Analysis and User Evaluation of a Musical-Visual System: Does music make any difference?

Suresh K. Lodha **Doug Whitmore** **Marc Hansen** **Erik Charp**
Department of Computer Science
University of California, Santa Cruz
Santa Cruz, CA 95064
{lodha,whitmore,mhansen,echarp}@cse.ucsc.edu

ABSTRACT
We describe, evaluate and analyze an integrated musical visualization system for assessing protein structural alignments. Superpositions of protein structures in three-dimensional space are visualized using the molecular graphics program Rasmol. Four environment parameters are examined: secondary structure, polarity, exposure, and goodness-of-fit. It is difficult to assess these parameters using visualization alone. Therefore, we employed melodic components with unique characteristics to convey these four parameters to the user. We used basic music theory (arranging, voice leading, development of melodic phrases, etc.) as the basis for sonification parameters. We attempted to maximize the individuality of the sonification elements by employing sound effects such as panning a voice to the left or right speaker and parameter adjustment such as changing its volume.

To validate the utility of our system, we conducted an experiment to evaluate the performance of users in estimating the value of these four variables under three distinct modes: visual, musical, and visual+musical presentation. The preliminary raw results of our experiments were reported in an earlier work [4]. We also conducted experiments when all the four variables were played together in a symphony-like fashion to assess the impact of presenting several variables simultaneously. Raw results seemed to indicate that the visual+musical delivery is more effective than the visual delivery alone in most cases.

In this work, we present the results of statistical tests and their implications. We found that the accuracy performance was statistically significantly better in the Audio Only mode in comparison to the Visual only mode both in the case of single variable presentation and the multiple variable presentation.

Keywords: audio, evaluation, statistics, sonification, visualization.

INTRODUCTION
Protein structural alignment is a useful application for multimodal information perceptualization because researchers in this area typically use several independent measures to assess their alignments. A system that allows the user to simultaneously evaluate different criteria will therefore be quite useful. Proteins are composed of amino acids connected somewhat like beads on a string. In protein structural alignments, protein folds are superimposed in three-dimensional space in such a way as to maximize their overlap. The protein structure to which others will be aligned is termed the “parent”, and the aligned structures are termed “children”. In this work, we identified foru variables that measure how well the two proteins can be aligned. The four variables are:

1. *Secondary structure or local folding pattern.* This pattern can be one of the three types – helix, loop, or beta sheet. This variable is a nominal variable. Visually the secondary structure is represented by a cartoon depiction similar to the traditional Richardson ribbons. Helices are shown as corkscrews, sheets as flat ribbons with arrowheads, and loops as thin spaghetti–like squiggles. This variable is usually easy to detect visually.

2. *Amount of exposure or nearness to the outside.* This variable is an ordinal variable and can take one of the three values – buried, partially buried, and exposed. In a visualization environment, the value of this variable is typically assessed by rotating the molecule in the three dimensions and estimating how near the surface a position is.

3. *Polarity or amount of charge in the local environment.* This variable is also an ordinal variable and can take one of the three values – low, medium, and high polarity. Polarity is not shown explicitly. More exposed positions are more polar because water is a polar molecule, so polarity and exposure are related. There may, however, be positions on the inside of the molecule that are polar even though they are buried. This task is difficult using vision only.

4. *Goodness of fit.* This variable is continuous and can take any value between 0 and 100. For the purposes of this study, we classified the goodness of fit into one of the three categories – poor, medium, and good. Goodness of fit is also not

shown explicitly. Regions in the middle of ordered secondary structure sections would probably have better goodness of fit scores than loop regions or positions on the edges of ordered secondary structure. This task is extremely difficult using vision alone.

The visualization system used for visualizing sequence alignment of two protein structures RASMOL is quite popular in the molecular graphics community. However, we felt that this system alone could be improved by conveying some of these variables through sonification. Although we could have used synthesized sounds to convey these variables, we wanted to use musical sounds to make the presentation more pleasant and acceptable to the users. The key challenge was whether we could devise musical mappings that could still convey the information accurately to the user. We were also interested in assessing the performance of the delivery of musical information when all the four variables were played simultaneously.

In order to assess the effectiveness of our musical+visual system, we conducted a user evaluation study, the results of which are discussed later in this report.

PREVIOUS WORK

Sonification presents several opportunities to researchers. Kramer [6] presents a good discussion of pros and cons of utilizing sound for conveying scientific information. Our motivation for using sound was the hope that the data represented visually may be enhanced or disambiguated by adding sound to the presentation. This aspect of data representation is particularly important for showing features that are difficult to represent visually, due to occlusion or other factors. Another feature of our system is that by representing some variables through sound and others visually, the amount of information that may be represented simultaneously may be extended. Our tool aims to augment the power of data visualization rather than replace it.

There is a long history of utilizing synthesized sounds for augmenting visualizations. The collection of articles edited by Kramer [5] provides a good sampling of early work in this area. In our work, we employed musical voices and melodic components with unique characteristics. We also used sound effects such as panning a voice to the left or right speaker and changing its volume to maximize the individuality of the sonification elements. By making the sonification parameters distinct, we allow the user to focus on those portions of the sonification necessary to resolve possible ambiguities in the visual display.

In contrast to several efforts on utilizing synthesized sounds including our own [10], efforts for using musical sounds are much more limited. Musical sounds have been utilized in a number of applications by the NCSA audio development group at the University of Illinois, Urbana Champaign. They have used musical sounds for studying dynamical systems [8], exploration of biological data sets in CAVE [3], and understanding alpha-shapes [2]. We utilized musical sounds by mapping data to timbre, rhythm, tempo, melody and harmony and applied it for the understanding of uncertainty in scientific data [7]. An early version of this work was presented elsewhere [4], which also contains details on previous work on the utilization of sound and music in representing bio-sequences, including DNA. Musical sounds have also been used by Alty et al. in debugging software and in presenting complex graphical information [1, 9].

MUSICAL-VISUAL SYSTEM

The visualization of protein alignments is done by using the RasMol molecular graphics program. The protein is displayed in cartoon mode to explicitly show secondary structure For structural alignments, the cartoon display mode is useful because it allows the overall features of the protein to be shown with minimal clutter. We use red highlighting to indicate which section of the protein is being examined. The remainder of the protein is colored light gray.

Each residue environment is mapped to a musical interval of eight measures (lasting about 20 seconds). The overall tempo of the music is constant. We use the musical qualities of melody, rhythm, timbre and dynamics to create mappings of music to the values of local environment variables. The auditory mappings are based on the idea of musical parts. Since we are sonifying four parameters, we based our arrangements on a typical jazz quartet consisting of a solo instrument to play the melody line, a drum part, a bass line, and a harmonic comp part (i.e., a rhythmic accompaniment consisting of the chords of the piece played on a keyboard instrument or guitar).

The mapping of the four data parameters – secondary structure, polarity, exposure and goodness-of-fit – are mapped as described in Table 1. We made these choices with careful consideration of the association between the sound and the data semantics.. For example, an exposed environment produces a brighter, sharper and busier drum part. In contrast, partially buried and buried environment sounds more dull. Similarly, an environment with low polarity results in a dull, softer, and more sparse piano part. Higher polarity sounds energetic and charged. We chose the melody line for goodness-of-fit since this is our primary variable and we expected the users to pick this out easily in multi-variable presentations. We mapped the bass line to secondary structures because secondary structure appears for several contiguous residue positions and the ear can tolerate a repeated bass pattern easily.

Considering that one can present none, one, two, three or all four of the variables at any time, and that there are 4 possible values (including absence of that variable) for each of the four variables, there are 256 different combinations possible. We sampled 8 of these combinations and decided on appropriate parameter levels for volume and pan for each univariate sonification in order

parameter	*data value*	*part*	*voice*
Secondary Structure	Sheet Helix Loop	Bass	acoustic bass twangy bass slap bass
Polarity	Low Medium High	Comp	electric piano marimba electric guitar
Exposure	Buried Partially Buried Exposed	Drums	brush cymbals full kit
Goodness of Fit	Poor Medium Good	Melody	trumpet saxophone synthesizer

Figure 1: Musical Mapping [4]

to make each sound distinguishable in a symphony-like presentation when all the four sounds are played simultaneously. We found that even at the same volume level, some voices sounded louder than others. To compensate for this we set individual volume levels for all the instruments. Some instruments (such as the saxophone) entered too softly, causing us to increase their note attack rates as a countermeasure. We also noticed that it was difficult to pick out individual parts when all the voices were sent to both the left and right audio channels. To correct this problem we moved the bass line to the left speaker and the comp part to the right. The melody tended to get lost in the comp part, so we moved it to the left where it could stand against the bass line. The drum part was relatively easy to hear, so we let it come out of both channels equally. To reiterate, we placed an emphasis on making the musical nature of these patterns both strongly distinguishable, and suggestive of the values they represent. The univariate sounds were then superimposed over each other to generate all 256 combinations.

The musical-visual system is running on a Silicon Graphics Octane running the IRIX6.4 operating system. The workstation contains a 195 MHz MIPS R10000 processor and 128 MB of RAM. Attached to the workstation is a 19 inch color monitor also from SGI. The headphones we use are Sony Digital headphones, model MDR-V6. To create our data visualization we have used RasMol version 2.6.

EXPERIMENTAL EVALUATION

In order to verify the utility of our program for enhancing visual representations of data, we conducted a controlled experiment on a total of 18 subjects. Subjects were given headphones. Subjects pushed a "play" button to cause the program to simultaneously present a data sonification and its corresponding visualization. Each of the four sonification parameters: bass, drums, comp, and melody, were presented to each subject in random order using a latin square design. The use of a latin square allowed us to minimize any systematic learning effects that might have otherwise appeared (for example if subjects consistently perform better on the drums tests only after having undergone the bass tests immediately prior). For each of the four sonification parameters the three allowable data values were presented in order from first to last. The appropriate radio buttons were lit to indicate which data values were currently being presented. The entire process was repeated twice for each of the four sonification parameters. The subject responses were recorded, as was the length of time taken to answer.

The structural alignment consisted of the G chain of lobster D-glyceraldehyde-3-phosphate dehydrogenase (1gpd-G) superimposed on the salmonella typhimurium strain LT2 galactose-binding protein (1gca). The alignment was obtained from the FSSP database. This particular alignment was chosen because the parent protein is an example of an alpha/beta structure, thereby allowing us to test all three of our secondary structure mappings. Also, since the sequences only have 56% identity, the alignment contained a nice spread of goodness-of-fit scores.

Each subject trial took about 45 minutes. Appropriate training was provided before the actual administration of tests. The test consisted of two parts– univariate musical presentation and multivariate musical presentation. The univariate musical presentation consisted of 12 stimuli in which each of the four variables could assume any one of the three values. These stimuli were presented in all the three modes – audio, visual and audio+visual. The multivariate musical presentation consisted of 8 stimuli in which all the four variables were sonified simultaneously. The 8 chosen combinations were: 1223, 3112, 3111, 2332, 3211, 1313, 3221, and 3333, where each digit in the above combination shows the value associated with the four variables in the following order – secondary structure, polarity, exposure and goodness of fit. An exit questionnaire was given to the subjects to determine whether the mappings were acceptable and intuitive.

ANALYSIS and DISCUSSION

Although a summary of raw results of the user evaluation tests was presented in [4], no statistical analysis was done. Here we present results of statistical analysis of the various results obtained and discuss their implications.

We have analyzed the results of our experiments by doing pairwise t-tests. We decided against using an ANOVA test as the results were boolean in nature. Since each trial resulted in either an accurate response or an inaccurate one, it is not meaningful to compare the variances of these tests. Therefore, we did pairwise comparisons of the tests both within each experiment and between experiments for both single and multi–variable experiments. In all of our analyses, we used a confidence value of 0.005 because of our large sample size. All sample sizes were greater than 144 samples.

Single Variable Experiments

We begin by reproducing the average accuracy scores for each of the four variables in each of three modes – visual, audio and audio+visual. A cursory glance indicates that the accuracy scores are much higher for the audio and audio+visual modes than for the visual modes (see Figure 2). The raw scores are presented in Figure 3

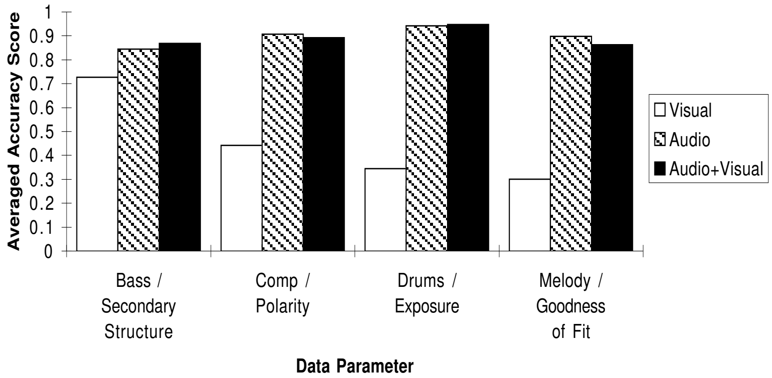

Figure 2: Accuracy results for the four data parameters in visual, audio, and audio+visual modes when parameters were presented one at a time [4]

Now, we describe the results of statistical tests for each of the four data variables separately. For each of the four parameters – secondary structure, polarity, exposure, and goodness of fit – we conducted three standard pairwise t-tests: (i) between audio and visual modes, (ii) between visual and audio+visual modes, and (iii) between audio and audio+visual modes.

For each of the four variables, we found that the accuracy of Audio Only mode was better than the Visual Only with a p-value $>$ 99.5%. Although this result may be expected from Figure 2, the significance of result even in the case of secondary structure is somewhat surprising. We also found that Audio+Visual mode was also significantly better than the Visual only mode in each of the four cases. Moreover, in each of the four cases, there was no significant difference between Audio Only and Audio+Visual. From these results, one can safely conclude that (i) the musical mappings did not interfere in data comprehension, and (ii) the musical mappings enhanced the accuracy of data comprehension in comparison to the visual display both by itself and in the presence of the visual cues.

We also conducted the pairwise t-tests on the accuracy scores after summing the results for all the four individual parameters. In this case as well, we found that the differences in accuracy were significant between Audio only versus Visual only, with Audio only performing better with a p-value $>$ 99.5%. Visual Only versus Visual + Audio tests showed that Audio+Visual accuracy was again significantly higher with a p-value $>$ 99.5%. The differences in accuracy between Audio only and Audio+Visual were not significant with a p-value $<$ 99%.

Parameter	Visual	Audio	Audio+Visual
Secondary Structure	98	128	134
Polarity	96	140	136
Exposure	32	144	134
Goodness of Fit	38	132	122
Total	264	540	536

Figure 3: Raw Accuracy Results for each of the modes – Visual Only, Audio Only, and Audio + Visual for the Single Variable Cases

Pair Compared	alpha	P-value
Visual Only vs Audio Only	22.51	> 99.5%
Visual Only vs Visual + Audio	21.48	> 99.5%
Audio Only vs Visual + Audio	-2.24	< 99%

Figure 4: Overall significance of differences for each of the experiment types – Visual Only, Audio Only, and Audio + Visual for the Single Variable Cases

Multi-Variable Experiments

We now present and analyze the results for the multi-variable experiments when all the four data variables were sonified simultaneously.

We begin by reproducing the results of our experiments [4]. One would expect some degradation in accuracy performance due to more information being condensed during the same time. Without the benefit of statistical analysis, we observed that "We were pleasantly surprised to see no such drop for the melody parameter. The drop in accuracy for the other parameters were not nearly as steep as we had feared." Here, we present the results of our statistical tests and discuss their implications.

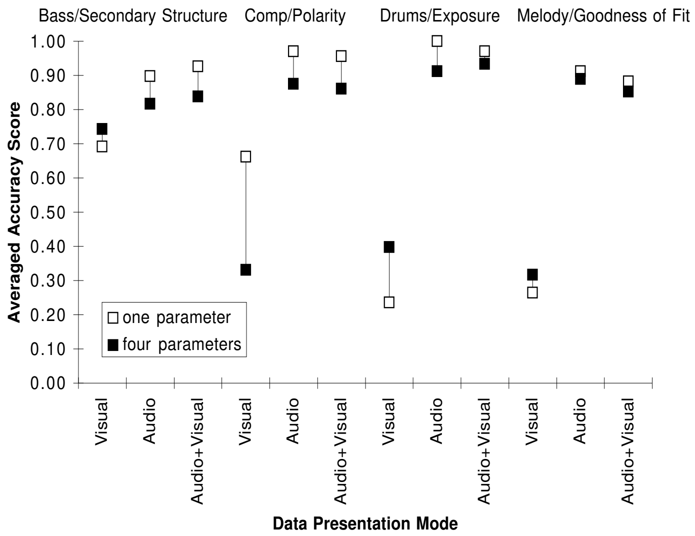

Figure 5: Accuracy results for the four data parameters in visual, audio, and audio+visual modes: Single and Multi-Variable Experiments Compared [4]

Interestingly, even in the multi variable case, the results are again statistically significant. When comparing Audio only versus Visual only, Audio only accuracy was significantly better with a p-value $>$ 99.5%. Again, when comparing Audio+Visual versus Visual only, the Audio+Visual accuracy was also significantly better with a p-value $>$ 99.5%. The difference in the accuracy between the Audio only and Audio+Visual was even less in multi variable presentations with a p-value of $<$ 80%. These results again suggest that musical mappings not only did not interfere with the visualization process, but did better than the visual mode in accurate data comprehension even when four parameters were sonified simultaneously (Figure 6.)

Pair Compared	alpha	P-value
Visual Only vs Audio Only	23.42	> 99.5%
Visual Only vs Visual + Audio	22.84	> 99.5%
Audio Only vs Visual + Audio	-0.65	< 80%

Figure 6: Overall significance of differences for each of the experiment types – Visual Only, Audio Only, and Audio + Visual for the Multiple Variable cases

Experiment type	alpha	P-value
Visual Only	1.95	$< 97.5\%$
Audio Only	-8.89	$> 99.5\%$
Audio + Visual	-6.84	$> 99.5\%$

Figure 7: Overall significance in differences between Single Variable experiments versus Multi-Variable experiments

Single Variable vs. Multi-Variable Experiments

Interesting results were obtained when comparing results between single and multi-variable case for the three modes. As one would expect, for Visual only, there was no significant difference between single and multi-variable case. However for both Audio and Audio+Visual, there was a significant decrease in the ability to correctly visualize the data with a P value $> 99.5\%$ (see Figure 7.) This points to the problem of overloading the information through the audio channel.

However, it is still worthy of note that the multi-variable sonifications still did better than Visual only for single variables (or Visual only for multi-variables). So users are still able to gather more accurate information from the musical-visual multi-variable environment than from the visual environment alone.

Surprisingly, the non-significance of the Audio-only vs Audio+Visual in the summed tests suggests that audio may be a good replacement for the visual in this case. Our system and experiment was designed and intended to use sound not as a replacement but as a means of providing supplementary information to the user.

Based on these results, musical sonification appears to have a useful role in disambiguating data which may be unclear if only presented visually. More experimentation is needed to determine the number of variables that can be presented through audio channel without significant degradation from univariate to multivariate case. Our results also indicate that musical sonifications may work effectively as a replacement for visual environment in some applications.

ACKNOWLEDGMENTS

We want to thank Alex Pang and Doanna Meads for helping us with the visualization component. We would like to than James Bowie and the UCLA-DOE lab of Structural Biology and Molecular Medicine for allowing us to use their protein environment analysis software: *Environments*. This research is partly supported by DARPA grant N66001-97-8900, ONR Grant N00014-96-1-0949, NSF grant IRI-9423881, NASA grant NCC2-5281, and DOE Grant B347879.

REFERENCES

1 J. Alty and D. Rigas. Communicating graphical information to blind users using music: The role of context. *Proceedings of CHI*, pages 574–581, 1998.

2 U. Axen and I. Choi. Investigating geometric data with sound. In *Proceedings of the Third International Conference on Auditory Display, Palo Alto*, pages 25–28. 1996.

3 R. Brady, R. Bargar, I. Choi, and J. Reitzer. Auditory bread crumbs for navigating volumetric data. In *Proceedings of the Late Breaking Hot Topics of IEEE Visualization '96*, pages 25–27. IEEE Computer Society Press, 1996.

4 M. Hansen, E. Charp, S. K. Lodha, D. Meads, and A. Pang. Promuse: A system for multi-media data presentation of protein structural alignments. *Proceedings of the Pacific Symposium on Biocomputing*, January 1999.

5 G. Kramer. *Auditory Display, Sonification, Audification, and Auditory Interfaces*. Addison-Wesley, 1994.

6 G. Kramer. An introduction to auditory display. In Gregory Kramer, editor, *Auditory Display, Sonification, Audification, and Auditory Interfaces*, pages 1–78. Addison-Wesley, 1994.

7 S. K. Lodha, T. Heppe, J. Beahan, A. Joseph, and B. Zane-Ulman. Muse: A musical data sonification toolkit. *Proceedings of the International Conference on Audiotry Display (ICAD)*, pages 36–40, November 1997.

8 G. Mayer-Kress, R. Bargar, and I. Choi. Musical structures in data from chaotic attractors. In Gregory Kramer, editor, *Auditory Display, Sonification, Audification, and Auditory Interfaces*, pages 341–368. Addison-Wesley, 1994.

9 P. Vickers and J. Alty. A musical program auralisation tool to assist novice programmers with debugging. *Proceedings of the International Conference on Audiotry Display (ICAD)*, November 1996.

10 C. M. Wilson and S. K. Lodha. LISTEN: a data sonification toolkit. In *Proceedings of the Third International Conference on Auditory Display, Palo Alto*, pages 35–40. 1996.