A visual interface to a music database

Robert St. Amant, James E. Blair, Patrick Barry, Yinon Bentor, Christopher G. Healey
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-7534
stamant@csc.ncsu.edu, healey@csc.ncsu.edu

ABSTRACT

This paper describes a system for exploring and selecting entries from a music database through a visualization interface. The system is designed for deployment in situations in which the user's attention is a tightly limited resource. The system combines research topics in intelligent user interfaces, visualization techniques, and cognitive modeling. Informal evaluation of the system has given us useful insights into the design tradeoffs that developers may face when building visual interfaces for off-the-desktop applications.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; I.2.0 [**Artificial Intelligence**]: General—*Cognitive simulation*

General Terms

Human Factors

Keywords

Cognitive Modeling, Visualization, Driving

1. INTRODUCTION

This paper describes a system, called Lola, that displays entries in a music database for exploration and selection. Lola is designed for environments that involve the user selecting musical pieces from a database while engaged in an unrelated primary activity, where the primary activity makes significant demands on the user's perception, attention, cognition, and even motor activity. This paper focuses on the task of driving a car, but this is not the only possibility; tasks in many other environments, such as working in an office, can be described in these terms. In such environments, the task of visualization and selection must be integrated into the primary activity without degrading performance.

.

By considering an interactive system in a context larger and possibly more complex than that of desktop applications, we face a number of less conventional issues in user interface design, including constraints on hardware input and output capabilities, environmental uncertainty and time pressure, limitations on the cognitive and attentional resources users can devote to the task, and development and deployment issues. The environment for Lola is a departure from the desktop in several ways:

- Large, high-resolution output devices may be impractical in the environment, for physical and cost reasons.
- Full keyboards and high-resolution pointing devices (e.g. the mouse) may be inappropriate for task reasons.
- Users have limited time in which to complete their selections.
- Users' attention, cognition, and motor resources carry a significant load, which narrows the possibilities among appropriate visualization and interaction techniques.

In general, building an effective system for this problem means carrying out a close analysis of the relationship between interaction techniques and the task and environment in which they are applied. In the remainder of this paper we describe Lola, its operating characteristics, and the environment in which it is used. We discuss a cognitive modeling framework and multidimensional visualization techniques we have applied in the continuing development of Lola. The paper concludes with a description of an early formative evaluation we have carried out.

2. COGNITIVE FACTORS IN LOLA

Although our goals are more general for the work than suggested by the driving domain, as discussed above, our discussion will focus on driving as the main environment for Lola. Models of the driving task have been the target of research for decades (the analysis of Gibson and Crooks in 1938 provides one of the earliest examples [3]; Bellet and Tattegrain-Veste [2] give a concise historical overview from a cognitive ergonomics perspective.) The hierarchical risk model of van der Molen and Bötticher is a representative example of recent models [14]. Activities are structured into strategic, tactical and operational levels. At the strategic level, planning activity takes place, such as the choice of route and travel speed. At the tactical level decisions encompass more concrete, situation-dependent actions, such as lane changing, passing, and so forth. The operational level describes skilled but routine activities, such as steering and acceleration. Moving up the hierarchy, each level describes an increasingly abstract set of behaviors that govern choices at the level below it.

The different levels of abstraction represent different demands on the cognitive, perceptual, and motor abilities of the driver. For example, feedback from assistive technology such as ABS or power steering is provided at the operational level through haptic channels, often imperceptibly. Feedback for travel speed, in contrast, requires some cognitive activity at the strategic level, to interpret speedometer readings. If the feedback channels from these different activities were reversed (e.g., if the driver had to interpret a numerical value to determine power steering assist), their usability would be seriously impaired. For our purposes, the difference between abstraction levels has implications for integrating a new secondary task into the primary task. A key issue is that the new task will inevitably compete for perceptual, attentional, and cognitive resources with the existing task. Because of time pressure and uncertainty in the environment, this conflict has serious potential consequences at the tactical and operational levels.

One heuristic for integrating tasks is to minimize the duration of the secondary task, thereby reducing the chances for conflict between them. The secondary task should be structured such that its resource demands are consistent with those of the primary task, and that the aggregate demands do not exceed environmental and user limitations. For example, a secondary task during driving might be searching for information on roadway signs, a strategic level activity. The designers of roadway signs can ameliorate the attentional demands of the search with easily recognized graphical icons and sign shapes, and short words and phrases. The environment can be further structured so that the visual search does not conflict with tactical or operational activities (e.g., signs indicating required lane changes are posted well in advance of the point at which the action becomes necessary.)

Some techniques in visualization are explicitly designed to operate at lower levels of abstraction in an activity hierarchy; these techniques are targeted at the exploration of datasets in which the attributes of data elements are of interest to the user individually and in combination. Data exploration in the musical domain has this characteristic. Data items are musical tracks whose properties include album, artist, release date, popularity, familiarity, and so forth; for example, a user might wish to find a piece by a specific artist within a range of years. The challenge in a visualization approach to Lola comes from the amount of data that must be handled, and from the significant demands that a driving task places on the visual channel. A variety of techniques for multidimensional visualization address this search and selection problem by presenting the data for multiple attributes simultaneously.

The multidimensional visualization techniques in which we are particularly interested attempt to leverage the built-in capabilities of human vision. Some variations in visual features, including hue, luminance, size, density, and regularity, can be detected very quickly, without the need for a sequential search of the field of view [5, 6, 7]. Processing of these features is sometimes called preattentive, because it precedes focused attention in the low-level human visual system [13, 15]. Visual interfaces built around these features can support the efficient, accurate performance of key tasks in data exploration. Examples of such tasks include searching for data elements with a unique feature, identifying the boundaries between groups of elements with common features, and estimating the number of elements with a specific visual feature. These tasks are a natural part of exploration and selection in many domains, including weather tracking [6, 7], and scientific simulations [4]. If the tasks associated with music exploration during driving can be accomplished without the need for extended shifts in attention, the promise is that they can be pushed down to the operational level, limiting interference with strategic activities.

Figure 1: The Lola system unit

3. THE LOLA SYSTEM

Our approach to designing interaction with Lola treats visualization as a cognitive tool that can be tailored to the capabilities of the user and the requirements of the task environment. While physical tools are often considered amplifiers of action or behavior, cognitive tools can better be viewed as mechanisms that translate a problem representation such that solutions are immediately apparent [8, 11]. The implication for Lola is that a specific type of visual interface might transform cognitive tasks, such as filtering, reading, and selecting musical tracks, into perceptual tasks, such as identifying properties of tracks and sets of tracks by visual means. This transformation will push activities in Lola down to the tactical and operational levels, where they can more easily be managed.

An initial hardware and software version of Lola, which we call β-Lola, was completed early in the year 2000. β-Lola allowed selection from a music database through a text-oriented interface, with no visualization functionality. It nevertheless acted as a basic proof of concept and was tested informally under actual driving conditions.

The usability limitations of the early system led to a new hardware and software platform, as given below and shown *in situ* in Figure 1. (β-Lola was installed and functional in the car cockpit, though the new platform has not reached that point.)

Intel Celeron processor at 433MHz, 64MB RAM
10GB laptop hard drive
6.4 inch TFT LCD, digital video input, at 640×480
3M Dynapro resistive touch screen, at 912×870
Dimensions $8.75 \times 6 \times 1$ inches, display

In software, Lola datasets contain personalized collections

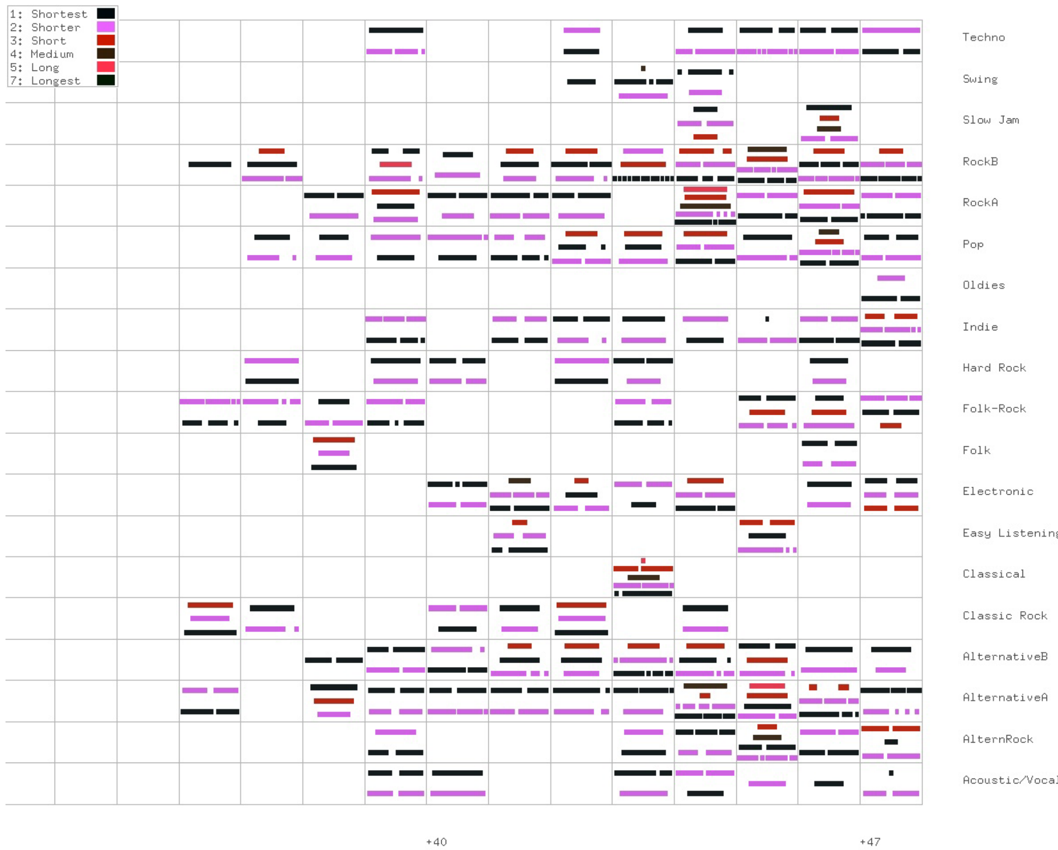

Figure 2: A visualization in Lola: position → genre and year; width → popularity; hue → song length.

of MP3-formatted musical tracks for different users. Data in a Lola dataset are compiled from a number of sources. Genre entries are assigned to tracks by a partially automated process. Categories include rock, jazz, swing, classical, and spoken. Artist, album, year, track name, and related information are taken from FreeDB.[1] The length of each track is calculated directly from the MP3 file. Popularity is measured by the Amazon.com sales rank of the album. Familiarity represents the number of times the user has recently listened to the track, and is recorded and maintained by the running system.

Figure 2 shows an example Lola visualization, minus the standard audio controls (stop, fast forward, rewind, etc.) This visualization is from the desktop-based development environment, rather than the target platform, but interaction is the same. The user selects tracks from this display (via the touch screen on the target platform, with a mouse in the development environment); results are queued and played in order. The design of the visualization was based on our experience with β-Lola and our analysis of comparable systems for music selection. Genre and year were judged to be the most salient properties of a musical track, and are represented as the x- and y-axes of the display. Within a cell, the popularity of the piece is represented by the size, or length, of a glyph, and its length is coded into several distinct colors. This is not the only visualization possible; in earlier work users worked with a visualization in which position (on both axes) mapped to genre, hue to popularity, brightness to familiarity, size to song length.

3.1 Formative evaluation

Our observations are based on a formative evaluation consisting of self-reported experiences of five users of the system. Three of the test users were involved with the development of the system, and two were new to the system. All users were familiar with the project goals. The evaluation was carred out on a 3D version of the visualization, and our findings have since been incorporated into the 2D visualization shown in Figure 2. We will not discuss all of our findings, rather only those relevant to both the 3D and newer 2D versions of the display.

Users commented positively on several aspects of the system. These dealt mainly with the selection capabilities facilitated by the visual display:

Selection range: The number of elements is a few orders of magnitude greater than can be presented simultaneously in a text-based system. Figure 2 shows about 500 entries from a music database.

Selection by property: Users choose a track by the imprecise selection of its component properties, rather than its identity. While this is a drawback in many domains, it suits the music domain well: by analogy, the user changes radio stations instead of selecting a specific piece of music. The effectiveness of this selection depends critically on the intuitive nature of the structure that over-

[1] `http://www.freedb.org/`

lays the data. In this case, organization by genre and year means that within specific regions, nearby tracks in the visualization are also conceptually near each other.

Element properties: Rather than seeing artist and title, as is common in most music selection interfaces, the user sees other (potentially interesting) properties as well as their combinations. The user can, for example, select a popular piece recorded in a particular year in a specific genre.

Touch interaction: Interaction via the touch screen was considered effective, using the 3D visualization. Although not as precise as a mouse or other input device, it was a good match for the constraints of the environment. We expect this property to extend to the improved 2D interface.

Users also encountered some difficulties in interacting with the system; the problems are simply the drawbacks of the advantages users identified above. As with many problems, solutions are dominated by difficult tradeoffs.

Imprecise selection: In some cases users know exactly which track they want to play; selection is hit or miss through the visualization, in which songs are not identifiable explicitly. This suggests that alternative interfaces might be useful, which provide different modes for different types of selection, precise (by track identity) or imprecise (by track properties.) Part of the problem may be due to known limitations of touch screen use. Lola does not filter or translate user selections, but this might improve selection.

Uninformative properties: Some track properties in the visualization are of limited value. Width, for example, is important if the user is looking for, say, long radio dramas, but less interesting otherwise. This suggests that we should explore alternatives in mapping visualization features to track properties.

3.2 Current development

Our evaluation has identified some distinct weaknesses in the visualization and interaction components of Lola. In our current research, we are applying work in artificial intelligence to improve the system at design time and run time.

For design time improvements, we draw on work in the area of automated assistance for generating visualizations. We have built a prototype system called ViA, a Visualization Assistant, with the goal of helping the user traverse the exponential space of possible mappings between visualization parameters and datasets to be visualized [12]. ViA performs an initial evaluation of candidate mappings and presents them to a user familiar with a domain; the user judges and refines the candidates, to produce a small collection of high quality visualizations. ViA does not automatically generate visualizations at the implementation level, but rather operates at an earlier stage in the visualization process, to produce specifications.

To improve the run time behavior of the system, we are extending Lola to automatically select musical tracks and adapt selections to the user's preferences and the surrounding environment. Our research in this area is preliminary, but draws on more mature work elsewhere in AI, specifically in the areas of collaborative and content-based filtering systems [1] and automated adapation. In addition, some work on music selection has already been carried out [9, 10], which has to some extent influenced our design decisions in this domain.

4. ACKNOWLEDGMENTS

This effort was partially supported by the National Science Foundation under award 9988507. The information in this paper does not necessarily reflect the position or policies of the U.S. government, and no official endorsement should be inferred.

5. REFERENCES

[1] Marko Balabanovi and Yoav Shoham. Fab: Content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72, March 1997.

[2] Thierry Bellet and Hélène Tattegrain-Veste. A framework for representing driving knowledge. *International Journal of Cognitive Ergonomics*, 3(1):37–49, 1999.

[3] J. J. Gibson and L. E. Crooks. A theoretical field-analysis of automobile-driving. *American Journal of Psychology*, 51:453–471, 1938.

[4] Christopher G. Healey, Kellogg S. Booth, and James T. Enns. Real-time multivariate data visualization using preattentive processing. *ACM Transactions on Modeling and Computer Simulation*, 5(3):190–221, 1995.

[5] Christopher G. Healey, Kellogg S. Booth, and James T. Enns. High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction*, 3(2):107–135, 1996.

[6] Christopher G. Healey and James T. Enns. On the use of perceptual cues and data mining for effective visualization of scientific datasets. In *Proceedings Graphics Interface*, pages 177–184, 1998.

[7] Christopher G. Healey and James T. Enns. Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE Transactions on Visualization and Computer Graphics*, 5(2), 1999.

[8] Edwin Hutchins. *Cognition in the Wild*. MIT Press, Cambridge, MA, 1995.

[9] Louis Lapat, Jared Dunne, Marc Flury, Mustafa Shabib, Tom Warner, Jay Budzik, Kris Hammond, and Lawrence Birnbaum. mpme!: Music recommendation and exploration. In *Proceedings of IUI*, 2002.

[10] Joseph F. McCarthy and Theodore D. Anagnost. Musicfx: an arbiter of group preferences for computer supported collaborative workouts. In *Proceedings of CSCW*, 1998.

[11] Malcolm McCullough. *Abstracting Craft: The Practiced Digital Hand*. MIT Press, Cambridge, MA, 1996.

[12] Robert St. Amant and Christopher G. Healey. Usability guidelines for interactive search in direct manipulation systems. In *Proceedings of IJCAI*, 2001.

[13] A. Triesman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[14] H. H. van der Molen and M. T. Bötticher. A hierarchical risk model for traffic participants. *Ergonomics*, 31(4):537–555, 1998.

[15] J. M. Wolfe. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1(2):202–238, 1994.