



Facultat de Filosofia i Lletres Departament d'Art i Musicologia

INTEL·LIGÈNCIA ARTIFICIAL I PATRIMONI MUSICAL:

UN CAS D'ESTUDI SOBRE LA VIABILITAT PRÀCTICA DE L'OMR.

Tutors

Carles Badal Perez Alarcon

Alicia Fornés

Alumne

Edgar Zamora Leal

Universitat Autònoma de Barcelona, Juny 2026

Resum: Aquest treball avalua la viabilitat pràctica de diferents eines de Reconeixement Òptic de Música (OMR) aplicades a la transcripció de manuscrits històrics. Mitjançant un estudi comparatiu entre la transcripció manual, un sistema d'OMR clàssic (Audiveris) i un model multimodal d'intel·ligència artificial (Gemini), s'analitza fins a quin punt aquestes tecnologies poden reduir el temps i l'esforç de treball del musicòleg.

L'estudi es basa en tres documents amb característiques diferenciades de degradació, complexitat gràfica i irregularitat cal·ligràfica. Els resultats mostren que l'eficàcia dels sistemes depèn de les característiques de la font documental. Mentre que els models multimodals ofereixen una major flexibilitat en determinats contextos, cap de les eines analitzades elimina la necessitat de revisió i validació especialitzada.

La recerca conclou que les tecnologies actuals poden actuar com a eines d'assistència en la transcripció musical, però encara requereixen una intervenció significativa per garantir la fiabilitat dels resultats.

Paraules clau: Reconeixement Òptic de Música (OMR), Musicologia Digital, Intel·ligència Artificial Multimodal, Transcripció Musical, Patrimoni Arxivístic, Postedició, MusicXML, Models de Llenguatge (MLLM).

Resumen: Este trabajo evalúa la viabilidad práctica de distintas herramientas de Reconocimiento Óptico de Música (OMR) aplicadas a la transcripción de manuscritos históricos. Mediante un estudio comparativo entre la transcripción manual, un sistema de OMR clásico (Audiveris) y un modelo multimodal de inteligencia artificial (Gemini), se analiza hasta qué punto estas tecnologías pueden reducir el tiempo y el esfuerzo de trabajo del musicólogo.

El estudio se basa en tres documentos con diferentes niveles de degradación, complejidad gráfica e irregularidad caligráfica. Los resultados muestran que la eficacia de los sistemas depende en gran medida de las características de la fuente documental. Aunque los modelos multimodales ofrecen una mayor flexibilidad en determinados contextos, ninguna de las herramientas analizadas elimina la necesidad de revisión y validación especializada.

La investigación concluye que las tecnologías actuales pueden actuar como herramientas de apoyo en la transcripción musical, pero todavía requieren una intervención significativa para garantizar la fiabilidad de los resultados.

Palabras clave: Reconocimiento Óptico de Música (OMR), Musicología Digital, Inteligencia Artificial Multimodal, Transcripción Musical, Patrimonio Archivístico, Posedición, MusicXML, Modelos de Lenguaje (MLLM).

Abstract

This study evaluates the practical viability of different Optical Music Recognition (OMR) tools for the transcription of historical manuscripts. Through a comparative analysis of manual transcription, a traditional OMR system (Audiveris), and a multimodal artificial intelligence model (Gemini), it examines the extent to which these technologies can reduce the time and effort required in musicological work.

The study focuses on three documents presenting different levels of degradation, graphical complexity, and handwriting irregularity. The results show that system performance strongly depends on the characteristics of the source document. Although multimodal models provide greater flexibility in certain contexts, none of the evaluated tools eliminate the need for expert review and validation.

The research concludes that current technologies can support musical transcription workflows, but they still require significant human intervention to ensure reliable results.

Keywords: Optical Music Recognition (OMR), Digital Musicology, Multimodal Artificial Intelligence, Music Transcription, Archival Heritage, Post-Editing, MusicXML, Large Language Models (MLLMs).

Índex de contingut:

1. Introducció.....	4
1.1. Motivacions.....	5
1.2. Objectius del treball.....	5
2. Estat de la qüestió.....	6
2.1 Marc Teòric.....	6
2.2 Eines Individuals.....	10
3. Metodologia i Cas d'Estudi Pràctic.....	12
3.1. Eines tecnològiques avaluades	12
3.2 Elecció de corpus	15
3.2.1. Fragment del fons CEDOC (Signatura: 1.5.1_0186)	15
3.2.2. Fragment del fons XAC (ACUR_TagFAu128_014)	16
3.2.3. Manuscrit de P. I. Txaikovski: El Trencaous (Op. 71)	17
3.3. Procediment d'extracció i postedició	18
3.3.1. Fases del protocol experimental.....	18
3.3.2. Anàlisi del cost temporal i l'experiència musicològica	18
4. Resultats.....	19
4.1. Anàlisi per document.....	19
4.1.1 CEDOC_CMM_1.5.1_0186.073	19
4.1.2 XAC_ACUR_TagFAu128_014	21
4.1.3 Txaikovski (Obertura “El Trencaous”)	22
4.2. Síntesi dels resultats	23
5. Conclusions generals	24

1. Introducció

La conservació, l'estudi i la difusió del patrimoni musical depenen en gran mesura de la possibilitat d'accedir als manuscrits que el transmeten. Arxius, biblioteques i centres de documentació custodien una gran quantitat de manuscrits que, malgrat el seu valor històric i musical, continuen presentant dificultats d'accés i anàlisi. Factors com la degradació del suport, les particularitats cal·ligràfiques dels copistes o la complexitat de la notació fan que la seva transcripció continuï requerint una inversió considerable de temps i coneixements especialitzats.

En aquest context, el Reconeixement Òptic de Música (Optical Music Recognition, OMR) s'ha consolidat com una de les principals línies de recerca orientades a facilitar la digitalització i l'explotació d'aquest patrimoni documental. Durant les darreres dècades, la disciplina ha evolucionat des de sistemes basats en regles geomètriques i processos de segmentació seqüencial fins a arquitectures d'aprenentatge profund capaces d'analitzar la partitura de manera global. Paral·lelament, l'aparició dels grans models multimodals ha obert noves possibilitats per a la interpretació automàtica de materials complexos.

Malgrat aquests avenços, bona part dels resultats publicats provenen d'entorns controlats o de conjunts de dades específicament preparats per a l'entrenament i l'avaluació dels sistemes. Aquesta situació planteja una qüestió rellevant per a la pràctica musicològica: fins a quin punt les eines actualment accessibles poden reduir de manera efectiva el temps i l'esforç necessaris per transcriure manuscrits històrics reals?

Aquest treball parteix d'aquesta pregunta i analitza el rendiment de diferents eines de reconeixement musical a partir d'un conjunt de manuscrits amb característiques diverses. Més enllà de la precisió tècnica dels sistemes, l'estudi posa l'accent en un aspecte sovint poc present en les avaluacions tecnològiques: el temps de post edició i la càrrega de treball que assumeix l'investigador per convertir una transcripció automàtica en un resultat musicològicament fiable.

1.1. Motivacions

Aquest Treball de Final de Grau neix de l'interès per analitzar la distància existent entre els avenços descrits a la literatura especialitzada i les possibilitats reals de les eines disponibles per a l'investigador. Tot i la millora constant dels sistemes d'OMR, la seva utilitat pràctica no depèn únicament de la capacitat de reconèixer símbols musicals, sinó també del temps necessari per revisar, corregir i validar les observacions obtingudes.

Des d'aquesta perspectiva, resulta especialment rellevant comparar l'esforç requerit per una transcripció manual amb el temps necessari per corregir una transcripció generada automàticament. Aquesta comparació permet valorar si les tecnologies actuals constitueixen una ajuda real en l'estudi de fonts històriques o si, en determinades situacions, continuen exigint una intervenció humana equivalent a la d'una transcripció tradicional.

1.2. Objectius del treball

L'objectiu principal d'aquesta recerca és avaluar la viabilitat pràctica de diferents eines de reconeixement musical aplicades a manuscrits històrics.

Per assolir aquest objectiu general, es plantegen els objectius específics següents:

- Comparar el temps necessari per realitzar una transcripció manual amb el temps requerit per revisar i corregir transcripcions generades automàticament.
- Analitzar el comportament de diferents eines davant documents amb característiques diverses, com la degradació física del suport, les irregularitats cal·ligràfiques o l'elevada densitat de la notació.
- Identificar les principals limitacions i patrons d'error associats a cada tecnologia analitzada.
- Valorar el paper actual de la intel·ligència artificial com a eina d'assistència en la recerca musicològica i la seva influència en els processos de transcripció documental.

2. Estat de la qüestió

2.1 Marc Teòric

Comprendre una partitura va molt més enllà de detectar taques de tinta sobre un paper; un repte monumental que ha marcat tota l'evolució de la tecnologia de l'OMR (Reconeixement Òptic de Música) ha culminat un llarg procés de maduració tecnològica: des d'una informàtica clàssica i rígida que només cercava patrons geomètrics, fins a una intel·ligència artificial capaç de llegir i comprendre la notació amb la flexibilitat d'un músic expert.

Durant les primeres dècades, la disciplina es va centrar exclusivament en la "transcripció automàtica de partitures impreses" ¹. El procediment inicial consistia en programar regles geomètriques per cercar patrons visuals predefinits, com ara les mides exactes dels caps de nota o la longitud de les pliques. No obstant això, tal com recullen Blostein i Baird en la seva revisió de 1992, la identificació de característiques geomètriques esdevenia "inadequada per a imatges degradades" ². En la pràctica, es va observar que qualsevol variació o soroll visual derivat de l'escaneig deixava sense efecte la regla programada, fent evident la necessitat d'abandonar l'heurística fixa per avançar cap a un processament d'imatge més adaptatiu.

Per gestionar aquest augment de complexitat, la comunitat científica va establir el "pipeline modular". Aquest model estructura l'OMR en etapes seqüencials i independents: preprocessament, detecció de pentagrames, reconeixement de símbols i assemblatge semàntic ³.

Aquesta estandardització va respondre amb la necessitat, subratllada per Choudhury et al. (2000), d'integrar els sistemes d'OMR en projectes de digitalització a gran escala per abastir arxius de biblioteques digitals i facilitar la cerca per contingut ⁴. Tanmateix, aquest model seqüencial va xocar amb un obstacle tècnic principal: l'eliminació de pentagrames (*staff removal*).

¹Nicholas J. Pruslin, *Automatic Recognition of Sheet Music*. Cambridge: MIT, 1966, p. 12.

²Dorothea Blostein i Henry S. Baird, "A Critical Study of OMR", *Structured Document Image Analysis*. Berlín: Springer, 1992, p. 406.

³Nicholas J. Bainbridge i Timothy C. Bell, *The Challenge of Optical Music Recognition*. Londres: Computers and the Humanities, 1997, p. 95.

⁴Sayeed Choudhury et al., "Optical music recognition system within a large-scale digitization project". Plymouth: ISMIR, 2000, p. 1.

Estudis com el de Bainbridge i Bell (2001) van demostrar que l'eliminació de les línies del pentagrama suposava un procés destructiu, ja que provocava la fragmentació dels símbols i una pèrdua d'informació important per a les etapes posteriors ⁵. Aquest problema va ser confirmat per revisions exhaustives de l'estat de l'art, com la de Rebelo et al. (2012), que van identificar l'eliminació de pentagrames com el principal coll d'ampolla de l'OMR clàssic, ja que la barreja de tinta feia impossible una separació neta⁶.

Com a alternativa a aquest mètode autors com Bellini, Bruno i Nesi (2001) van proposar mètodes alternatius de segmentació basats en perfils de projecció, capaços de detectar símbols bàsics sense requerir l'eliminació prèvia i destructiva de les línies del pentagrama⁷

Ben aviat es va constatar que programar manualment totes les excepcions i variacions de la notació musical era una tasca inabastable. Això va accelerar el salt definitiu cap al *Machine Learning*: el sistema ja no obeïa normes rígides, sinó que aprenia a través de classificadors estadístics. Per aconseguir que aquests models fossin robustos davant qualsevol cal·ligrafia o impressió, es van fer servir tècniques d'augmentació de dades. Mitjançant estratègies com la deformació elàstica de les imatges, es va entrenar la màquina per incrementar exponencialment la seva flexibilitat en la interpretació⁸. En definitiva, es tractava de deixar enrere les normes i patrons definitius per avançar cap a un model capaç d'entendre a través de masses d'exemples, aconseguia així la capacitat de generalització que li permetia transcriure formes noves gràcies a la deducció de regularitats implícites.

Tot i disposar d'un marc d'avaluació estable, l'anàlisi de Rebelo et al. (2012) indicava que el reconeixement de manuscrits històrics continuava sent un dels principals reptes de la disciplina⁶. Altres investigacions assenyalaven que els errors de classificació rítmica seguien sent "massa freqüents per fer de l'OMR una eina pràctica" en entorns operatius reals⁸.

⁵ David Bainbridge i Timothy C. Bell, "The Problem of Staff Line Removal", *IJDAR*. Berlín: Springer, 2001, p. 45.

⁶ Ana Rebelo et al., "OMR: A comprehensive survey", *Expert Systems with Applications*. Amsterdam: Elsevier, 2012, p. 5275.

⁷ P. Bellini, I. Bruno i P. Nesi, "Optical Music Sheet Segmentation", *Proceedings of the First International Conference on WEB Delivering of Music*. Florència: IEEE, 2001, p. 12

⁸ Ana Rebelo et al., "Optical recognition of music symbols: A comparative study", *IJDAR*. Berlín: Springer, 2010, p. 20

Com a millora a aquestes limitacions, Padilla et al. (2014) van introduir estratègies de *Multi-OMR*, basades en la combinació de diversos sistemes de reconeixement en paral·lel per creuar evidències i minimitzar falles en la interpretació⁹.

No obstant això, el veritable punt d'inflexió tecnològic va arribar amb la consolidació de l'aprenentatge profund (*Deep Learning*). Davant d'aquest escenari, el desenvolupament de xarxes neuronals combinades, com les proposades per Wen (2014), va aportar una resistència fins aleshores inèdita en la classificació de símbols, reduint la incidència de les diferències en la forma i el traçat de la notació¹⁰. Al mateix temps, la introducció de tècniques d'aprenentatge no supervisat, com l'algoritme K-means aplicat per Yang i Teng (2015), va millorar de manera significativa la fase de segmentació inicial. Aquestes innovacions van permetre distingir amb més precisió els elements musicals dels components externs, com elements de fons o les anotacions textuais, en fonts de gran complexitat visual¹¹.

Aquests avenços van preparar el camí per a la transformació més radical de la disciplina l'any 2019, Baró et al. Va proposar un dels primers models de referència (baseline) per al reconeixement de música manuscrita basat en arquitectures CRNN. Com a sistema End-to-End, aquesta xarxa processava la partitura de manera global, llegint la imatge d'esquerra a dreta i evitant la pèrdua d'informació derivada de les segmentacions clàssiques¹². Ara bé, l'anàlisi de manuscrits històrics en alta resolució comportava un cost computacional elevat, ja que gran part dels píxels corresponien al fons degradat del exemplar.

⁹ Victor Padilla et al., "Improving OMR for Digital Music Libraries with Multiple Recognizers". Londres: ACM, 2014, p. 3.

¹⁰ J. Wen, "A robust symbol classification method for optical music recognition", Proceedings of the ICPR. Estocolm: IEEE, 2014, p. 1.

¹¹ B. Yang i M. Teng, "Unsupervised learning for music symbol segmentation", Journal of New Music Research. Londres: Taylor & Francis, 2015, p. 1.

¹² Arnau Baró et al., "From Optical Music Recognition to Handwritten Music Recognition: A baseline," *Pattern Recognition Letters* 123 (2019)

Per reduir aquest problema, es van introduir els Selectional Auto-Encoders (SAE)¹³. Aquest model actuava com una etapa prèvia de selecció visual, separant la tinta del fons abans del reconeixement. D'aquesta manera, la xarxa només treballava amb la informació rellevant, fet que va permetre reduir considerablement el temps d'anàlisi de les imatges¹⁴.

Malgrat aquesta millora en eficiència, el sistema continuava depenent principalment de la informació visual i presentava dificultats davant de taques, degradacions o grafies ambigües. Per abordar aquesta limitació, Torras (2021) va incorporar arquitectures Sequence-to-Sequence (Seq2Seq) combinades amb Models de Llenguatge (LM)¹⁵. Mitjançant tècniques com el *Shallow Fusion*, el sistema podia tenir en compte el context musical durant la transcripció. Això permetia inferir la nota o el ritme més probable quan la informació visual era insuficient, reduint els errors de reconeixement¹⁶.

Finalment, les arquitectures basades en Transformers van ampliar aquesta capacitat contextual. A diferència dels models recurrents Seq2Seq, limitats per la dependència seqüencial i la dificultat per mantenir relacions a llarg termini, els Transformers utilitzen mecanismes d'atenció global que processen simultàniament tots els elements de la pàgina¹⁷. Això facilita la detecció de relacions estructurals complexes —com relacionar l'armadura inicial amb les notes finals— i resulta especialment útil en partitures polifòniques i visualment denses.

¹³ Francisco J. Castellanos et al., "Selectional Auto-Encoders for OMR of Handwritten Manuscripts," *Expert Systems* 35, núm. 4 (2018)

¹⁴ Arnau Baró Mas, *Reading Music Systems: From Deep Optical Music Recognition to Contextual Methods* (Tesi doctoral, Universitat Autònoma de Barcelona, 2022)

¹⁵ Pau Torras Coloma, *Integration of Language Models in Sequence to Sequence Optical Music Recognition Systems* (Treball de Fi de Grau, Universitat Autònoma de Barcelona, 2021)

¹⁶ Pau Torras et al., "On the Integration of Language Models into Sequence to Sequence Architectures for Handwritten Music Recognition" (Centre de Visió per Computador - Universitat Autònoma de Barcelona, 2021)

¹⁷ Jorge Calvo-Zaragoza i Cuihong Wen, "A Sequence-to-Sequence Framework Based on Transformer with Masked Language Model for Optical Music Recognition," *IEEE Access* 10 (2022)

2.2 Eines Individuals

No obstant això, aquesta mateixa sofisticació algorísmica i el viratge cap a l'atenció global generen una paradoxa crítica en la disciplina: com més complexa i potent és l'arquitectura teòrica, més s'allunya de la realitat material del musicòleg.

Després de revisar les principals línies d'investigació, cal preguntar-se fins a quin punt aquests avenços són accessibles per als musicòlegs en la seva pràctica quotidiana ja que molts investigadors continuen tenint dificultats per accedir a aquestes tecnologies. Segons Ríos-Vila, Fuentes-Martínez i Calvo-Zaragoza, els sistemes més avançats sovint requereixen recursos computacionals elevats i infraestructures que excedeixen les possibilitats d'un investigador individual¹⁸. Per aquest motiu, gran part de la pràctica musicològica continua depenent de programaris de codi obert com Audiveris, considerat una de les principals eines d'OMR de pàgina completa¹⁹.

Diversos estudis han analitzat el rendiment d'Audiveris com a suport per a la transcripció musical. Torras, Biswas i Fornés assenyalen que el sistema reconeix correctament el 76,9% de les mesures en una pàgina tipogràfica estàndard, però presenta dificultats davant documents més complexos²⁰. Els autors també observen que el programa pot ometre aproximadament un 20% dels caps de nota negres²¹. Aquestes omissions afecten la resta del procés de reconeixement, ja que alteren la mètrica i la relació temporal entre els esdeveniments musicals. Com a conseqüència, l'usuari ha d'invertir un temps considerable en la correcció manual de les dades generades²².

¹⁸ Antonio Ríos-Vila, Francisco J. Fuentes-Martínez i Jorge Calvo-Zaragoza, "Towards Sheet Music Information Retrieval: A Unified Approach Using Multitask Transformers", a *Proceedings of the 6th International Workshop on Reading Music Systems* (2024), 7.

¹⁹ Ibídem

²⁰ Pau Torras, Sanket Biswas i Alicia Fornés, "On Designing a Representation for the Evaluation of Optical Music Recognition Systems", a *Proceedings of the 6th International Workshop on Reading Music Systems* (2024), 23

²¹ Ibídem

²² Ibídem

Davant aquestes limitacions, diversos investigadors han començat a explorar l'ús de grans models de llenguatge multimodals (MLLMs), com Claude o Gemini. Calvo-Zaragoza et al., en l'estudi “*Can multimodal large language models read music score images*”, van analitzar la capacitat d'aquests sistemes per interpretar partitures²³. Les conclusions experimentals indiquen que encara tenen dificultats per produir transcripcions simbòliques completament precises. Tot i així, identifiquen amb força fiabilitat característiques generals de la música, com la tonalitat, la textura o determinades relacions contextuais²⁴. Aquesta capacitat els permet afrontar millor alguns problemes derivats de la degradació visual dels registres històrics.

La situació actual mostra dues aproximacions amb fortaleeses i limitacions diferents. D'una banda, els sistemes tradicionals d'OMR ofereixen una transcripció estructurada, però continuen sent sensibles als errors de segmentació. De l'altra, els models multimodals poden interpretar millor el context global del manuscrit, encara que no garanteixen una transcripció musical exacta. En aquest escenari, el rendiment final depèn tant de la qualitat de les eines com de la intervenció de l'usuari durant la post edició.

²³ Jorge Calvo-Zaragoza et al., "Can multimodal large language models read music score images?", a *Proceedings of the 6th International Workshop on Reading Music Systems* (2024), 4.

²⁴ Calvo-Zaragoza et al., "Can multimodal large language models", 4.

3. Metodologia i Cas d'Estudi Pràctic

Després d'analitzar la profunda evolució teòrica de l'OMR a l'estat de la qüestió, he volgut traslladar aquest debat al terreny pràctic. Com a musicòleg, el meu objectiu amb aquest experiment no és avaluar la capacitat computacional abstracta de les màquines, sinó mesurar l'impacte d'aquestes tecnologies en el meu propi flux de treball. Per fer-ho, he dissenyat un cas d'estudi empíric on poso a prova la viabilitat de dues èpoques tecnològiques diferents (el *pipeline* clàssic vs. els *Transformers*) enfront de la transcripció manual tradicional.

En primer lloc, es realitza una transcripció completament manual que serveix com a valor de referència (*baseline*). En segon lloc, s'avalua el temps necessari per revisar i corregir les transcripcions produïdes automàticament pels diferents sistemes analitzats.

La comparació entre aquests dos processos permet valorar l'eficiència real de les eines estudiades i determinar si la reducció del temps de treball compensa l'esforç de postedició que requereixen els resultats generats automàticament

3.1. Eines tecnològiques avaluades

Abans de descriure el corpus documental, cal presentar les eines utilitzades durant la fase pràctica d'aquest treball. En la metodologia comparo tres tipus de sistemes que representen aproximacions diferents al reconeixement i la transcripció musical.

3.1.1. MuseScore

MuseScore ha constituït l'entorn de treball principal de la recerca. Aquest editor de partitures de codi obert s'ha utilitzat amb dues funcions. En primer lloc, ha servit per elaborar les transcripcions manuals de referència (*baseline*), introduint les notes directament des dels manuscrits. Aquest procés ha permès establir el temps necessari per a una transcripció completament manual. En segon lloc, MuseScore ha funcionat com a eina de validació i correcció dels fitxers generats per Audiveris i pels models d'intel·ligència artificial, facilitant les tasques de post edició i revisió musical.

3.1.2 Audiveris

Audiveris representa l'enfocament clàssic de l'Optical Music Recognition (OMR). Segons la documentació oficial del projecte, el sistema està dissenyat per processar partitures escrites en Notació Musical Occidental Comuna (CWMN) i orienta el seu funcionament principalment cap a documents impresos. La mateixa documentació indica que no ofereix suport específic per al reconeixement de manuscrits però internament, Audiveris divideix el procés de reconeixement en diverses etapes consecutives, com la detecció de pentagrames, compassos i símbols musicals. Per fer-ho, genera internament un graf de dependències (associant, per exemple, una plica amb el seu cap de nota i amb l'armadura prèvia) i ho tradueix a l'estàndard MusicXML. Aquest llenguatge de marcatge obert és el pont universal entre l'OMR i els editors de partitures (com MuseScore). L'algorisme llegeix les coordenades espacials dels elements reconeguts a la pantalla i els assigna etiquetes semàntiques (<note>, <pitch>, <duration>). És precisament en aquest pas de traducció on es produeix l'efecte de "validació destructiva" observat en l'experimentació: si Audiveris no pot tancar correctament l'estructura de l'arbre XML per manca de dades gràfiques (per exemple, un compàs que no suma els temps correctes), l'exportació fallarà o ometrà el fragment, traslladant un fitxer incomplet a l'editor. Aquest funcionament segueix el model de *pipeline* seqüencial descrit anteriorment a l'estat de la qüestió. La seva incorporació a l'estudi permet analitzar fins a quin punt aquest tipus d'arquitectura manté el rendiment quan treballa amb manuscrits històrics degradats.²⁵

3.1.3 Models multimodals (Gemini)

Gemini representa una aproximació diferent basada en grans models de llenguatge multimodals (MLLMs). Aquests sistemes utilitzen arquitectures Transformer processant simultàniament informació visual i textual.

Per obtenir les transcripcions, l'usuari proporciona instruccions textuales (*prompts*) que guien la interpretació de la imatge. En demanar a Gemini la transcripció d'un fragment, el model analitza la imatge i infereix la sintaxi musical, traduint-la directament a codi estàndard de representació musical, a causa de la seva forta presència en les dades d'entrenament de la màquina. El model actua com un intèrpret que redacta, etiqueta a etiqueta, l'estructura de la partitura.

²⁵ Audiveris, "Audiveris Handbook: OMR Engine for CWMN"

A continuació, s'exposa un breu exemple d'escriptura MusicXML generada pel model, il·lustrant com la IA tradueix un element visual (una nota Do negra) en una jerarquia d'etiquetes lògiques interpretables per MuseScore:

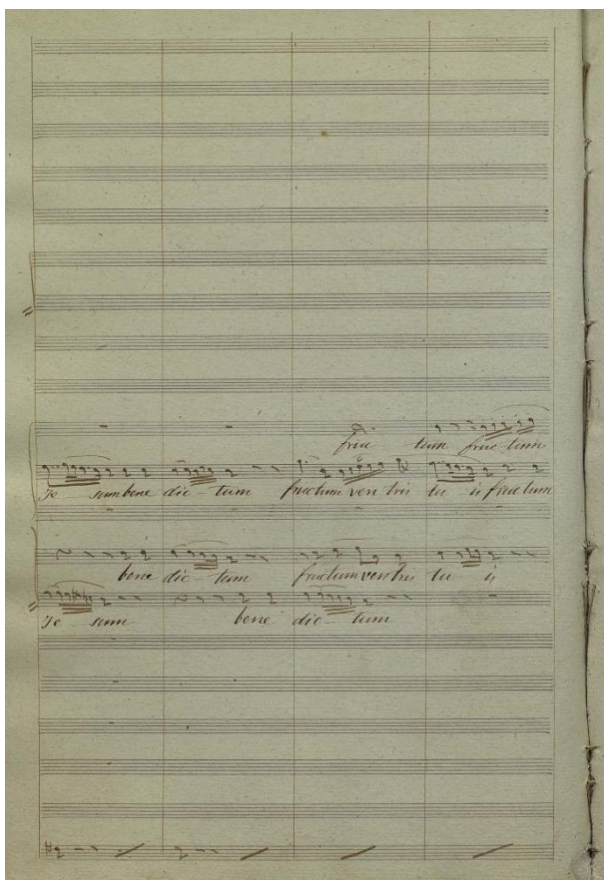
```
<measure number="1">
  <attributes>
    <divisions>1</divisions>
    <key>
      <fifths>0</fifths>
    </key>
    <time>
      <beats>4</beats>
      <beat-type>4</beat-type>
```

La seva inclusió en aquest treball permet avaluar fins a quin punt un model multimodal generalista pot afrontar la complexitat gràfica dels manuscrits històrics i generar resultats útils per a la transcripció musical

Per garantir que l'experiment reproduïxi els reptes reals del treball d'arxiu, he delimitat la prova a tres fragments de música manuscrita amb diferents nivells de degradació i complexitat cal·ligràfica. Aquestes són les tres obres que he escollit:

3.2 Elecció de corpus

3.2.1. Fragment del fons CEDOC (Signatura: 1.5.1_0186)



Imatge 1: CEDOC_CMM_1.5.1_0186.073

Origen i font: L'original d'aquest document es localitza físicament a la secció de Col·leccions de Partitures del Centre de Documentació de l'Orfeó Català (CEDOC), catalogat sota la signatura 1.5.1_0186 i el codi topogràfic C01.03.02.11. La llicència de la imatge digitalitzada prové d'aquest fons d'arxiu.

Característiques formals: Es tracta d'una peça de música manuscrita datat l'any 1844. L'obra és una *Salve*, una peça de música vocal religiosa de caràcter polifònic escrita per al seu ús a cor pel compositor Josep Barba i Benda. A la pàgina objecte d'estudi, la textura musical s'organitza en un sistema polifònic on les diferents veus corals requereixen una estricta alineació vertical i una lectura simultània dels pentagrames. Repte tecnològic: El document presenta un nivell elevat de *bleed-through* (traspàs de tinta) i un deteriorament general que introdueix una quantitat considerable de soroll visual en la imatge digitalitzada.

3.2.2. Fragment del fons XAC (ACUR_TagFAu128_014)



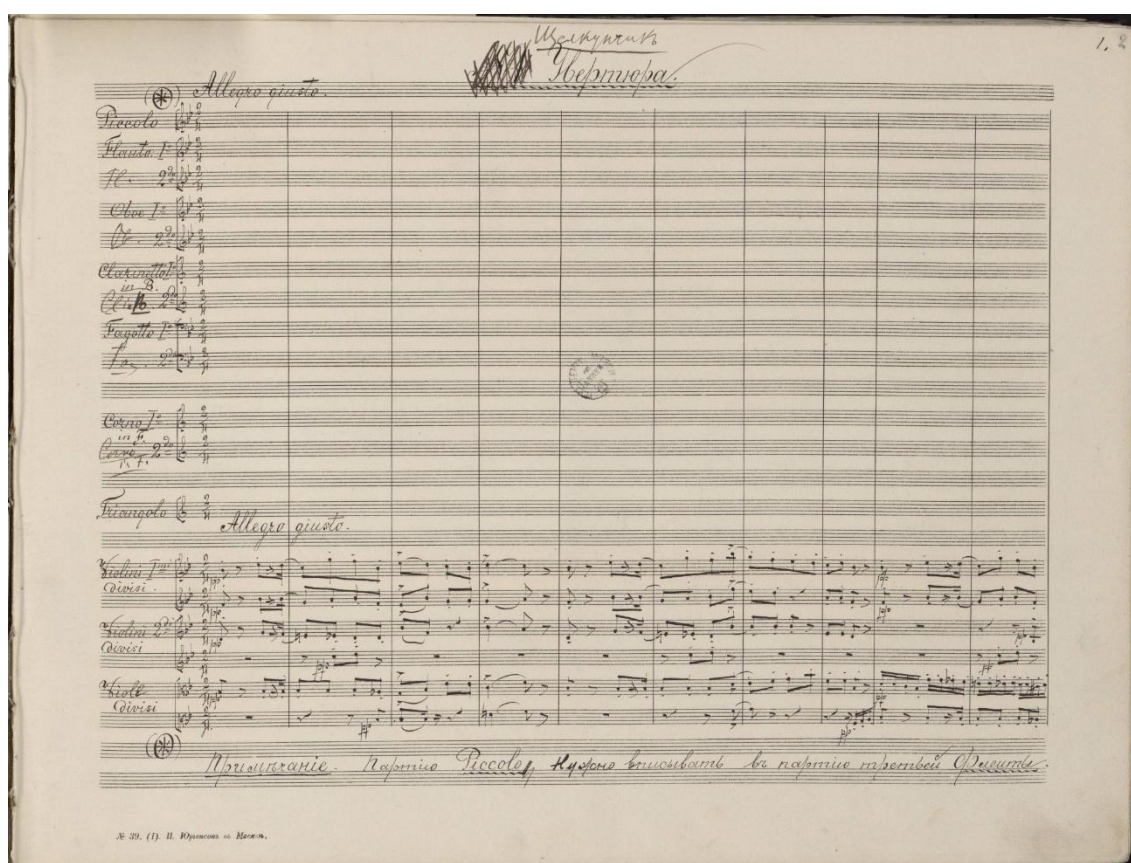
Imatge 2: XAC_ACUR_TagFAu128_014

Origen i font: Aquest document pertany a la Xarxa d'Arxius Comarcals (XAC), concretament a l'Arxiu Comarcal de l'Urgell (ACUR), d'on s'ha obtingut la reproducció digital a partir del portal d'accés en línia del fons d'arxius de la Generalitat de Catalunya.

Característiques formals: L'obra és un manuscrit musical de copista que conté una peça polifònica instrumental/vocal on es distribueix un contrapunt de veus independents sobre un sistema de pentagrames. El traç clarament humà denota una intenció de còpia de conservació o d'ús pràctic immediat.

Repte tecnològic: El principal obstacle d'aquesta font no deriva de l'estat de conservació del suport, sinó de l'asimetria cal·ligràfica del copista. Les inclinacions irregulars de les pliques dificulten que els sistemes d'OMR basats en regles geomètriques associïn correctament els diferents elements de la nota, fet que provoca omissions i errors de reconeixement.

3.2.3. Manuscrit de P. I. Txaikovski: El Trencauous (Op. 71)



Imatge 3: Txaikovski (Obertura “El Trencauous”)

Origen i font: El tercer document correspon a una font manuscrita de l'Obertura del cèlebre ballet *El Trencauous* (Op. 71, TH 14), compost per Piotr Ílitx Txaikovski l'any 1892. La imatge digitalitzada per a l'experiment s'ha obtingut a través del catàleg de la biblioteca digital pública Petrucci Music Library (IMSLP).

Característiques formals: La pàgina pertany al primer gran bloc de l'obra (Obertura). A nivell morfològic, la capçalera presenta una armadura de dos bemolls (tonalitat de Si bemoll major) i indica explícitament el tempo d'*Allegro giusto*. A nivell de textura, la partitura concentra una gran densitat d'informació pròpia de l'escriptura orquestral de finals del segle XIX, amb una superposició elevada de veus i elements notacionals.

El manuscrit constitueix un cas especialment exigent per als sistemes d'OMR a causa de l'efecte cascada. La gran densitat d'acords i la disposició gràfica molt comprimida augmenten la dificultat de segmentació i de reconeixement dels símbols.

Un altre repte important apareix en la presència de text en cal·ligrafia cursiva ciríl·lica, utilitzat per a indicacions d'instruments i dinàmiques. Tant per al lector humà com per als sistemes automàtics, aquest tipus de cal·ligrafia genera ambigüitats visuals per la seva forma i la seva integració amb la notació musical.

3.3. Procediment d'extracció i post edició

3.3.1. Fases del protocol experimental

- Creació del *Baseline* (Transcripció Manual): S'ha realitzat la transcripció manual directa a MuseScore de cada fragment. Aquesta etapa ha servit per fixar el temps de referència (*benchmark*) necessari per a una transcripció acurada des de zero.
- Processament automatitzat: Introducció de les imatges a l'entorn d'Audiveris i als models MLLM (Gemini) per a l'extracció de dades en format MusicXML o codi de sortida.
- Post edició i Correcció: Importació de les dades generades per la màquina a MuseScore i correcció manual fins a obtenir una partitura idèntica a l'original. Aquesta fase és on s'ha mesurat la "càrrega cognitiva" i l'eficiència real del procés.

3.3.2. Anàlisi del cost temporal i l'experiència musicològica

Manuscrit / Fons	Transcripció Manual	OMR Clàssic (Audiveris)	IA Multimodal (Gemini)
CEDOC (Salve)	13m 57s	+25m (25m 40s)	19m 03s
XAC (Contrapunt)	12m 45s	18m 07s	10m 25s
Txaikovski (Orquestral)	14m 59s	+25 m (27m 35s)	23m 48s

4. Resultats

4.1. Anàlisi per document

4.1.1 CEDOC_CMM_1.5.1_0186.073 (Imatge 3): Aquest fragment presenta una degradació física severa. L'Audiveris, en intentar interpretar el soroll del paper com a informació musical, provoca un col·lapse en la segmentació. La post edició ha estat extremadament costosa (25 minuts) perquè ha calgut corregir gairebé compàs per compàs. Gemini ha millorat el temps, però ha requerit una supervisió constant (*mitjançant prompts*) per corregir elements gràfics que interpreta com a notes, però que en realitat són taques o arrugues del suport.

The image shows a musical score for four parts: Trombon, Tiple 1º (Coro 1), Tiple 2º (Coro 1), and Tialto 1º (Coro 1). The score is in 6/8 time and features the lyrics 'Je-sum be-ne-dic-tum fruc-tum ven-tris tu-i fruc-tum'. The Tiple 2º part is highlighted with a red box, and the Tialto 1º part is also highlighted with a red box.

Imatge 4 Lletre i Instruments Transcrits; CEDOC/ Gemini

The image shows a handwritten musical score for four parts: Trombon, Tiple 1º (Coro 1), Tiple 2º (Coro 1), and Tialto 1º (Coro 1). The score is in 6/8 time and features the lyrics 'Je-sum be-ne-dic-tum fruc-tum ven-tris tu-i fruc-tum'. The score is heavily degraded with noise and artifacts.

Imatge 5 Combinació binària, Soroll visual; CEDOC/ Audiveris

Es mostra una diferència clara entre el reconeixement del text i la interpretació de la notació musical. A la Imatge 4 es pot observar com el model multimodal Gemini identifica i transcriu amb precisió tant el text en llatí situat sota els pentagrames com els noms dels instruments presents al marge esquerre. Aquest resultat coincideix amb les conclusions de Calvo-Zaragoza et al. (2024), que assenyalen una bona capacitat dels grans models multimodals per reconèixer informació contextual i elements textuais presents a la partitura, fins i tot en documents històrics manuscrits²³.

La situació canvia quan el sistema ha d'interpretar la informació musical en un document amb un estat de conservació deficient. La Imatge 5 mostra el resultat del preprocessament aplicat per Audiveris. El *bleed-through* (traspàs de tinta), les taques d'humitat i altres alteracions del suport introdueixen una gran quantitat d'elements visuals que dificulten la separació entre els traços musicals i les marques pròpies del manuscrit. Com a conseqüència, el sistema genera errors durant les fases posteriors del reconeixement.

Aquesta observació reforça les conclusions de Blostein i Baird (1992), que ja advertien de les limitacions dels mètodes basats en característiques geomètriques davant d'imatges degradades². Igualment, les dades obtingudes coincideixen amb l'anàlisi de Rebelo et al. (2012), que identifica la separació correcta dels elements gràfics com un dels principals reptes dels sistemes d'OMR tradicionals⁶.

4.1.2 XAC_ACUR_TagFAu128_014 (Imatge 2): És l'únic escenari on MLLMS s'han mostrat eficient (10m 25s). Aquí, la netedat de la imatge ha permès que el model *Transformer* (Gemini) entengués la lògica del contrapunt i proposés una transcripció gairebé completa. En aquest cas, la meua tasca s'ha desplaçat de la "còpia" cap a la "revisió", reduint dràsticament el temps total. Aquest cas il·lustra clarament com la tecnologia pot actuar com una eina de suport si el document té una qualitat cal·ligràfica mínima.



Imatge 6 Encerts de la transcripció, XAC/ Gemini

Tal com es mostra a la Imatge 6, la baixa presència de degradació facilita el reconeixement de l'estructura rítmica per part del model *Transformer*. El sistema genera una transcripció amb una disposició de compassos coherent, fet que redueix el temps de post edició fins als 10 minuts i 25 segons. Aquest resultat s'alinea amb el canvi de paradigma descrit per Baró et al. (2019), on els sistemes End-to-End processen la imatge de manera global i eviten la segmentació prèvia, fet que millora el rendiment en documents amb bona qualitat visual¹².

Malgrat aquest bon resultat general, la revisió a MuseScore ha permès detectar errors en la separació i agrupació de notes. En alguns casos, el sistema representa notes individuals com a grups simultanis, probablement a causa d'una associació imprecisa de les pliques i la seva direcció. ("grups simultanis" – Annex: Imatge12)

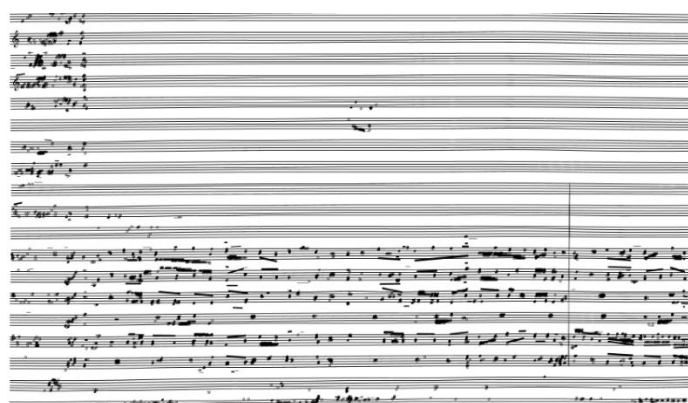
Aquest tipus d'errors es relaciona amb les observacions de Torras, Biswas i Fornés (2024), que assenyalen que les alteracions en els elements verticals de la notació poden afectar la reconstrucció correcta de l'estructura rítmica²⁰. En aquest cas, les petites irregularitats del traç manuscrit dificulten la separació entre veus independents, fet que limita la precisió del reconeixement en contextos polifònics.

Aquesta limitació coincideix amb les conclusions de Rebelo (2010), que destaca les dificultats persistents dels sistemes d'OMR en l'anàlisi de notació manuscrita lliure⁸.

4.1.3 Txaikovski (Obertura “El Trencaous”) : L'obra orquestral és on s'ha produït l'al·lucinació més crítica. Tant Audiveris com Gemini han estat incapaçs de gestionar la densitat. Audiveris s'ha quedat bloquejat en el procés de reconeixement de les pliques, mentre que Gemini ha caigut en una "al·lucinació de patró". El temps de post edició (23m 48s) reflecteix l'increment del temps de revisió necessari: corregir la lògica musical inventada per la màquina és més lent que escriure la nota des de zero, ja que cal primer "desaprendre" l'error induït.



Imatge 7 Repetició de patrons rítmics; Txaikovski/ Gemini



Imatge 8 Etapa caps de nota- (pipeline) Txaikovski/ Audiveris

A la Imatge 7, es fa evident el fenomen de l'al·lucinació rítmica i formal per part de Gemini. Incapaç de desxifrar la densitat dels acords i la compressió de l'espai gràfic, el sistema multimodal interromp la transcripció literal i s'apropia d'un patró d'acompanyament repetitiu i homogeni extret dels seus propis vectors d'entrenament, inventant-se una solució musical que no existeix en l'hològraf. Aquesta reacció gràfica és una conseqüència directa de l'arquitectura del model: tal com analitzen Torras et al. (2021) en l'aplicació de models de llenguatge (Language Models) a l'OMR, quan la informació visual de la imatge és greument ambigua o atapeïda, el sistema prioritza la probabilitat del context musical i la versemblança sintàctica apresada sobre de la realitat diplomàtica de la tinta original¹⁶.

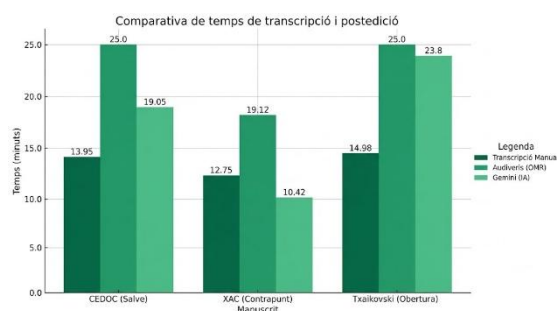
La fallada estructural esdevé encara més clara en examinar el comportament de l'OMR clàssic. La Imatge 8 ens mostra una captura del procés intern del pipeline d'Audiveris en un estadi intermedi del reconeixement. Resulta de gran interès per a la recerca comprovar com, abans de passar a la fase de transcripció de pliques i lligadures, el programa té perfectament senyalats i localitzats una gran quantitat de caps de nota. No obstant això, un cop el sistema avança cap a la següent capa seqüencial, pràcticament la totalitat d'aquests encerts desapareix de la pantalla, lliurant un fitxer final buit. (Annex: Figura 9)

Aquest comportament visualitzat a la Imatge 8 ofereix un exemple especialment il·lustratiu de la validació destructiva pròpia dels sistemes modulars definits per Bainbridge i Bell (1997)³. Com que Audiveris opera sota regles geomètriques tancades per a notació impresa moderna (CWMN), cada element reconegut ha de ser validat per la presència de la seva plica o unió corresponent; si l'amuntegament de l'escriptura de Txaikovski impedeix establir geomètricament aquesta connexió, l'algorisme assumeix que el cap de nota és soroll o un error i l'esborra de manera irreversible. Es confirma així l'efecte cascada teòric: un error de segmentació estructural en les capes inicials del pipeline destrueix els encerts parcials previs i provoca el col·lapse total de la transcripció automàtica.

4.2. Síntesi dels resultats

Els resultats suggereixen limitacions significatives d'Audiveris en el context dels manuscrits històrics analitzats ja que la seva rigidesa impedeix el post edició àgil. Per contra, els models multimodals són eines de gran potencial, però actualment presenten un risc semàntic: la seva capacitat per semblar "encertats" fins i tot quan al·lucinen els fa perillosos. El musicòleg, avui dia, no pot delegar la transcripció a la màquina; ha de passar a ser un auditor d'al·lucinacions.

Taula 1 Comparativa de temps, resum final



5. Conclusions generals

L'anàlisi del procés de transcripció i postedició m'ha permès avaluar el rendiment de diferents eines de reconeixement musical en manuscrits històrics. Els resultats mostren que la utilitat d'aquestes tecnologies depèn en gran mesura de les característiques materials del document, especialment del seu estat de conservació, de la qualitat de l'escriptura i de la densitat de la notació musical.

Les dades obtinguts amb Audiveris indiquen que els sistemes basats en arquitectures seqüencials continuen presentant dificultats importants davant manuscrits històrics. La degradació del suport, les irregularitats cal·ligràfiques i la complexitat gràfica afecten el procés de segmentació i provoquen errors que sovint requereixen una intervenció manual extensa. En els casos analitzats, el temps necessari per corregir aquests resultats redueix notablement els beneficis potencials de l'automatització.

Els models multimodals analitzats han mostrat una major capacitat d'adaptació davant fonts complexes. En particular, han obtingut bons resultats en el reconeixement de text i han reduït el temps de treball en aquells casos amb una qualitat visual favorable. Malgrat tot, les dades recollides també evidencien limitacions importants. Davant partitures amb una elevada densitat gràfica o estructures musicals complexes, aquests sistemes poden generar informació incorrecta que requereix una revisió detallada per part del musicòleg.

L'experiment em confirma que molts dels problemes descrits a la literatura especialitzada continuen presents en l'ús quotidià d'aquestes eines. La separació entre els elements musicals i les alteracions pròpies del document, així com la interpretació correcta de la notació manuscrita, continuen representant alguns dels principals obstacles per al desenvolupament de sistemes de reconeixement més fiables.

La informació recol·leix que les eines actuals poden actuar com a suport en determinades fases del procés de transcripció, però encara no permeten prescindir de la supervisió especialitzada. Per tant, la tasca del musicòleg evoluciona progressivament des de la transcripció manual cap a la revisió, verificació i correcció dels resultats generats automàticament. Per aquest motiu, el coneixement de la notació històrica, la capacitat analítica i el criteri musicològic continuen essent elements essencials per garantir la qualitat i la fiabilitat de les transcripcions.

BIBLIOGRAFIA

- Audiveris. "Audiveris Handbook: OMR Engine for CWMN." Consultat el 18 de juny de 2026.
<https://audiveris.github.io/audiveris/>.
- Bainbridge, David, i Timothy C. Bell. "The Challenge of Optical Music Recognition." *Computers and the Humanities* 31, núm. 2 (1997): 95-136.
- Bainbridge, David, i Timothy C. Bell. "The Problem of Staff Line Removal." *International Journal on Document Analysis and Recognition (IJDAR)* 4 (2001): 43-45.
- Baró, Arnau, et al. "From Optical Music Recognition to Handwritten Music Recognition: A baseline." *Pattern Recognition Letters* 123 (2019).
- Baró Mas, Arnau. *Reading Music Systems: From Deep Optical Music Recognition to Contextual Methods*. Tesi doctoral, Universitat Autònoma de Barcelona, 2022.
- Bellini, Pierfrancesco, Ivan Bruno, i Paolo Nesi. "Optical Music Sheet Segmentation." *A Proceedings of the First International Conference on WEB Delivering of Music*. Florència: IEEE, 2001.
- Blostein, Dorothea, i Henry S. Baird. "A Critical Study of OMR." *A Structured Document Image Analysis*. Berlín: Springer, 1992.
- Calvo-Zaragoza, Jorge, Jan Hajič Jr., i Alexander Pacha. "Understanding Optical Music Recognition." *Computing Research Repository (arXiv)*, 2020.
- Calvo-Zaragoza, Jorge, et al. "Can multimodal large language models read music score images?" *A Proceedings of the 6th International Workshop on Reading Music Systems*. Alacant: WoRMS, 2024.
- Calvo-Zaragoza, Jorge, i Cuihong Wen. "A Sequence-to-Sequence Framework Based on Transformer with Masked Language Model for Optical Music Recognition." *IEEE Access* 10 (2022).
- Castellanos, Francisco J., et al. "Selectional Auto-Encoders for OMR of handwritten manuscripts." *Expert Systems* 35, núm. 4. Amsterdam: Elsevier, 2018.
- Choudhury, Sayeed, et al. "Optical music recognition system within a large-scale digitization project." *A Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*. Plymouth, 2000.
- Fornés, Alicia, et al. "CVC-MUSCIMA: a ground truth of handwritten music score images." *International Journal on Document Analysis and Recognition (IJDAR)* 15, núm. 3. Berlín: Springer, 2012.

- Google. "Gemini: Multimodal Large Language Model." Versió 1.5 Pro. Model d'IA generativa. Consultat el 18 de juny de 2026.
- Padilla, Victor, et al. "Improving OMR for Digital Music Libraries with Multiple Recognizers." A Proceedings of the ACM/IEEE Joint Conference on Digital Libraries. Londres: ACM, 2014.
- Pruslin, Nicholas J. Automatic Recognition of Sheet Music. Cambridge: MIT, 1966.
- Rebelo, Ana, et al. "Optical recognition of music symbols: A comparative study." International Journal on Document Analysis and Recognition (IJDAR) 13, núm. 1. Berlín: Springer, 2010.
- Rebelo, Ana, et al. "OMR: A comprehensive survey." Expert Systems with Applications 39, núm. 5. Amsterdam: Elsevier, 2012.
- Ríos-Vila, Antonio, Francisco J. Fuentes-Martínez, i Jorge Calvo-Zaragoza. "Towards Sheet Music Information Retrieval: A Unified Approach Using Multitask Transformers." A Proceedings of the 6th International Workshop on Reading Music Systems. Alacant: WoRMS, 2024.
- Torras Coloma, Pau. Integration of Language Models in Sequence to Sequence Optical Music Recognition Systems. Treball de Fi de Grau. Bellaterra: Universitat Autònoma de Barcelona, 2021.
- Torras, Pau, Sanket Biswas, i Alicia Fornés. "On Designing a Representation for the Evaluation of Optical Music Recognition Systems." A Proceedings of the 6th International Workshop on Reading Music Systems. Alacant: WoRMS, 2024.
- Torras, Pau, et al. "On the Integration of Language Models into Sequence to Sequence Architectures for Handwritten Music Recognition." Centre de Visió per Computador - Universitat Autònoma de Barcelona, 2021.
- Wen, J. "A robust symbol classification method for optical music recognition." A Proceedings of the ICPR. Estocolm: IEEE, 2014.
- Yang, B., i M. Teng. "Unsupervised learning for music symbol segmentation." Journal of New Music Research. Londres: Taylor & Francis, 2015.

Annex:

Trancipcions

- Imatge 9: (CEDOC Transcripció Manual a Musescore)

CEDOC_CMM_1.5.1_0186.073

Transcripció manual

Compositor / arreglista

Violín 1

Violín 2

Viola

Oboe

Flauta

Corneta en Sib 1

Corneta en Sib 2

Fagot

Trombón

Soprano 1

Mezzo-soprano

Contralto 1

Tenor 1

Bajo 1

Soprano 2

Contralto 2

Tenor 2

Bajo 2

Violonchelo

Contrabajo

- Imatge 10: (CEDOC Transcripció Gemini)

Violines

Violas

Flautas

Oboes

Clarinetes en Sib

Trompas en Sib

Fagotes

Trombon

Tiple 1º (Coro 1)

Tiple 2º (Coro 1)

Contralto 1º (Coro 1)

Tenor 1º (Coro 1)

Bajo 1º (Coro 1)

Tiple 3º (Coro 2)

Contralto 2º (Coro 2)

Tenor 2º (Coro 2)

Bajo 2º (Coro 2)

Violoncelo

Bajo Continuo

tris

bene tumtum ven

- Imatge 11: (CEDOC Transcripció Audiveris)

Data

1

2

3

4

2

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

- Imatge 12: XAC_ACUR (Transcripció manual a Musescore)

OMRXAC_ACUR_TagFAu128_014

Transcripció Manual

Compositor / arreglista

The image displays a musical score for a piece titled "OMRXAC_ACUR_TagFAu128_014". The score is written for a large ensemble, including Violín 1, Violín 2, Flauta 1, Flauta 2, Clarinete en Do 1, Clarinete en Do 2, Fagot, Corneta en La 1, Corneta en La 2, Trompa en Re, Trombón 2, and Bajo. The key signature is one sharp (F#) and the time signature is 2/4. The Violín 1 and Violín 2 parts feature a rhythmic melody of eighth notes. Flauta 1 and Flauta 2 have more complex melodic lines, with Flauta 2 including a trill marked with an "8". The Clarinete en Do 1 and Clarinete en Do 2 parts are mostly rests. The Fagot part has a melodic line in the lower register. The Corneta en La 1 and Corneta en La 2 parts are mostly rests. The Trompa en Re part is mostly rests. The Trombón 2 part is mostly rests. The Bajo part has a rhythmic line of eighth notes.

- Imatge 13: XAC_ACUR (Transcripció Gemini)

XAC_ACUR_TagFAu128_014

Trancripció Gemini

V. 1
 V. 2
 Flau 1
 Flau 2
 Cla 1
 Cla 2
 Fagcon
 Cor 1
 Cor 2
 Tromb
 B.

-Imatge 14: XAC_ACUR (Transcripció Audiveris)

7

The image displays a musical score for a piece titled 'XAC_ACUR (Transcripció Audiveris)'. The score is written for a single melodic line, likely a voice or a flute, in treble clef with a key signature of one flat (B-flat) and a common time signature (C). The notation is spread across seven staves. The first staff begins with a treble clef, a key signature of one flat, and a common time signature. The subsequent staves continue the melodic line, featuring various note values including quarter notes, eighth notes, and sixteenth notes, as well as rests. The notation is clean and professional, typical of a musical score generated by software like Audiveris.

-Imatge 16: Txaikovski (Transcripció Gemini)

Txaikovski

Transcripció Gemini

The musical score is arranged in 14 staves, each representing a different instrument or section of the orchestra. The notation is as follows:

- Piccolo:** Treble clef, 2/4 time signature, six measures of whole rests.
- Flauto 1mo:** Treble clef, 2/4 time signature, six measures of whole rests.
- Flauto 2do:** Treble clef, 2/4 time signature, six measures of whole rests.
- Oboe 1mo:** Treble clef, 2/4 time signature, six measures of whole rests.
- Oboe 2do:** Treble clef, 2/4 time signature, six measures of whole rests.
- Clarinetto 1mo in B:** Treble clef, 2/4 time signature, six measures of whole rests.
- Clarinetto 2do in B:** Treble clef, 2/4 time signature, six measures of whole rests.
- Fagotto 1mo:** Bass clef, 2/4 time signature, six measures of whole rests.
- Fagotto 2do:** Bass clef, 2/4 time signature, six measures of whole rests.
- Corno 1mo in F:** Treble clef, 2/4 time signature, six measures of whole rests.
- Corno 2do in F:** Treble clef, 2/4 time signature, six measures of whole rests.
- Triangolo:** Treble clef, 2/4 time signature, six measures of whole rests.
- Violini 1mi divisi:** Treble clef, 2/4 time signature. Measures 1-5 contain eighth notes with stems up and down. Measure 6 contains a sixteenth-note triplet. Measures 7-14 contain a continuous sixteenth-note pattern.
- Violini 2di divisi:** Treble clef, 2/4 time signature. Measures 1-5 contain eighth notes with stems up and down. Measure 6 contains a sixteenth-note triplet. Measures 7-14 contain a continuous sixteenth-note pattern.
- Viole divisi:** Bass clef, 2/4 time signature. Measures 1-5 contain eighth notes with stems up and down. Measure 6 contains a sixteenth-note triplet. Measures 7-14 contain a continuous sixteenth-note pattern.

- Imatge 17: Txaikovski (Transcripció Audiveris)

The image displays a musical score for Tchaikovsky's 'The Nutcracker', specifically the 'Snowflake Dance' (Act II). The score is presented in a multi-staff format, with each instrument having its own staff. The key signature is one flat (B-flat major/D minor) and the time signature is 3/4. The score shows the beginning of the piece, with measures 1 through 7. The instruments listed on the left include Violins I, Violins II, Violas, Cellos, Double Basses, Flutes, Oboes, Clarinets, Bassoons, Horns, and Trombones. The notation includes various musical symbols such as notes, rests, and dynamic markings.