Clique: Task-based Auditory Display for GUI Applications

Department of Computer Science University of North Carolina at Chapel Hill March 2007

Introduction

The purpose of the Clique project is to explore a new way of adapting applications with graphical user interfaces (GUIs) for user in audio. The state-of-the art, screen reading, retains the visual components and metaphors in the resulting single-stream audio display. At least two studies have shown that this straightforward transform from a widget on the screen to speech describing it results in usability problems. Clique, on the other-hand, presents the user with a multi-channel audio environment mapping application tasks to virtual assistants. The user interacts solely with this display while Clique takes charge of inspecting and controlling the underlying programs via their GUIs. In effect, the graphical nature of program interfaces is hidden from the listening user. We hypothesize that audio displays produced in this manner will prove more effective and satisfying for common office computing tasks than existing solutions. Furthermore, we believe such an interface may benefit sighted users during brief departures from environments conducive to visual desktop computing.

Tasks, Not Widgets

For the typical user, the purpose of interacting with a computer is to complete a set of tasks. The user does not care how the computer supports the execution of his or her tasks, as long as it does so in an effective, unobtrusive manner. When the user is working in audio, how common tasks are manifested visually is often irrelevant to their successful completion. For instance, a user does not need a visual mental model to write an email. He or she merely requires a way to state the recipient, subject, and body of the message, and indicate the message should be sent. Only tasks tied to vision (e.g. designing a GUI, drawing an image) require intimate knowledge of visual presentation.

Under this premise, the goal of audio adaptation is not to mimic visual interfaces, but rather to best aid the completion of user tasks. To meet this goal, an audio display must do more than provide a superficial layer between the listening user and the screen. It must interpret what is on the screen and expose the meaning of the visual components, relationships, and metaphors to the user in forms appropriate to audio. Likewise, it must take charge of the menial work of controlling applications via their native GUIs. Such an audio display, one that frees the user from dealing with visual concepts, is likely to avoid many of the usability problems associated with screen reading.

Clique accomplishes this goal by using scripts that pair auditory representations for tasks with information exposed by the platform accessibility API (e.g. Microsoft Active Accessibility). For instance, the script for Microsoft Outlook Express contains a task definition for browsing email. This definition contains auditory views for browsing a hierarchy of mailboxes, a list of message headers, and the text of an email body. The script associates these views with adapters for the GUI tree view, multi-column list, and text area widgets showing the relevant information on the screen. The auditory views draw information from these adapters and present it to the user. This model-view separation allows Clique to create a consistent auditory experience across applications by reusing the auditory views to represent equivalent tasks that have differing visual representations.

A Perceptually-Based Audio Environment

Traditional screen reading has relied on a single stream of speech with an occasional auditory icon inserted to convey information about the visual desktop. The large mismatch between the bandwidth of the visual domain and this single stream causes information loss for the listening user. For instance, the user is completely unaware of new emails arriving, Web downloads completing, and system notification popups outside the narrow application focus.

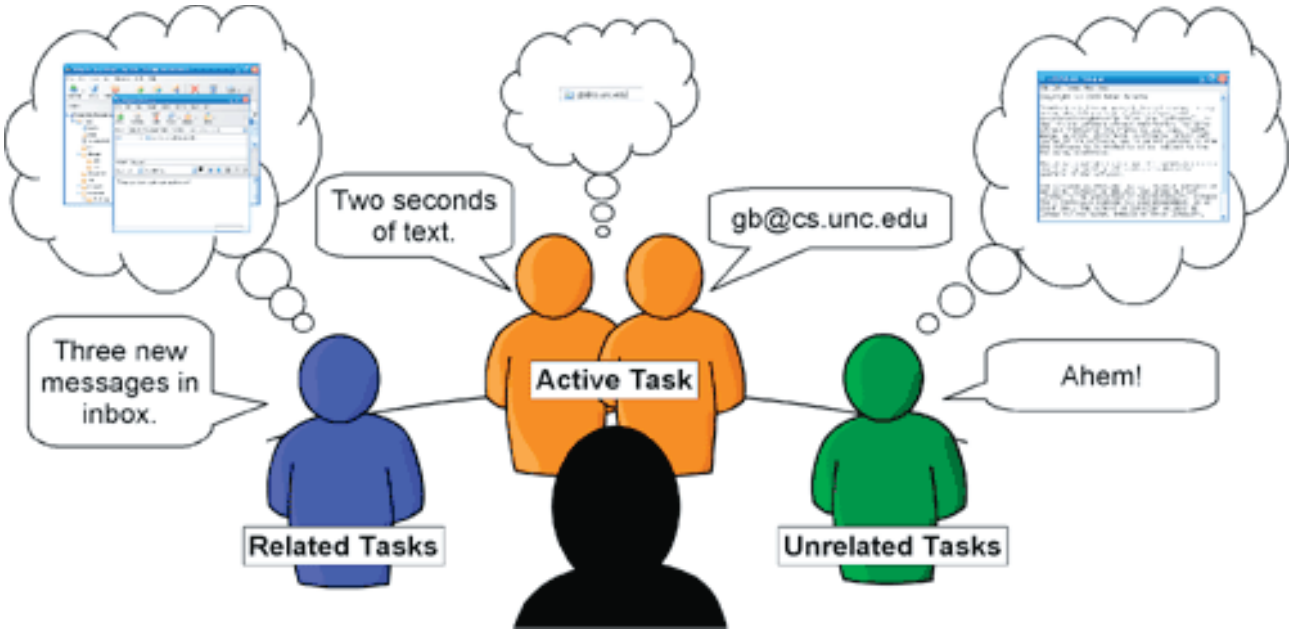
Clique builds on the theory of auditory scene analysis in an attempt to increase the amount of information available to the user at any given moment. Multiple audio streams representing a number of virtual assistants are positioned around the user in a spatial sound environment. Each assistant speaks with a unique voice originating from a unique position. A given assistant plays a specific role in the conversation: reporting content in the current task, summarizing the state of the current task, reporting events in related tasks, reporting events in other tasks, and echoing user commands. The purpose of this design is not to force the user to pay attention to multiple speakers at once. Rather, the objective is to allow the user to quickly attend to information of importance to his or her current task by focusing on a single audio stream.

All assistants follow familiar rules of conversation such as referencing, pacing, turn-taking, and interruptions. For instance, an assistant who wishes to inform the user that a web page has finished loading outside the active task will either speak immediately, play an audio icon, or wait for the floor. The assistant will make this decision by considering who else is speaking and if the user is giving input.Such behavior avoids having so many active speech streams at once that all become unintelligible.

Finally, Clique uses continuous auditory ambience and intermittent auditory icons to provide reminders of the user's current working context. For instance, if the user is busy working in a calendar application, the calendar program may be represented by a continuous sound of rainfall, the task of browsing the current week may be indicated by occasional claps of thunder, and the task of adding a new appointment may be represented by the sound of footsteps splashing in puddles. The user may use these sounds as hints for remembering what he or she is doing, or simply ignore them while working.

Related Publications

Parente, Peter. Clique: A conversant, task-based audio display for GUI applications. Doctoral consortium participant in Proceedings of the Seventh international ACM conference on Assistive technologies. October, 2005.



A depiction of the virtual assitants in Clique.