Composing Music on Paper and Computers: Musical Gesture Recognition

Bobby Owolabi
Department of Computer Science
Human Computer Interaction Lab
University of Maryland
College Park, MD 20742
rowolabi@umd.edu

ABSTRACT

Paper is preferred and utilized far more than computers in the composer's music creation cycle because it is the natural medium in which music notation convention is learned. Current music notation software utilizes only WIMP interfaces (Windows, Icon, Menus and point and click). Our system enables users to create musical compositions by utilizing digital pen technology and having their work captured and recognized into music notation in the digital world. This recognized pen gesture data has the potential of being imported into a popular musical notation composition program for editing purposes.

INTRODUCTION

Despite technological breakthroughs in computing technology in the last few decades, paper is being used increasingly [8]. This is especially true in artistic and creative disciplines, such as music, where paper tends to be utilized a great deal. This state of affair exists because of the complementary set of affordances that paper and digital documents provide [4]. Paper is portable, flexible and inexpensive [8]; while, digital documents are easy to transmit, store and modify. For instance, when a composer is in the early stages of creating a musical piece, it is far easier to take out a sheet of paper on the spark of inspiration, rather than a technological device, because one can express themselves on paper quicker than waiting for a device to load. At the same, there are reasons a composer would want to use a computer such as instant playback or the ability to quickly transpose a piece to a new key. Using the traditional written approach as well as computer software in conjunction often presents challenges.

Software applications such as Finale and Sibelieus [2][9] enable users to create the same musical notation they would on paper in the digital realm; however, WIMP (Windows, Icon, Menus and point and click) interfaces are not serving composers well in their composition activities because they bear little resemblance to the handwritten techniques learned by many musicians [3]. This shortcoming can cause a hindrance to the music creation cycle.

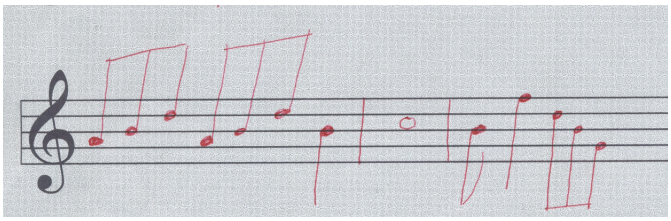


***Figure 1**: Music created on anoto paper with a digital pen.*

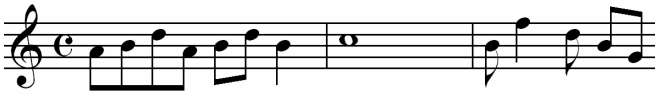


***Figure 2:** Written music as it would appear entered in Finale software [2].*

Software, such as The Music Notepad [3], designed for the tablet PC, is a move towards a more natural interface for composers because of the familiarity of its pen based input; however, affordances offered by paper, such as flexibility and portability are often lost.

Our system attempts to take the affordances offered by both paper and digital documents and unify them into one system. Users will be able to create traditional handwritten music and have their creations captured and recognized into music notation in the digital world. Our system will better support the musician's music creation cycle by reducing the time composers spend synchronizing their works in both the physical and digital realms.

RELATED WORK

Investigations have been performed to address many of the difficulties that composers experience when utilizing today's musical notation software and also understand the interactions composers engage in with

paper and computers. Three categories of work that are relevant to our system include: The Role of Paper and Computers, Tablet PC Interfaces and Digital Pen Technology.

The Role of Paper and Computers

In trying to develop an interface that will support composers in their activities, it is important to understand how composers create music and their interactions with computers and paper.

In the Paperroles Project [6], the music creation cycle of composers was investigated. The description provided by the Paperroles Project leans more towards classical composers and culture; however, it provides an overview of the interactions with paper and computers by people in the musical domain and can be summarized into three basic parts.

In the beginning, composers prefer to work on paper because of the freedom of expression that it offers. They are not bound by the limitations of a software program such as slow input, hardware loading time and portability issues such as weight and size of hardware.

In the middle of the cycle, there is a mixture of paper and computer use. Computers provide composers with ease of modification. Often, composers experience difficulties during the middle stage because they are slowed down by slow user input speed.

In the end, most composers in the classical genre prefer paper for archival purposes.

The Paperroles Project offers a design scenario that proposes actually taking the handwritten document of a user, utilizing digital pen technology to recognize the written gestures and provide the ability for the musical notation to be opened up in an application in digital form for further editing. Our system aims to demonstrate the recognizing framework of this interface.

Tablet PC Based Interfaces

A major trend that is currently being investigated is pen based interfaces over traditional mouse interfaces. Such examples include Presto [5] and The Music Notepad [3].

There are a lot of complexities in detecting and recognizing musical gestures because the vast array of possible handwriting styles of users and interpretations of music notation convention. Proposed systems have offered simplified gestures that correspond to music notation objects. With these mappings, designers wish to offer a quick learnable interface to users, while being able to more accurately detect the true intention of a user's gestures. The Presto and Music Notepad [5][3] systems have taken this route.

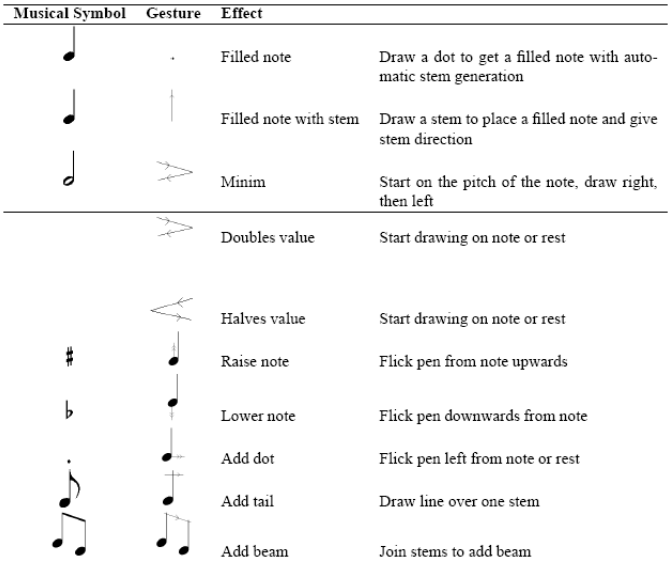


Figure 3: *Main gestures in the Presto system, reproduced from original text [5]*

The researchers of the Presto system performed an investigation of users' hand-written music composition habits and styles and developed a proposal to replace standard music notation with simplified versions to make gesture recognition easier and more accurate. The collection of gestures is presented in a way that encourages building a note rather than providing gestures that are directly mapped to specific music gestures (Figure 3).

The Music Notepad is an application that supports music notation entry and also replaces the standard music notation with simplified versions that are different from the Presto system. The Music Notepad system offers a larger set of gestures that map to each music notation object (Figure 4). The Music Notepad also utilizes a special stylus pen that has buttons that correspond to different classes of gestures.

The Presto system offers a smaller table of gestures for the user to learn while allowing multiple ways to draw gestures. This characteristic would most likely enable the user to adapt to the system quicker than the Music Notepad. However some of Presto's gestures do not resemble the gestures they are mapped to; for instance, the half note (Figure 3). Unfamiliar gestures increase the learning curve of the system. The Music Notepad's gestures more closely resemble the gestures they are mapped to and can be drawn in one stroke for the most part; whereas, Presto's gestures may take multiple strokes to complete; consequently resulting in slower gesture creation.

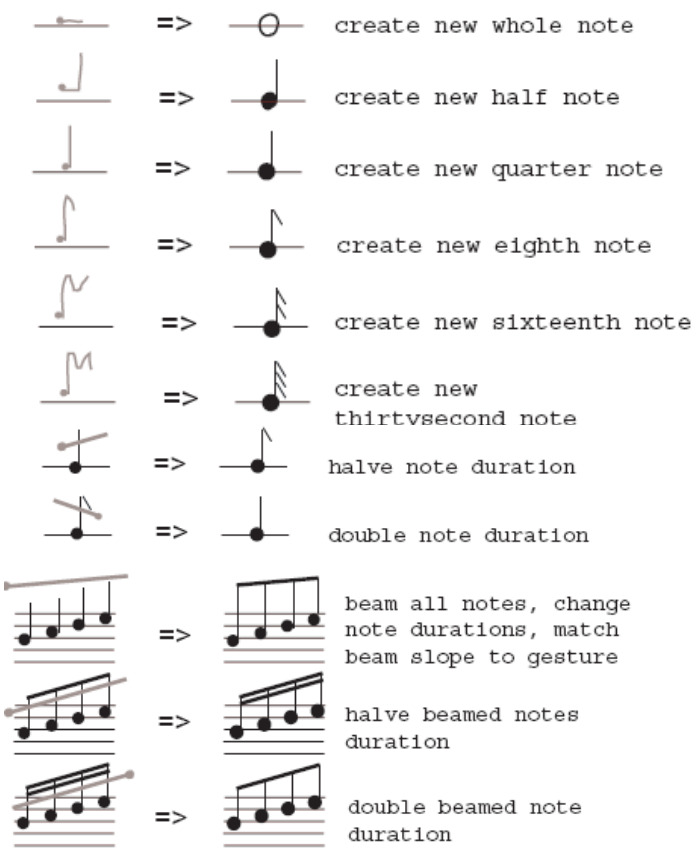


Figure 4: *Gestures from the Music Notepad reproduced from original text [3]. Gestures on the left are drawn by the user. Gestures on the right are resulting gestures drawn by the computer.*

The system we are proposing will directly recognize standard music notation rather than providing a mapping of simplified gestures to music notation objects.

Digital Pen Technology

In moving towards pen based interfaces for music notation software, the emergence of digital pen technology has opened up exciting possible applications for software interfaces. Such work includes Paper Augmented Digital Document (PADD) [4]. A PADD is a digital document that can exist and be manipulated in both the digital and physical realms. The framework of our system is built off this concept.

MUSIC BACKGROUND

Written music notation has gestures that convey both pitch and rhythm called notes. There are gestures that convey silence over time, called rests. Lastly, there are gestures that provide information about how a note should be played, called articulations. During this stage of system development, our main focus will be on the detection of notes.

Staff

Music is written on a staff. A staff (Figure 5) consists of a series of horizontal lines, most commonly five, that are stacked, equally spaced, on top of each other that run across a page.

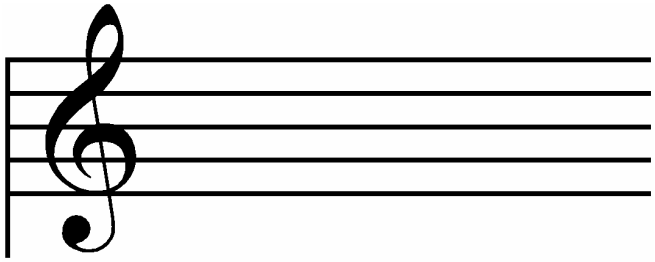


Figure 5: *A standard 5 line staff with treble clef symbol for pitch orientation purposes.*

Notes

Notes consist of four fundamental parts (Figure 6, Appendix A for examples): (1) *Head*, which is an empty or filled circle, its position on the staff specifies pitch. (2) *Vertical beam*, which is connected to the head. (3) *Stem,* which intersects vertical beams and adds meaning to the duration of a note. (4) *Horizontal beams* are extended stems that connect several notes.

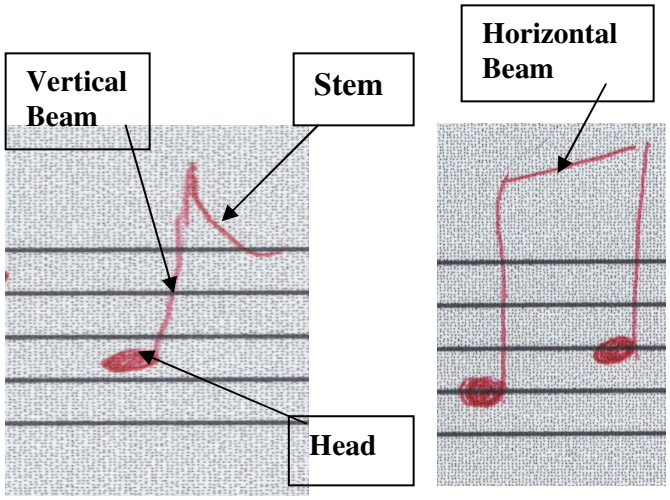


Figure 6*: Parts of notes.*

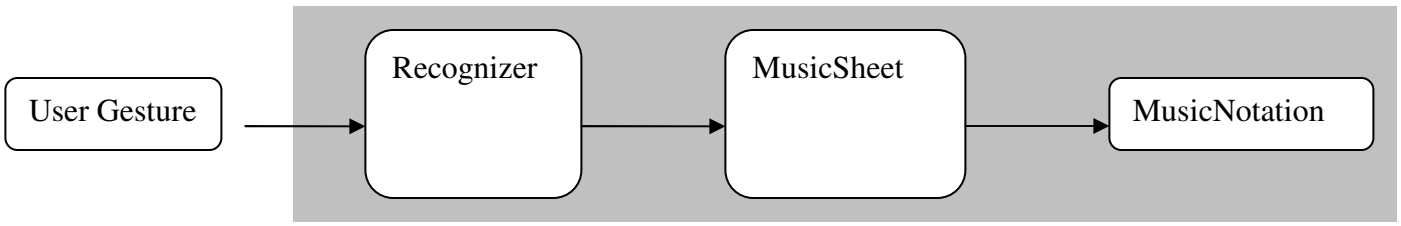


Figure 7: *Overview of detection process: The user makes a gesture with the digital pen, once the pen is lifted off the paper, it is sent to the Recognizer to get fundamental shape (ex: line or circle). It is then sent to the MusicSheet object to get its location on the sheet as well as its spatial relationship with the other gestures on the page. Based on that information, it is recognized as a MusicNotation object and stored.*

Articulations

Articulations are symbols drawn around notes that provide information about how the note should be played. For instance, an articulation may express that a series of notes should be played smoothly and connected without minimal space of rest between them. Where as another articulation may express that a series of notes should be play very short and fast with more rest space between them (For example, see accent in Appendix A).

SYSTEM

Our system enables users to write musical compositions utilizing digital pen and paper technology and have their work captured and recognized in the digital world with the potential of importing the data into a popular musical notion composition program.

There are three main components of our system: (1) *Preprinted Paper*, (2) *Digital Pen* and (3) a *Recognizer*.

Preprinted Paper

In our system, we preprinted staves on paper and stored the location of the staves of in our digital model of the page. The staves were printed on Anoto [1] paper. This special paper, as described below, enables the digital pen to record the user's strokes.

Digital Pen

The digital pen has the properties of a regular pen; however, it has a small camera at the bottom that records the coordinates of the user's gesture. The coordinates can either be stored in the pen's memory and sent to the computer at a later time or streamed in real-time with Bluetooth technology. The digital pen is able to know its location on the sheet of paper and record the coordinate points of user drawn gestures because of special properties of the paper. Every sheet of anoto paper has a unique id number and pattern of small dots that enables the pen to know the exact paper, its location on it; thus, enabling it to record user gestures.

System Components

Once a user draws a gesture, it goes through two stages: the initial recognition and spatial analysis on the MusicSheet object (Figure 7).

The main elements of the systems are:

- **MusicSheet -** A MusicSheet object is a digital representation of a physical sheet of music. It is created based on key parameters of the physical sheet music such as the y-coordinate values of the staff line and the distance between staff lines. It consists of a series Staff objects. Staff objects consist of a group of five horizontal lines. Each Staff object contains gestures that the user has drawn on it. A MusicSheet object allows for ease for common queries such as finding which staff a user's gesture falls on. In order to query the pitch of a given note, the MusicSheet object can query its Staff objects and discover the pitch of the note, based on where it falls on the lines.
- **MusicNotation** – Represents one of the standard musical symbols (Appendix A). Recognition is determined by the spatial relationships of fundamental shapes that the user draws to build a musical gesture.
- **Recognizer** – Detects fundamental shapes that the user draws in one pen stroke. Its main component is an implementation of the $1 Recognizer [10]. The $1 Recognizer consists of a four step algorithm where a candidate user gesture is compared with pre-defined templates of various expected gestures.

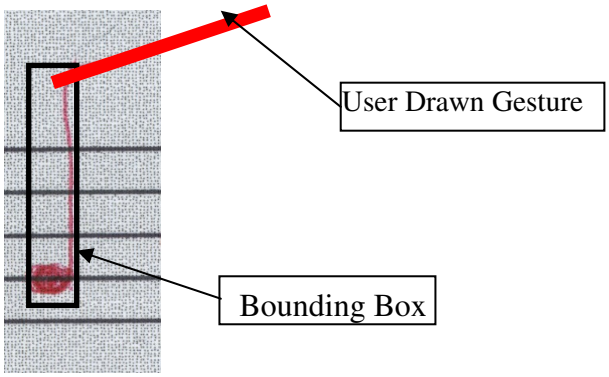


Figure 8: *Example of how gestures are linked together to form one object. If a user draws a gestures that intersects or is reasonably close to a bounding box, then the gestures are merged together to form a new object.*

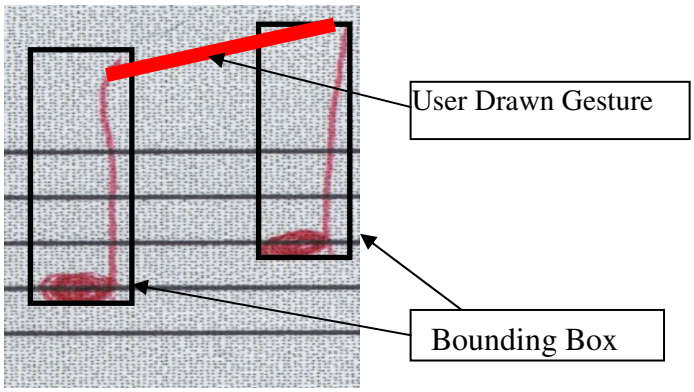


Figure 9: *Example of a situation where a user draws a gesture that intersects more than one previously drawn gesture that results in the formation of a beamed note.*

Recognizing fundamental shapes

Written music notation has gestures that convey both pitch and rhythm called notes. It was found that notes could be represented with 3 fundamental gestures: filled circle, empty circle and line segments with various orientations.

Our system requires that the user draws the fundamental gestures of music notation, circles and lines, one at a time. This means after each fundamental gesture that the user draws, the user must pick up their pen from the page. Given this constraint, we are able to detect single notes through parsing the sequence of fundamental gestures that intersect other previously drawn gestures.

There are situations where the result of the Recognizer is not sufficient and semantic information of the gesture, such as its relation to other gestures must be taken into account. For instance, for detecting filled circles, it is not practical to create templates and expect that the points of a template will match up with a user candidate filled circle. Therefore, it is necessary to do secondary detection checks to see if the candidate gesture is a filled circle. Things such as the number of points that are on the border of a gesture versus the number of points that are contained within the gesture can be investigated. The ratio of the length and height of the gesture can also be taken into account to determine if it is a filled circle.

Detecting Single Notes

For our purposes, single notes are defined as notes that have only one head; or more specifically, notes that are not beamed together with other notes. They include whole, half, quarter, eighth, sixteenth and thirty-second notes (See appendix A).

Each object that is created by the user has a bounding box. The recognition state of the bounding box is governed by an automaton shown Figure 10.

As a user draws gestures that intersect previously drawn gestures, which are determined by querying the MusicSheet object, the group of gestures becomes one object (Figure 8). A new object and automata is started when the user draws a gesture that does not intersect another gesture.

The MusicSheet can also be queried to find information such as the pitch of the note.

Detecting Beamed Notes

Our approach for detecting beamed notes was to identify key scenarios where a beamed note is likely to occur. For example, if a user draws a gesture that intersects more than one previously drawn gesture, then there is a high chance that the user drew or is going to draw a beamed note (Figure 9). Further analysis, such as looking at the recognition status of the intersected gestures, can verify the recognition.

Detecting Articulations

Detecting articulations currently relies primarily on creating articulation templates for the $1 Recognizer [10] that is utilized in the system. The system detects two articulations: ties and accents (Appendix A). Currently, detected articulations are not linked to notes; however, as the system is developed, spatial information of the articulations will be used connect articulations with their corresponding notes.

EDITING

If a user wishes to delete a note, they can draw a "scribble" on top of the note and it will be deleted from memory.

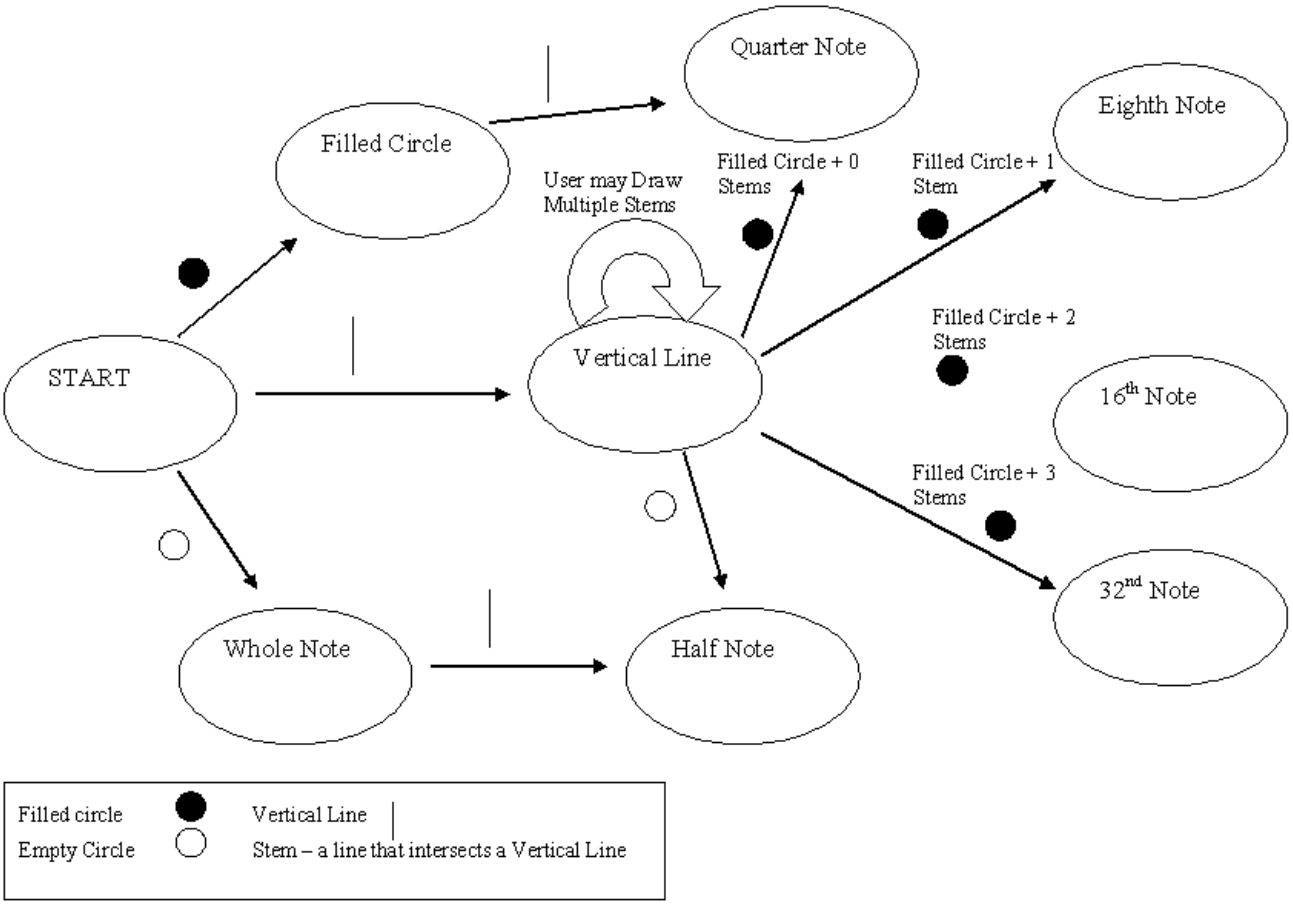


Figure 10: *Parser for Single Notes. As a user draws a gesture that does not intersect another gesture, the parser for that object starts at the "START" state. As the user draw lines and circles that intersect the given gesture, the state for the object is updated according to this diagram.*

FUTURE WORK

The system is still in its early stages of development. We would like to study how various composers actually write compositions and investigate ways to enable our system for more "natural" usage. Currently, the system puts the constraint on the user that they must lift up the pen after creating key gestures such as lines and circle. We would like to look at methods that would focus more on the end result of a user's gesture rather than recognizing the steps a user takes to create the gesture. This idea is demonstrated in Simu-Sketch [7].

With regards to modification of notes, we would like to investigate with users what works well and what does not. What is practical and does it work well with the music creation cycle?

We would like to also investigate the performance of this system compared with traditional methods and systems such as Presto [5]. What is the speed improvement of this system over the traditional pen and paper method, which is to compose music on paper then manually transfer manually to software. The Presto system [5] offers some figures on time improvements over traditional methods; it would be interesting to see how ours compares.

Lastly, we would like to create a plug-in for an application such as Finale, and import recognized pen data into the program.

CONCLUSION

The system presented in this paper demonstrates the recognition portion of the design scenario mentioned in the Paperroles Project [6]. It attempts to demonstrate an interface that support a more natural music creation cycle of composers. We described previous related work, an overview of the system and future work.

The interface of our system better supports composers in their activities because it provides a more natural interface to music entry and moves away from mouse-point-click interfaces.

ACKNOWLEDGMENTS

The author would like to thank Dr. Dr. François Guimbretière for his support throughout the project. Hyunyoung Song, Nick Chen and Chunyuan Liao for their generosity in time and help. Dave Levin for his advice, time and offered experience. This work was supported in part by a REU for grant IIS #0447703 and by the Louis Stokes Alliance for Minority Participation Undergraduate Research Program.

REFERENCES

1. Anoto AB, Anoto Technology. http://www.anoto.com

2. Finale, Coda Music Technology, http://www.codamusic.com/.

3. Forsberg, Andrew, Mark Dieterich, and Robert Zeleznik. "The Music Notepad", Proceedings of UIST '98, ACM SIGGRAPH.

4. Guimbretière, François, Paper augmented digital documents, Proceedings of the 16th annual ACM symposium on User interface software and technology, p.51-60, November 02-05, 2003, Vancouver, Canada

5. J Anstice, T Bell, A Cockburn and M Setchell. The Design of a Pen-Based Musical Input System. OzCHI'96: The Sixth Australian Conference on Computer-Human Interaction. Hamilton, New Zealand. 24-27 November, 1996. pages 260-267. IEEE Press.

6. Letondal, C. and Mackay, W. (2007) Paperoles: The Paperoles Project: An analysis of paper use by music composers. In Proceedings of CoPADD, Collaborating over Paper and Digital Documents, London, U.K.

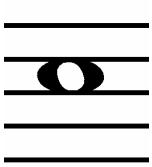
7. Levent Burak Kara , Thomas F. Stahovich, Hierarchical parsing and recognition of hand-sketched diagrams, Proceedings of the 17th annual ACM symposium on User interface software and technology, October 24-27, 2004, Santa Fe, NM, USA

8. Sellen, A.J. and R.H.R. Harper, The Myth of the Paperless Office. 1st ed. 2001: MIT press.

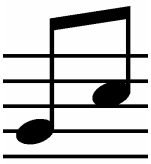
9. Sibelius, http://www.sibelius.com/.

10. Wobbrock, J.O., Wilson, A.D. and Li, Y. (2007) Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes. Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '07). Newport, Rhode Island (October 7-10, 2007). New York: ACM Press, pp. 159-168.

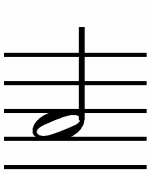
Appendix A:
Recognized Musical Notes



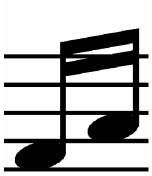
Whole Note



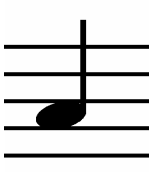
Beamed Eighth Note



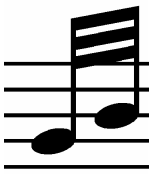
Half Note



Beamed Sixteenth Note

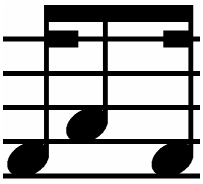


Quarter Note

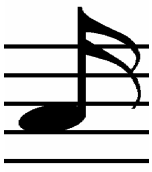


Beamed Thirty-Second Note

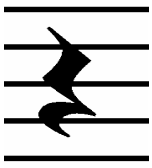
Eighth Note



Mixed Beamed Note

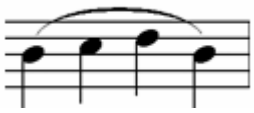


Sixteenth Note

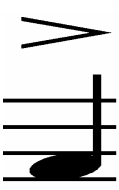


Quarter Rest

Thirty-Second Note



Tie



Accent