IMUTUS
AN INTERACTIVE MUSIC TUITION SYSTEM

D.Fober, S.Letz, Y.Orlarey
Grame - Centre national de création musicale
{fober, letz, orlarey}@grame.fr

A. Askenfeld, K. Falkenberg Hansen, E. Schoonderwaldt
KTH - Royal Institute of Technology
{andersa, hansen, schoondw}@speech.kth.se

ABSTRACT

IMUTUS is a European project that aims at the development of an open platform for training students on non-MIDI musical instruments, as well as to acquire theoretical music knowledge. The project involves many components oriented towards a new approach of music learning. After a brief overview of the system, the performance evaluation module and the music score processing components are described to show how they enforce the pedagogical approach.

1. INTRODUCTION

Computers have been used in music education for over 20 years. Applications are covering fundamentals of music, instrument performance skills, analysis of music as well as composition skills. Pedagogical approaches use a range of techniques from Computer Assisted Instruction (CAI) to Intelligent Tutoring System (ITS) in conjunction with different instructional strategies [2].

The place of the IMUTUS project is crosswise to these categories: it is primary focused on performance skills but since it addresses beginners, it also includes fundamentals of music and games which are combined to practicing sessions to create a complete learning setting. Exercises include various instructional strategies[1] such as *programmed learning*, *drill and practice*, *monitoring* and *exploratory* strategies.

Although designed to be autonomous, the system is expected to be most successful when used as a supplement to traditional music lessons with a teacher. It also provides support for distance learning: the students will be provided with updated content and the ability to communicate with a specially mandated distance learning teacher group to obtain electronic guidance, feedback and assistance whenever they require it, through asynchronous Internet communication. Finally, the recorder is chosen as the target instrument because it is a traditional instrument widely taught in European schools, providing no MIDI support.

Existing works concerning performance skills are limited in number. The most accomplished system is probably the Piano Tutor [3] [4] designed in 1990. More recent works explore the visual feedback of student mistakes [1] [6] but are more oriented towards expressive performance than basic performance skills. A project named *Virtual Music Teacher* [5] has been conducted since IMUTUS start; results are essentially pointing the project limitations, notably in error detection. Some commercial applications claim to succeed in teaching instrument practice, including non-MIDI instruments: the more advanced and closest to the IMUTUS system are applications designed by the Musicalis company[2] but the results are disappointing. Thus, the IMUTUS object is fully justified in the context of existing research or applications: it aims at improving both the technologies to support acoustic instruments as well as those to evaluate the student performance and provide him with an efficient and constructive aural and visual feedback.

The next section will present an overview of the system, including global architecture and components. Section 3 presents the objective performance evaluation module. Sections 4 and 5 are dedicated to the score processing modules : we'll first present the score viewer since it is at the center of the feedback process; we'll finally show how a simplified score editor supports fundamentals of music learning as well as exploratory strategies. Although the project is currently still in progress and results are not available, we'll draw some conclusion and prospects.

2. OVERVIEW OF THE SYSTEM

Overview of the system will be presented according to the main tutoring directions i.e. performance skills and fundamentals of music. Additional components that are not

[1] we use the terminology as defined in [2]

[2] Musicalis: http://www.musicalis.net

directly related to pedagogical issues will only be mentioned without more details; it represents:

- Optical music recognition (OMR): mainly intended for teachers as part of the authoring tools.
- The distance learning platform: it is the lessons repository and also includes components dedicated to the communication among the system users (students - teachers).

The curriculum design is similar to [3]: it is based on skills attached to a lesson and granted to the user when he/she succeed in performing the lesson and prerequisites skills, also attached to a lesson but required to start this lesson.

The partners involved in the project are ILSP (Institute for Language and Speech Processing - Greece), EXODUS S.A. (Greece), Systema Informatics S.A. (Greece), DSI (Department of Systems and Informatics - Univ. of Florence - Italy), Music School of Fiesole (Italy), Grame, Centre National de Création Musicale (France) and KTH, Royal Institute of Technology (Sweden).

2.1. Performance skills

From performance point of view, the system may be viewed as the interaction of a user and a virtual teacher by the mean of a score viewer (figure 1). When the user is playing, the virtual teacher may point the next note to be played, provide automatic turn/scroll page capabilities and may play the metronome. When the user performance ends, the virtual teacher may evaluate the performance on request and shows the results on the score. It may also add visual helpers on top of the score and show the correct fingering. The score is at the heart of the user and virtual teacher communication and interaction: visual helpers or feedback from the virtual teacher are provided as annotated music scores; aural feedback may be requested by the user using the score as user interface. All these mechanisms are detailed in section 4.

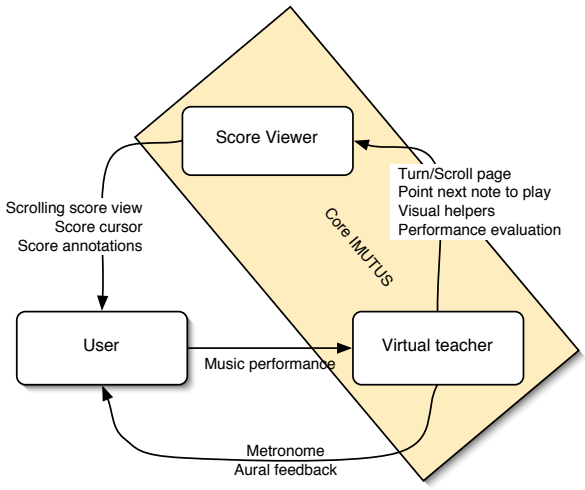

Figure 1. *Interactions during performance.*

The virtual teacher may be viewed as a combination of a virtual listener, a virtual player, a virtual analyst and virtual fingering viewer (figure 2):

- The virtual listener is in charge of the music performance low level analysis and recording; it includes a pitch tracker, a tempo tracker, a score follower, MIDI and audio recorders. It transforms raw audio data to higher level MIDI information including a tempo map. It generates synchronization information critical to the real-time services like the turn/scroll page strategies and cursor indication.
- The virtual analyst evaluates the performance based on extracted features from the audio signal and higher-level information produced by the virtual listener. The output is a grading of the performance and a list of mistakes with corresponding feedback. It may also provide positive feedback when a student succeed in performing a section identified as difficult.
- The virtual hands includes a VRML engine to display fingering.
- The virtual player includes audio and MIDI players. The audio player operates on a reference performance or a student past performance. It may play a whole performance or a specific section only. The MIDI player is used when no audio performance is available.

2.2. Fundamentals of music

Since the IMUTUS system addresses the needs of beginners, it includes theory lessons with associated exercises, mainly targeted towards ear training and music reading and writing. Some of these exercises are intended to make sure that essential music notions are well understood: they are designed as question like exercises (multiple-choice, true-false or fill-in questions). These kind of exercise may require the score viewer operations to display the question.

Exercises intended for ear training are music dictation or variations on the dictation principle. *Fill score* exercises are an example of such variation: the student listens to a melody; the corresponding score is presented to the student but some notes are missing; the student must complete the score. These kind of exercise requires a score editor to be available. This score editor operates in conjunction with the virtual analyst mentioned above.

Finally, the score editor is also provided as an exploratory tool: it is available to the student to freely write music and experiment the corresponding aural space.

3. OBJECTIVE PERFORMANCE EVALUATION

The IMUTUS system aims at developing the young recorder student's performance skills. The main target group is children between 9 and 14 years age, from beginners up to intermediate level. A basic idea is to make the practising sessions more efficient by providing structured feedback

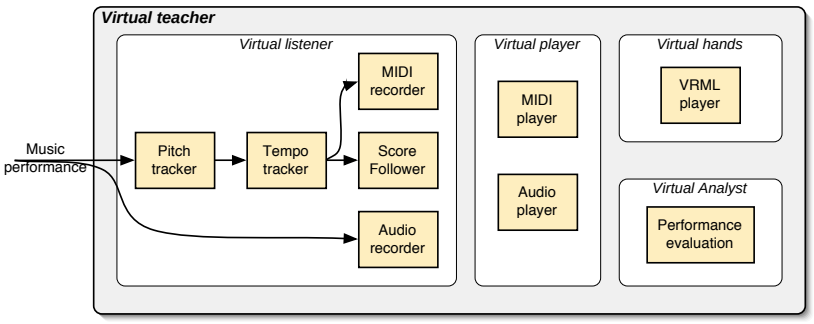

Figure 2. *The virtual teacher components.*

immediately after each performance, adapted to the student's current skill level. A central part of the system is the *virtual analyst* which makes an automatic evaluation of the performance, makes a prioritized list of all detected errors/mistakes, and gives specific feedback on the top-ranked errors/mistakes as output.

3.1. Basic performance errors

The virtual analyst is designed to recognise a set of nine *basic errors/mistakes*, which are typical of beginners' performances. The set of basic errors covers aspects of instrument control (e.g. *attack*, *airflow*) as well as musical performance (e.g. *rhythmic performance*), see table 1. Instrument control, such as control of air flow, fingering, and attack quality, takes an important role in the early stages of learning (first four terms). Higher-level musical performance aspects like intonation, phrasing, tempo and articulation (staccato legato) are ranked lower at these stages of learning. In between are the basic musical performance abilities; to perform the notated rhythm correctly, and play the correct notes (pitches).

The selection and ranking of the basic error categories is based on the expertise of a group of more than 40 experienced recorder teachers in three countries, collected via questionnaires and supplemented with interviews. More specific knowledge and criteria for error recognition were obtained from detailed analyses of recorded student performances, made by five recorder teachers. These recordings and analyses provided a valuable reference material for developing and testing the virtual analyst.

The analysis of the performance uses the score, and data from an audio-to-MIDI conversion module and a score matcher, as input. After identification of all basic errors/mistakes in the performance, the error list is sorted in order of importance by a prioritising scheme. Only a small number of errors, typically three, and their associated descriptions are displayed to the student via the score viewer. These three are considered as the most important errors on the students current level of performing. The restriction in the feedback is made in order not to overwhelm the student with information and to make the student concentrate on the most important aspects at the moment.

3.2. Support in error detection

The content author has the possibility to indicate specific difficulties in the piece (e.g. a complex fingering transition, or a difficult rhythm) by adding *score annotations*. These annotations, which are hidden for the student, can help the virtual analyst to find and interpret certain basic errors in the performance. By way of score annotations, the knowledge of the content author (typically an experienced recorder teacher) is included in the system and used to guide the virtual analyst. An additional advantage is that they offer IMUTUS a possibility to provide well-founded positive feedback. If a certain passage, indicated as difficult in the score annotations, is played correctly, praise can be given and the student is stimulated to go on.

Another way to tailor the virtual analyst to specific needs of performance evaluation is by use of *projections*. In a projection, the aim is to practise a specific element in the performance, e.g. a pitch sequence, or a rhythmic figure. The evaluation is limited to one (or a few) type of basic errors, and the feedback will be more concentrated, thereby increasing the efficiency when practising a particular performance aspect.

3.3. Feedback display

The output of the virtual analyst that is made visible to the student is a short list with the three top-ranked basic errors, including type of basic error/mistake, position in the score, and a short description. The score viewer displays the information on the screen, one error at a time, when the student clicks an error button. The corresponding note (or bar, passage) is indicated, and the description is displayed to the student (see figure 3). The student continues by choosing between: getting more information on the error and how to correct it, repeat the phrase containing the error, practise on a special exercise focused on the current type of error, or play the whole piece again.

4. A SCORE VIEWER AS A MIRRORING USER INTERFACE

The role of the score viewer is to facilitate the learning process by providing innovative and enhanced forms of

No	Basic error/mistake	Average ranking	Instrument control (IC) Musical Performance (MP)
1	Air flow	1.7	IC
2	Fingering	1.7	IC
3	Rhythmic perf.	2.0	MP
4	Attack	2.0	IC
5	Melodic perf.	4.2	MP
6	Tempo	5.0	MP
7	Intonation	5.3	IC
8	Phrasing	6.0	MP
9	Articulation	7.5	IC/MP

Table 1. *Basic error categories, average ranking, and performance aspects. The ranking refer to the relative importance of the errors/mistakes during the first four terms of playing the recorder. Aspects of instrument control as well as musical performance are represented. Instrument control is considered most important to develop in the early stages of learning.*

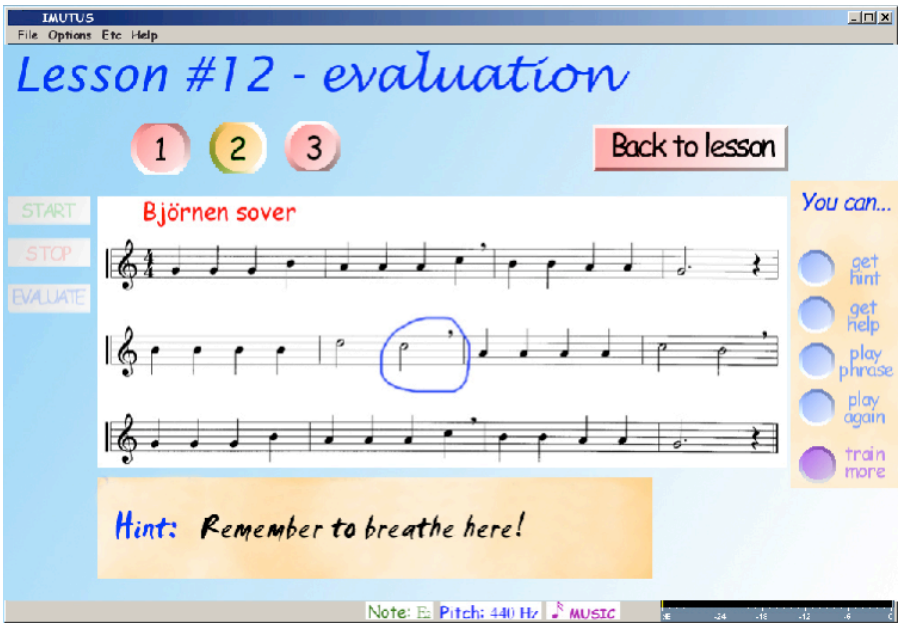

Figure 3. *Screen view example. The student has pressed Error button 2 and the position in the score and the descriptive message "Phrasing!" are displayed. The student has then asked for "Get hint" and a feedback message "Remember to breathe here" is displayed.*

feedback. To use a metaphor, the score viewer acts as a visual and sonic mirror that helps the student to become aware of his own performance. The interaction between the student and the score viewer can be divided in three main phases:

- In a first phase the student will use the score viewer to read the score to be performed, to listen to the reference performance and take note of the difficulties as well as any indications on how to perform the score correctly: this is the preliminary phase.
- In a second phase the student will use the score viewer to perform the piece by possibly using the metronome, the automatic cursor following and the turn-page facilities of the score viewer: this is the practice phase.
- In a third phase the student will use the score viewer to become aware of his own performance by the means of the automatic performance evaluation annotations (figure 3) as well as the possibility to listen to its own performance and to compare it with the reference performance. In particular he will be able to directly click on a specific note or selected section on the score to listen to the corresponding audio recording: this is the feedback phase.

While almost any evolved sequencer with common music notation capabilities[3] or score editors with MIDI playback facilities[4] provide automatic score scrolling and cursor movement, they usually lack two essential features needed by the IMUTUS score viewer :

- an automatic way to map the audio recording of the performance to the corresponding signs on the graphic score,
- an efficient look ahead feature allowing the performer to anticipate the next notes to be played when the end of the page is reached.

Electronic or digital music stands on the other hand, beeing designed to read music during a performance, provide look ahead features. For example Freehand MusicPad Pro[5] incorporates a half turn page feature. The MUSE system [7] that was designed (but never implemented) as part of the 1995 Apple Design Project at Carnegie Mellon University, proposes an automatic turn page system with look ahead. The MOODS system [8] provides look ahead using progressive vertical or horizontal replacement of the already played music. Unfortunately it seems that very few empirical studies have been made to evaluate the various look ahead options. The only publication we are aware of is a paper by John McPherson [9] but the empirical data are limited to 6 adult musicians.

[3] Emagic Logic (http://www.emagic.de), Mark of the Unicorn Digital Performer (http://www.motu.com)

[4] Finale (http://www.finalemusic.com), Sibelius (http://www.sibelius.com)

[5] http://www.freehandsystems.com

4.1. The score viewer global architecture

To provide the required features the score viewer is organized in several submodules :

- a set of players submodules: in charge of the audio, MIDI, metronome rendering. They can be triggered to play the whole piece or a specific time range. They can synchronize the graphic submodule to display the current position in real time.
- a graphic module: in charge of producing and merging several graphic representations in order to draw what is actually seen on the screen according to the current position in the score. It handles also the mouse clicks and selections on the score and may trig the audio or MIDI playback.

The graphic module is at the heart of the student/virtual teacher communication and interaction. All the information linked to pedagogical issues are related to the corresponding musical material and therefore displayed on the score: it represents pedagogical annotations, tips for a correct performance or for improvements, graphical signs to highlight a specific section. This information is attached to a given exercise or dynamically computed by the virtual analyst. Additional information a student may require is also related to the music in the score and therefore is obtained by interacting with the score: for example, a student may listen to a given note or ask for the fingering by clicking on the note. The score viewer may therefore be viewed as the user interface of the players mentioned above.

4.2. Segment mappings to relate graphic and sound

The rôle of a *segment mapping* is to relate *time based resources* defining correspondencies between *segments* of resources. In the framework of the IMUTUS project such mappings are typically used to link graphical positions, musical positions and audio positions. They are necessary to provide aural feedback mechanisms.

A *segment* is a portion of a time based resource defined by a starting date and an ending date. The *segmentation* is the process by which, according to some criteria, we divide a time based resource into a set of contiguous segments that cover the whole duration of the resource without overlappings. Audio segments for example, could correspond to recognized notes, extra notes, silence or noise.

IMUTUS time based resource use different time representations: absolute time in frames for the audio player, musical time expressed in bar/beat/unit for the MIDI players, graphic time in page number and x, y coordinates for the graphic module. Note that the musical time actually covers two different representations:

- the *unrolled* musical time: corresponds to the linear time of the performance; all the repeats and jumps (dal segno, coda) are *unrolled*.

- the score musical time (or *rolled* musical time): corresponds to the structured time of the score; repeats and jumps are *rolled*.

A *segment mapping* provides the required information to define *time convertion* functions. Appropriate links to synchronize all the related time based resources are created by combining such mappings (figure 4).

4.3. Real-time score position feedback

During the playback or recording phase the score viewer is informed in realtime by the score follower or the players of the current position in the score. It displays the corresponding page as well as a cursor on the next note to be played. When the cursor is on the last system of the current page it may also display the first system of the next page to allow the student to anticipate what is to be played next.

5. A SIMPLE SCORE EDITOR FOR AN EXPLORATORY APPROACH

The IMUTUS system includes a simple score editor intended for the ear training as well as for exploratory strategies. This editor provides a very simple way to write a short score and includes all the automatic layout capabilities to help in writing a well formatted score. The score complexity that is supported extends to monophonic scores (soprano clef, time and key signatures), notes and rests values down to 1/32, triplets, basic expression markings (legato, staccato, accents, breath marks) and tempo indications. Due to the complexity of the music language, most of the notation programs are not intuitive and the time necessary to learn and to master them represents an important investment. The score editor has been designed to avoid this complexity as well as to support the kind of feedback provided by the score viewer.

5.1. Underlying concepts

A score is viewed as a collection of measures, which is a collection of notes and/or rests. For simplification, we'll later use the term *symbol* to refer to notes and rests. The score editor basic concepts are reduced to two simple ideas:

- symbols are considered similarly to alphabetical character in a text editor: they are carrying attributes that may be compared to characters styles.
- time space within a measure is always consistent, it pre-exists to symbols and plays a role similar to magnetic grids in drawing applications.

5.2. Measures time space

Let's consider a 4/4 measure which grid size is the quarter note: such an empty measure is including 4 virtual quarter notes and will take the place of 4 real quarter notes. Since the time space pre-exists to symbols, a symbol may be drawn at the place of any of the virtual quarter notes (figure 5). This approach has some important repercussions for the editor design as well as from pedagogical viewpoint:

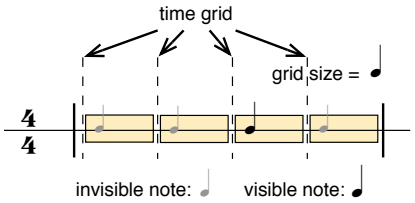

Figure 5. *A measure time grid.*

- a measure is internaly always complete and must remain so: wether visible or virtual, it always contains the required number of beats,
- since a measure is always complete, an *insert* operation doesn't exist: new symbols can only replace existing virtual symbols. For the same reason, a *delete* operation doesn't exist: symbols to be deleted are tranformed into virtual symbols.
- since a new symbol can only replace a virtual symbol, it's time location will be that of the virtual symbol. Therefore, places and size of the virtual symbols operate like a magnetic grid.

This approach has some important pedagogical and cognitive implications. The idea of magnetic grid may be used to guide the student in the music writing process. It allows to write scores with "*holes*" (i.e. scores with measures that appear to be incomplete). It simplifies the editor cognitive approach: writing a score consists in filling a pre-existing static time space with symbols, there are no side effects such as those introduced by insertion or delete operations for example. It has also the effect to visually stabilize the score through editing operations since it minimizes the space to be possibly created by these operations.

5.3. Symbols attributes

Duration, accidentals and expression marks are symbols attributes; they may be viewed like text styles (bold, italic...) and are edited similarly:

- new notes carry the current selected attributes,
- when selected, an attribute is assigned to or removed from the current selection of symbols.

Special attributes that only concern group of notes (tie, slur, triplets) are also edited similarly: the group should be selected first and the attribute is applied or removed in a second step.

5.4. Example of supported activities

Fundamentals of music are approached by game-like exercises. It includes *music dictation*, *fill score* and *simple composition* exercises. All of them require the support of

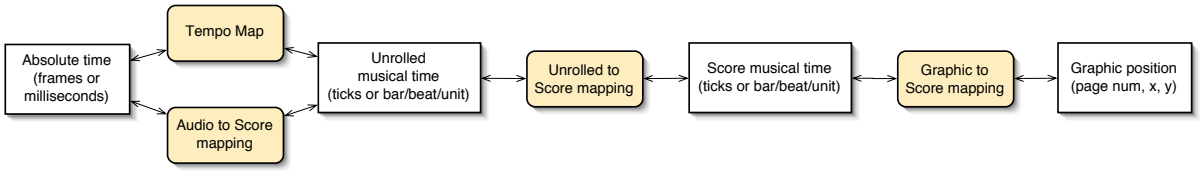

Figure 4. *Mappings between different time representations.*

the score editor. The *fill score* exercise for example consists in writing notes that are missing in an incomplete score that the student can listen to: the score editor provides the possibility to write such scores (with missing notes) while preserving a consistent time/space layout.

6. CONCLUSION

A computer can provide a feedback that adds to human teacher practise, by incorporating visual and aural feedback; it may be used as a mirror reflecting certain aspects of a student performance. IMUTUS will provide interesting tools for exploring the possibilities of computer-assisted teaching. The combination of performance-specific feedback, visually presented score-viewer could enhance the efficiency of practising. This consideration has been notably investigated in the aural domain with the extended score viewer presented above. At first use, interesting new possibilities appear such as comparing how different separated notes have been played. We have also proposed a music score editor based on reduced and simplified concepts. First experiments of the editor shows that these concepts are easy to understand and intuitive enough to be mastered by young students. In the future they may be applied to more complex score.

7. REFERENCES

[1] MICHAEL J. BAKER, *Design of an intelligent tutoring system for musical structure and interpretation.* In K. E. M. Balaban and O. Laske, editors, Understanding Music with AI: Perspectives on Music Cognition, pages 467489. The MIT Press, Cambridge, MA, 1992.

[2] BRANDÃO, PAIN, WIGGINS, *Computers in Music Education.* Proceedings of the AISB'99 Symposium on Musical Creativity, 1999

[3] ROGER B. DANNENBERG, MARTA SANCHEZ, ANNABELLE JOSEPH, PETER CAPELL, ROBERT JOSEPH, RONALD SAUL, *A Computer-Based Multi-Media Tutor for Beginning Piano Students.* Interface - Journal of New Music Research, 19(2-3), 1990, pp. 155-173

[4] ROGER B. DANNENBERG, MARTA SANCHEZ, ANNABELLE JOSEPH, ROBERT JOSEPH, RONALD SAUL, PETER CAPELL, *Results from the Piano Tutor Project.* Proceedings of the Fourth Biennial Arts and Technology Symposium, Connecticut College, (March 1993), pp. 143-150

[5] GILLES MATHIEU, *Virtual Music Teacher.* Projet de fin d'étude - Rapport de synthèse, INSA Lyon, 2003, (http://gmwork.free.fr/Vmt/)

[6] S. W. SMOLIAR, J. A. WATERWORTH, P. R. KELLOCK, *A system for piano education beyond notation literacy.* In Proceedings of ACM Multimedia 95, pp. 457-465. San Francisco, CA, 1995.

[7] CHRISTOPHER GRAEFE, DEREK WAHILA, JUSTIN MAGUIRE, ORYA DASNA, PITTSBURGH, *Designing the muse: A Digital Music Stand for the Symphony Musician.* Proceedings of CHI 96 - Conference on Human Factors in Computing Systems - ACM, 1996

[8] PIERFRANCESCO BELLINI, FABRIZIO FIORAVANTI, PAOLO NESI, *Managing Music in Orchestras.* in Computers - IEEE, September 1999 - pp.26-34

[9] J. MCPHERSON, *Page turning — score automation for musicians.* Honours thesis, University of Canterbury, New Zealand, 1999