Towards a Quantitative Analysis of Audio Scrolling Interfaces

Eric Lee

Media Computing Group

RWTH Aachen University

52056 Aachen, Germany

eric@cs.rwth-aachen.de

Abstract

We present the results of a user study inspired by previous work in document navigation comparing rate and position control for navigating an audio timeline. Although interfaces for controlling playback speed (rate) are favored over playback position, we found that position control is, on average, 15—19% faster than rate control when searching for targets 90 to 100 seconds away in the audio timeline. Additional studies are being planned to further characterize audio scrolling performance with position and rate controls.

Keywords

Audio interfaces, audio scrolling, empirical study.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems.

Introduction

Recent technology advancements have led to a corresponding increase in popularity of digital audio as a communication medium. Radio, for example, has in recent years been reborn in digital form as "podcasts"; with today's technology, it is possible for even the average home user with a computer to *create* such

content, *distribute* it to a global audience via the internet, where it is *consumed* using portable media players such as the iPod (www.apple.com). According to Apple, there were over one million podcasts subscriptions in the first two days after the iTunes podcast directory became available [1].

Unfortunately, most commonly available interfaces for navigating the timeline of audio do not differ significantly from the **"tape recorder"** metaphors of *play*, *stop*, *fast-forward* and *rewind* from the 1950's. In contrast, a variety of novel interfaces and metaphors [7,8,9] have been developed for quickly navigating text documents. Recent versions of QuickTime Player (www.quicktime.com) and Windows Media Player (www.microsoft.com) support audio playback at variable rates, but this is controlled using an **"advanced"** setting that is neither accessible, nor visible, in the main user interface.

There is, moreover, little work that studies the various methods to interpret user input in an audio navigation interface. Research in document navigation, in contrast, has examined in detail both rate controls (where the user controls the scrolling speed) and position controls (where the user controls the scrolling position) [10], and both input types are used in practice. Hinckley et al. found that scrolling performance with rate and position controls exhibit a crossover effect; while rate control is better for long searches, position control is better for short searches [4].

Rate control is often employed in existing interfaces that aim to improve audio navigation performance [2,5], often with the implicit assumption that rate is

superior to position control. The shuttle control on certain DVD players for changing the playback speed, the speed sliders on QuickTime Player and Windows Media Player, and even the ubiquitous *fast-forward* and *rewind* are all rate controls.

Our work is inspired by previous work on document navigation, and we begin our discussion by comparing spatial (document) navigation and temporal (audio) navigation. We then present the results of a user study that compares scrolling performance using position and rate controls, and describe the design implications of our results in the context of our ongoing work.

Spatial vs. Temporal Navigation
Navigating through the timeline of continuous time-based media, such as an audio recording, is in many ways similar to navigating through a document. Despite the spatial nature of documents, as opposed to the temporal one of audio, the input techniques used for scrolling through a document often apply to audio as well.

A common software interface widget for scrolling through audio is the timeline slider, analogous to a scrollbar in a document window (see Fig. 1).
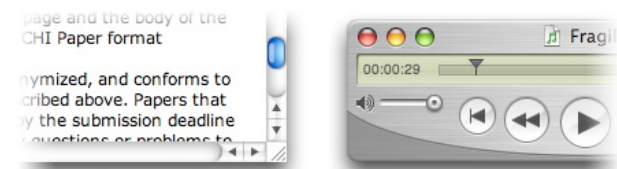

Figure 1. The scrollbar for document navigation (left) is analogous to the timeline slider for audio navigation (right).
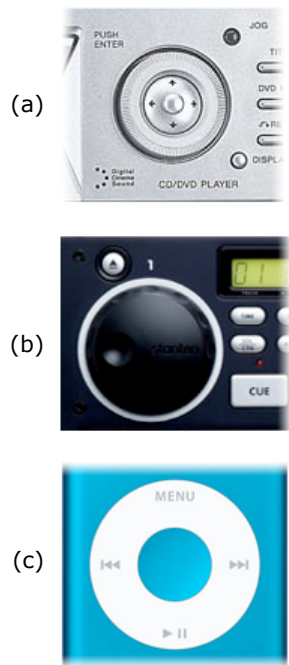
(a)

(b)

(c)

Figure 2. (a) Scroll ring in the Sony DVP-NS700P DVD player. (b) Jog dial in the Stanton C.500 CD player. (c) Touch wheel on the iPod portable media player.

Zhai et al. [10] discuss the drawbacks of using the scrollbar interface for navigating through a document. In particular, time is required to acquire the wiper (the box that represents the currently visible area in the document), scrollbars are ill-suited for continuous scrolling with precision, and navigating to the scrollbar shifts the user's locus of attention away from the target. The timeline slider uses the same mappings, and thus suffers from similar drawbacks; moreover, the playhead in a timeline slider is usually small, and thus acquiring it is even more difficult than with a scrollbar.

As a result, alternative interfaces to more efficiently navigate a document have been studied [4,7,8,9,10]. These interfaces typically control *rate* (user input maps to scrolling velocity), or *position* (input maps to viewing area position). For audio navigation, a rate control would similarly control the play rate, and a position control the current playhead position.

Position & Rate Control In Audio Navigation

The following usage scenarios illustrate temporal navigation with position and rate controls:

1) Eva has just returned from a Dave Matthews Band concert. She recorded the entire concert onto her iPod, which the band permits for personal use. Alas, the recording has no track marks, and she must scroll through the audio using the "Click Wheel" to find the start of her favorite song.

2) David and Victor are arguing over the name of the ensign killed off in Episode 213 of Star Trek: TOS*, and they decide to refer to the DVD to resolve their dispute. David's DVD player, a Sony DVP-NS700P (www.sony.com), has a shuttle ring control to*

interactively adjust play rate. Using the DVD menu, David jumps to the scene where the ensign first appears, and uses the shuttle ring to navigate to the part where his name is mentioned.

We hypothesized that, similar to spatial navigation, position control can be faster for audio navigation for closer target distances. We designed an audio navigation experiment where users were asked to locate a target between 90 and 100 seconds from the current playhead position using both rate and position controls. We measured and compared the targeting times for these devices.

Setup
We included three input devices in this study:

Scroll ring: A rate control consisting of a spring-loaded ring. This device is found on certain DVD players, such as David's DVP-NS700P in the second usage scenario (see Fig. 2a). The play rate increases in the forward direction as it is rotated clockwise, and backward when rotated counter-clockwise. When released, the ring "snaps" back to its original position. Previous research has shown that this self-centering mechanism is an important characteristic for rate control [4,7].

Jog dial: A position control using a solid dial. The audio will advance forwards as the dial is rotated clockwise, and backwards when rotated counter-clockwise. Some CD players used by DJ's, such as Stanton's C.500 (www.stantondj.com, see Fig. 2b) offer this type of control.

Touch wheel: A position control that operates similarly to the jog dial, except that it is touch-sensitive and
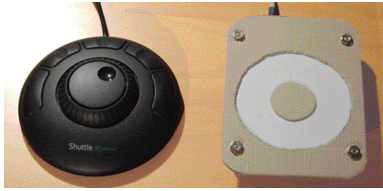
Figure 3. The ShuttleXpress (left) has an outer self-centering ring for rate control and an inner dial for position control. The Phidgets CircularTouch (right) has a touch-sensitive surface that we used for position control, repackaged to resemble an iPod Click Wheel.

thus lacks any haptic feedback. The touch wheel is the primary control on the iPod, where it is used for menu navigation in addition to audio timeline navigation (see Fig. 2c).

In our experiments, we used the ShuttleXpress (www.contourdesign.com, see Fig. 3), a device that includes both a scroll ring and a jog dial. Its self-centering scroll ring offers 7 unique positions in either direction, which we mapped to play rates from $\pm\frac{1}{4}\times$ to $\pm16\times$ on an exponential scale. Audio playback stops when the ring is in its rest position. The ShuttleXpress jog dial snaps lightly into place on 10 unique positions per revolution.

For the touch wheel, we used the Phidgets CircularTouch (www.phidgets.com); it is touch sensitive, with a diameter twice as wide as the ShuttleXpress jog dial. The CircularTouch advertises an ability to report 128 unique positions per revolution; in practice, however, we found that the effective resolution was much lower (roughly 32 positions), due to the uncertainty of a finger making contact with a region on the wheel. We repackaged the CircularTouch to resemble a large iPod Click Wheel (see Fig. 3).

We wrote a software tool that accepts input from all three devices. Hinckley et al. found that incorporating an acceleration function into a position control significantly improves scrolling performance [4], and thus we used a similar function in our implementation. The user was presented with a slider that shows the current playhead in the audio timeline; however, the position was not reported numerically to the user. Audio feedback was provided using a pitch-preserving time-stretching algorithm [6].

Procedure
16 volunteers (10 male, 6 female) in their 20's and 30's participated in the experiment, including office workers and students from various disciplines. Each trial consisted of locating a cut point placed 90 to 100 seconds into a roughly three minute-long audio recording consisting of either speech or music.

Users searching through audio in a realistic situation would know their search target, and to simulate this, we used a relatively easy target: the cut point was either a speaker change (male to female), or a music genre change (classical to pop). Each user was tested with either music or speech, but not both. Users performed two trials for each of the three devices; both the cut point location and device order were pseudo-randomized to minimize learning effects.

Results and Discussion
The data points for a specific user and device were averaged together, and a repeated measures ANOVA revealed significant differences between the mean search times for the three devices ($F(3, 16) = 5.6$, $p < 0.01$; see Fig. 4). The Tukey HSD post-hoc test revealed that the average search time with the scroll ring is significantly higher than the average search times with both the jog dial and touch wheel ($p < 0.05$). The difference between the jog dial and touch wheel is not significant.

Users were also asked to subjectively rate each of the three devices on a scale from 1 ("very poor") to 5 ("very good") (see Fig. 5). The Wilcoxon signed-rank test showed that the touch wheel was rated significantly higher than the scroll ring ($p < 0.05$); this is the only significant difference. Feedback from users
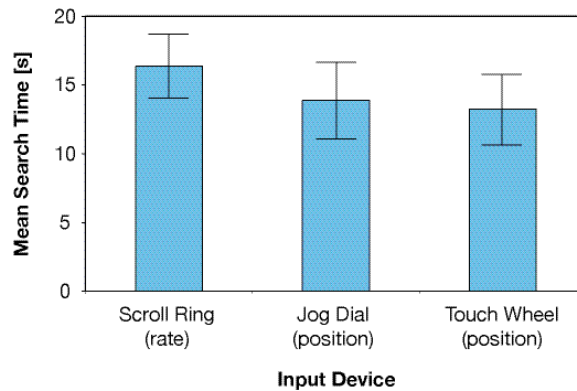
Figure 4. Mean search times using the scroll ring, jog dial, and touch wheel with 95% confidence bars. The mean search time for the scroll ring is significantly higher than for both the jog dial and touch wheel.

was consistent with previous studies comparing rate and position controls. Users observed, for example, that it was easier to play through audio at constant rates with the scroll ring, and operating it did not require as much physical movement compared to the dial and wheel. On the other hand, most people felt the dial and wheel were easier to control for precise position changes.

Our results confirm our initial hypothesis that the jog dial and touch wheel, both position controls, allow users to locate a target in a continuous audio stream significantly faster than the scroll ring, a rate control, for search targets between 90 and 100 seconds from the current playhead position. The mean search times are 16.4, 13.9, and 13.3 seconds for the scroll ring, jog dial and touch wheel, respectively, and thus, for our
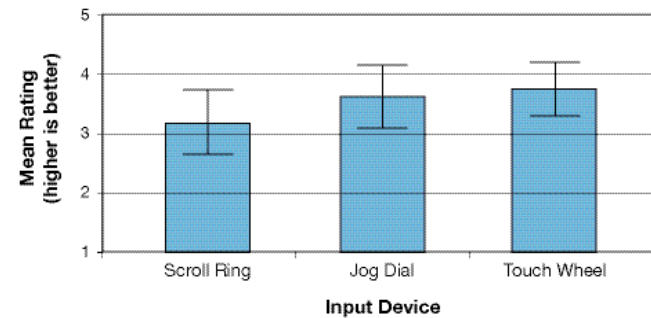

Figure 5. Mean subjective ratings for the scroll ring, jog dial, and touch wheel with 95% confidence bars (1 = very poor, 2 = poor, 3 = average, 4 = good, 5 = very good).

users, the jog dial was 15% faster than the scroll ring, and the touch wheel was 19% faster.

Hinckley [4] showed that for document navigation, position control is faster than rate control for closer search targets, but as the search target distance increases, the performance difference becomes less significant until a "crossover point", when rate control becomes superior (see Fig. 6). We believe this same crossover effect applies to audio navigation as well, in which case 100 seconds is a minimum upper boundary, below which position control is significantly faster than rate control. We are currently planning further studies to characterize this effect for audio navigation.

Conclusions and Future Work

Interfaces to navigate through an audio timeline can be constructed using a variety of techniques, which can typically be classified as position or rate controls. While rate controls are frequently used in audio navigation interfaces, we discovered that position control using the
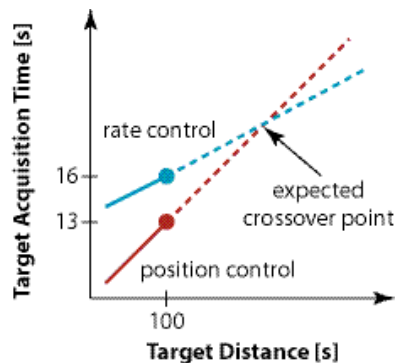
Figure 6. Hypothetical crossover effect. Position control is faster when acquiring close targets, and rate control is expected to be superior for more distant targets.

jog dial or touch wheel is, on average, 15—19% faster than rate control with the scroll ring for targets 90 to 100 seconds away in the audio timeline. We are currently planning further user tests with a larger variety of target distances to verify that Hinckley's "crossover effect" for document navigation also applies to audio navigation.

Determining a numerical value for this crossover point will have interesting implications for the design of audio navigation devices. For example, audio tracks are typically less than 15 minutes in length. If the crossover point is in the minutes range, then the use of a position control like an iPod Click Wheel to "fast-forward" through tracks would be empirically justified over the rate controls found on other brands of portable media players.

As continuous, time-based media and devices become increasingly ubiquitous, we hope our work will inspire alternative interfaces to improve audio navigation, and furthermore increase their adoption.

Acknowledgements

References

[1]   Apple Computer. iTunes podcast subscriptions top one million in first two days. Press release (June 2005).

[2]   Arons, B. SpeechSkimmer: A system for interactively skimming recorded speech. *ACM Transactions on Computer-Human Interaction*, 4, 1 (1997), 3—38.

[3]   Card, S., Mackinlay, J., and Robertson, G. A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems*, 9, 2 (1991), 99—122.

[4]   Hinckley, K., Cutrell, E., Bathiche, S., and Muss, T. Quantitative analysis of scrolling techniques. In *Proc. of CHI 2002 Conference on Computer-Human Interaction*. Minneapolis, USA, 2002, 65—72.

[5]   Hürst, W., Lauer, T., Bürfent, C., and Götz, G. Forward and backward speech skimming with the elastic audio slider. In *Proc. 19th British HCI Group Annual Conference*. Edinburgh, Scotland, 2005.

[6]   Karrer, T., Lee, E., and Borchers, J. PhaVoRIT: A phase vocoder for real-time interactive time-stretching. *Proc. ICMC 2006 Computer Music Conference.* New Orleans, USA, 2006, 708—715.

[7]   Moscovich, T. and Hughes, J. F. Navigating documents with the virtual scroll ring. In *Proc. of UIST 2004 User Interface Software and Technology Symposium*. Santa Fe, USA, 2004, 57—60.

[8]   Smith, G. and schraefel, m.c. The radial scroll tool: Scrolling support for stylus- or touch-based document navigation. In *Proc. of UIST 2004 User Interface Software and Technology Symposium*. Santa Fe, USA, 2004, 53—56.

[9]   Wherry, E. Scroll ring performance evaluation. In *Extended Abstracts of the CHI 2003 Conference on Human Factors in Computing Systems.* Ft. Lauderdale, USA, 2003, 758—759.

[10] Zhai, S., Smith, B. A., and Selker, T. Improving browsing performance: A study of four input devices for scrolling and pointing tasks. In *Proc. of INTERACT 1997 Conference on Human-Computer Interaction*. Sydney, Australia, 1997, 286—292.