Listen to the Music:
Audio Preview Cues for Exploration of Online Music

m.c. schraefel
Department of Electronics and Computer Science
University of Southampton
Highfield, Southampton, UK
E-mail: mc at ecs.soton.ac.uk

Maria Karam, Shengdong Zhao
Department of Computer Science,
University of Toronto,
Toronto, Canada
E-mail:{mkaram | sszhao} at dgp.toronto.edu

ABSTRACT

This paper presents a novel mechanism that seeks to allow people to explore large collections of loosely structured audio. The approach provides a lightweight preview mechanism that allows people to explore the audio collection by providing supporting information (analogous to the use of tooltips in visual interfaces) We present an evaluation of these “preview cues” towards developing a design heuristics for their deployment.

KEYWORDS: Multimodal interfaces, information access, data exploration, audio interfaces

INTRODUCTION

Standard techniques of keyword search and browsing are insufficient in cases where users do not have the domain expertise to express their query in terms that will produce a meaningful result. If we imagine the person who may say, “I don’t know anything about Classical Music but know what I like when I hear it,” keyword search engines like Google will fail if the person wishes simply to find some music they would enjoy hearing: they do not have the lexical expertise to express something for pattern match, like “baroque serenades on period instruments.” Similarly, browsing can be equally unhelpful: unless the person already knows what Mahler or Baroque sounds like, seeing a complete listing of same will not help the user make a selection.

To address this problem, we present a novel mechanism that seeks to allow people to *explore* large collections of loosely structured audio without relying on previous expertise about that domain. The approach, called “preview cues” provides a lightweight preview mechanism that allows people to explore an audio collection by providing supporting information (analogous to the use of tooltips in visual interfaces or Earcons in audio interfaces) at the point of interest.

In the following sections, we situate related work and describe the evaluations we carried out to test the efficacy of the cues in a variety of interfaces. We conclude by presenting several design heuristics which stem from this study.

AUDIO PREVIEW CUES CONTEXT

Audio preview cues have two components, the audio cue, and the information representation behind the cue. We describe each in turn, and then situate the cues in terms of related work.

The Audio Component

The design goal for an audio preview cue is to help associate what a user already knows or can assess (represented in the preview cue) with domain specifics (domain lexicon/organization). We associate audio preview cues with domain-specific labels. For instance, Agricola would be associated with a representative piece of music by Agricola; like wise the Romantic period would be associated with an indicative Romantic piece. We will return to how these associations are made.

Information Representation

Audio preview cues presuppose an organized domain: in an information space using preview cues, users might encounter the term Baroque as part of a category like Period or Style. The audio preview cue associated with that label is to assist users to refine their selections, that they can determine whether or not they wish to continue to pursue that part of the domain or not. Part of our ongoing research is to determine whether or not the combination of cues with domain organization helps users not only make decisions about an instance (I like this sound: yes or no) but to begin to develop an implicit expertise about the domain itself (I do not care for Mahler symphonies, but do enjoy the concertos).

RELATED WORK

Auditory preview cues are related to but distinct from both Tool Tips and Earcons. Tool tips are generally employed as text, which appears when a user brushes (mouses over) an unlabelled icon for a specific command in the tool bar of an application. The text describes what the icon command invokes when clicked. Similarly, earcons are highly structured non-speech auditory cues in which the associated auditory cue represents one specifically defined UI event, such as the selection of a particular tool [1]. More recently, Terry and Mynatt have proposed Sideviews [5] previews for graphics applications in which an artist can preview multiple versions of a filter on an image, rather than a seeing a preview of only one filter setting, as is common now. Preview cues are similar to tool tips and Earcons in that they provide additional information about a UI marker, but they are also broader – they do not need to be so semantically specific. They are more "intensional," in the Montague semantics sense of the term [6]. That is, rather than defining a specific command or reflecting a specific state, or as with Sideviews a set of explicit states, preview cues, suggest a potential *range* of values associated with a given area of a domain.

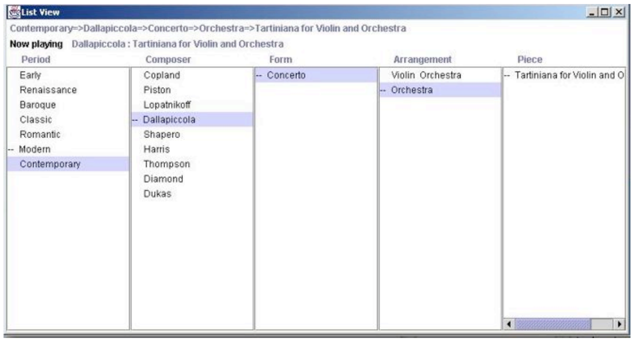
EXPERIMENTAL DESIGN

Visualization

There are many ways to represent information hierarchies for preview cues, from lists to hyperbolic trees. We have been experimenting with various visualization models, and have carried out a large scale, gender and age balanced study of a variety of interface treatments with preview cues. In this paper, we report on our use of single and multicolumn visualizations. While these types of interfaces have have been evaluated in document design [3], more specifically, they reflect two current approaches for Web-based information display. This approach would let us see if our technique would afford improvements for information access, and if so, whether or not these were significant. The results will help us refine principled design heuristics for audio exploration interfaces.

Protocol

We used a two-within participants repeated measures ANOVA design. We evaluated two interfaces types, and tested two conditions in each type, counterbalancing 24 participants. The study was gendered balanced and ranged in ages from 18-54. The interface types compared a single column (temporal) view with a multiple column (spatial) view of the domain hierarchy; the audio condition compared when in the hierarchy a cue is available (at each point in the hierarchy; only at the final level of the hierarchy). This yielded a total of 4 interface conditions.



List View

Contemporary=>Dallapiccola=>Concerto=>Orchestra=>Tartiniana for Violin and Orchestra

Now playing Dallapiccola : Tartiniana for Violin and Orchestra

Period	Composer	Form	Arrangement	Piece
Early	Copland	-- Concerto	Violin Orchestra	-- Tartiniana for Violin and O
Renaissance	Piston		-- Orchestra	
Baroque	Lopatnikoff			
Classic	-- Dallapiccola			
Romantic	Shapero			
-- Modern	Harris			
Contemporary	Thompson			
	Diamond			
	Dukas			

Figure 1: Spatial Multi-Column Layout view (with Labels).

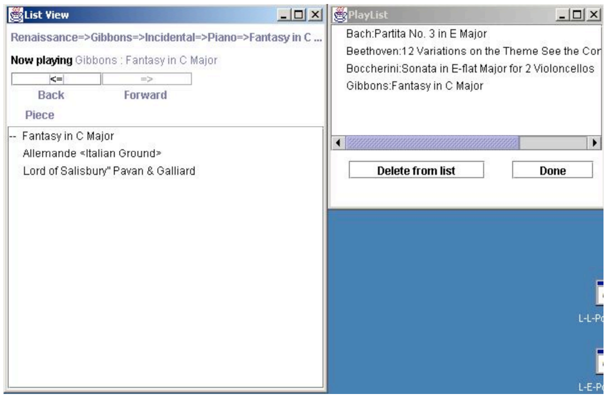
Persistent Attributes Across Conditions. In each interface we represented a classical music data hierarchy, organized as Period, Composer, Form, Arrangement and Piece. Beneath each category header, users see elements of the hierarchy. For instance, Period shows Early, Renaissance, Baroque, Classic, Romantic, Modern and Contemporary elements. Also, each interface displayed the path to the current element in the upper left of the window, as per Figure 1, showing the spatial version of the layout.

Column Condition. The single column view simulated Web-based exploration of hierarchies, where clicking on one level of a hierarchy takes a user to a new page representing the next level of the hierarchy (the Internet Directory, Yahoo.com). Context of where one has been is largely maintained temporally, but for textual information that describes the current path. We refer to this view throughout as the Temporal interface. We use a multicolumn view to maintain spatial context: the previous part of the path remains persistently available.

Extant research would suggest that the multicolumn view would be preferred to the single column view. Such work has not considered the inclusion of path information in the single column view. By reevaluating single column views with path information (the Web model) against multicolumn representations (the simple shift of variable from temporal to spatial views) we can assess two qualities: first, if preview cues enhance the exploration experience for the Web-like, single column model, then this will be a cheap mechanism for improving existing page designs. If the column or spatial views with preview cues afford significantly greater improvements in experience and efficiency, then this would provide a compelling heuristic for interaction designers to consider an effective means for improving content delivery.

Audio Condition. We also tested when audio cues would be available. In the Early condition, preview cues were available for each label in each level of the hierarchy. In the

Late condition, preview cues were only available when the user reached the final "Piece" level of the hierarchy. The late case simulates the manner in which Web sites such as Amazon.com present audio: only at the selection of the final level of the hierarchy. The early/late audio conditions would help us confirm when audio cues are best suited to be available. Our hypothesis was that Early would be an easy, obvious win, but this condition proved to return the most surprising result.



Renaissance=>Gibbons=>Incidental=>Piano=>Fantasy in C ...

Now playing Gibbons : Fantasy in C Major

-- Fantasy in C Major
Allemande <Italian Ground>
Lord of Salisbury" Pavan & Galliard

Bach:Partita No. 3 in E Major
Beethoven:12 Variations on the Theme See the Cor
Boccherini:Sonata in E-flat Major for 2 Violoncellos
Gibbons:Fantasy in C Major

Figure 2, Column View (left) with Collection Window (right).

RESULTS

Quantitative

The most preferred interface style was the spatial layout. (Table 1).

Interface \/	Pref ->	Most	Least
Spatial Early (SE)		21	0
Spatial Late (SL)		17	1
Temporal Early (TE)		1	20
Temporal Late (TL)		3	21

Table 1: Most and Least Preferred Interfaces

Duration overall that each user spent per interface was significant: women spent less time overall per interface than men (F=6.776, p=.015). While 80% of participants had less than two years of any kind of formal music training, and forty percent of those had no training, there was no correlation among musical training, size of audio collection, or general level of education and either performance or preference.Looking at the number of actions per interface, and time between adding selections, however, did yield significant results.

	SE	SL	TE	TL
clicks:	.45	.80	.62	.65
brushes	.67	.79	.73	.64

(all values are significant at $p < .01$)

Table 2: Brush and Click correlation per interface

Looking specifically at actions in the interface and times between adding selections, we see that participants who took longer to make their selections clicked/expanded more elements in the interfaces (Table 2).

There were significant effects of layout and early/late audio on how many actions participants took, both brushes and clicks. There were more actions taken in the spatial interfaces and in the late interfaces. There was one gender effect for actions, a layout*gender interaction: men clicked and brushed more than women in the spatial interfaces, and in fact women clicked and brushed about equally in both layouts. There was a significant difference between the early and late temporal interfaces for how much people used the “back” action (F=8.276, p=.008). It was used more in the “late” interface. In the temporal layouts there was a negative correlation between age and both duration of use and number of actions (clicks, brushes, adds), which was not present in the spatial layouts (Table 3).

	spatial	temporal
actions	-.08	-.45**
duration	-.15	-.34*

$*p < .05$, $**p < .01$

Table 3. Age correlation for Action/Duration in Temporal Layouts.

Qualitative

All participants reported that preview cues made the process of discovering music enjoyable. Many participants commented on how the preview cues made finding new music "easy." Comments like they "wished [a certain music store] used this to let shoppers find new tunes," or "I want to take this [software] home and use it," were common. Evaluators noted that participants were frequently reluctant to stop playing with the spatial interfaces in particular, checking for "new" pieces to audition. Participants who had no previous experience of this domain, and reported having had "no way" of accessing it before, reported that they discovered new music to enjoy.

Participant comments made clear that they did not enjoy the temporal interfaces. Participants suggested several problems with this approach, but the main one was perceived lack of context: despite the fact that the path to their current location in the hierarchy was consistently available at the top of the interface window – an attribute brought to participants' attention repeatedly through the training process – many users reported that they felt lost. Comments like "I didn't know where I was…I couldn't see the whole path" were common. However, in all cases, users said they preferred having audio cues available. The results show a higher percentage of users (50%) preferred the Early Spatial interface to the Late Spatial version (40%); user comments indicate what were perceived as the strengths and weaknesses in each of these approaches. Users who preferred the Early Spatial interface recorded enjoying being able to get a sense of each area before exploring it. Those who preferred the Late Spatial interface reported enjoying starting one cue playing and having that play while they moved through the domain, until they picked another cue.

ANALYSIS

Preview Cues and Spatial Layouts

That users brushed less in the Late condition than the Early is not surprising: brushing in the Late condition only triggered a preview cue in the final level of the hierarchy. That users clicked less in the Early versions, however, and indeed, spent less time between selection additions, suggests that the preview cues were working as designed: they allowed participants to assess *quickyl*[i] whether or not they *wanted* to click/expand/explore an area of the domain.

While 50% of participants preferred the Early Spatial, 40% preferred the Late spatial. This is a curious finding, since the only information available by clicking through the hierarchy in the Late condition were names or terms that "meant nothing" to the majority of those participants. Such users seemingly preferred to click/hack randomly through the rest of the interface paths just to get to brushing over a set of pieces at the end of the path. This approach seems counter-intuitive. Many who preferred the Late Spatial condition, however, suggested that they would have preferred the Early Spatial if they could have controlled (1) *when* or *whether* a preview cue played and (2) that they could make a preview cue continue to play, uninterrupted, as a selection. This suggests that preview cues effectively enable explorations of unfamiliar domains, but that for preview cues to be strongly effective, users need to be able to control when and how preview cues are available.

Gender Effect and Task Focus

It is not clear at this point how the gender effect on overall shorter use and overall fewer clicks in each interface by women relative to men may be interpreted or leveraged at this time. It does seem to be a finding in concert with studies of women and computer use that show women spend less time "playing" with computers than men, treating them more frequently as tools rather than toys [2], [4]. Thus, even though women reported enjoying using the interface, they gave themselves perhaps less opportunity to go off task.

CONCLUSION

This paper presents the findings on the use of audio preview cues for exploration of a structured hierarchical representation of the classical music domain. We tested audio preview cues comparing several factors of Web-like clients. The study tested two hypotheses: that audio preview cues would improve effectiveness and efficiency of standard, temporal Web-like presentations of audio, and that a spatial layout with audio preview cues would significantly improve user experience for exploring the domain. From the results, we can see that both hypothesis have been validated. Indeed, with respect to the second hypothesis, we saw a significant negative effect of age with the temporal layout that was not evident at all with the spatial layout. Several design heuristics fall out from the work:

- exploration of structured domains representing music can be improved by adding preview cues to the elements of the domain, whether the hierarchy is represented temporally or spatially
- this effect can be significantly enhanced if a spatial layout is used.
- the negative effect between age and temporal representation of hierarchy can be nullified by using a spatial layout.

ACKNOWLEDGMENTS

Tom Rodden, Andy Cockburn, David de Roure and Richard Beale for their comments. This work was funded by CITO: Communications and Information Technology Ontario Center of Excellence.

REFERENCES

1. Brewster, S.A., Wright, P., Edwards,, A.D.N. An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces. In *Proc of CHI '93*, (1993), 222-227.
2. Cherny, L., Weise, E.R., eds. (ed.), *Wired Women: Gender and New Realities in Cyberspace*. Seal Press, NY, 1996.
3. Chimera, R. and Shneiderman, B., An Exploratory Evaluation of Three Interfaces for Browsing Large Hierarchical Tables of Contents. In *ACM Transactions on Information Systems (TOIS)*, (1994), 12.
4. Margolis, J. and Fisher, A. *Unlocking the Clubhouse: Women in Computing*. MIT Press, Mass., USA, 2001.
5. Terry, M. and Mynatt, E.D., Side Views: Persistent, on-Demand Previews for Open-Ended Tasks. In *UIST*, (2002), 71 - 80
6. Thomason, R. (ed.), *Formal Philosophy, Selected Papers of R. Montague*. Yale University Press, 1974.
7. Tzanetakis, G., Automatic Musical Genre Classification of Audio Signals. In *Proc of ISMIR 2001*, (2001).

[i] Humans require only 250ms of audio to determine genre. A similar effect may operate for preference. [7]