7% ($17674 \rightarrow 8719\text{ms}$), além de uma redução substancial do “excesso de cauda” (diferença entre $P95$ e $P50$).

Em termos práticos, isso indica que o sistema não apenas ficou mais rápido, mas também mais previsível, aumentando a probabilidade de o feedback verbal ocorrer dentro de um intervalo temporal compatível com a prática deliberada, sem interrupções longas que comprometem o encadeamento entre ação, correção e repetição.

Esses ganhos tornam-se ainda mais relevantes quando se considera o perfil de uso do componente de IA. Registros de telemetria do serviço indicam 765 requisições totalizando 106,174 milhões de *input tokens* no período de testes, o que corresponde a aproximadamente 138 mil tokens de entrada por requisição em média. Esse volume é consistente com um regime de *prompting* multimodal e “carregado” de evidências (por exemplo, representações visuais da trajetória e referências), no qual o custo computacional e, consequentemente, a latência passam a ser fortemente sensíveis ao desenho do *prompt* e ao contrato de saída.

Nesse contexto, a queda acentuada nos percentis superiores ($P95/P99$) é compatível com refinamentos que aumentam o determinismo do processo (redução de ambiguidade, formatação estruturada e concisa da resposta e/ou otimização do contexto fornecido), mitigando eventos raros de atraso prolongado associados a variações na geração, instabilidades de rede ou reprocessamentos.

Na etapa de análise dos gráficos gestuais, o estudo assumiu deliberadamente uma leitura que ultrapassa a interpretação estritamente computacional dos resultados. Vale registrar que o autor do trabalho atua simultaneamente como docente, pesquisador e performer na área de regência, além de desenvolvedor do software empregado. Essa confluência de papéis permitiu conduzir a coleta e, sobretudo, interpretar os registros com um olhar técnico-musical rigoroso, articulando critérios de legibilidade gestual, clareza do ictus e consistência espacial — dimensões que, na regência, constituem o próprio objeto de correção pedagógica.

A questão central que emergiu nessa fase foi distinguir duas fontes possíveis para as incorreções residuais observadas:

1. limitações do sistema de reconhecimento;

2. variações reais da execução do estudante que, do ponto de vista pedagógico, demandariam refinamento técnico (trajetória, clareza, estabilidade do plano e regularidade entre tempos).

Como o software gera automaticamente os gráficos da trajetória, tornou-se possível avaliar, em cada tentativa, se a inconsistência apontava mais para a necessidade de ajuste do aluno ou para a necessidade de recalibração do sistema. Em termos metodológicos, isso transforma o erro de classificação em um “sinal” interpretável: não apenas um número, mas uma evidência associada a uma geometria concreta do gesto.

Os gráficos capturam, com fidelidade suficiente, a delineação espacial da prática: ao rastrear um ponto de referência da mão, o sistema armazena séries temporais de coordenadas (x, y) que revelam o “desenho” do gesto ao longo de cada repetição. Isso permite observar amplitude, direção, estabilidade do plano e regularidade entre batidas.

Nas Figuras 6 e 7, torna-se explícita a documentação visual do que foi executado em cada tentativa — e, portanto, viabiliza uma interpretação pedagógica mais precisa do desempenho técnico dos usuários. Um achado recorrente nessa leitura foi a identificação, no 2/4, de uma tendência à horizontalização do trajeto do primeiro para o segundo tempo (indicada pelas setas em verde), delineação que se aproxima do “abrir” característico do 3/4 (do primeiro para o segundo tempo) e do 4/4 (na transição entre tempo 2 e 3), o que ajuda a explicar por que o binário pode “parecer” geometricamente próximo de outros padrões. Em contraste, a técnica esperada para o 2/4 exige que o primeiro tempo seja predominantemente vertical e que a abertura em direção ao segundo tempo se organize, metaforicamente, como um “cabo de guarda-chuva”.

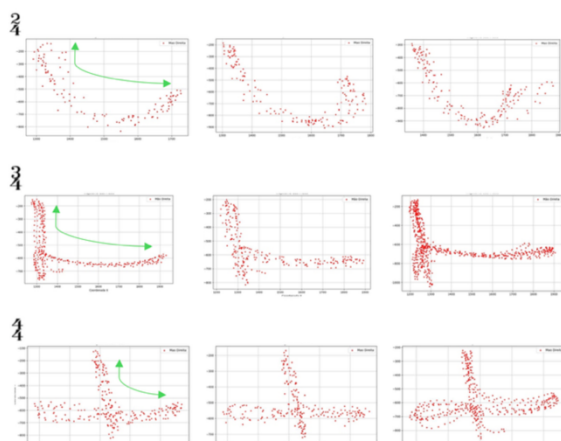


Figura 6. Gráficos gestuais e comparação da abertura do gestual referente ao compasso binário simples.

Essa diferenciação geométrica observável nos gráficos orienta diretamente a mediação por IA: ao invés de produzir apenas um rótulo, o sistema passa a estruturar a análise e a correção a partir de evidências do traço. A partir dos desvios detectáveis (por exemplo, a horizontalização na abertura do 2/4), a IA pode ser solicitada a gerar feedback verbal corretivo focado em objetivos técnicos claros (direção do movimento técnico, estabilidade do plano gestual, e proporção de amplitude, por exemplo).

Diferentemente das Fases 1 e 2 (centradas em reconhecimento, estabilidade e tempo de resposta), foi realizada uma terceira etapa de coleta com foco explícito no efeito do feedback sobre a execução.

Como o compasso binário simples (2/4) permaneceu como o padrão mais sensível a confusões e, simultaneamente, revelou maior necessidade de correção técnica a partir da visualização dos gráficos, ele foi tomado como referência única nessa etapa. Os estudantes praticaram por 30 minutos a 80 bpm, registrando um gráfico ao final de cada intervalo de 10 minutos. A análise foi qualitativa, baseada em inspeção comparativa de dispersão, contorno e consistência do traço ao longo do tempo.

Em geral, observou-se que as nuvens de pontos tendem a se organizar em contornos recorrentes, com maior concentração nos pontos de inflexão (mudanças de direção) e segmentos alongados nas fases de deslocamento contínuo; quanto mais condensada e consistente essa configuração se torna entre repetições, maior parece ser a estabilidade do gesto o que é compatível com um processo de refinamento técnico durante a prática guiada.

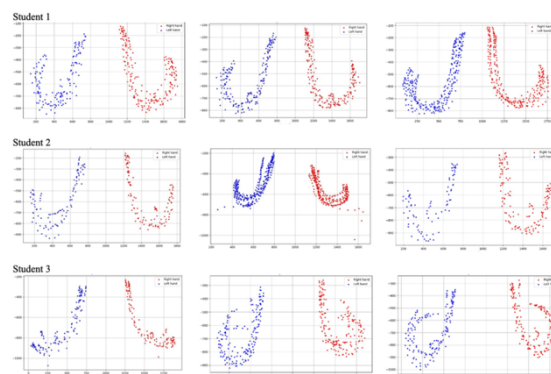


Figura 7. Gráficos gestuais demonstram a progressão do gestual referente ao compasso binário simples.

Os gráficos gestuais funcionam como a “evidência” que dá lastro ao *prompt* multimodal: eles mostram que o que chega ao modelo não é uma descrição abstrata do gesto, mas um registro espacial observável, com contorno, dispersão, pontos de inflexão e diferenças entre mãos. Isso permite que o



prompt seja escrito para ancorar a interpretação em propriedades relativamente estáveis do gesto (forma global, direção entre batidas, regularidade e estabilidade do plano), em vez de depender de inferências livres do modelo.

Ao mesmo tempo, a variabilidade entre estudantes — por exemplo, nuvens mais condensadas *versus* mais dispersas, assimetrias entre mão direita e esquerda e mudanças de densidade nos pontos de inflexão — evidencia a necessidade de o *prompt* estar estruturado para explicitar invariâncias (“priorizar contorno e inflexões; não usar posição absoluta; tratar as mãos como canais distintos”), reduzindo ambiguidade e aumentando consistência tanto na classificação quanto no diagnóstico técnico.

Além de sustentar a decisão classificatória, os gráficos também justificam o desenho do *prompt* como instrumento de mediação pedagógica acessível. Como a saída final é verbal, o *prompt* precisa traduzir padrões visuais em orientações acionáveis: quando a trajetória é estável e recorrente, o modelo pode gerar correções mais específicas; quando a nuvem está excessivamente dispersa, a própria visualização sugere que o melhor feedback é solicitar repetição guiada e ajustes de estabilidade antes de uma correção fina.

Os gráficos conectam diretamente a proposta de acessibilidade sobre prompting: eles mostram por que o *prompt* deve ser estruturado, restritivo e baseado em evidência, para evitar respostas “convincentes, porém pouco úteis” e para garantir retorno verbal consistente que sustente o ciclo execução–correção na prática autônoma.

Adicionalmente, os gráficos gestuais extrapolam sua função analítica e tornam-se um artefato pedagógico de acompanhamento, útil tanto para o estudante quanto para o docente. Ao registrar cada tentativa como uma trajetória comparável, eles constituem um histórico material da prática — um “diário técnico” do gesto — que pode ser revisitado e compartilhado. Isso permite que o aluno documente sua progressão ao longo do tempo — por exemplo, observando maior consistência do contorno, menor dispersão e redução de assimetrias entre mãos — e envie esses registros ao professor entre aulas. Para o docente, esse material funciona como apoio diagnóstico: oferece evidências objetivas do que foi praticado fora do contexto presencial, facilitando intervenções mais precisas e personalizadas e ampliando a continuidade do acompanhamento sem exigir supervisão visual constante durante o estudo individual.

CONSIDERAÇÕES FINAIS

Retomando a problematização da introdução, o ponto de partida deste trabalho não é a suposta “limitação artística” do estudante cego, mas a desigualdade

estrutural produzida por uma pedagogia da regência centrada na observação visual do gesto (Scherchen, 1989; Schuller, 1997). Quando a validação do ictus, da trajetória e da legibilidade técnica depende do olhar constante do docente, reduz-se a possibilidade de prática deliberada com correção imediata — exatamente o mecanismo que sustenta autonomia, confiança e progresso entre aulas. É nesse vazio que o estudo posiciona sua contribuição: propor um ciclo execução–análise–correção em que o feedback verbal não é um “extra”, mas a própria interface acessível da aprendizagem.

À luz do que foi apresentado nesse texto, para a IA generativa operar como mediadora educacional, o *prompt* precisa ser tratado como parte do método: ele define o que conta como evidência, que tipo de resposta é aceitável e como a saída pode ser validada. Protocolos de *prompting* centrados em contexto, estrutura e avaliação reforçam a ideia de que a qualidade do resultado não decorre apenas da capacidade do modelo, mas do contrato de interação estabelecido (clareza de tarefa, formatação, critérios e verificação).

Em paralelo (Yu et al., 2023; Hao et al., 2026), estudos críticos mostram que respostas podem ser convincentes e ainda assim incorretas, o que exige mecanismos de restrição, validação e desenho híbrido (IA como suporte, não substituto). Finalmente, abordagens multimodais reforçam que, quando o fenômeno é corporal/espacial, a ancoragem em sinais estruturados reduz deriva: a multimodalidade não é “efeito visual”, mas estratégia de controle e coerência.

A melhora pronunciada da responsividade (P50: 6898 → 4999 ms; P95: 13423 → 7060 ms; P99: 17674 → 8719 ms) não indicam apenas “um classificador melhor”, mas o amadurecimento de um regime de prompting mais consistente: a redução da cauda e dos outliers sugere aumento de previsibilidade e menor ocorrência de comportamentos raros e custosos (respostas longas, instáveis ou fora do formato esperado). Isso se torna ainda mais significativo quando se observa a alta carga de entrada do sistema (≈ 138 mil *input tokens* por requisição, em média), compatível com *prompting* multimodal ancorado em evidências; nesse cenário, pequenas decisões de design (estrutura e concisão da saída, delimitação do escopo, redução de ambiguidade) tendem a impactar diretamente custo, latência e estabilidade.

Os gráficos gestuais, por sua vez, oferecem a ponte metodológica entre a dimensão técnica e a pedagógica, permitindo discutir o “erro” sem reduzi-lo a falha algorítmica. Ao materializar a trajetória em registros comparáveis, os gráficos gestuais evidenciam propriedades relevantes para a regência



(direção entre batidas, estabilidade do plano, dispersão e regularidade) e tornam interpretáveis as confusões residuais no binário simples. A recorrência de horizontalizações no 2/4, por exemplo, ajuda a explicar por que o padrão pode se aproximar geometricamente de aberturas típicas de 3/4 e 4/4; ao mesmo tempo, define um alvo pedagógico concreto de correção (verticalidade do primeiro tempo e organização da abertura). Nesse sentido, o *prompt* multimodal funciona como mecanismo de mediação: ele transforma evidência do traço em uma saída estruturada (diagnóstico) e em feedback verbal acionável (orientação), reduzindo a dependência de julgamento externo durante o estudo individual.

O fechamento deste estudo articula três níveis que se reforçam mutuamente:

1. a problemática de acessibilidade exige que o feedback verbal seja confiável e imediato para sustentar prática deliberada;
2. a literatura indica que essa confiabilidade depende de *prompting* estruturado, validável e, quando necessário, conservador;
3. os resultados mostram que, ao tratar o *prompt* como componente central — multimodal, restritivo e orientado a metas técnicas — é possível aproximar desempenho, previsibilidade temporal e utilidade pedagógica.

As limitações (amostra reduzida e recorte exploratório) permanecem, mas apontam uma agenda clara: estudos futuros podem realizar ablações de componentes do *prompt*, medir consistência de saída (falhas de formato, necessidade de reprocessamento), e avaliar sistematicamente a qualidade pedagógica do feedback verbal com especialistas e usuários, fortalecendo a contribuição do *prompting* multimodal como estratégia de acessibilidade no ensino e preparação técnico-performativa da música.

REFERÊNCIAS

- CASAS, Llogari; MITCHELL, Kenny J. Structured Teaching Prompt Articulation for Generative-AI Role Embodiment with Augmented Mirror Video Displays. In: THE 19TH ACM SIGGRAPH INTERNATIONAL CONFERENCE ON VIRTUAL-REALITY CONTINUUM AND ITS APPLICATIONS IN INDUSTRY — VRCAI '24, 2024, Nanjing, China. *Proceedings [...]*. New York: ACM, 2024. p. 1-7. DOI: <https://doi.org/10.1145/3703619.3706049>.
- DAHÒ, Margherita; CACI, Barbara. Exploring AI-assisted design of executive function rehabilitation programs for individuals with ADHD: a mixed-methods evaluation of prompts and ChatGPT outputs. *BMC Psychology*, v. 14, artigo 25, 2026. DOI: <https://doi.org/10.1186/s40359-025-03729-2>.
- ENDLER, Michael; RAHMANI, Nazmul; KHAN, Faisal. Analysis of the AI prompt used to improve safety management system accessibility. *Journal of Loss Prevention in the Process Industries*, v. 99, artigo 105834, 2026. DOI: <https://doi.org/10.1016/j.jlp.2025.105834>.
- FARBERMAN, Harold. *The art of conducting technique: a new perspective*. Warner Bros., 1997.
- GARBUNOVA, Irina. Music computer technologies in informatics and music studies at schools for children with deep visual impairments: from the experience. In: *Lecture Notes in Computer Science*. Cham: Springer, 2018. v. 11169.
- HAO, S.; TENG, F.; HOU, R. et al. Explainable AI and echo state networks calibrate trust in human-machine interaction. *Scientific Reports*, v. 16, artigo 1189, 2026. DOI: <https://doi.org/10.1038/s41598-025-30899-1>.
- HIRANCHAN, Sansanee; KIATTISIN, Supaporn. The Current State of Generative AI Prompt Framework Design for Enhancing Utility in Organizational Decision-Making. In: TECHNOLOGY INNOVATION MANAGEMENT AND ENGINEERING SCIENCE INTERNATIONAL CONFERENCE — TIMES-ICON, 5., 2024. *Proceedings [...]*. IEEE, 2024. DOI: <https://doi.org/10.1109/TIMES-ICON61890.2024.10630713>.
- LI, M. Intelligent music recognition system based on neural network algorithm. In: INTERNATIONAL CONFERENCE ON MOBILE NETWORKS AND WIRELESS COMMUNICATIONS — ICMNWC, 3., 2023, Tumkur, India. *Proceedings [...]*. IEEE, 2023. p. 1-6. DOI: <https://doi.org/10.1109/ICMNWC60182.2023.10435795>.
- LIMA, Erickinson. Can blind people conduct musical ensembles? The deconstruction of visual dependency in conducting teaching through Maestro v0.1. *Música Hodie*, v. 24, e.76883, 2024. DOI: <https://doi.org/10.5216/mh.v23.76883>.
- QI, Xin; ZHOU, Jianlong; DONG, Yifei; CHEN, Fang. Enhancing collaborative learning through AI-powered feedback systems. *International Journal of Data Science and Analytics*, v. 21, artigo 20, 2026. DOI: <https://doi.org/10.1007/s41060-025-00882-0>.
- ROBERTSON, Jeandri; FERREIRA, Caitlin; BOTHA, Elsamari; OOSTHUIZEN, Kim. Game



changers: A generative AI prompt protocol to enhance human-AI knowledge co-construction. *Business Horizons*, v. 67, p. 499-510, 2024. DOI: <https://doi.org/10.1016/j.bushor.2024.04.008>.

RUDOLF, Max. *The grammar of conducting: a comprehensive guide to baton technique and interpretation*. Schirmer/Thomson Learning, 1997.

SCHERCHEN, Hermann. *Handbook of conducting*. Oxford: Oxford University Press, 1989.

SCHULLER, Gunther. *The compleat conductor*. New York: Oxford University Press, 1997.

YU, X.; MA, N.; ZHENG, L.; WANG, L.; WANG, K. Developments and applications of artificial intelligence in music education. *Technologies*, v. 11, n. 2, artigo 42, 2023. DOI: <https://doi.org/10.3390/technologies11020042>.

FINANCIAMENTO

Este trabalho é financiado por fundos nacionais através da FCT — Fundação para a Ciência e a Tecnologia, I.P., no âmbito do Estímulo ao Emprego Científico — Apoio Individual — (2022.02280.CEECIND).