Usability Test Report :: Working with Scores

Authors: Inna Kouper and Mark Notess

Last Edited: March 30, 2005
Test dates: February 23, 2005 – March 7, 2005
Test facilitator: Inna Kouper
Observers: Mark Notess, Don Byrd
Executive Summary

After the release of Variations2 annotation tools set, a usability test was conducted to identify any possible problems with using this set of tools to work with the scores.

During the last week of February 2005 and first week of March 2005 eight participants individually worked through a series of tasks and filled out background and satisfaction questionnaires. The task list consisted of three tasks: highlighting the melody on the score, replicating the analysis of a music piece, choosing correct answers for the key and harmonies in the particular piece. Tests were conducted in the School of Library and Information Science (SLIS) usability lab; each session lasted for about an hour.

The results indicate that participants were interested in working with annotation tools and eager to learn and apply as many tools as possible. Via comments, questionnaires and debriefing sessions participants indicated above-average satisfaction with the system as a whole. Some specific ratings such as ratings related to the buttons in the annotation score window and to the navigation within this window are lower on average than others. Participants also expressed doubts over the general usefulness of the annotation tools set because it requires additional time for learning and working comparing to the hand-written analysis. Nevertheless, most of participants were enthusiastic about using annotation scores in classes. They liked the possibility to create professionally looking annotations and make multiple copies of the annotated score as well as the availability of scores online.

One apparent failure of the study is the failure to print the score. When first 4 participants tried to print an annotated score, nothing came out of the printer. As it turned out to be a bug in the system associated with highlighting, this subtask was removed from the test. Another problematic issue that caused frustration among almost all participants was a general interface issue. Participants expected the tool buttons to stay selected after they were used to be able to draw the same figure again. Other issues raised in the study were related to specific tools and buttons associated with it. Thus, participants had problems with changing the style of lines and arcs to add arrows, with ending the polygon figure, with using Connection and Elbow Connection tools and with turning the bracket figure upwards.

Partially, the problems encountered are due to the learning curve associated with using new software. However, some improvements can be made to facilitate navigation and use of annotation tools. Recommendations for resolving the issues mentioned above as well as other issues described below are provided based on the results of the study and the suggestions of the participants.
# Table of Contents

Executive Summary ................................................................................................................. 2  
Table of Contents .................................................................................................................... 3  
Purpose of Study ....................................................................................................................... 4  
Participants ............................................................................................................................. 4  
Method ..................................................................................................................................... 5  
Findings ...................................................................................................................................... 6  
  Quantitative measurements ................................................................................................. 6  
  Observations ......................................................................................................................... 7  
  Satisfaction ratings ............................................................................................................... 9  
Recommendations ................................................................................................................... 12  
Limitations of the study and future work .............................................................................. 16  
Appendix .................................................................................................................................. 18  
  Demographic Questionnaire ............................................................................................... 18  
  Tasks ..................................................................................................................................... 19  
  Satisfaction Survey ............................................................................................................ 20  
  Debriefing Questions ........................................................................................................... 20  
  Examples of the Task Decisions ......................................................................................... 21
**Purpose of Study**

This test is part of ongoing usability testing for Variations2: IU Digital Music Library project (V2 hereafter). Primarily the V2 project is a research project aimed at providing an arena for research in such areas as usability, copyright, metadata, system design, networking and music instruction. As functional software, V2 is a digital library system, which provides access to a multimedia collection of music in a variety of formats and styles. Using V2, students, educators and librarians may search for and listen to sound recordings, browse and display scores (scanned or encoded), as well as work closely with musical compositions using timeline and bookmark tools.

In fall 2004 a new set of tools called “Annotation Tools” has been added to V2 functionality. With the annotation tools, one can annotate existing scores with text, basic figures such as circles or rectangles, arrows, etc. and special markers. The tools can be used for teaching an interactive lecture, writing/creating papers and presentations based on scores as well as for personal studies and assignment work. The modified scores can be then saved or printed out. In order to identify problems users may experience with V2 annotation tools and provide recommendations for improvement it is necessary to conduct a usability study of annotation tools set.

The overall goal of usability testing is to identify problems users experience with the product and provide recommendations for further improvement. Three major concerns defined the goal of this test:

- Will a user be able to understand the functionality presented in annotation tools menu options and buttons?
- Will a user be able to use the annotation tools set effectively (e.g. navigate between pages, create annotations, edit / save annotations, find help if necessary)?
- How helpful are contextual help pages and the user guide (if consulted)?

To address these concerns, the test was designed to accomplish the following:

- register sequences of actions necessary to perform certain tasks
- measure time users spend performing certain tasks
- register errors users encounter while working with annotation tools
- identify confusing steps or messages or wrong actions users experience while working with annotation tools
- measure overall user satisfaction with annotation tools as well as V2

**Participants**

Eight students were recruited by the facilitator from graduate and undergraduate music classes. During recruitment, the facilitator visited classes, explained the purpose of the study and passed around a sign-up sheet. Those students who were interested in participating wrote their names and email addresses on the sheet. The facilitator contacted them at a later time and set up session times. Participants worked individually during all sessions. A total of eight sessions were completed.

As shown in the table below, three females and five males participated in the study. Six participants were undergraduate students, two – graduate. Only one participant claimed to use computers more than 21 hours a week, three participants said they used computers 11-20 hours per week, two claimed 6-10 hours a week and last two claimed 0-5 hours a week. On average those who use computers between 0 and 20 hours per week spend about 1.5 hours a day, which is quite understandable for music majors but can be insufficient to become an experienced user.

Platform experience ratings reflected medium experience with PCs and essentially more moderate experience with Macs. On a scale from 1 to 5, with 1 being Novice and 5 being Expert, the mean result for PC use was 3.3 ranging from 2 to 4 and for Mac use was 2.1 ranging from 1 to 4.
Table 1. Participant Characteristics.

<table>
<thead>
<tr>
<th>#</th>
<th>Gender</th>
<th>Current academic level</th>
<th>Computer use (hrs per week)</th>
<th>Computer experience (1=novice - 5=expert)</th>
<th>Used Variations before</th>
<th>Used Variations2 before</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>male</td>
<td>undergraduate</td>
<td>11-20</td>
<td>4</td>
<td>1</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>male</td>
<td>undergraduate</td>
<td>0-5</td>
<td>3</td>
<td>1</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>male</td>
<td>undergraduate</td>
<td>0-5</td>
<td>2</td>
<td>2</td>
<td>yes</td>
</tr>
<tr>
<td>4</td>
<td>female</td>
<td>masters</td>
<td>21+</td>
<td>4</td>
<td>4</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>female</td>
<td>masters</td>
<td>11-20</td>
<td>3</td>
<td>3</td>
<td>yes</td>
</tr>
<tr>
<td>6</td>
<td>male</td>
<td>undergraduate</td>
<td>11-20</td>
<td>3</td>
<td>1</td>
<td>yes</td>
</tr>
<tr>
<td>7</td>
<td>male</td>
<td>undergraduate</td>
<td>6-10</td>
<td>3</td>
<td>3</td>
<td>yes</td>
</tr>
<tr>
<td>8</td>
<td>female</td>
<td>undergraduate</td>
<td>6-10</td>
<td>4</td>
<td>2</td>
<td>yes</td>
</tr>
</tbody>
</table>

All participants reported previous experience with Variations; two participants reported previous experience with Variations2. One participant had participated in usability tests before.

**Method**

All test sessions were coordinated by the facilitator and completed at the SLIS Usability Lab. The usability lab consists of two rooms. In one, the testing room, furniture and technology are laid out to simulate a typical office environment. Here the user interacts with the computers. In the observation and control room, video monitors and computer monitors are set up and evaluators can watch user’s hand movements on the keyboard and mouse as well as perceive any pertinent facial expressions. The testing room is monitored by several videos and through a one-way window so that observers can view the proceedings in real-time. Test facilitators can also communicate with users from the control room via a table microphone connected to a one-way speaker in the testing room.

A current version of Variations2 (version 3.2) was installed on the PC in the testing room. The screen resolution of the computer was changed to 1024x768 to match the most common resolution.

Upon each participant’s arrival at the lab, the facilitator greeted a participant, asked him or her to be seated at the PC and read the orientation script explaining the purpose of the test, asking to perform as they usually would in similar real-life contexts, and briefing on the session procedure. Each participant filled out a receipt for the $15 Borders gift card awarded for performing the test.

Once done with that, each participant filled out a demographic questionnaire (see Appendix, p. 18 for the list of questions) and proceeded with the task list (see Appendix, p. 19.) Tasks included identifying melody in a score, replicating annotations drawn on paper, identifying the key, harmonies, and non-chord tones by choosing correct answers from drop-down boxes on the score. Participants had to search for required pieces and complete the given tasks while articulating aloud all actions. Initially, the second task asked participants to print out the score with annotations. However, after first 4 tests where printing completely failed, this part was removed from the task.

Audio as well as computer screen interactions were videotaped for future analysis. After the scenarios and tasks were completed to the best of the participant’s ability, each participant filled out a V2 satisfaction survey (see Appendix, p. 20 for the list of questions.) After a participant completed the survey, the facilitator then asked a few debriefing questions based upon observations of the participant’s task completion activities and comments (see Appendix, p. 20 for approximate set of questions.)
Finally, the facilitator thanked the participant for his or her participation and answered any questions about the testing procedure that the participant might have had. During the test sessions and subsequent analysis the facilitator measured time to complete tasks and its parts and registered success or failure in completing a task.

**Findings**

Findings below are organized into three sections. The first section provides quantitative measurements of time for completing the tasks and discusses failure/success rates. The second section provides observations on each task. Finally, satisfaction measures for this study and a comparison of satisfaction measures across several studies are provided in the third section.

In cases where specific qualitative comments are presented, each comment is designated by a number which references the associated user. For example, a comment made by participant number 4 would be represented by “[4]”.

**Quantitative measurements**

As illustrated in the table below searching did not take much time, ranging from 10 seconds in the first task to 3 minutes in the second task. All participants except one were able to understand how the search window worked and find the required pieces.

**Table 2. Time in min:sec to complete actions.**

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean time</th>
<th>Range (minimum – maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highlighting the melody ($V2$ already open; from typing text to click on the link)</td>
<td>0:58</td>
<td>0:10 – 2:10</td>
</tr>
<tr>
<td>Annotating ($from$ click on “Show Tools” $to$ going to File menu to save)</td>
<td>6:45</td>
<td>3:10 – 9:08</td>
</tr>
<tr>
<td>Replicating existing score annotations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Searching for the piece ($V2$ already open; from typing text to click on the link)</td>
<td>1:11</td>
<td>0:35 – 3:00</td>
</tr>
<tr>
<td>Annotating ($from$ click on “Show Tools” $to$ going to File menu to save)</td>
<td>18:00</td>
<td>5:30 – 25:00</td>
</tr>
<tr>
<td>Choosing correct answers</td>
<td>4:18</td>
<td>1:24 – 7:26</td>
</tr>
</tbody>
</table>

Before starting to work on each task, participants usually spent 1-2 minutes exploring $V2$ annotation tools interface by clicking through menu options and moving over various buttons. When a score is opened, the annotation tools are not visible at the start so this short exploration helped participants to find “Show Tools” button to start annotating. Average time to highlight the melody on the first five pages of Debussy “Prelude to the afternoon of a faun” (task 1) is 6:45 (m:ss),
ranging from 3:10 to 9:08. Average time to complete the second task is 18:00, ranging from 5:30 to 25:00. The participant who spent the least time on this task used only the pencil tool to draw various shapes and lines. The third task turned to be the easiest one and took only 4:18 on average, ranging from 1:24 to 7:26.

As for the rate of failures and successes in tasks completion, 7 participants succeeded in all tasks without intervention. Even though participants couldn’t print the scores, they didn’t recognize it as a failure and moved on, assuming that the printer was in the other room and that the printing was successful. One participant needed help with searching and was able to complete all annotations after that. Even though annotations made by participants didn’t match the model examples developed for the testing, all annotations including the one made only with the pencil tool were indicative enough to recognize the tasks and the analysis. The details of completing the tasks will be discussed in the observations section and in limitations section.

Observations

Searching and exploring Variations2

1. Most participants found the search interface easy to understand except one participant who expected the scores to be already opened and had troubles figuring out that this is a search interface.
2. Some participants couldn’t find the pieces due to misspellings or typing in the wrong field. Thus, few participants typed work titles in the second field “Performer/Conductor.” All of them were able to correct themselves from the second or third try.
3. One participant, who couldn’t find the piece for the first task without facilitator’s help, had several failures (mixed French and English spellings in the work title, got results without scores, etc.) and couldn’t find necessary information in the user guide. After facilitator’s intervention, this participant was able to continue.
4. After opening the score, participants explored the interface of AT by clicking on various menu options and examining buttons. All of them were able to find how to make the AT toolbar visible and used basic functions such as drawing and editing annotations, turning pages and saving files quite well.

Working with annotation tools

1. All participants except one were interested in exploring the annotation tools and tried to apply them properly.
2. In the first task participants used different colors and text annotations to highlight melodies of the piece. In the second task they used as many tools as possible to replicate the existing analysis. During the third task one participant reported a missing option for the Cadence label – the deceptive cadence.
3. 6 out of 8 participants used the same tool repeatedly and expected tool buttons to stay selected after they were used to be able to draw the same figure again. The comments were “… I would just make the tools more permanent; having to re-click the same tool for every annotation got tedious” [3], “It was frustrating that I had to remember to select highlight every time” [4], “Selecting this [highlighter - IK] every time was a bit frustrating” [6].
4. Almost all participants expected annotations to remain selected after they’re drawn to be able to change shape or color: “It’s weird, I draw it and then I have to highlight it again when I want to change the color or size” [1].
5. Participants had difficulties creating final points of polygons. One participant figured it out quickly: “I forgot but then remembered that I used something similar in other drawing
programs [1]”. Another one went to help and found instructions. A third one managed to end polygons by clicking on other tools and left them as is (see the picture below):

Figure 1. Example of improperly drawn polygons.

The other three participants, who tried polygons and failed, simply abandoned them and used other tools such as rectangles and brackets.

6. Participants had difficulties understanding Connection and Elbow connection tools. Seven participants tried to draw without other objects present and assumed these are the only tools for drawing arrows. Partially it happened because most participants also had difficulties finding arrows for line/arc tools (“it’s [the interface] is quite easy since you figure out some things. Like I couldn't figure out the arrows” [3]) Participants like this one didn’t notice that straight lines above the score were tools for changing arrow styles (see the picture below):

Figure 2. Example of confusing arrow styles boxes.

Those who found how to change arrow styles commented that they did it accidentally by clicking on every drop-down box on the top: “I finally decided to click through all of them just to see what’s in there” [7].

7. Four participants had difficulties with the Bracket tool. As brackets are downward by default, they couldn’t find how to point them upward. One participant consulted help, another two found a way by clicking and dragging on various places, one left them downward-oriented.

8. Two participants commented that labels (e.g. NCT or PAC labels) were too big and blocked the other labels (“When label boxes are close, I can't see the next one when another is clicked on” [4], “Labels are too big, I couldn't see other chords” [5]).

9. Some participants mentioned that they could e-mail both scores and annotations to their instructors and one of them explicitly said he thought both the score and annotations are saved on the hard drive. Even though it can’t be confirmed with sufficient data, it seems that there is confusion about how annotations and scores are stored (e.g. separately or together, on the server or on the hard drive, what can be saved and how, what would be available later, etc.)
Navigating Variations2 and annotation tools

1. Mostly participants used the “trial-and-error” method, clicking and double-clicking through menu options and buttons. This allowed some of them to accidentally discover useful things, e.g. how to draw polygons or change arrow styles. But sometimes participants were puzzled by results. Thus, one participant accidentally switched the drawing toolbar and the label toolbar (moving the former to the bottom and the latter to the top) and didn’t know how to switch them back. Another participant added bookmarks without any purpose (“I just clicked their; I suppose it does something but I don’t know what… bookmarking scores?” [2]).

2. Two participants consulted the user guide. One participant who couldn’t figure out how to search started looking through user guide pages, went to the sections not related to search and abandoned reading as not helpful. Another participant who needed help for specific tasks was able to find it easily and commented that it was very helpful.

3. Some participants expected be able to make the left panel narrower to enlarge the score viewing area (“It’d be nice to have them larger but to fit into the screen” [3], “I wanted to make it narrower because I don’t need it and it takes quite a bit of the screen” [6]).

4. Some participants expected to change colors/fonts from the menu: “I didn't notice it [the font box] on the toolbar and assumed there must be a way to do it from the menu” [1].

5. One participant commented that he expected the next page to automatically scroll to the top of the page. When a score is zoomed in and is being annotated sequentially, it may be annoying to scroll up at every page.

Printing

1. First four participants sent annotated score from the second task to the printer but nothing was printed out.

Satisfaction ratings

Participants were asked to fill out a satisfaction survey after completing their tasks (see Appendix, p. 20 for questions.) Each question asked users to rate their experience with the system by circling a number on a scale from 1 to 7. The questions asked about general system satisfaction as well as about satisfaction with score annotation tools. Negative adjectives were placed at the low end of the scale (1) while positive adjectives were placed at the high end (7). Thus, a lower mean score for a particular question indicates a lower satisfaction rating. Consolidated results of the satisfaction ratings are presented in the table below. Each response is marked with a “•” in its corresponding rating score column. Means and standard deviations are also shown.
### Table 3. Satisfaction ratings.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Ratings</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, I found Variations2:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terrible</td>
<td>••••</td>
<td>5.63</td>
<td>0.74</td>
</tr>
<tr>
<td>Difficult</td>
<td>• •••</td>
<td>4.75</td>
<td>1.49</td>
</tr>
<tr>
<td>Frustrating</td>
<td>• ••••</td>
<td>4.75</td>
<td>1.91</td>
</tr>
<tr>
<td>Dull</td>
<td>• ••••</td>
<td>5.63</td>
<td>1.06</td>
</tr>
<tr>
<td>Slow</td>
<td>• • •••</td>
<td>5.88</td>
<td>1.13</td>
</tr>
<tr>
<td>Navigating Variations2 was:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficult</td>
<td>•••••</td>
<td>4.88</td>
<td>1.64</td>
</tr>
<tr>
<td>Tasks could be performed in a straight-forward manner:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>• • •••</td>
<td>5.13</td>
<td>1.25</td>
</tr>
<tr>
<td>My location within Variations2 at any given moment was:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never apparent</td>
<td>• ••••</td>
<td>5.5</td>
<td>0.93</td>
</tr>
<tr>
<td>Overall, I found working with scores in Variations2:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terrible</td>
<td>• •••</td>
<td>6.13</td>
<td>0.83</td>
</tr>
<tr>
<td>Difficult</td>
<td>• ••••</td>
<td>5.5</td>
<td>1.69</td>
</tr>
<tr>
<td>Frustrating</td>
<td>• ••••</td>
<td>5.5</td>
<td>1.77</td>
</tr>
<tr>
<td>Dull</td>
<td>• ••••</td>
<td>6.13</td>
<td>0.99</td>
</tr>
<tr>
<td>Slow</td>
<td>• • •••</td>
<td>5.63</td>
<td>1.77</td>
</tr>
<tr>
<td>Navigating within the score annotation window was:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficult</td>
<td>• ••••</td>
<td>4.63</td>
<td>1.69</td>
</tr>
<tr>
<td>The buttons in the score annotation window are:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confusing</td>
<td>• • ••</td>
<td>4.25</td>
<td>1.98</td>
</tr>
</tbody>
</table>

As can be seen from the table, most participants rated Variations2 as satisfying, stimulating and relatively easy. Overall ratings tend to be on the higher half of the scale. However, some participants rated their experience with navigation and tasks as 3; some ratings were as low as 1 or 2. Buttons in the score annotation window were the most dissatisfying, rated by participants on average as 4.25 (ranging from 2 to 7).

To compare some of the overall ratings of Variations2 with the previous studies, see the table below.
Table 4. Mean satisfaction ratings for the current study (February-March 2005) and previous usability studies (conducted in November 2002, July 2003, November 2003 and June 2004.)

<table>
<thead>
<tr>
<th>Mean</th>
<th>Nov-02</th>
<th>Jul-03</th>
<th>Nov-03</th>
<th>Jun-04</th>
<th>Feb-March-05</th>
</tr>
</thead>
</table>

Overall, I found Variations2:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>terrible - wonderful</td>
<td>5.05</td>
<td>6</td>
<td>5.57</td>
<td>4.6</td>
<td>5.63</td>
</tr>
<tr>
<td>difficult - easy</td>
<td>4.6</td>
<td>5.43</td>
<td>5</td>
<td>4</td>
<td>4.75</td>
</tr>
<tr>
<td>frustrating - satisfying</td>
<td>4.1</td>
<td>5.86</td>
<td>4.57</td>
<td>3.6</td>
<td>4.75</td>
</tr>
<tr>
<td>dull - stimulating</td>
<td>5.3</td>
<td>5.86</td>
<td>5.43</td>
<td>5.3</td>
<td>5.63</td>
</tr>
<tr>
<td>slow - fast</td>
<td>4.8</td>
<td>5.57</td>
<td>5.57</td>
<td>2.6</td>
<td>5.88</td>
</tr>
</tbody>
</table>

Navigating Variations2 and its components was:

<table>
<thead>
<tr>
<th></th>
<th>Nov-02</th>
<th>Jul-03</th>
<th>Nov-03</th>
<th>Jun-04</th>
<th>Feb-March-05</th>
</tr>
</thead>
<tbody>
<tr>
<td>difficult - easy</td>
<td>4.3</td>
<td>5.43</td>
<td>5.86</td>
<td>4.4</td>
<td>4.88</td>
</tr>
</tbody>
</table>

Tasks could be performed in a straight-forward manner:

<table>
<thead>
<tr>
<th></th>
<th>Nov-02</th>
<th>Jul-03</th>
<th>Nov-03</th>
<th>Jun-04</th>
<th>Feb-March-05</th>
</tr>
</thead>
<tbody>
<tr>
<td>never - always</td>
<td>4.7</td>
<td>5.57</td>
<td>4.86</td>
<td>5.3</td>
<td>5.13</td>
</tr>
</tbody>
</table>

My location in Variations2 at any given moment was:

<table>
<thead>
<tr>
<th></th>
<th>Nov-02</th>
<th>Jul-03</th>
<th>Nov-03</th>
<th>Jun-04</th>
<th>Feb-March-05</th>
</tr>
</thead>
<tbody>
<tr>
<td>never apparent - always apparent</td>
<td>5.8</td>
<td>6.14</td>
<td>6.57</td>
<td>4.8</td>
<td>5.5</td>
</tr>
</tbody>
</table>

The level of frustration with Variations2 declined in comparison to the studies of June 2004 and November 2003. Overall, it seems that there is no significant change in ratings across the studies.

Via comments, questionnaires and debriefing sessions participants indicated relatively high level of satisfaction with the system as a whole. All of them except one were interested in exploring Variations2 and enthusiastic about learning and using annotation tools. They considered these tools to be useful for music education and asked about availability of the tools in the future. Comments like “I think this is a very exciting program…” or “It’s neat” were quite common. However, four participants doubted that V2 annotation tools have significant advantages compared to hand-and-pencil annotations (“I like it overall. Just not sure what the advantage is to annotating a score on the computer as opposed to doing it by hand. I guess you can send it electronically to your teacher, etc.” [1], “It could be useful but I prefer do it by hand. I used Sybelius at some point and it was very confusing sometimes. This one is better but by hand would be faster and easier” [2])

When asked to compare annotating by hand and using V2, participants mentioned that hand work would be definitely faster but V2 annotations would be more clean and nice-looking. Among other advantages of annotation tools were the possibility to e-mail, print one’s work and make copies as well as the availability of scores online. The main disadvantage was that it required learning and practicing to master.
Recommendations

Recommended improvements for annotation tools component of V2 are listed below and are organized by the importance (with the most important issues being at the beginning of the list). The importance is also indicated as follows:

<table>
<thead>
<tr>
<th>Importance</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[high]</td>
<td>Importance is high. The issue prevented users from making progress or led to severe mistakes</td>
</tr>
<tr>
<td>[medium]</td>
<td>Importance is medium. The issue caused confusion, inefficiency, or minor mistakes</td>
</tr>
<tr>
<td>[low]</td>
<td>Importance is low. The issue causes slight amounts of confusion or dissatisfaction</td>
</tr>
<tr>
<td>[bug]</td>
<td>Technical problem that needs to be fixed</td>
</tr>
</tbody>
</table>

1. **Issue: scores doesn’t print [bug]**
Annotated scores from the task 2 sent to the printer were not printed at all.

**Recommendations**

Fix the bug with printing.

2. **Issue: items don’t stay selected after they’re drawn [high]**
Frustration was caused by necessity of selecting the annotation again after it was drawn to change color or line style.

**Recommendations**

Make the annotation remain selected after it has been drawn. The annotation will become unselected after a user clicks somewhere on the screen (beyond the annotation itself) or chooses another tool.

3. **Issue: tool buttons don’t stay selected after use [high]**
Frustration was caused by the necessity of selecting the same tool every time when using this tool repeatedly.

**Recommendations**

Provide a more visible way to lock tools on (e.g. by providing a special button for locking/unlocking tools or by adding more tool tips for buttons or by providing an appearing/disappearing tip after a figure was drawn.)

– or –

Make the tool remain selected after it has been clicked. The tool can become unselected after a user chooses the selection tool or another tool on the toolbar. However, following this recommendation introduces a new frustration where the user may have a hard time figuring out how to get out of drawing mode and back into selection mode.

---

1 Changing the setting in the previous issue (2) so that the annotation remains selected after it has been drawn can make this issue less problematic (“switching” between the annotation area and the toolbar will become more apparent).
Make the cursor change more dramatically when a tool is selected (e.g. make it look like an arrow contour or like an enlarged “plus” sign or imitate the annotation form). This will allow a user to recognize the mode (drawing or selecting) by looking at the cursor.

4. **Issue: confusing arrow styles boxes [high]**

Straight lines in the boxes for changing line styles prevented some participants from finding ways to create arrows.

**Recommendations**

Use graphical buttons instead of selection boxes to designate functions (the interface will similar to MS Office drawing bar – see the picture below for a mock up). The trade-off, however, is that the current style in use won’t be visible. This lack of visibility heightens the ambiguity over whether changing the style for a drawn object also changes the style for future objects of that (or other) type.

**Figure 3. Mock-up 1 for the “arrow style” boxes.**

– or –

Combine Left arrow style and Right arrow style into one box to save space and provide labels for boxes (see the picture below for a mock-up.) Moreover the line color, highlight color, and line style boxes could all be shortened to make more room for labels.
5. **Issue: confusing connection tools (Connection and Elbow connection) [medium]**
Confusion was caused by the misunderstanding of the functions of these tools².

**Recommendations**
Provide the longer tool tip (e.g. “connection between the two objects”).

– or –

Change icons to look more like a connector (see the picture below, from the MS Office drawing tools, for possible variants).

**Figure 5. Connector icons.**

– or –

Provide a longer description on the status bar at the bottom of the page³ (e.g. “this tool connects two previously drawn objects.”)

---

² Selecting a connection tool doesn’t indicate changes in arrow styles boxes, which is probably a bug. Another potentially confusing issue is that one has to actually draw (by dragging) the connection from inside one annotation to another. If possible, provide visual feedback by highlighting the object the mouse is over and allowing two clicks (one on each object) as an alternate mechanism.

³ This recommendation can be combined with other recommendations. Providing longer explanations in the status bar might also help to reduce the confusion about other tools such as brackets or arrows as well as explain how to select annotations and lock buttons. However, we do not know whether users notice the status bar.
6. Issue: difficulties finding final points of polygons [medium]
Confusion was caused by the misunderstanding of how to end this action.

Recommendations
Provide an appearing/disappearing tool tip after first edge of a figure was drawn (e.g. “double-click to finish” – see the picture below for a mock-up). To make the tool tip less annoying, make it appear once for a short period of time when mouse button is up.

Figure 6. Mock-up for the polygon tool tip.

– or –

If users click on another tool while in polygon-drawing mode, terminate the drawing (as is currently done) but subtract the last change, i.e., don’t assume that the newly-clicked tool location is relevant for the intended polygon shape.

– or –

Provide a right-click menu when in polygon-drawing mode, with one item, “Finish”.

– or –

Allow the Escape key on the keyboard to terminate polygon-drawing mode.
7. **Issue: large labels block other labels or parts of a score [low]**
When two labels are placed close together, the one that is selected blocks another one (see the picture below).

![Figure 7. Labels.](image)

**Recommendations**
Make the label boxes smaller, perhaps by creating a custom control that is less bulky. (?)

– or –

Make drop-down arrows of label boxes half-transparent (again, probably requires a custom control).

8. **Potential Issue: users may be confused by what annotations really are, how to create them, and how to save them [unknown—suspect high]**
This study didn’t address the issue of confusion with “New / Open” options in the menu. All participants searched for items and opened them by clicking on the link or button or by double-clicking on the existing .v2a file. Therefore patterns of using these options and problems associated with it can’t be derived from this study.

The overall model of creating annotations needs more investigation. Some participants mentioned that they could e-mail both scores and annotations to their instructors and one explicitly said he thought both the score and annotations are saved on the hard drive. It seems that there is confusion about how annotations and scores are stored (e.g. separately or together, on the server or on the hard drive, what can be saved and how, what would be available later, etc.) Further investigation is needed to address these issues.

**Limitations of the study and future work**

The limitations of this study worth mentioning are the following.

Even though all participants were asked to articulate their thoughts and feelings during testing, most of them were silent while performing the tasks. During debriefing sessions not all of them were able to remember their motivations and actions and explain it. Due to the lack of sufficient data about cognitive decisions and learning strategies a comprehensive model of using AT couldn’t be reconstructed.

The sample of this study is slightly skewed towards males (5 out of 8 participants) and towards undergraduate students (6 out of 8). This may reduce the ability to generalize results and require further testing of participants with corresponding characteristics.
As mentioned in Issue 8, above, further work is needed to explore users’ mental model of score annotations and their connection to the score, what is saved in an annotation file, and how to create a new annotated score when a score is not already open.
Appendix

Demographic Questionnaire

Please answer the questions below

1. Are you male / female? (Circle one)
2. What is your native language? ___________________
3. What is your current academic involvement at IU (e.g. undergraduate, master, PhD, non-student)? ______________
4. Approximately, how many hours per week do you spend using a computer?
   [ ____ 0-5]  [ ____ 6-10]  [ ____ 11-20]  [ ____ 21 or more]
5. Rate your computer experience on the following systems by circling 1-5 below:
   a. PC: Novice 1 2 3 4 5 Expert
   b. Macintosh: Novice 1 2 3 4 5 Expert
6. Have you used Variations in the IU music library (circle one)?  YES   NO
7. Have you used Variations2 (circle one)? YES   NO
8. Have you participated in any prior Variations2 usability tests (circle one)
    YES   NO
Tasks

This is a computer-based assignment for the music theory class. Complete the tasks described below using Variations2.

1. *Melody identification in a score*

   Piece: Claude Debussy’s “Prélude à l'après-midi d'un faune” (“Prelude to the afternoon of a faun”).

   Highlight the melody (the main theme) in the score using Variations2. You may use different colors for different motives and add text annotations if you wish. If more than one instrument is playing the melody, highlight the one that is most prominent. Mark the first 5 pages.

   When you are done, save the document you created on the desktop as “Debussy_username” (substitute *username* with your username.)

2. *Presentation preparation*

   During the semester we worked on an analysis of Beethoven’s piano sonata Op. 2, no. 1. You will be presenting this analysis in class next week and you need to make it look nice. Use the hand-written analysis (attached) to replicate the first two pages of the annotated score in Variations2. Use different colors. When you are done, save it on the desktop as “Beethoven_username” (substitute *username* with your username).

3. *Choose the correct answers on the score*

   On the desktop of your computer there is a file “Example3.v2a”. You need to look at the score and identify the key, harmonies, and non-chord tones in the piece by selecting the correct options. Make sure you select the answers for all of the missing elements.

   When you are done, save the document on the desktop as “Local_username” (substitute *username* with your username).
**Satisfaction Survey**

For each question below, circle a number from 1-7. The number should best represent your feelings about the Variations2 experience during testing. Feel free to write additional comments in the space provided below.

1. **Overall, I found Variations2:**
   - Terrible
   - Difficult
   - Frustrating
   - Dull
   - Slow
   - Wonderful
   - Easy
   - Satisfying
   - Stimulating
   - Fast

2. **Navigating Variations2 was:**
   - Difficult
   - Easy

3. **Tasks could be performed in a straight-forward manner:**
   - Never
   - Always

4. **My location within Variations2 at any given moment was:**
   - Never apparent
   - Always apparent

5. **Overall, I found working with scores in Variations2:**
   - Terrible
   - Difficult
   - Frustrating
   - Dull
   - Slow
   - Wonderful
   - Easy
   - Satisfying
   - Stimulating
   - Fast

6. **Navigating within the score annotation window was:**
   - Difficult
   - Easy

7. **The buttons in the score annotation window are:**
   - Confusing
   - Clear

**Additional Comments (use back of page if you need more space):**

**Debriefing Questions**

1. How do you feel about this experience?
2. Tell me about the tasks, — how do you feel about them?
3. What do you think about working with the score in Variations2? Do you think such tools can be useful?
4. Comment on the annotation tools interface.
5. Comment on help (if used).
6. Would you like to use Variations2 for annotating scores in the future? Why (and for what purposes) or why not?
7. What do you see as the advantages and disadvantages of working with scores on the computer versus working with them on paper? Which do you prefer, and why?
8. Do you have any other questions / comments about Variations2 or your experience with it?
Examples of the Task Decisions

Figure 8. Task 1, page 1

![Screenshot of the View and Details of a Digital Sheet Music Score](image-url)
Figure 10. Task 1, page 3
Figure 13. Task 2, page 1
Schmücke dich, o liebe Seele.