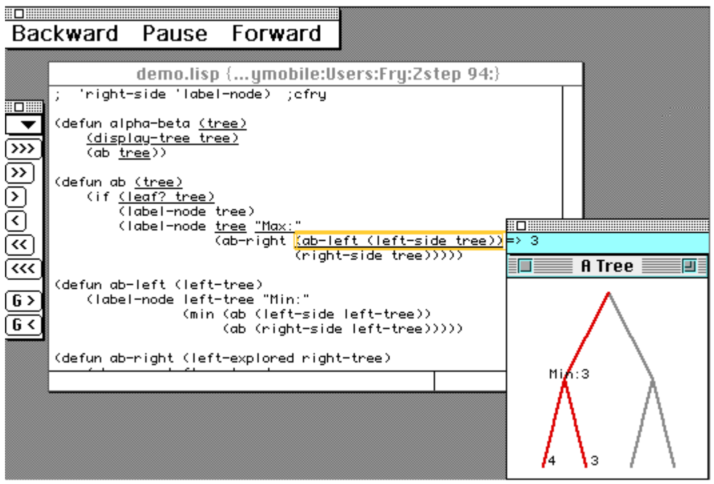
Bridging the Gulf Between Code and Behavior in Programming

Henry Lieberman
Media Laboratory
Massachusetts Institute of Technology
Cambridge, Mass. USA
lieber@media.mit.edu

Christopher Fry
Harlequin, Ltd.
1 Cambridge Center
Cambridge, Mass. USA
cfry@harlequin.com



ABSTRACT

Program debugging can be an expensive, complex and frustrating process. Conventional programming environments provide little explicit support for the cognitive tasks of diagnosis and visualization faced by the programmer. *ZStep 94* is a program debugging environment designed to help the programmer understand the correspondence between static program code and dynamic program execution. Some of ZStep 94's innovations include:

- An animated view of program execution, using the very same display used to edit the source code
- A window that displays values which follows the stepper's focus
- An incrementally-generated complete history of program execution and output
- "Video recorder" controls to run the program in forward and reverse directions and control the level of detail displayed
- One-click access from graphical objects to the code that drew them
- One-click access from expressions in the code to their values and graphical output

KEYWORDS: Programming environments, psychology of programming, debugging, educational applications, software visualization

INTRODUCTION

Debugging accounts for about half of the estimated $100/line cost of a programmer's time [13], a major expense in the $92 billion US software market [6]. While the interface community has directed much attention toward improving interfaces for end users of applications, surprisingly little attention has gone toward improving the human interface of program debugging tools. Applying widely recognized human-computer interface principles to the problem can result in dramatic improvements in the effectiveness of debugging interfaces.

Donald Norman [12] refers to the *Gulf of Evaluation* and the *Gulf of Execution* as the gaps that often occur between the user's intent, the actual effect of a command given to the computer, and the result. In programming, this manifests itself as the difficulty of understanding the dynamic behavior of a program from the static source code of the program. The primary cognitive task in debugging is forming a mental model of the correspondence between code and behavior.

SUPPORTING THE COGNITIVE TASKS IN PROGRAMMING

Programming is the art of constructing a static description, the program code, of a dynamic process, the behavior which results from running the program. In that sense, it is analogous to composing music. The program code is like a musical score, whose purpose it is to cause the performer [in the programming case, a computer] to perform a set of actions over a period of time.

What makes programming cognitively difficult is that the programmer must imagine the dynamic process of execution while he or she is constructing the static description, just as a composer must "hear the piece" in his or her head, while composing. This puts a great burden on the programmer's short term memory. What makes programming even more difficult than composing is that a musical composition usually specifies a single performance, whereas a program may be executed in a wide variety of conditions, with different resulting behavior.

Many, if not most, routine program bugs result from a discrepancy between the programmer's imagining of the desired behavior in a given situation, and the actual behavior of the program code in that situation. [This would not be true only in the case of deep conceptual misunderstandings, where the program may actually perform as originally intended, but the programmer realizes that the original intention does not solve the problem.]

For the programmer, the problem of translating intent into program code corresponds to what Norman calls the *Gulf of Execution*. Interactive tools such as on-line context-sensitive help systems and syntax-directed editors can provide intelligent assistance in bridging that gap by relieving reliance on the programmer's memory of programming language details.

Once program code is written, the problem remains of verifying that the code written actually expresses the programmer's intent under all circumstances of interest. This is the *Gulf of Evaluation*. Interactive tools such as debuggers and program visualization systems can be invaluable in bridging that gap. Instead of trying to imagine how the events in a program unfold over time, why not have the machine show them to you?

We have designed a program debugging environment to explicitly support the problem solving methodology of matching the expectations of a programmer concerning the behavior of code to the actual behavior of the code. This environment is called *ZStep 94*, a descendent of the stepper described in [8].

PROBLEM SOLVING PROCESSES IN DEBUGGING: LOCALIZATION AND INSTRUMENTATION

Two principal activities in debugging that can be assisted by tools in the programming environment are *instrumentation a*nd *localization*. Instrumentation is the process of finding out what the behavior of a given piece of code is, the software analog of attaching oscilloscope probes to a hardware component. Traditional tools that assist instrumentation are trace, breakpoints, and manually inserted print statements: trace instruments all calls to a function, a breakpoint or print statement instruments a specific function call. The problem with using trace and breakpoints in debugging is that they require some plausible hypothesis as to where the bug *might* be, so you know where to place the instrumentation. They are not of much help when you have no idea where a bug might be, or there are too many possibilities to check individually.

Localization is the process of isolating which piece of code is "responsible" for some given undesirable behavior of a buggy program, without any prior knowledge of where it might be. Among traditional tools, a *stepper* is potentially the most effective localization tool, since it interactively imitates the action of the interpreter, and the program can in theory be stepped until the error is found.

However, traditional steppers have a fatal interface flaw: they have poor control over the level of detail shown. Since the programmer's time is too valuable to make looking at every step of an evaluation feasible, only those details potentially relevant to locating the bug should be examined. Typical steppers stop before evaluation of each expression and let the user choose whether or not to see the internal details of the evaluation of the current expression. But how can the user make the decision about whether to see the details of an expression if he or she doesn't know whether this expression contributes to the bug or not? This leaves the user in the same dilemma as the instrumentation tools -- they must have a reasonable hypothesis about where the bug might be before they can effectively use the debugging tools!

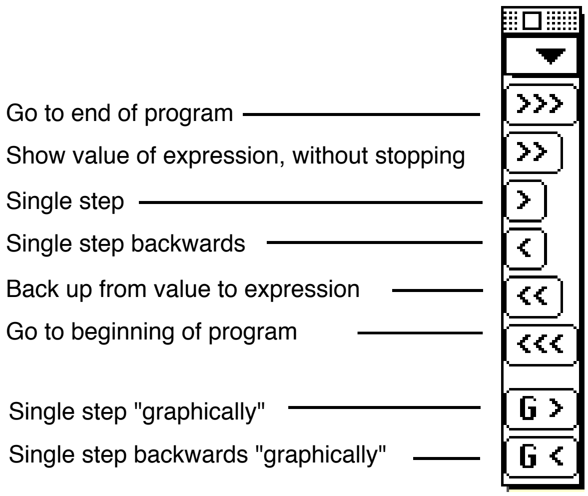
20-20 HINDSIGHT: REVERSIBLE CONTROL STRUCTURE

The solution adopted by ZStep 94 is to provide a reversible control structure. It keeps a complete, incrementally generated history of the execution of the program and its output. The user can confidently choose to temporarily ignore the details of a particular expression, secure in the knowledge that if the expression later proves to be relevant, the stepper can be backed up to look at the details. Thus, ZStep 94 provides a true localization tool.

There has been a considerable amount of past work on reversible program execution [1,3,8,9,11,15], but this work has concentrated on the details of minimizing the space requirements of the history and tracking side effects to data structures, both of which are important, but secondary to the user interface aspects which are the emphasis of this paper. It is important not only to "back up" variables to their previous values, but also to "back up" a consistent view of the user interface, including static code, dynamic data, and graphical output so that the user "backs up" their mental image of the program execution.

The reversible control structure aspects of ZStep 94 are discussed in more detail in [9]. We will address the issue of the computational expense of the history-keeping mechanism later.

ZStep 94's main menu uses a bi-directional "video recorder" metaphor. The single-arrow "play" and "reverse" correspond to single-step in a traditional stepper, and the "fast-forward" and "rewind" operation go from an expression to its value and vice versa, without displaying details.

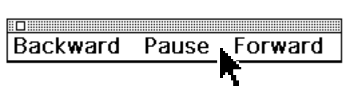


ZStep 94's "control menu"

It is important to note that ZStep 94's expression-by-expression stepping is *not* the same as statement-by-statement or line-by-line stepping found in many steppers for procedural languages. Individual lines or statements may still contain complex computations, and it is a severe limitation on the debugger if the granularity cannot be brought down finer than a statement or line.

The two "graphic step" operations G> and G< are an innovation that lets the user step the *graphic output of the program* back and forth, rather than control the stepper in terms of the expressions of the code. This will be discussed below.

ZStep 94 also has a "cruise control" mode, in which the stepper can run continuously, in either direction, without user intervention. The distance of the cursor from the center of the panel controls the speed. The user can stop it at any point when an interesting event appears, and run the stepper in either direction at that point.



ZStep 94's "cruise control"

KEEPING THE DEBUGGING PROBLEM IN CONTEXT

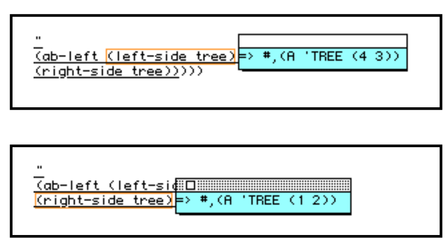
A crucial problem in designing an interface for program debugging is *maintaining the visual context*. Because programming is an activity in which many items of interest have complex temporal and spatial relationships to other items, it is important to present each item with its context clearly identified.

Items such as the expression currently being evaluated, the value of a variable, or graphics drawn by the code may have almost no meaning outside their proper context. The programmer wants to know *where* in the code that expression was evaluated, *which* instance of the code was it, *when* did the variable have that value, *how* did that graphic appear on the screen?

If the item and its visual context are spatially or temporally separated, a new cognitive task is created for the user -- matching up the item with its context. This new cognitive task creates an obstacle for debugging and puts additional burden on the user's short term memory. Linear steppers or tracers that simply print out the next expression to be evaluated create the task of matching up the expression printed to the place in the code where that expression appears. "Follow the bouncing ball" interfaces that point to an expression and print out a value in another window lead to "ping-ponging" the user's attention between the code display and the value display.

FOLLOW THE BOUNCING WINDOW

Because most programmers input their code as text in a text editor, the primary mental image of the program becomes the text editor's display of the code. Thus, to preserve the WYSIWYG property, ZStep 94 always use the text editor's display to present code during debugging. To maintain visual continuity, it is important that the exact form of the user's input be preserved, including comments and formatting details, even if they are not semantically significant. As the interpreter's focus moves, the source code of the expression that is about to be evaluated, or has just returned a value, is highlighted.



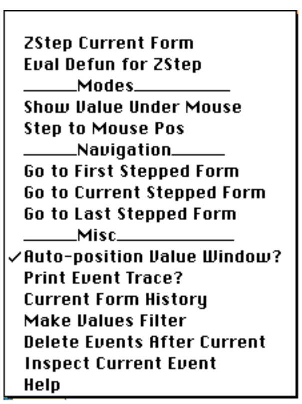
The value window moves through the code

Steppers always have the problem of how to show code expressions and their values simultaneously. In ZStep 94, as the editor's focus moves from expression to expression, we use a floating window to display the value. The display of the value is always exactly aligned with the expression to which it corresponds, so that the visual association between an expression and its value is apparent. The floating value window is colored light green to indicate if the expression is about to be evaluated, light blue to indicate a return value, or yellow if it has caused an error.

Earlier versions had used the idea of *substituting* the value for the code in place. This keeps the user's attention focused on the point of execution, but loses the original expression. Another version maintained two windows, one with the original code, the other with the substituted values, but this was affected by the "ping-pong" problem. All these are valid approaches, but we prefer the floating value window as the best compromise between visibility and maintaining visual continuity. We might also explore making the values window *translucent*, to reduce the effect of obscuring the code underneath.

WHAT DID THAT CODE DO?

Our approach is to integrate instrumentation tools directly into the stepper. We provide two facilities for pointing at a piece of code and inquiring about the behavior of the code. Rather than inserting breakpoints or print statements into the actual code and resubmitting the code to the interpreter or compiler, we let the user simply point at the desired expression, then run the stepper until that expression is reached. This is called *Step to Mouse Position*. This is like a breakpoint, but an advantage is that the stepper is runnable both forward and backward from the point where the program stops, and all information about the computation remains available.

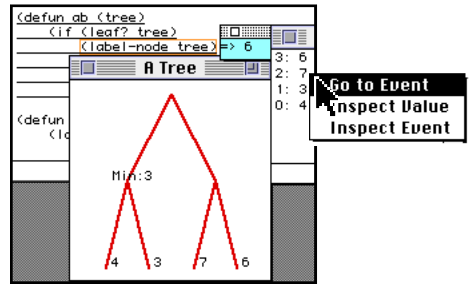


ZStep 94's pull-down menu

Even more dynamic is *Show Value Under Mouse*, which is like a continuously updated Step to Mouse Position. The user simply waves the mouse around the code, without clicking, and the expression currently underneath the cursor displays its value window. Unlike Step to Mouse Position, this works only for values that have been previously computed and are quickly retrievable, and does not run the stepper past the current execution point.

WHAT HAS THAT CODE DONE?

We also provide a facility to track the behavior of a given expression over different execution histories. The operation *Current Form History* allows the user to point at an expression and bring up a menu of the past values of that expression.



The history of values of an expression

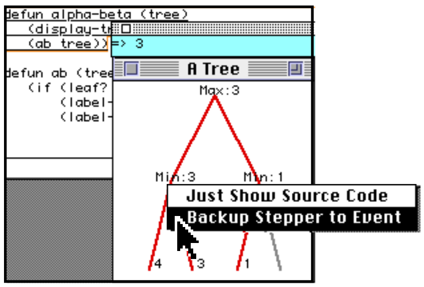
Another history facility is the *Values Filter*, which brings up a menu of all returned values up to that point satisfying a condition. Clicking on one of the values returns the stepper to the corresponding event. We could also provide a filter on the expression executed, which would correspond to a traditional trace.

WHAT CODE DID THAT?

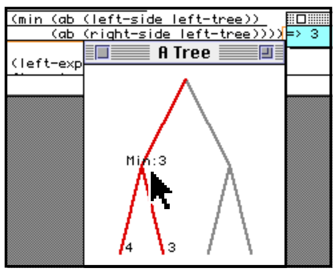
One of the most essential, but also most difficult, tasks in debugging is being able to reason backward from the manifestation of some buggy behavior to the underlying cause. This is especially problematic when the program in question itself has a graphical user interface. The programmer must work backwards from an incorrect user interface display to the code responsible. Traditional tools do not make any special provision for debugging programs with graphic output; worse, the user interface of the debugger often interferes with the user interface of the target program itself, making it impossible to debug!

ZStep 94 maintains a correspondence between events in the execution history and graphical output produced by the current expression. Considerable care is taken to assure that the graphic output always appears consistent with the state of execution. When the stepper is run forward or backward to a certain point in the execution, the graphic display is also moved to that point.

Furthermore, individual graphical objects on the display also are associated with the events that gave rise to them. We allow the user to click on a graphical object, such as a tree node in our example, and the stepper is automatically positioned at the event which drew that node. Just as in our other operations like Step to Mouse Position, the stepper is active at that point, and the program can be run forward and backward from that point.



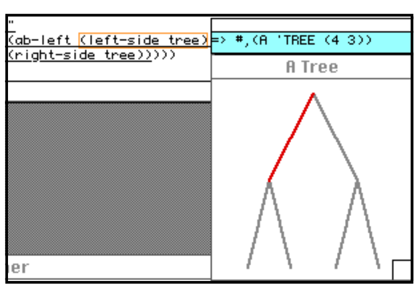
Clicking on a graphical object backs up the stepper to the event which drew it

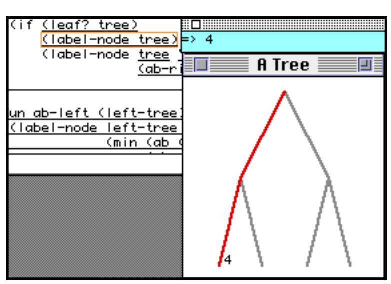


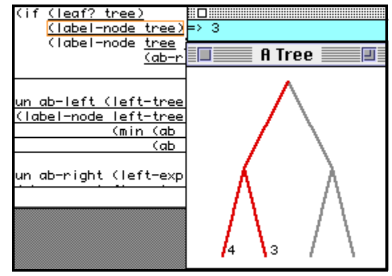
LET'S SEE THAT AGAIN, SLOWLY

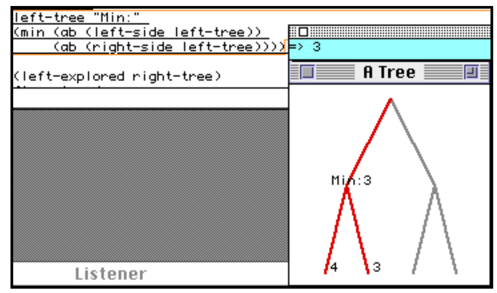
In reasoning from the behavior of the program to the code, it is useful to be able to *step the behavior* rather than step the code. The user conceptualizes the behavior of the program as a set of graphic states that unfold over time, as the frames of an animation. The increments of execution should be measured in terms of the animation frames rather than execution of code, since events that happen in the code may or may not give rise to graphic output.

ZStep 94 provides two operations, Graphic Step Forward and Graphic Step Backward, that run the stepper forward or backward, respectively, until the next event happens that results in significant graphic output. Below, each graphic step results in an exploration of the next branch of the tree.









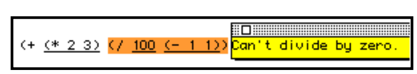
Four successive "graphic steps"

Each graphic step runs the stepper forward or backward until it is pointing at the event which was responsible for the graphic output, and the stepper remains live at all times. While stepping non-graphic code, the effects of previous graphic operations remain visible, just as they do in a normally-running program.

We could also provide graphic step operations analogous to "graphic fast forward" and "graphic rewind". Because the stepper can be run from either the code or the graphics at any point in time, the user can easily move back and forth between the different points of view.

ERROR CONDITIONS

ZStep 94 has a unique approach to dealing with execution errors. Traditionally, errors during program execution are disruptive. They either print an error message and halt execution, or put the user into a breakpoint loop, from which special commands can examine the error, and the stack can be inspected. In either case, errors disrupt execution and often lose information about partial computations which may have finished correctly.



ZStep 94 displays an error message

In ZStep 94, an expression which results in an error simply displays the error message in place of the value of the subexpression most relevant to the error. The value window is colored yellow to indicate the error condition. The stepper remains active, all intermediate values are preserved, and the program can be run backward to examine the history that led up to the error.

The current version of ZStep 94 has no independent display of the stack, though the user can flip through events representing containing computations manually, whether or not an error has occurred. Previous ZStep versions had a stack display that was updated continuously with each event, animated in tandem with the source code animation. Each stack frame was itself a menu item, and clicking it would return you to that frame.

What happens once an error is found? ZStep 94 facilitates the repair phase, since it leaves you in the text editor with the cursor pointing to exactly the code in need of repair. Further, you are then just one click away from restarting the entire computation after the edit. However, we cannot support restarting the computation from any point in ZStep's history after the edit, because editing a running program cannot guarantee consistency of the event data structures. However, we could imagine techniques that would allow conservative restart of the code, or at least warn you of potential difficulties.

THERE'S NOTHING SLOWER THAN A PROGRAM THAT DOESN'T WORK YET!

By now, many readers will be thinking: isn't all this history-keeping and use of special-purpose interpreters ridiculously expensive? Histories eat up enormous amounts of storage, and interpreting code is slow. The answer is: yes, it can be expensive. But does it matter?

First of all, we have to keep in mind that the purpose of the stepper is to debug programs that don't work yet, and so worrying about optimizing execution is silly. Even in large programs, judicious testing can frequently quickly isolate a buggy example involving short executions that can be feasible to debug using complete history-keeping.

Nevertheless, we admit that there may be bugs that appear only after long runs involving large amounts of code, and so our techniques may be inappropriate in these cases. However, we conjecture that the vast majority of bugs are relatively shallow, and the productivity improvements from finding simple bugs quickly will far outweigh slower execution during testing.

Second, the key to making these techniques feasible over a wider range of programs are tools for selectively turning on and off history-keeping mechanisms. A simple way to automate selective processing of history in a common case would be to run the program normally until an error occurs. Then a program could use the stack inspector to determine what functions were involved in the error, and history-keeping would be turned on selectively for those functions. The problem would then be run again from the beginning.

Another common objection to our approach is that it is not guaranteed to work in the presence of programs that have side effects. First, we should observe that certain kinds of shallow side effects do in fact work without any special provisions. Incrementing a variable will, because of the history, preserve the values both before and after the operation. More complex side effects involving shared data structures will not work, but one is no worse off than with conventional debugging techniques. More elaborate history-keeping mechanisms such as those studied in the simulation literature could alleviate this problem.

One of the principal uses of our stepper could be in educational applications, where execution efficiency is not of much concern, but interactive control over speed is important. One of the best ways to teach programming to a beginner would be to have the student step through example programs. In general, a stepper is an excellent tool for understanding code written by others.

IMPLEMENTATION

ZStep 94 was implemented in Macintosh Common Lisp 2.0. It is a prototype, and not a production implementation in several respects. First, it works with only a subset of Common Lisp. Second, no attempt was made to optimize speed or space constraints. Third, we did not perform any formal user testing, besides getting feedback from both experienced and novice programmers. Our main goal was to experiment with novel interfaces to dynamic program visualization.

Adapting ZStep 94 concepts to C would be possible, but challenging. A complete parser and unparser for C syntax would be required and care would have to be taken to assure the C interpreter or compiler kept enough type and run-time information.

The lesson for design of languages and environments is to consider debuggability as a primary criterion. If the goal of a new environment is to make programmers more productive, nothing could contribute to this goal more than introspective features that provide the foundation for sophisticated debuggers.

RELATED WORK

We're sad to report that there is not as much related work in this area as there should be. Even recently implemented programming environments seem to provide only the same set of tools that have been around for the past 30 years: trace, breakpoints, stack inspection, and perhaps a line-by-line stepper.

A notable exception has been the Open University in England, which has long been a source of innovative programming environments. The Transparent Prolog Machine [5] provides an innovative graphical view of program execution, and an interface carefully designed with Prolog's more complex execution model in mind. An innovative stepper for Lisp which shares some of the principles described here was recently implemented by Watt [14].

Many of the elements of our approach do have a long history. Reversible debuggers have been explored as far back as 1969 [3], and more recently by Moher [11].

However, these debuggers did not provide reversible animations of both the code and its graphical output, nor connections between individual code expressions, values and individual graphical objects.

The field of *visual programming* [7] uses graphical objects to represent the elements of the program, such as variables, functions and loops. The best of these environments also provide some animation of the graphical representation during execution. The pioneering work on animated visualization of program code in a single stepper was done by Ron Baecker [2].

Animation of visual representations of data manipulated by programs often appears under the name *algorithm animation* [4] [or *scientific visualization* if the algorithm represents a physical process]. Animations of data help a programmer visualize the dynamic behavior of a program as it runs. But most visual programming and algorithm animation systems confine themselves to visualizing and animating either the code or the data, but not both.

No one of these approaches -- reversiblity, animation of code, or animation of data -- by themselves will lead to a satisfactory set of debugging tools. ZStep 94's contribution is to integrate reversibility, animation of code, and correspondence between code expressions, values and graphic output, all under unified interactive control. Using the control structure of a stepper to control visualization of data helps solve one of the fundamental problems of software visualization: establishing the correspondence between data that looks faulty and determining the code that corresponds to the error. Adding data visualization facilities to the code visualization that steppers provide solves the problem of determining the effect of a particular piece of code upon the [sometimes complex] program state.

ACKNOWLEDGMENTS

Support for Lieberman's work comes in part from research grants from Alenia Corp., Apple Computer, ARPA/JNIDS, the National Science Foundation, and other sponsors of the MIT Media Lab, and for Fry's work from Harlequin, Inc. The authors would like to thank Marc Brown, John Stasko and Blaine Price, who ran last year's CHI Workshop on Software Visualization, and John Domingue for helpful suggestions.

REFERENCES

[1] Hiralal Agrawal, Richard deMillo, and Eugene Spafford, An Execution-Backtracking Approach to Debugging, IEEE Software, May 1991.

[2] Ron Baecker, Two Systems Which Produce Animated Representations of the Execution of Computer Programs, SigCSE Bulletin, February 1975.

[3] Robert Balzer, EXDAMS, EXtensible Debugging and Monitoring System, Spring Joint Computer Conference, 1969.

[4] Marc Brown, Perspectives on Algorithm Animation, Proceedings of ACM CHI'88 Conference on Human Factors in Computing Systems, p. 33-38.

[5] Marc Eisenstadt and Mike Brayshaw, The Transparent Prolog Machine: An Execution Model and Graphical Debugger for Logic Programming, Journal of Logic Programming 5(4), p.1-66, 1988.

[6] W. Wayt Gibbs, Software's Chronic Crisis, Scientific American, Sept 1994.

[7] E.P. Glinert, ed, Visual Programming Environments: Applications and Issues, IEEE Press, 1991

[8] Henry Lieberman, Steps Toward Better Debugging Tools for Lisp, ACM Symposium on Lisp and Functional Programming, Austin, Texas, August 1984

[9] Henry Lieberman, Reversible Object-Oriented Interpreters, First European Conference on Object-Oriented Programming, Paris, France, Springer-Verlag, 1987.

[10] Henry Lieberman, A Three-Dimensional Representation for Program Execution, in [7]

[11] Thomas Moher, PROVIDE: A Process Visualization and Debugging Environment, IEEE Transactions on Software Engineering, Vol. 14, No. 6, June 1988.

[12] Donald Norman, Cognitive Engineering, in Donald Norman and Stephen Draper, eds. User Centered System Design, Lawrence Erlbaum Assoc. 1986.

[13] Dennie van Tassel, Program Style, Design, Efficiency, Debugging and Testing, Prentice-Hall, 1974

[14] Stewart Watt, Froglet: A Source-Level Stepper for Lisp, Human Cognition Research Laboratory, Open University, Milton Keynes, England, 1994.

[15] M. V. Zelkowitz, Reversible Execution, Communications of the ACM, Sept. 1973.