**DOI: https://doi.org/10.15276/ict.01.2024.01**
**UDC 004.9; 004.8**

# Optical music recognition: challenges and future directions

**Khrystyna O. Melnychuk**[1)]
Bachelor, Department of Artificial Intelligence Systems
ORCID: https://orcid.org/0009-0006-4549-649; khrystyna.melnychuk.kn.2021@lpnu.ua
**Solomiia E. Liaskovska**[2)]
Ph.D, Department of Artificial Intelligence Systems
ORCID: https://orcid.org/0000-0002-0822-0951; solomiya.y.lyaskovska@lpnu.ua. Scopus Author ID: 57204561106
[1)] Lviv Polytechnic National University, 12,Str.St.Bandera. Lviv, 79000, Ukraine
[2)] Kingston University, London, Friars Avenue London SW15 3DW United Kingdom

## ABSTRACT

Optical music recognition (OMR) as a branch of computer vision has deep roots dating back to the sixties, but has been actively developing only in the last few decades. The main goal of OMR is to automate the process of converting a musical score into a digital format. Despite the advances in image processing, there are still some difficulties, caused by the field's specifics, described in the work. Defining the concept of OMR is problematic, as there are numerous definitions ranging from task-specific to general. A comprehensive definition is proposed in the work, which allows more clearly outlining the semantic boundaries of the studied concept. The peculiarities of the contextuality of musical notation in comparison with text systems of writing are discussed. The range of sizes of musical symbols as a separate feature of notation is mentioned. The importance of the impact of text marks on the recognition difficulty is noted. The importance of visual differences between musical symbols and their influence on recognition accuracy is explained. The difficulty of recognizing sheets with several voices within one staff and with multiple staves is highlighted. The classification of sheet music types depending on the presence of several voices and staves is reviewed. The impact of score format on recognition difficulty is discussed. The impact of musical notation types on the OMR process is noted. The work considers the general structure of the OMR system, proposed by D. Bainbridge and T. Bell, and the main stages of the musical notation recognition process, according to the structure. The «bottom-up» structure of the OMR system, according to A. Pacha, is considered. The difficulties of OMR systems evaluation are discussed, examples from the literature are provided. Currently available software for OMR, its capabilities and limitations are also reviewed. The results of testing one of them, the Audiveris module built into the MuseScore platform for converting sheet music into digital format, on specific musical compositions are described and summarized.

**Keywords:** Optical music recognition; optical music recognition; computer vision; musical notation; music complexity; image processing; musical scores; sheet music; optical music recognition evaluation

**Relevance.** Optical Music Recognition (OMR) technologies are becoming increasingly significant in the world of digital music, especially in the context of growing demand for process automation and digitalization of cultural heritage. Given the diversity and complexity of musical notation, OMR plays a key role in the preservation, analysis, and dissemination of musical works, enabling efficient use and processing of scores in modern contexts. Despite significant advancements in computer vision technologies, OMR remains one of the most challenging fields due to the vast number of variations in musical notation, requiring continuous refinement of algorithms and models to achieve high recognition accuracy.

**The aim** of the work is to analyze existing approaches to musical notation recognition and to identify the key challenges related to achieving high accuracy and reliability in recognition across various musical score formats.

Optical music recognition as a branch of computer vision originated in the 60s of the last century, but has been actively developing during the last decades [1]. Its primary goal is to automate the process of converting printed and handwritten musical scores into digital format. Its primary goal is to automate the process of converting printed and handwritten musical scores into digital format. Despite considerable advancements in image processing, many aspects of this task remain challenging due to the complexity and peculiarity of musical notation. These challenges create difficulties for developers and researchers, complicating the accurate and reliable recognition of musical scores [2, 3].

Presently, defining the boundaries of OMR is not an easy task. Researchers in the field generally agree on the intuitive concept of "a computer capable of reading musical notation".

However, most of the literature focuses on solving specific problems, and as a result, the task formulation and definition boundaries in such studies are often tailored to correspond to the proposed solution or highlight the relevance of the research for a specific target audience (e.g., the field of computer vision or music information retrieval). Therefore, OMR definitions are often either too narrow, tailored to a specific task (e.g., "converting images of musical scores into MIDI files"), or overly general, such as "optical character recognition for music" (or "OCR for music"), which do not provide a clear understanding of OMR. To avoid ambiguity, J. Calvo-Zaragoza et al. propose a more comprehensive definition that would at the same time clearly delineate the essence of the field: "Optical Music Recognition is a field of research that investigates how to computationally read music notation in documents", illustrating how the various definitions of OMR in the literature associated with this definition and how it covers them [4].

The following is a mathematical model that describes the main stages of converting a music score image into a digital format.

Input Image ( $I$ ): This is an image of a musical score containing musical symbols.

$$I : R^{h \times w} \Rightarrow R \ , \tag{1}$$

where $h$ – image height; $w$ – image width; $R$ – is a matrix of pixel intensities.

Detection of musical symbols ( $D_N$ ): This stage involves detecting all the note symbols in an image using techniques such as pattern matching. A model is built to detect note positions based on the patterns.

$$D_N(I) = \{s_1, s_2, ...s_n\} \ , \tag{2}$$

where $s_i \in S$ – the set of all detected symbols, $S$ – the set of all possible musical symbols.

Recognition of each note: After each note is detected, its characteristics are analyzed.

Note position recognition ( $P(s_i)$): The position of the note relative to the note lines. For each detected note, its vertical position is determined, which corresponds to the pitch of the note.

$$P(s_i) = note\_position \in N \ , \tag{3}$$

where $P(s_i)$ – the specific position of each symbol $s_i$ on the staff.

Note duration recognition ( $D(s_i)$): The duration of a note is determined based on the shape of the symbol. Possible durations include quarter notes, eighth notes, etc.

$$D(s_i) = note\_duration \in \{1,1/2,1/4,1/8...\} \ , \tag{4}$$

Output: Each musical symbol generates output data after position and duration recognition. The output is presented as a pair of values: recognized position and duration.

$$R = \{(P(s_1), D(s_1)), (P(s_2), D(s_2)), ...(P(s_n), D(s_n))\} \ , \tag{5}$$

As mentioned earlier, Optical Music Recognition is often considered quite similar to text recognition; however, this field has certain features that are distinct enough from related areas to be recognized as a separate one [5]. One of the key differences is that musical notation is a contextual writing system: the meaning of symbols or elements in such a system depends not only on their appearance but also on their placement and interaction with other elements [4]. Fig. 1 provides an example of the importance of context in music notation recognition. In contrast, text uses a fixed set of characters, so whether the letter «a» appears at the beginning, middle, or end of a word, it remains the same character for a character recognition system, meanwhile the placement of the note impacts the result pitch.

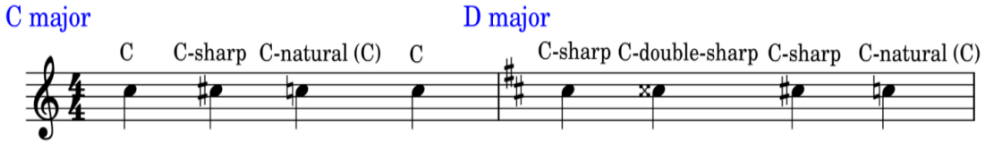


*Fig. 1:* **The importance of context in music notation recognition: the same note can have different pitches depending on the key signature and alterations**

The peculiarity of musical notation in terms of its recognition lies not only in the variety of graphical symbols but also in the range of their sizes: while some writing systems are relatively complex, such as Chinese characters, the size of the character (glyph) is fixed. In music, however, we can have both small notations (e.g., a dot that increases note duration or a dot indicating staccato) and large ones (for example, a bracket connecting four separate instrumental parts, extending vertically across half a page) [4]. Furthermore, scores usually contain text annotations [5]: about dynamics, tempo, or more specific instructions, such as the mark «Swing», which indicates that the piece should be played with a particular rhythm. Additionally, the appearance of most glyphs in text is quite varied, whereas in music, many forms are graphically similar, and even minor differences convey important information. For instance, there is a dot that extends the duration of the basic note by half of its original value. So if the note is misrecognized, the rest of the musical piece will also be rhythmically incorrect. Special attention should be paid to articulation marks [3, 5]. Other examples of the relative similarity of musical symbols compared to text glyphs include indications of note length, the use of beams, ties, note groups (e.g., triplets), other special marks, and the variation of some symbols' shapes (e.g., slurs and glissando), as shown in Fig. 2.

The significance of visually small differences in musical notation (a) a dot extends a quarter note by half its duration (thus, after the first note there are two eighth rests, after the second note there is one, as the duration of the other is included in the note); (b) variety of articulation marks (staccato, marcato, accent, tenuto); (c) eighth notes connected by a horizontal beam, while with two notes (d) the beam can take any angle; (e) a tie connecting two heads of the notes of the same pitch; (f) a triplet – a group of three eighth notes, which lasts as long as two eighth notes (g); (h) a grace note (ornamentation); (i) the dependence of the length and angle of a wavy line for indicating glissando on the range of notes it covers. Another challenge of OMR is the presence of multiple voices within a single staff and/or the presence of multiple staves in a musical score. It is important to note that when two voices are present on one staff, an additional complexity arises because intervals and chords with different note durations may appear, as shown in Fig. 3.

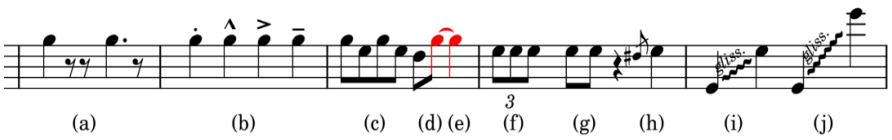


*Fig. 2.* **The significance of visually small differences in musical notation**

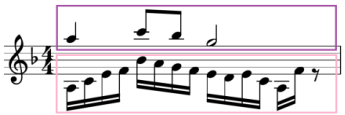

*Fig. 3.* **Example of a staff with two voices (notes of each voice are highlighted with rectangles in different colors; highlighting is ours) [4]**

Byrd and Simonsen define musical scores with several voices within one staff as the most complicated type of musical compositions, referring to it as “pianoform” [3]. In turn, J. Calvo-Zaragoza et al. extended this classification by dividing musical scores into four types: monophonic – scores have only one staff, and all notes are single; homophonic – scores have only one staff, but chords may occur; polyphonic – when there are multiple voices within a single staff; and pianoform – scores with multiple staves [4]. It follows that the more voices and staves present, the more challenging it becomes to recognize the score. A good illustration of this is the work by A. Ng and A. Khan. The program developed by the researchers achieved the best accuracy after recognizing the music sheet for “Jingle Bells”, which is a simple score with a single staff. At the same time, it makes more errors when processing more complex works, such as Bach’s “Allemanda”. Additionally, with seemingly simple but pitch-varied sheet music for the song “Twinkle Twinkle Little Star”, the system finds it harder to analyze and correctly recognize musical notation, leading to mistakes in note and pitch recognition [2].

The musical scores format occupies a special place among the peculiarities of OMR field: digital sheets are more suitable to its tasks compared to those that were printed on paper and then photographed/scanned. Moreover, in addition to the printed form, sheet music may be handwritten. Then, as with handwritten text, it is important to take into account that each musician has his or her own handwriting style and the same elements may look different, especially when it comes to recognizing ancient handwritten scores. The type of notation also affects the complexity of solving the problem of OMR. Fig. 4 illustrates the variety of musical notations.

The main body of work in the field of music score recognition focuses on specific tasks and their solutions. Thus, it can be said that the essence of the field is not to solve a single task, but rather a multitude of subtasks [4]. Nevertheless, the general structure for building an OMR system remains the same across all works. Fig. 5 demonstrates the general structure of a music notation recognition system, proposed by D. Bainbridge and T. Bell (with the labels on the block diagram translated by us), includes: detection of staff lines in the image, identification of the location of musical symbols, classification of musical symbols (which occurs in two stages: detecting the primitives that make up the musical symbols, and combining them into musical objects), and restoration of the musical semantics of the score. After the identification of staff lines and the determination of the location of musical symbols, additional image processing is possible [5].

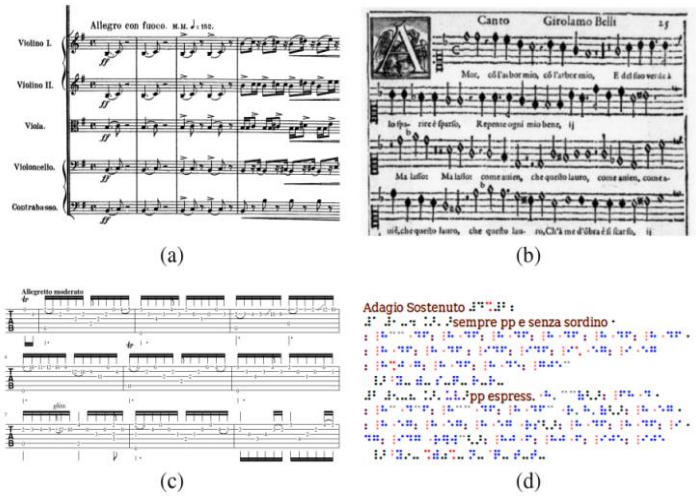


*Fig. 4*. **Variety of musical notations:**
**a – modern musical notation; b – mensural notation;**
**c – guitar tablature; d – Braille [4]**

To understand the structure of an OMR system “from the bottom up”, A. Pacha proposes five levels of questions [6].

1) Can the computational device distinguish notes from arbitrary information?

2) Can it understand the structure of a musical score (staff lines, system – a set of parts to be played simultaneously) and differentiate basic musical symbols from each other and from the background?

3) Can it detect and locate musical symbols (notes, rests, ornaments, accidentals, barlines, articulations, etc.) within scores?

4) Can it understand the relationships between the objects in the score (for example, the connection between a note and the staff lines, an accidental to the left of the note it applies to, etc.)?

5) Can it fully understand the syntax and semantics of musical scores (determine the actual note from its relative position, shape, and preceding symbols such as key signatures or accidentals)?

It is also challenging to evaluate the performance of OMR systems: the issue arises of determining the weight of each error (for example, an incorrectly detected pitch of a single note may only slightly affect the melody, while an incorrect note duration could throw the rest of the piece off completely). Currently, there is no universal tool that is suitable for every proposed development overall [4]. This is due to both the variety of used approaches and methods and the fact that studies typically have limited datasets. For example, A. Nyati uses only digital format scores, monophonic (according to J. Calvo-Zaragoza et al. [4]), which contain only three types of note durations and a limited number of time signatures, while many musical symbols are not considered at all, and the main method used is the Otsu method [1]. In A. Pacha’s work, a dataset consisting of notes and non-notes (tables and/or text) is used for the first experiment (image classification into scores and non-scores), and the Handwriting Online Musical Symbols (HOMUS) dataset is used for the second, which consists of images of individual handwritten notes and other musical signs. To test the system’s performance, musicians were given the same task separately to compare the accuracy of symbol classification by neural networks and humans. [6] Y. Li et al. have a dataset that is half digital half printed and photographed images of scores, including polyphonic scores in pianoform format (according to J. Calvo-Zaragoza et al. [4]). Their system includes a

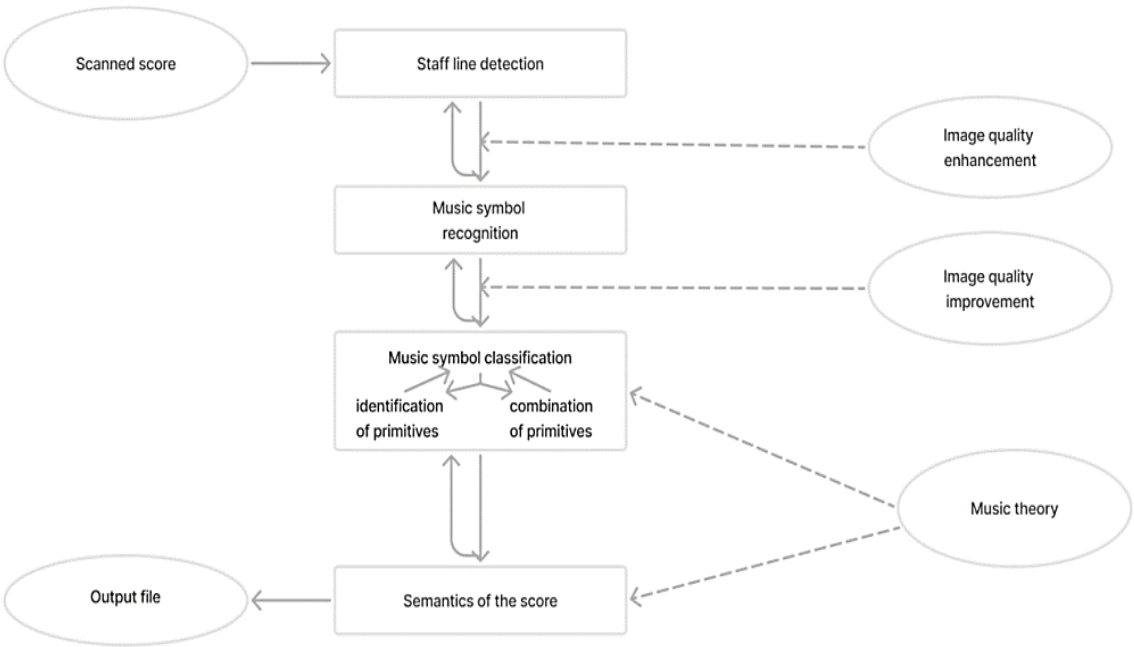


*Fig. 5.* **The general structure of a music notation recognition system, proposed by D. Bainbridge and T. Bell**

separate module for determining staff lines and score type based on the number of staves, while the main part of the system is transformer-based [7]. J. Calvo-Zaragoza and D. Rizo, in their research, use only printed, monophonic scores (according to J. Calvo-Zaragoza et al. [4]), but the goal of their work is to find a way to apply deep learning systems to solve the OMR task comprehensively – end-to-end, without the need to break the problem into smaller stages [8]. Unlike most studies that use modern five-line notation scores, M. Alfaro-Contreras and J. J. Valero-Mas work with mensural notation scores, using two datasets: handwritten notes and scanned printed ones [9].

One of the main challenges in Optical Music Recognition is the difficulty of recognising musical notation. Traditionally, methods such as the **Otsu method** [1] have been used to binarize images. However, modern approaches more often use neural networks. **Convolutional neural networks (CNNs)** allow for efficient feature extraction from images and improve recognition accuracy [8]. **Residual neural networks (ResNets)** solve the problems of performance degradation in deep learning, which results in better accuracy [6]. **Transformers**, due to their self-learning mechanism, provide efficient processing of context and relationships, which makes them promising for music notation recognition [7].

Nowadays, there is a number of software available for average users that convert sheet music images into digital music formats, including MusicXML and midi: **SmartScore**, **ScanScore**, **OMeR**, **PhotoScore**, **Arusprix** and others. However, all of these programs are paid, and they do not guarantee accurate conversion, especially when considering the current state of the field.

Currently, the only free tool for recognising music notation is **Audiveris**. MuseScore, a well-known platform among musicians that has its own music notation editor, allows you to convert score files from .pdf to .mscz (the format used as the main one for storing scores in the editor) using this module. The webpage with the converter states that it is an «experimental service» and, in case of unsuccessful conversion, notes that the file may not be compatible, which "could be due to various reasons such as the quality of the PDF, the complexity or various other reasons" [10]. We chose and tried to convert on this platform several pieces by the known modern American ragtime pianist Tom Brier, whose music is notable for its complexity, unique style, musical experimentation, and, despite the tragic fate of the composer, will have a special place in the circles of sincere fans of the genre for a long time ahead [11]. All of the chosen scores were in printed format. During our brief study, we discovered that in some cases the Audiveris module recognized the notes absolutely accurately (usually when it was a relatively monotonous notation, for example, when the notes were eighths and sixteenths and there were minimal alterations), even the articulation and dynamics markings were correct. But in other cases it was the other way round: the module either made mistakes in the alteration notation, did not recognize chords completely correctly with a large number of notes, or failed in the rhythm (especially when there were several notes of different lengths, pauses, and dotted notes). In addition, the software failed to recognize some notation, such as octave transposition (fully) and volta (partially), as well as inserted false legato slurs everywhere and could add additional voices within the same staff when in fact there was only one voice. Overall, the tool made frequent errors, so, in our opinion, it needs further improvement and is not yet effective enough for the recognition and digitalisation of relatively complex music scores.

Example of sheet music to digital format conversion by the Audiveris module on the MuseScore platform we can see on Fig. 6 where 1 – the first bars of the second part of Brier's "Just Peachy" (a – original, b – conversion result), the notes were recognized absolutely accurately, the only mistake was in a wrong legato slur (red highlight); 2 – the first bars of the third part of Brier's "Razor Blades" (a – original; b – conversion result), there are errors in recognizing alternative notation (red highlight); 3 – the first bar of the second part of "Razor Blades" by T. Brier (a – original, b and c – results of converting this bar in two places in the piece), where two voices are placed in the bottom staff, written with four groups of triplets for each of them; this fragment turned out to be the most difficult for recognition, so the module failed with the task in both cases: (b) and (c).

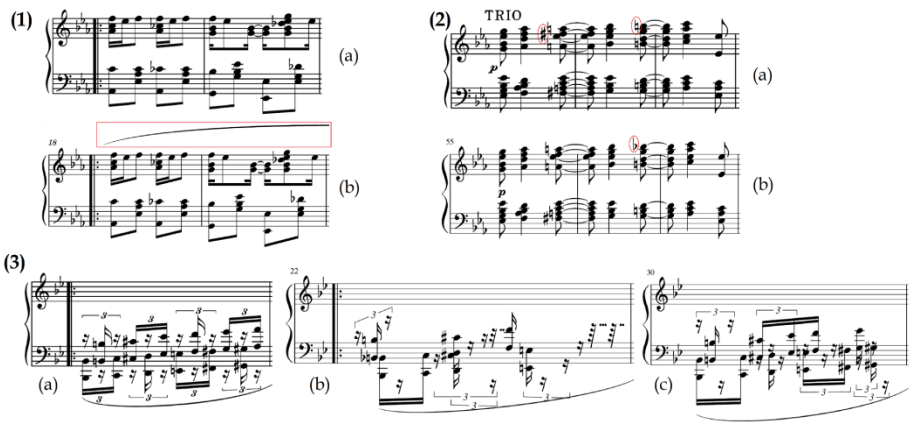


*Fig. 6.* **Example of sheet music to digital format conversion by the Audiveris module on the MuseScore platform**

**Summary.** Despite the fact that music notation recognition remains a relatively young field of research that faces numerous challenges and has a number of peculiar features, it has been actively developing and achieving noticeable success. Modern technologies, including deep learning, significantly contribute towards OMR development. Innovations in computer vision and pattern recognition help to reduce errors and improve the accuracy of converting sheet music into digital formats. Although there are still many challenges ahead, the pace of development and novelties in this field is promising for the future. Thanks to ongoing advances in technology and research efforts, we can expect to see significant advancements in the development of efficient and reliable OMR systems, bringing new opportunities for musicians and researchers in the future.

## REFERENCES

1. Nyati A. "*cadenCV: An Optical Music Recognition System with Audible Playback"*. *Github.* 2024. – Available from: https://github.com/afikanyati/cadenCV.

2. Ng A., Khan A. "Automated Instructional Aid for Reading Sheet Music". 2012.

3. Byrd D., Simonsen J. G. "Towards a standard testbed for optical music recognition: Definitions, metrics, and page images". *Journal of New Music Research*. 2015; 44 (3): 169–195. DOI: https://doi.org/10.1080/09298215.2015.1045424.

4. Calvo-Zaragoza J., Hajic Jr J., Pacha A. "Understanding optical music recognition". *ACM Computing Surveys (CSUR)*. 2020; 53 (4): 1–35.

5. Bainbridge D., Bell, T. "The challenge of optical music recognition". *Computers and the Humanities*, 2001; 35: 95–121.

6. Pacha A., Eidenberger H. "Towards self-learning optical music recognition". *In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2017. p. 795–800.

7. Li Y., Liu H., Jin Q., Cai M., Li P. "TrOMR: Transformer-Based Polyphonic Optical Music Recognition". *In: IEEE International Conference on Acoustics, Speech and Signal Processing.* 2023. p. 1-5. DOI: https://doi.org/10.1016/j.eswa.2024.123664.

8. Calvo-Zaragoza J., Rizo D. End-to-end neural optical music recognition of monophonic scores. *Applied Sciences*, 2018; 8(4): 606.

9. Alfaro-Contreras M., Valero-Mas J. J. "Exploiting the two-dimensional nature of agnostic music notation for neural optical music recognition". *Applied Sciences*, 2021;11 (8): 3621.

10. "The world's largest free sheet music catalog and Community". *MuseScore*. 2024. – Available from: https://musescore.com.

11. “Tom Brier: The Story and Tragic Ending of a Piano Prodigy”. *DanHon Music*. 2024. – Available from: https://danhon.substack.com/p/tom-brier-the-story-and-tragic-ending.

**DOI: https://doi.org/10.15276/ict.01.2024.01**
**УДК 004.9; 004.8**

# Оптичне розпізнавання нотного запису: виклики та перспективи

**Мельничук Христина Олегівна**[1)]
Бакалавр каф. Систем штучного інтелекту
ORCID: https://orcid.org/0009-0006-4549-6494; khrystyna.melnychuk.kn.2021@lpnu.ua
**Лясковська Соломія Євгенівна**[1), 2)]
Д-р техніч.наук, доцент каф. Систем штучного інтелекту
ORCID: https://orcid.org/0000-0002-0822-0951; solomiya.y.lyaskovska@lpnu.ua. Scopus Author ID: 57204561106
[1)] Національний університет «Львівська політехніка», вул. С. Бандери, 12. Львів, 79000, Україна
[2)] Kingston University, London, Friars Avenue London SW15 3DW Велика Британія

## АНОТАЦІЯ

Оптичне розпізнавання нотного запису (англ. Optical Music Recognition, OMR) як галузь комп'ютерного зору має глибоке коріння ще з шістдесятих років, але активно розвивається лише в останні кілька десятиліть. Основна мета OMR – автоматизувати процес перетворення музичної партитури у цифровий формат. Незважаючи на прогрес в обробці зображень, все ще існують певні труднощі, викликані специфікою галузі, описані в роботі. Визначення поняття OMR є проблематичним, оскільки існує безліч формулювань, починаючи від конкретизованих, що відповідають чітким завданням, і закінчуючи більш узагальненими. У роботі запропоновано всеосяжне визначення, що дозволяє чіткіше окреслити семантичні межі досліджуваного поняття. Обговорено особливості контекстуальности музичної нотації в порівнянні з текстовими системами письма. Обговорено діапазон розмірів музичних позначень як окрему особливість нотного запису. Зазначено про важливість впливу текстових позначень у партитурах на складність при розпізнаванні нот. Пояснено важливість візуальних відмінностей між музичними символами та їхній вплив на точність розпізнавання. Висвітлено складність розпізнавання музичних творів з кількома голосами в межах однієї партії та творів з кількома партіями. Розглянуто класифікацію типів музичних творів в залежності від наявности кількох голосів та партій. Обговорено вплив формату партитур на складність розпізнавання. Зазначено про вплив різних типів музичної нотації на процес OMR. Крім цього, у роботі розглянуто загальну структуру OMR-системи, запропоновану Д. Бейнбріджем та Т. Беллом, та основні етапи процесу розпізнавання нотного запису відповідно до цієї структури. Розглянуто структуру OMR-системи «знизу-вгору» за А. Пахою. Обговорено труднощі в оцінці працездатности OMR-систем, наведено приклади з літератури. Також оглянуто навне на сьогодні програмне забезпечення для OMR, його можливості й обмеження. Описано та узагальнено результати тестування одного з них, вбудованого в платформу MuseScore модуля Audiveris для перетворення нот у цифровий формат, на конкретних музичних творах.

**Ключові слова:** розпізнавання нотного запису; OMR; комп’ютерний зір; музична нотація; складність нотного запису; обробка зображень; ноти; партитури; оцінка OMR-систем