



The Challenge of Optical Music Recognition

DAVID BAINBRIDGE¹ and TIM BELL²

¹*Department of Computer Science, University of Waikato, Hamilton, New Zealand (E-mail: d.bainbridge@cs.waikato.ac.nz);* ²*Department of Computer Science, University of Canterbury, Christchurch, New Zealand (E-mail: t.bell@cosc.canterbury.ac.nz)*

Abstract. This article describes the challenges posed by optical music recognition – a topic in computer science that aims to convert scanned pages of music into an on-line format. First, the problem is described; then a generalised framework for software is presented that emphasises key stages that must be solved: staff line identification, musical object location, musical feature classification, and musical semantics. Next, significant research projects in the area are reviewed, showing how each fits the generalised framework. The article concludes by discussing perhaps the most open question in the field: how to compare the accuracy and success of rival systems, highlighting certain steps that help ease the task.

Key words: optical music recognition, musical data acquisition, document image analysis, pattern recognition

1. Introduction

Optical Music Recognition (OMR) – a computer system that can “read” printed music – has much promise: a clarinetist could scan a tune and have it transposed automatically; a soloist could have the computer play an accompaniment for rehearsal; a music editor could make corrections to an old edition using a music notation program; or a publisher could convert a piece to Braille with very little work. OMR has been the focus of international research for over three decades, and while numerous achievements have been made, there are still many challenges to be faced before it reaches its full potential.

OMR addresses the problem of *musical data acquisition*, the key impediment to many computer music applications. It is not, however, the only data entry method for music. The most common method for music data entry in current use combines synthesiser keyboard entry and computer keyboard entry. The musical keyboard is typically used to enter the notes by playing each voice in isolation, either in time with a metronome or using the computer keyboard to enter rhythmic information. The computer keyboard and mouse are then used to correct any mistakes and to add other notation such as lyrics, slurs, and dynamics. Music data entry in this form demands a level of skill from the keyboard player, and adding the remaining notation is time-consuming.

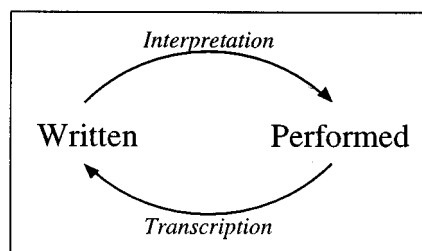


Figure 1. Forms of music representation.

Even if current commercial software could be refined to record information such as dynamics from the synthesiser keyboard (and the performer was able to play these accurately), the computer keyboard stage is still required, since there are many more features to printed music than the way it is finally played – for instance, clefs, key signatures, time signatures, titles, and lyrics. Also, deciding on line breaks, page breaks, even the spacing and grouping of notes within a bar, is a skilled task, and is generally seen as a craft rather than some process that can be defined by a set of rules and automated (Ross, 1970; Read, 1974; Heussenstamm, 1987).

OMR greatly simplifies the task of musical data acquisition; however we are not arguing that OMR should replace musical keyboard entry. In the particular circumstance where the music is already available in printed form, OMR can substantially accelerate the process of musical data acquisition. OMR, therefore, is an attractive supplement. Not only does the technique reduce the chance of human error made in the transcription, but it can also capture much of the “extra” information that the former method requires the user to laboriously add after the notes have been played.

The most likely scenario is one where OMR is used to process the majority of symbols on the page, followed by an editing stage using a standard music editor, where the musical and computer keyboards are used to correct mistakes and omissions. This has the added benefit of greatly reducing the musical keyboard skills required. Of course, OMR is not perfect, and the editing can be a significant part of the time taken for OMR-based data entry; however, OMR systems are steadily improving, and the number of music-based projects utilising OMR steadily increasing (Carter, 1992a). Given the vast body of printed music, OMR could radically reform computer applications in music.

To help clarify the limits in musical data acquisition by computer, let us consider the larger context of music representation. Figure 1 shows the relationship between written and played music. Written music is converted to performed music by *interpretation*, and performed music is converted to written music by *transcription*. Both are non-trivial operations. A performer may study a written work for months before finally presenting their interpretation. The performer will draw upon a range of information, such as knowledge about the composer, and the technical limits