BABBLE ONLINE: APPLYING STATISTICS AND DESIGN TO SONIFY THE INTERNET

M. H. Hansen
Bell Laboratories
Murray Hill, New Jersey
cocteau@bell-labs.com

B. Rubin
EAR Studio
New York City, New York
benrubin@earstudio.com

ABSTRACT

A statistician (Hansen) and a media artist (Rubin) investigate the application of statistical methods and sound-design principles to the real-time sonification of Internet communications. This paper presents results from two applications: the sonification of browsing activity on Lucent's Web site, and the sonification of a large number of Internet chat sites in real-time. These experiments suggest new ways to experience the diverse and dynamic data streams generated by modern data networks. As an art-technology collaboration, the project outcomes range from the creation of art installations to the development of practical monitoring platforms. This paper discusses the interplay between these two perspectives, and suggests that each is motivated by a common interest in generating meaningful experiences with dynamic data.

1. INTRODUCTION

Modern work in sonification emerged in the literature on computer-human interfaces and over the last decade has matured into its own field of scientific inquiry. The use of sound in exploring the information hidden in data, the principles and broad application of auditory displays are eloquently described in [10]. Early application areas included real-time monitoring of financial data, medical diagnostics, and even air traffic control systems. Computer simulations also provided extensive data for sonification. Since that point, a virtual explosion has taken place in our ability to collect data relating to human communication and social systems.

Today, almost every aspect of our lives can be "rendered" digitally. Advances in data collection technologies have made commonplace continuous, high-resolution measurements of our physical environment (weather patterns, seismic events, ecological indicators). Equally open to observation are our routine movements through and interactions with our physical surroundings (automobile and air traffic, large-scale land use). In computer-mediated settings, our activities either depend crucially on or consist entirely of complex digital data (financial transactions, accesses to global information systems, Web site and Internet usage). As a reflection of the diversity and variety of the systems under study, these data-based descriptions of our daily lives tend to be massive in size, dynamic in character, and repleat with rich structures.

The advent of these enormous repositories of digital information presents us with an interesting challenge: How can we represent and interpret such complex, abstract and socially important data? In a new collaboration, Ear to the Ground [4], we have begun an exploration into ways of creating experiential encounters with otherwise abstract data streams, especially through sound.[1] In [8], we discuss the broad goals of our collaboration and examine sonification from both an artistic as well as a data analytic perspective. In this paper, we examine the use of auditory displays in understanding large-scale Internet communications.

2. EXAMPLES

2.1. Web site traffic

Every day, large Web sites like Yahoo attract hundreds of thousands of visitors. During active periods, there can be thousands of people accessing data simultaneously from a Web site. For users of information portals like Yahoo, the speed of the servers (as reflected by rapid or sluggish responses) provides the only clue about the number of other people accessing information. While attempts have been made to visually assess browsing patterns in real-time [11, 1], the effectiveness of these displays deteriorates for high-traffic domains. For our first sonification example, we create an ambient display to characterize certain aspects of the activity on a busy Web site.

As you navigate the Web, your browser requests various kinds of data from one or more Web servers. As part of part of the process of delivering content, most Web servers will record information about the visitor and the items they requested. These items include HTML pages, images, Java class files and PostScript documents. The information available to the Web server about each request includes a timestamp, the IP address of the visitor's computer, the type of browser they are using, the URL of the requested item, the "referral page" (the URL that directed the visitor to the requested item), and perhaps a "cookie" to recognize returning visitors. These details are typically stored as a single line of a potentially enormous log file.[2] The data for this experiment came from the Lucent Technologies corporate site, `www.lucent.com`. On a typical day, 60K visitors to this site will generate a 15Mb (compressed) log file, consisting of 700K entries. Given that our interest is on how users navigate the content of a site, we restrict our attention to HTML files, PostScript and PDF documents. All other requests made to the Web server are ignored, reducing the data by a factor of 10. We then further process the data to extract user paths or "visits," where a visit is a contiguous sequence of requests made by a user while browsing the site. Over 70% of the visitors to `www.lucent.com` look at just three pages or less, and hence a minority of the visits exhibit "interesting" navigational patterns.

[1] Ear to the Ground is part of the Arts in Multimedia project cosponsored by Lucent Technologies and the Brooklyn Academy of Music (BAM). The authors gratefully acknowledge the help of project managers Wayne Ashley (BAM) and Marah Rosenberg (Lucent; now with Avaya Communications).

[2] Each line is commonly referred to as a "hit."

The Lucent Web site is built hierarchically, in the sense that pages deeper in the directory tree represent more detailed information than those at shallower levels. At its busiest, there can be as many as 300 people browsing `www.lucent.com`; while during the pre-dawn hours there can be as few as 5 simultaneous visitors. Our sonification is designed to convey qualitative information about site usage, answering questions like:

- Overall, is the site busy or quiet?
- What proportion of the visitors are delving for specific information deep within the site, as compared to those visitors who are "just passing through," glancing briefly at the home page and then moving on?
- How are users distributed across the various content areas of the site?
- Which portions of the site are visited together? What kinds of patterns do we find in user behavior?

We think of this sonification as one possible "background" information stream that can inform content providers, Web designers and even the visitors themselves.

2.1.1. *Sonification design*

Our audio display makes use of the hierarchical structure of the content offered by `www.lucent.com`. First, a unique pitch was used to identify each of five high-level subdomains within the site: `/micro`, representing Lucent's microelectronics design and manufacturing business (now Agere Systems); `/enterprise`, for the enterprise systems and software business (now Avaya Communications); `/minds`, a corporate introduction to Bell Labs research; `/press`, a collection of press releases and investor information; and `/search`, the local search engine for the site.

The total number of visitors accessing any information from a subdomain affects the loudness and tonal balance of a low-register drone at the associated pitch. Visitors requesting content deeper in the site are represented by higher-pitched pulsing tones (separated by one or two octaves from the base pitch for the subdomain): the faster the pulse, the more people are accessing that area, and the greater the proportion of high-register sounds, the more detailed the content. By assigning well-separated pitches to each subdomain, shifts in activity both within and between the areas can be heard. In Table 1 we present a simple mapping of data collected by the Lucent Web server to a continuously time-varying vector of usage statistics. In the category of Overall browsing, we count any visitor accessing content pages (HTML, PostScript or PDF) from the indicated subdomain. A Mid-Level access is a request for content two or more directories down. Simple examples are `/micro/K56flex/index.html` (information on a brand of 56K modem) and `/press/0101/010118.nsb.html` (a press release for January 18, 2001). The final category, Deep browsing, refers to pages that are four or more directories down in the tree. One example is a paper from the April/June 2000 issue of the Bell Labs Technical Journal, located at `/minds/techjournal/apr-jun2000/pdf/paper02.pdf`.

Then, the resulting 15 values in Table 1, A1–E3, were mapped to sound as follows:

Overall activity Measured by A1–E1, voiced with a low-register drone. The aggregate number of visitors accessing information within each of the five areas modulates the loudness of each of the five pitches.

	/micro	/enterprise	/minds	/press	/search
Overall	A1	B1	C1	D1	E1
Mid-Level	A2	B2	C2	D2	E2
Deep	A3	B3	C3	D3	E3

Table 1: Mapping used for Web site traffic example. Overall activity records the movements of all users; Mid-Level counts users 2 or 3 directories into the site; Deep browsing consists of users 4+ directories down.

Mid-Level browsing Measured by A2–E2 and assigned a rhythmic middle-register tone pulse; pulse loudness and repetition speed rises and the timbral brightness increases as the volume of mid-level browsing increases. There are five independent pulses, each at a different fixed pitch, representing the five content areas.

Deep browsing Measured by A3–E3 and made audible via rhythmic high-register "ting" sounds (plucked steel string samples). Loudness and repetition speed rises as the volume of deep browsing increases. Again, there are five independent "ting" sounds, each at a different fixed pitch, representing the five content areas.

We used pitch groups that were consonant, and for the sounds that incorporated rhythm (A2–E3), the phase and frequency of each pulse in the matrix varies independently, yielding a sound with a changing rhythmic texture but no fixed beat.

The purpose of this sonification is to make interpretable the activities of users on a Web site. Therefore, the stream of hits being processed by a Web server (reduced to include only the HTML, PostScript and PDF documents) needs to be transformed to extract meaningful user-level data. A real-time monitoring tool was developed that maintains a bank of active visits (recording separately the activities of all the people browsing the site at a given time) and updates various statistics with each user request. When cookies or some other authentication mechanism allows us to recognize returning visitors, the monitor will update a more complicated user profile that encapsulates previous browsing patterns. Our traffic sonification as described above takes as input the location of each visitor within a site at a given point in time. When constructing more elaborate sound displays, our design will continue to focus on user activities, drawing more heavily on the statistics culled by the monitoring tool. This emphasis distinguishes our approach from sonification methods that assess Web server performance by making audible statistics relating to server load, HTTP errors, and agent types [**?**].

2.1.2. *Impressions and extensions*

We have created three audio examples for the activity on the Lucent site. Our data were captured on November 11, 1999 and we created sonifications of the traffic at 6:00 am, an extremely slow period for the site; noon, a relatively active time; and 2:30 pm, the point at which the site was busiest. The samples are located at our project Web site [6]. Even with this relatively straightforward mapping, one finds compelling patterns. For example, the affinity between the `/enterprise` subdomain and the `/search` facility can be heard as the pulses for these areas rise and fall together.[3]

[3] While clearly audible, these shifts can really only be precisely associated with areas after a certain amount of experience with the mapping.

Also, when comparing moderately active to extremely busy periods, we find that the number of people digging deep into the site is not a fixed fraction of the total number of visitors. That is, the volume of the low-register drones exhibits much more variation than the components for the other two categories of accesses. Each of these effects can be verified by examining the logs, reinforcing the usefulness of our sonification as a tool for constructing hypotheses about site traffic.

As mentioned at the beginning of this section, Web browsers offer a rich set of data about the visitor when requesting data from a server. This display makes use of only the most basic information about a visit, namely the depth of pages accessed. In ongoing work, we are augmenting our sonification with extra features derived both directly from the server data as well as from statistical navigation models [12] fit for the Web site under study. So far, we have found that such extensions are most effective when developed in the context of a particular monitoring application. For example, an extended version of this ambient display can aid system architects of large, Web hosting services understand cache performance and can aid in server provisioning. Another extension will make greater user of our navigation models and can help designers and usability engineers better architect Web sites. We will report on these and other developments through the project Web site [4].

2.2. Chat rooms and bulletin boards

At any given moment, tens of thousands of real-time conversations are taking place across the Internet on public forums, bulletin boards and chat sites. To imagine making these conversations simultaneously audible evokes an image of uproarious babble. And yet, in the aggregate, this massive stream of live communication could exhibit rich thematic structure. Can we find a meaningful way to listen in to so many conversations, rendering them in a way that is comprehensible and not overwhelming?

In some sense, a byproduct of our Web traffic sonification is the creation of a kind of community from the informal gathering of thousands of visitors to a given Web site. Traditionally, informational Web sites like `www.lucent.com` have provided us with very little sense of the other people who are requesting data from the server. To attract and retain visitors, however, many commercial sites recognize the potential of the Web to form social as well as informational networks. As a result, Web-based forums, message boards and a variety of chat services are common components of current site designs. While Internet Relay Chat (IRC) has been a widely used standard since the inception of the Internet, the popularization of the Web has resulted in a virtual explosion of chat applications.[4] For example, `www.yahoo.com` (a US-based Web portal) offers hundreds of separate chat rooms attracting tens of thousands of visitors a day. Specialized sites like `www.style.com` (the homepage for Vogue magazine) or `www.audiworld.com` (an resource for Audi owners) have also found their message boards to be the most frequently accessed parts of their domains.

To get a sense of the amount of content that is available in these dynamic formats, we examined sites contained in the DMOZ Open Directory [3], an open source listing of over 2 million Web sites compiled and categorized by 33,000 volunteer editors. From the November 20, 2000 image of the directory, we counted 36,681 separate sites offering some kind of chat, bulletin board or other public forum. While we did not examine the activity on all of these sites, the number is staggering. If we include other peer-to-peer communication technologies like instant messaging,[5] the amount of dialogue taking place on the Web at any point in time is almost unfathomable. The goal of our second sonification is to make interpretable the thousands of streams of dynamic information being generated on the Web. In so doing, we attempt to characterize a global dialogue, integrating political debates, discussions of current events, and casual exchanges between members of virtual communities.

2.2.1. Content monitors and the statistics engine

Our starting point is text. Albeit diverse in style and dynamic in character, the text (or transcript) of these data sources carries their meaning. Therefore, any auditory display consisting only of generated tones would not be able to adequately represent the data without a very complex codebook. The design of our sonification then depends heavily on text-to-speech (TTS). As with the traffic example in the previous section, we think of the audio output as another background information stream. The incorporation of spoken components in the sound design poses new challenges, both practical and aesthetic. For example, simply voicing every word taking place in a single chat room can produce too much text to be intelligible when played in real-time and can quickly exhaust the listener. Instead, we build a hierarchical representation of the text streams that relies on statistical processing for content organization and summarization prior to display.

Before considering sonification design, we first had to create specialized software agents that would both discover new chat rooms and message boards, as well as harvest the content posted to these sites. (See Figure 1 for an overview of our system architecture.) Most bulletin boards and some chat applications use standard HTML to store visitor contributions. In many cases, a specific login name is required to gain access to the site. For these situations, we constructed a content agent in Perl, as this language provides us the most convenient platform for managing access details (like cookies). The public chat rooms on sites like `chat.yahoo.com` can be monitored in this way. For IRC we built a configurable Java client that polls a particular server for active channels. Web sites like `www.cnn.com` (a popular news portal) and `www.financialchat.com` (a financial community hosting chat services for day traders) offer several IRC rooms, some of which are tightly moderated.

In addition to collecting content, each monitoring agent also summarizes the chat stream, identifying basic topics and updating statistics about the characteristics of the discussion: What percentage of visitors are contributing? How often to they contribute and at what length? Is the room "on topic," or are many visitors posting comments on very different subjects? Topics are derived from the chat stream using a variant of *generalized sequence mining* [7] that incorporates tags for the different parts of speech. While the exact details are beyond the scope of this abstract, a generalized sequence is a string of words possibly separated by a wildcard, "*". For example, if we let A, B and C denote specific "contentful" words (say, nouns, adjectives and adverbs), then ABC, $A*BC$ and $A*B*C$ are all generalized sequences. The wildcard allows us to identify "Gore * disputes * election" from the sentences

[4] RC was developed by Jarkko Oikarinen in Finland in the late eighties, and was originally intended to work as a better substitute for `talk` on his bulletin board.

[5] AOL alone records tens of millions of people using their instant messaging service each month.

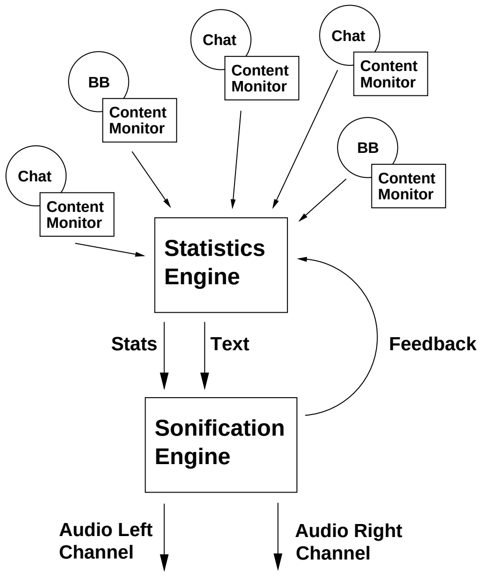

Figure 1: System architecture overview. A large number of content streams (Chat = chat rooms; BB = Bulletin boards) are gathered by specialized agents that transmit them in a homogenized format to the statistics engine. The statistics engine then distills the streams into a much smaller number of configurable text streams as well as a number of descriptive vectors. The sonification engine then "plays" these text and data streams. The entire systems operates in real-time.

"Vice President Gore filed papers to dispute the presidential election," "Aides for Gore indicated that he has every reason to dispute the election", and "Gore is still deciding whether or not to dispute the election".

As many posts to chat rooms contain spelling mistakes and incorrect grammar, assigning words to different parts of speech is error-prone. However, unlike most applications of statistical natural language processing, our content monitors update their summaries each time new material is posted and downweight older contributions. Because our sonification renders these sources in real-time, small mistakes have little effect on the power of the overall display to convey the ideas being discussed.

Each of the content monitors are periodically polled by the statistics engine (see Figure 1). This Java-application clusters the different chat rooms and bulletin boards based on their topic and numerical summaries. As the topic in a room changes over time, the statistics engine is constantly updating and reformulating cluster membership. Because a content stream can in fact support a number of simultaneous discussions (the threads of a bulletin board, say), we employ a soft-clustering technique. In our initial work, we have used a mixture-based scheme that determines the number of clusters with an MDL (Minimum Description Length) criterion [9]. Each room is then assigned a probability that it belongs to the different groups. This model also provides for topic summarization at the cluster-level. Next, a stochastic framework was developed to sample representative sentences posted to the chat or bulletin board. When a discussion is extremely unstructured, this selection is essentially random sampling from all the contributions added to the chat since the last polling point. In addition to textual data streams, the statistics engine is also responsible for communicating the various ingredients for the display to our sonification engine, Max/MSP [2] (see Figure 1). We have adopted the Open Sound Control [13] protocol from Center for New Music and Audio Technologies to transfer data between the statistics engine (running on a Macintosh with LinuxPPC) and the sonification engine (running on a Macintosh with OS/9).

2.2.2. *Sonification design*

As with the previous example (Section 2), our goal is to create a sonification that is both communicative and listenable. Here we face the additional challenge of incorporating verbal content. With TTS annotations, it becomes more difficult to intelligibly convey more than one layer of information through the audio channel. Our design incorporates spatialization, pitch and timbral differentiation, and rhythm to achieve clarity in the presentation of the hierarchically structured data coming from the statistics engine.

The auditory display cycles through topic clusters, spending relatively more time on subjects being actively discussed by the largest numbers of people. Each different topic is assigned a different pitch group, reinforcing subject changes when they occur. For each cluster, the statistics engine sends three streams of information to the sonification engine:

Topics A continuously updated list of up to ten "topics" (the most frequently appearing words and phrases – generalized sequences – mined from the multiple chat streams associated with the given cluster; the number of topics is configurable, but ten was chosen based on timing considerations);

Content samples A selection of sample sentences, identified by the statistics engine as typical or representative, in which these topics appear;

Content entropy A vector that represents the changing level of entropy in the source data.

The topics are spoken by the TTS system[6] at regular intervals in a pitched monotone, and are panned alternately hard left and hard right in the stereo field, creating a sort of rhythmic "call and response." The sample sentences are panned center, and rendered with limited inflection (as opposed to the pitched monotone of the topics). The tonal, rhythmic and spatial qualities of the topics contrasts sufficiently with the sample sentences to create two distinctly comprehensible streams of verbal information.

The entropy vector controls an algorithmic piano score. When entropy is minimal and the discussion in the chat room or bulletin board is very focused on one subject, chords are played rhythmically in time with the rhythmic recitation of the topics. As entropy increases and the conversations diverge, a Gaussian distribution is used to expand the number, range and dynamics of notes that fall between the chords. With this audio component, one can easily differentiate a well-moderated content source from a more free-form, public chat without distracting from the TTS annotations. The piano score also serves a secondary function as an accompaniment to the vocal foreground, enhancing the compositional balance and overall musicality of the sound design.

2.2.3. *Sample sonification and impressions*

On our project Web site [5], we have a sample chat room sonification that cycles through three topics. In this sound file, we are

[6] The built-in MacOS TTS capability controlled by Max/MSP.

listening to the output of only three content monitors. Hence, by design, each topic is confined to a single site. The first portion of this example (ending at 1:47 into the sample) concerns the recent recall of Bridgestone tires and was based on a `www.cnn.com` chat room. This discussion was heavily moderated and hence the backing piano score frequently reduces to a simple rhythm. For our second topic (from 1:47 to 3:21 of the sample) we recorded chat exchanges on `www.financialchat.com` one morning when Yahoo's stock opened low. In this example, we hear day traders frantically exchanging predictions about when Yahoo's stock will "bounce." The final topic in this sample (from 3:21 to the end) is again from `www.cnn.com` and treats a recent strike by the Screen Actor's Guild and the American Federation of Television and Radio Artists. This chat room was much less moderated than the previous CNN chat, and the backing piano score reflects that.

Although this example does not make full use of the clustering capabilities of the statistics engine, the essence of our sonification design is clear. The audio display provides an informative and accessible representation of dynamic, textual content. The topic and content sample streams are easy to separate, and when placed in the background, call our attention to important new subjects being discussed on the Web.

2.2.4. Applications and Extensions

Our sonification provides an audible interface to the (now) massive amount of dynamic content available on the Web. Given the preprocessing that takes place in the content monitors and the statistics engine, a simple extension is to provide search-like functionality. A user can register interest in a certain topic and "tune" our display to present only rooms where this subject is being discussed. The necessary ingredients to implement this feature are all currently available in the statistics engine. Similarly, one can easily restrict the sites that are used for the display. When a new subject appears that draws the user's interest, it is also trivial to add a feature that would direct the user's browser to one or more chats associated with the topic. As a final extension, we have provided the content monitors with a configurable list of Web sites that can be used to help disambiguate elements in the chat stream. For example, the day traders speak in ticker symbols. Providing the content monitor with the URL for the ticker symbol look-up service offered by Yahoo allows the content monitor to weave not only company names but also recent company-related headlines directly into the stream fed to the statistics engine.

While we have focused mainly on chat and bulletin boards, this technology can be applied in other settings. We have begun collaborating with the designers of a natural language interface for Web-based help systems. Here, we give voice to the hundreds of simultaneous conversations taking place between Web site visitors and the automated help system. A similar display can be imagined for other natural language interfaces, including search engines like AskJeeves (`www.jeeves.com`). In general, the practical applications of this summarization and auditory display tool abound.

3. CONCLUSION AND COMMENTS ON COLLABORATIVE RESEARCH

The two applications outlined in this paper are the first outcomes of a collaboration sponsored by Bell Laboratories and the Brooklyn Academy of Music under the Arts in Multimedia project (AIM). The goal of AIM is to bring together researchers (in this case a statistician) and artists (in this case a sound artist), with the objective of advancing our separate agendas through collaborative projects. Our work together is predicated on the notion that sophistication both in data treatment and aesthetics are crucial to the successful design of audio displays. Thus, in each of our examples, we have endeavored to create a result which communicates information clearly, yet at the same time sounds well composed and appealing. Moving forward, it is our intention to apply these techniques both to practical applications, and also to create a series of artworks. These artworks will use our sonification techniques to establish a series of real-time listening posts, both on the Web and in physical locations. The listening posts will tap in to various points of interest on the Internet, using sound to reveal patterns and trends that would otherwise remain hidden.

In terms of applications, we are exploring the use of sonification to support the design, provisioning and monitoring of communication networks. A network operations center (NOC), for example, routinely receives clues about the health of the system in the form of text messages generated by routers and switches. An audio display installed inside a NOC can act as an early warning system for approaching bottlenecks as well as aid in troubleshooting. By continued exposure to the sound of a "normally" functioning network, operators will be alerted to system changes that could signal problems.

Art emerges unexpectedly from experimentations with new statistical methods or considerations involving practical applications; and new tools for data analysis and modeling develop in response to artistic concerns. Each of us continues to be surprised by the connections that emerge from rethinking familiar problems in a new context. Through our project, we hope to illustrate both the value of art-technology collaborations as well as their necessity, especially when finding meaning in complex data.

4. REFERENCES

[1] Visual insights. `www.visualinsights.com`.

[2] Cycling74. Max/msp. `www.cycling74.com`.

[3] Open directory project. `www.dmoz.com`.

[4] Ear to the ground. `cm.bell-labs.com/stat/ear`.

[5] Ear to the ground, chat example. `cm.bell-labs.com/stat/ear/chat.html`.

[6] Ear to the ground, web traffic samples. `cm.bell-labs.com/stat/ear/samples.html`.

[7] W. Gaul and L. Schmidt-Thieme. Mining web navigation path fragments. In *Proceedings of the Workshop on Web Mining for E-Commerce – Challenges and Opportunities*, Boston, MA, August 2000.

[8] M. H. Hansen and B. Rubin. The audiences would be the artists and their life would be the arts. *IEEE MultiMedia*, 7(2), April 2000.

[9] M. H. Hansen and B. Yu. Model selection and the principle of minimum description length. *Journal of the American Statistical Association*. To appear.

[10] G. Kramer. An introduction to auditory display. In G. Kramer, editor, *Auditory Display*. Addison-Wesley, 1994.

[11] N. Minar and J. Donath. Visualizing the crowds at a web site. In *Proceedings of CHI 99*.

[12] R. Sen and M. H. Hansen. Predicting a web user's next request based on log data. Submitted to ASA Student Paper Competition.

[13] M. Wright. Open sound control. `cnmat.berkeley.edu`.