Rasika Ranaweera*
Michael Cohen
Spatial Media Group
University of Aizu
Aizu-Wakamatsu
Fukushima 965-8580
Japan

Michael Frishkopf
Dept. of Music, Faculty of Arts
University of Alberta
3-67 Fine Arts Building
Edmonton, AB T6G 2C9
Canada

Narrowcasting and Multipresence for Music Auditioning and Conferencing in Social Cyberworlds

Abstract

We describe a musical cyberworld, Folkways in Wonderland, in which avatar-represented users can find and listen to selections from the Smithsonian Folkways world music collection. When audition is disturbed by cacophony of nearby tracks or avatar conversations, one's soundscape can be refined since the system supports narrowcasting, a technique which allows information streams to be filtered. Our system supports two different kinds of sound sources: musical selections and avatar conversation (voice-chat). Narrowcasting for music enables aesthetic focus; narrowcasting for talk enables cognitive focus. The former is required for dense presentation of musical sound, the latter for virtual worlds in which many avatars are expected to be able to interact. An active listener can fork self-identified avatars using a novel multipresence technique, locating representatives at locations of interest, each clone capturing respective soundscapes, controlled using narrowcasting functions {`self, non-self`} × {`select (solo), mute, deafen, attend`}. Likewise one can participate in a conference and at the same time join a global tour of music. Our music browser is architected to use MX: IEEE 1599, a comprehensive, multilayered, music description standard. Using our cyberworld as a virtual laboratory, we evaluated the effectiveness of narrowcasting when auditioning music and conferencing. Experimental results suggest that narrowcasting and multipresence techniques are useful for collaborative music exploration and improve user experience. We also got positive feedback from the participants regarding narrowcasting representations, variously based on colors, symbols, and icons.

1 Introduction

As both immersive virtual environments and online music networks become increasingly popular, it behooves researchers to explore their convergence: groupware music browsers populated by figurative avatars representing distributed users. Collaborative virtual environments (CVEs) offer immersive experiential network interfaces to online worlds and media. Contemporary research and systems address a broad range of needs regarding searching,

*Correspondence to d8121104@u-aizu.ac.jp.
Supplementary content for this article can be found online at http://www.mitpressjournals.org/doi/suppl/10.1162 /PRES_a_00232

Presence, *Vol. 24, No. 3, Summer 2015*, 220–242
doi:10.1162/PRES_a_00232

locating and visualizing for enhanced music audition but generally lack VR-style interfaces.

Folkways in Wonderland[1] (hereafter FiW) is a novel application for listening to music (Ranaweera, Cohen, & Frishkopf, 2013; Ranaweera, Frishkopf, & Cohen, 2011), in which users can find and audition selections from the Smithsonian Folkways world music collection[2] inside Open Wonderland,[3] a pure Java framework (originally developed as Project Wonderland by Sun Microsystems, now supported by an independent foundation) for creating collaborative 3D virtual worlds (Kaplan & Yankelovich, 2011) like Second Life. Avatar-represented users can meet in its synthetic spaces and actively participate in conferences, not only listening and speaking but also performing shared object manipulation and other group tasks. One can explore music in several ways in FiW, as suggested by Figure 1. Wonderland features both audio conferencing and text-chat. An exotic multipresence feature (Cohen, 2000) allows forked presence for radically flexible avatar deployment. When audition of music is disturbed by cacophony of nearby tracks, narrowcasting operations can be invoked to refine one's soundscape (Alam, Cohen, Villegas, & Ahmed, 2009; Fernando, Adachi, Duminduwardena, Kawaguchi, & Cohen, 2006).

A typical conferencing configuration consists of several avatars, representing distributed users, moving around a shared space, with sources associated with each user's voice and sinks associated with each user's ears. Research has been conducted to improve audio quality in conferences (Yankelovich, Kaplan, Provino, Wessler, & DiMicco, 2006), address problems that impact effectiveness (Yankelovich et al., 2006, 2004), and explore how audio improves communicative capability for interesting and useful shared media systems (Ackerman, Starr, Hindus, & Mainwaring, 1997). Analyzing existing conferencing systems, we pose the following questions: How does one effectively listen only to a particular song when cacophony might distract, or listen to a particular speaker while excluding distracting session participants, or prevent one's voice from being delivered to other members? We explore how narrowcasting and multipresence operations can be effectively engineered to answer these questions.

1. www.youtube.com/watch?v=5Nmyc01qZmI
2. www.folkways.si.edu
3. openwonderland.org

2 Related Research

Our system integrates various functionalities that are typically offered only separately by more spatialized programs. In this section, we respectively consider several classes of such focused applications.

2.1 Ethnomusicology of, and through, Cyberworlds

Ethnomusicology can be defined as a branch of human sciences that studies music in its social-cultural contexts, especially the ways in which people interact through shared musical experience and discourse about music, and how music thereby facilitates the emergence of social groups and communities (Nettl, 2005). Methodologically, ethnomusicology centers on qualitative research, mainly ethnographic fieldwork relying upon participant-observation and informal interview techniques (Fine, 2001; Barz & Cooley, 2008). Variables typically cannot be controlled.

Cyberworlds open new avenues for ethnomusicological research. A cyberworld is a social space, with important ramifications for real social interaction and culture formation, and thus of tremendous concern to many scholars working in the social sciences and the humanities (Kong, 2001; Taylor, 1997). As social cyberworlds incorporating music become increasingly prominent, the task of studying them falls to ethnomusicology. The ethnomusicologist seeks to comprehend social dimensions of musical cyberworlds, to enhance their musical functions, and to further understand music in social-cultural contexts more generally, since cyberworlds are closely related to the real world, and impact it strongly.

Now it is not only possible to build a cyberworld as the focus for ethnomusicological research, but necessary as well, since cyberworlds represent contemporary musical reality. Musical cyberworlds can enable a new paradigm for ethnomusicology. Instead of observing

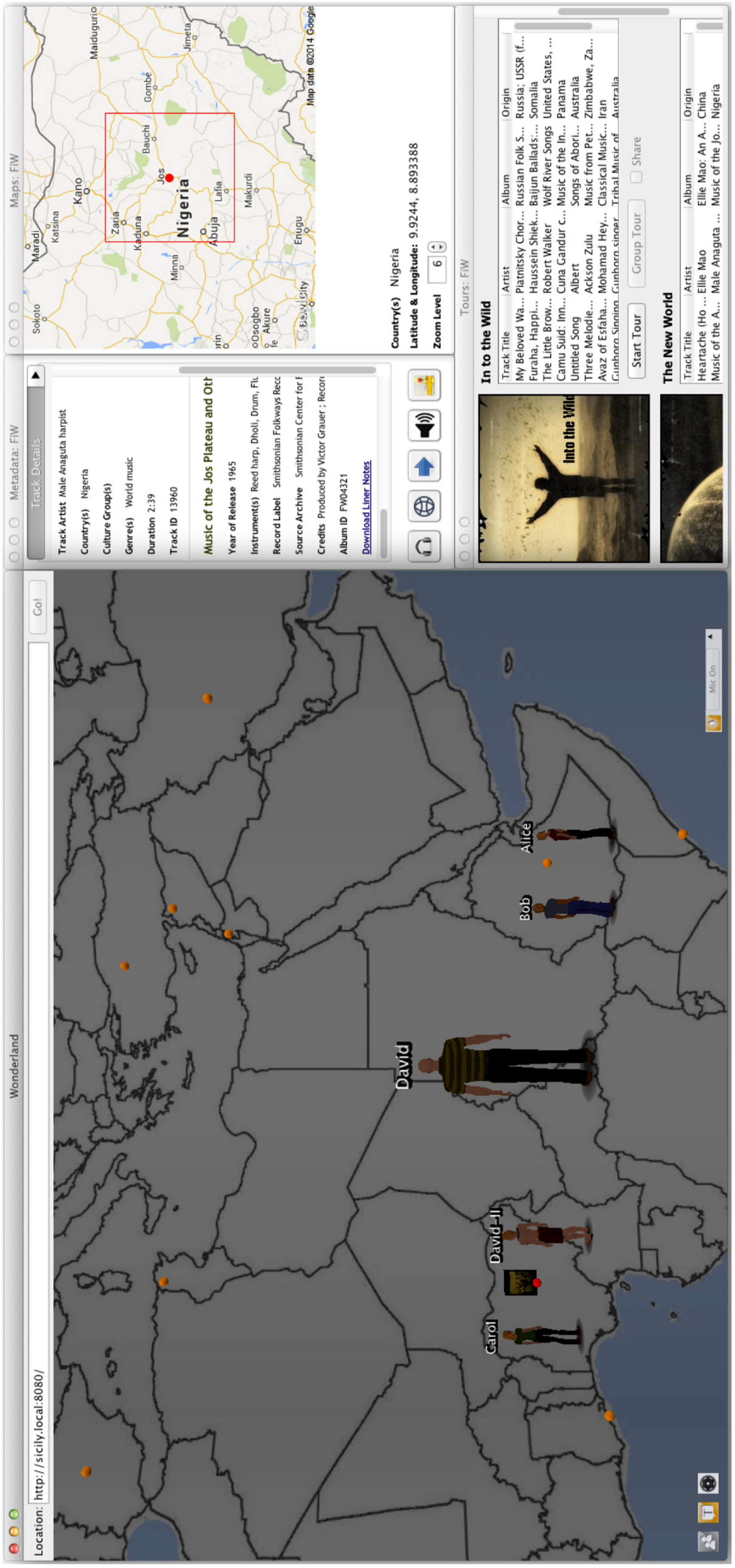


Figure 1. *A typical Folkways in Wonderland session: In the upper-center window, a user browses metadata for a selected track (located in Nigeria) in the left window. Buttons allow the user to view liner notes, listen via virtual headphones (excluding competing sounds), find the track on the Smithsonian Folkways website, teleport to the origin of the track, search for other tracks, or view the track location on a zoomable Google map (upper right). The user may also embark on a tour, using a window like that shown in the lower right. A user with multiple instances of self can audition music in different places at once (as by David in center and David-II next to red dot). The metadata window shows details of the musical track auditioned by the focal avatar.*

musical interactions in the world-as-encountered, one can study a virtual world whose parameters are, to a great extent, under the researcher's control. Such a cyberworld becomes a laboratory for ethnomusicological research, a means of better understanding other musical cyberworlds, and providing, for the first time, a controlled environment for the ethnomusicological study of virtual community.

2.2 Music Information Retrieval

Finding a particular recording is generally supported by traditional search interfaces via metadata (Hughes & Kamat, 2005), but there is a growing need for improving search techniques via different information retrieval strategies. Damm, Fremerey, Kurth, Müller, and Clausen (2008) introduced a novel user interface for multimodal (audio-visual) music presentation as well as intuitive browsing and navigation. Many music search engines exist. For instance, Musipedia[4] offers melody search functions. Similarly, the Music Ngram Viewer[5] encodes songs for look-up. The Folktune Finder[6] also has melody and contour search. MusicSim (Chen & Butz, 2009) uses audio analysis techniques and user feedback for browsing and organizing large music collections. Although most such applications and interfaces facilitate locating music and visualizing collections, it is also important to take into account what information is desired and how that information will be used after its retrieval (Downie, 2002). Kuhn, Wattenhofer, and Welten (2010)'s mobile music player incorporates several smart interfaces to access larger personal music collections and visualize content using similarity maps.

2.3 Spatial Sound Diffusers

Since FiW features immersive audition and spatial audio, it is also related to spatial sound diffusers, although most of the ones we know about are for virtual concerts, with at least logically collocated musicians. That is, even though the tracks might have been separately recorded, the pieces of music are presented as if in a concert venue. One could argue that by collecting global tracks into a single space as we have done, the various musicians are logically collocated, but the distinction is that the songs were compiled by us (or Smithsonian Folkways, or whomever) long after the tracks were recorded. Tracks in almost all integral songs are made to be listened to together, whereas collections of music are compiled "post facto," after the fact.

Mention must also be made of online games, which increasingly feature spatialized voicechat. Some games (for example, those in the Splinter Cell and Thief series) even use the amplitude of positional player sounds to alert NPCs (AI-driven non-player characters), further encouraging "stealth" operations.

Funkhouser, Min, and Carlbom (1999) describe an acoustic model to locate moving sources and receivers in a distributed VE. A virtual museum application demonstrated in Naef, Staadt, and Gross (2002) illustrates design principles and practical implementation issues for audio rendering.

2.4 Social (Distributed) Music Audition

Many research systems have been developed for music consumption, both stand-alone and distributed, of which perhaps Frank, Lidy, Peiszer, Genswaider, and Rauber's (2008) work is representative. Such groupware systems are instances of collaboration technology for synchronous but distributed (not collocated) sessions. Boustead and Safaei (2004) compare various architectures for delivery of streamed audio, including techniques for optimization based on similarity of distribution of avatars in a virtual space with that of human players in the real world.

The major commercial labels haven't yet capitalized on the way many people really consume, share, and experience digital music. Napster anticipated distributed music sharing, but presented basically an asynchronous experience. Many people, especially younger listeners, enjoy music through networked music audition services. Such systems often offer social media features,

4. www.musipedia.org
5. www.peachnote.com/info.html
6. www.folktunefinder.com

generalized as "groupware" among human–computer interaction researchers and scientists. For instance, Last.fm promotes "scribbling," publishing one's music-listening habits to the internet, to monitor when and how often certain songs are played, but such journaling is an asynchronous practice. SongPop[7] is a social multi-player online music identification game, in which players complete against others in real time to identify song snippets. (In 2012 it was the highest-rated game on Facebook.[8]) Both Shazam[9] and SoundHound[10] feature realtime maps of music neighbors and what other users are listening to as My Music and Explore, respectively.

Maybe in the future, online communities with avatars, currently mostly used for interactive 3D social interaction, will be used for browsing media. The main example of such a not-quite-mainstream environment is Second Life, which allows virtual concerts, and runs from a distributed network of 40,000 servers (but might eventually be eclipsed by its founder's subsequent venture, High Fidelity[11]). Although network and processing latency prevents totally satisfying realtime experience for globally distributed online musicians, prerecorded tracks (such as those served by Music in Wonderland) can be streamed for "concert-like" experience. Boustead, Safaei, and Dowlatshahi (2005) consider server-side optimization of compiled soundscapes, including accommodation of limited bandwidth and soundscape compilation distribution to clients for load-sharing. For a perfect network, running at the speed of light, packets would take about 100 ms to get halfway around the world ("worst best case"). This delay would be fine for conversations, but probably distractingly audible for distributed performance.

Folkways in Wonderland (FiW) has no music search features, besides text-based search on its tracks' metadata tags. What distinguishes Folkways in Wonderland from the aforementioned applications is its collaborative music audition, integrated text-chat, voice-chat, spatial music rendering, and figurative presence and natural spatial navigation, for real time, interactive, dynamic consultation for immersive experience.

Our system is an instance of social music browsing, or distributed music audition, allowing collaborative music exploration and ethnomusicological "music safaris." It realizes some of Alan Lomax's vision of a Global Jukebox (Lomax, 1997). Crossing groupware social audition with music information retrieval yields collaborative music information seeking, which is what FiW is intended to foster.

7. www.songpop.fm
8. en.wikipedia.org/wiki/SongPop
9. www.shazam.com
10. www.soundhound.com
11. highfidelity.io

3 Folkways in Wonderland ("FiW")

Our musical cyberworld is populated with track samples from Folkways Recordings, founded by Moses Asch and Marian Distler in 1948. Folkways was directed by Asch until his death in 1986, and thereafter published and curated by Smithsonian Folkways, the non-profit record label of the Smithsonian Institution, the national museum federation of the U.S., headquartered in Washington, D.C. Artistic, geographic, and generic information describing the Folkways music collection is curated (by us) in XML (Extensible Markup Language) format. Our music browser was architected to use MX: IEEE 1599 (Baggi & Haus, 2009; Baggi & Haus, 2013), a comprehensive, multilayered music description standard.

To enter the Folkways in Wonderland cyberworld, a user connects to a public server hosted over the internet using a web browser and downloads the extended Wonderland client. After authentication, one can explore music in multiple ways, including visually (dereferencing placemarks and bookmarks or browsing a map), auditorily (entering a track's "nimbus" or sonic sphere, as described in Greenhalgh & Benford, 1995), and socially (through discussions with other users), as seen in Figure 1. The system is collaborative: multiple avatars can enter a space, audition track samples, and contribute their own sounds (typically speech) to the mix via voice-chat. By default avatars can directionally hear within a space all sound sources (musical tracks and sounds produced by other avatars), attenuated for distance and mixed according to a spatial sound engine that emulates

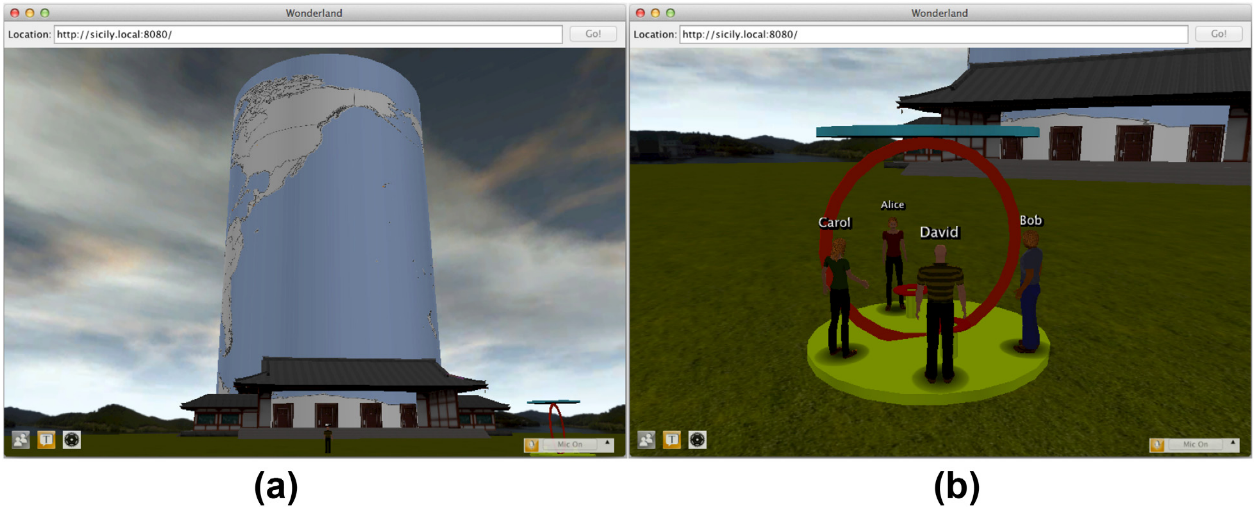


Figure 2. *Exploring Folkways in Wonderland: Image on the left (a) shows exterior view of the cylindrical map bounding the main cyberspace. A discussion park allows quiet conversations among avatars populating the cyberworld; the park is shown on the right (b).*

binaural hearing. Avatar-represented users are free to explore the cyberworld (as shown in Figure 2[a]), using keyboard and mouse/trackball/trackpad controls to navigate through the surrounding virtual environment (including a building and a verdant park, as shown in Figure 2[b]), while interacting with one another and listening to music. When tracks are near each other, overlapping nimbus projections create a dense mix, which is appropriate when exploring an entire collection by moving one's avatar among distributed songs. However, in order to listen to a particular track, an auditory focus function is available which causes other musical streams to be blocked. The `select` (`solo`) function (as seen among the controls in the panel at the bottom of Figure 1 top center and Figure 9), iconified by a headphone symbol, is auto-released when a different track is played or corresponding button is pressed. (Narrowcasting is explained more fully in Section 4, along with the related idea of multipresence.)

3.1 Architecture and Implementation

Wonderland uses a client-server model (Crisostomo, Safaei, & Platt, 2004) with various networking protocols for different data types (Kaplan & Yankelovich, 2011). TCP (Transmission Control Protocol) is used for communicating object properties and positions, while SIP (Session Initiation Protocol) and RTP (Real Time Protocol) are used for audio communication. The Wonderland suite integrates several services that can be distributed across multiple machines for scalability (Gardner, Gnem-Gutirrez, Scott, Horan, & Callaghan, 2011). The Darkstar game server (Waldo, 2008) provides a platform for Wonderland to track the frequently updated states of objects in a session. jVoiceBridge, a pure Java audio mixing application, communicates directly with the Darkstar server, providing server-side mixing of high-fidelity, immersive audio (Kaplan & Yankelovich, 2011).

As shown in Figure 3, an OpenWonderlandClient (OWLClient, near the lower right of the diagram) connects to a server for messaging with other clients, while audio mixing is performed at jVoiceBridge, which is built into the Wonderland server (top). The `FiWCellMO` server (top left) generates a list of track samples by parsing an XML database using a shared `FiWTrackInfoParser` when the server starts up. A `Spatializer` (top right) creates spatialized audio for the track list, and `VoiceManager` (also top right) is responsible for handling audio communication between clients. A `Softphone` (right upper center) at each client connects to the voice server. An `AudioManagerClient` (right lower center) controls stereo audio and provides narrowcasting operations for

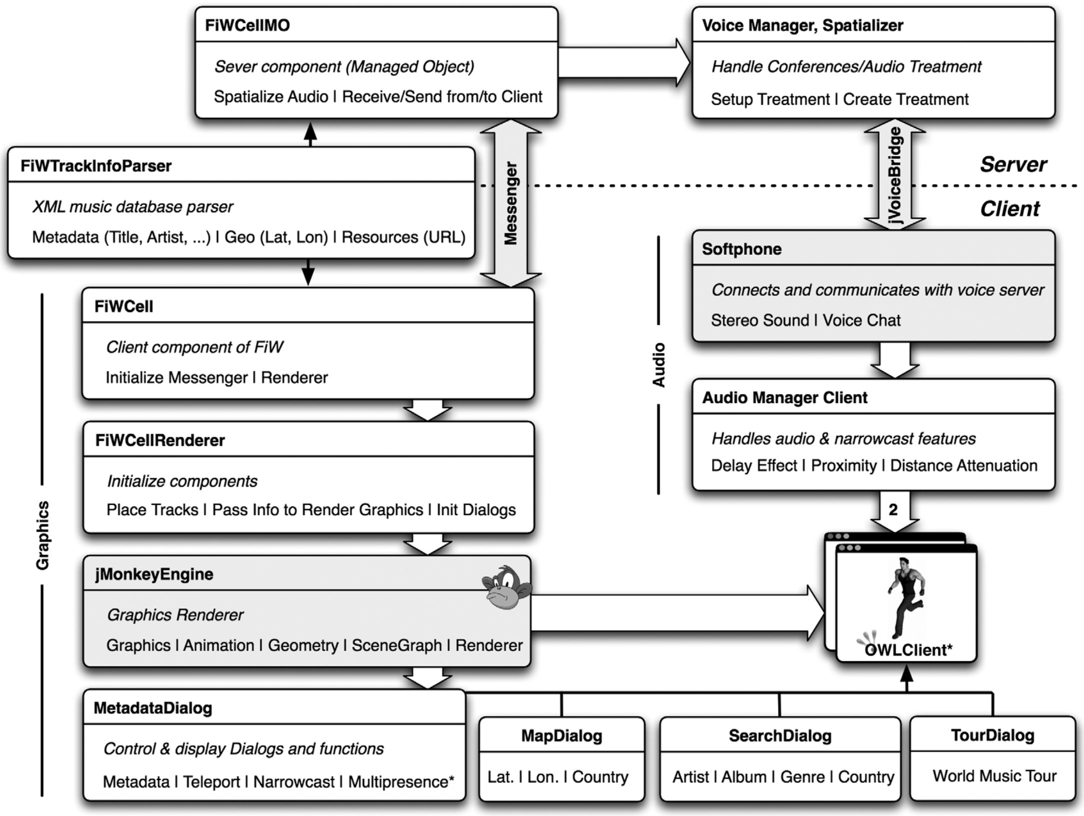


Figure 3. *FiW system schematic.*

each client. `FiWCellRenderer` (left lower center), also using `FiWTrackInfoParser` (upper left), generates a list of track samples at the client side. The generated track list is used for rendering track markers with jME, displaying metadata of a selected track, and searching for particular keywords. The `FiWCellRenderer` is also responsible for rendering Java2D Swing dialogs, including those for Metadata, Map, Search, and Tour. (Thin vertical black arrows in the schematic represent inclusions, thick horizontal white arrows indicate API invocations, and thick vertical gray arrows denote communication. The modules drawn with white background were extended or newly developed in FiW.)

3.2 MX: IEEE 1599

Artistic, geographic, audio-related, and generic information describing the Folkways music collection is curated in XML format conforming to MX: IEEE 1599 (Baggi & Haus, 2009), a comprehensive, multilayered music description standard. MX, standing for musical application using XML, inherits all the features of XML—including inherent human-readability, extensibility, and durability (Ludovico, 2009)—and unifies features of MML[12] (Music Markup Language, a syntax for encoding different kinds of music-related events) and MusicXML[13] (which is designed for the exchange of scores) with some additional features, including the concept of layers. The six MX layers, which allow integrated representation of several aspects of music, are general, logic, structural, notational, performance, and audio. Even though the Folkways curation has no information corresponding to the MX logical, structural, or performance layers, MX: IEEE 1599 allows empty layers (Ludovico, 2008), and there are no restrictions preventing browsing of

12. www.musicmarkup.info
13. www.musicxml.com

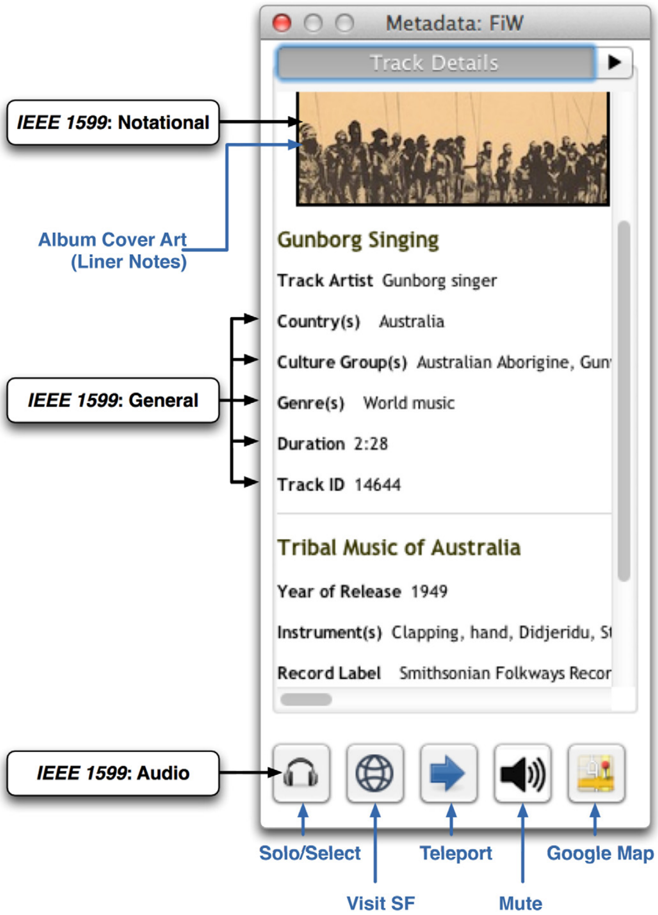


Figure 4. *Metadata display window: The Track Details tab shows song information, the Playlist tab displays the entire collection as an outline, and the History tab lists tracks visited by the user. Other operations, invoked by buttons at the bottom, allow exclusively auditioning a track, browsing selected track information at the Smithsonian Folkways site, teleporting to the origin of a track, muting a track, and opening map window with a Google map to provide detailed, zoomable, topographic information. Clicking on album art brings up liner notes, which may include scores, musician interviews, critical commentaries, etc.*

other music collections when such information is available (as shown in Figure 5 and reflected in the music browser as shown in Figure 4). Note that layers may contain URLs as well as directly accessed data, for extra flexibility and late binding.

4 Narrowcasting in Wonderland

Narrowcasting describes a technique which allows information streams to be filtered, for privacy, security, and user interface optimization in groupware solutions (Alam et al., 2009; Fernando et al., 2006). Traditional conferencing systems over the PSTN (public switched telephone network) have become almost obsolete, as contemporary telecommunications systems support teleconferencing by providing audio, video, and data services.

Even though narrowcasting operations have been implemented for workstations and mobile phones (Cohen & Győrbiró, 2009; Cohen & Kawaguchi, 2003; Fernando et al., 2006) or online chat systems such as Dolby Axon,[14] this project explores how narrowcasting operations can be effectively used when voice- or

14. axon.dolby.com

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ieee1599 SYSTEM "http://standards.ieee.org/downloads/1599/1599-2008/ieee1599.dtd">
<ieee1599>
        <track>
                <general>
                        <albumTitle>Folk Music of Ghana</albumTitle>
                        <trackTitle>Ataa oblanyo</trackTitle>
                        <trackArtist>Various Artists</trackArtist>
                        <instruments>Ashwa, Talking drum</instruments>
                        <cultureGroups>Ewe</cultureGroups>
                        <trackDetailURL>http://music/trackdetail.aspx?itemid=28728</trackDetailURL>
                        <country>Ghana</country>
                        <latitude>5.555717</latitude>
                        ...
                </general>
                <notational>
                        <linerNotes>file:/music/folkways/notes/FW08859.pdf</linerNotes>
                </notational>
                <audio>
                        <audioURL>file:/music/folkways/FW08859_03.30.mp3</audioURL>
                        <length>3:38</length>
                </audio>
                <logic/>
                <structural/>
                <performance/>
         </track>
        <track>...</track>
</ieee1599>
```

Figure 5. XML *stub corresponding to the* MX: IEEE 1599.

text-chatting and exploring music in enterprise-quality immersive virtual worlds. Most CVEs configure private conversations by just selecting a subset of connected members (Yankelovich et al., 2005). Full-featured narrowcasting can be invaluable when listening to collections of music, especially when the music is spatialized. An obvious use case is to avoid unwanted cacophony when multiple tracks are too closely located, causing multiple nimbi (Greenhalgh & Benford, 1995) to overlap. Our voice bridge basically implements distributed modulation of source → sink connectivities, as expressed by the predicate calculus expressions shown in Figure 6.

4.1 Conferencing Using jVoiceBridge

Wonderland natively provides individually adjustable audio channels for each in-world, recorded sound source and live avatar. jVoiceBridge,[15] an open-source conferencing module, handles voice over IP (VoIP) audio communication for Wonderland. The voice bridge supports a conference by receiving monaural audio streams from all session members, directionalizing the streams into individualized soundscapes, and streaming a personalized stereo mix back to each member. Users can adjust their virtual speakers and microphones as well as other participants' apparent intensities and sensitivities. A user mutes him/herself by effectively setting others' virtual source-wise sensitivities to zero. By extending such built-in capabilities, narrowcasting features were implemented in Wonderland for voice-chat and music audition, considered as separate modalities.

15. java.net/projects/jvoicebridge

4.2 Audio Transmission for Voice and Music

It is important to keep in mind the entire end-to-end pathway of sound in a teleconference, including careful use of often muddled jargon. A speaker's voice, causing compression and rarefaction of air, is sensed by a microphone that measures pressure, transducing acoustic energy into electrical. This measurement, expressed as a voltage, is sampled (in time) and quantized (in amplitude) by an audio interface, converting the analog acoustic phenomenon into a digital signal, encoding it perhaps uniformly (as in PCM, pulse code modulation)

The general expression of inclusive selection is

$$\texttt{active}(x) \;=\; \neg\texttt{exclude}(x) \;\wedge\; (\exists\ \mathrm{y}\ (\texttt{include}(\mathrm{y}) \;\wedge\; (\texttt{self}(\mathrm{x}) \;\Leftrightarrow \texttt{self}(\mathrm{y}))) \;\Rightarrow\; \texttt{include}(x).$$

So, for `mute` and `select` (`solo`), the relation is

$$\texttt{active}(\mathrm{source_x}) = \neg\texttt{mute}(\mathrm{source_x}) \;\wedge\; (\exists\ \mathrm{y}\ \texttt{select}(\mathrm{source_y}) \;\wedge\; (\texttt{self}(\mathrm{x}) \;\Leftrightarrow \texttt{self}(\mathrm{y}))) \;\Rightarrow\; \texttt{select}(\mathrm{source_x}).$$

`mute` explicitly turning off a source, and `select` disabling the complement of the selection (in the spirit of "anything not mandatory is forbidden"). For `deafen` and `attend`, the relation is

$$\texttt{active}(\mathrm{sink_x}) = \neg\texttt{deafen}(\mathrm{sink_x}) \;\wedge\; (\exists\ \mathrm{y}\ \texttt{attend}(\mathrm{sink_y}) \;\wedge\; (\texttt{self}(\mathrm{x}) \;\Leftrightarrow \texttt{self}(\mathrm{y}))) \;\Rightarrow\; \texttt{attend}(\mathrm{sink_x}).$$

Figure 6. *Simplified formalization of narrowcasting and selection functions in predicate calculus notation, where ¬ means "not," ∧ means conjunction (logical "and"), ∃ means "there exists," ⇒ means "implies," and ⇔ means "is equal to" (mutual implication). The duality between source and sink operations is tight, and the semantics are identical: an object is inclusively enabled by default unless, a) it is explicitly excluded (with* $\overbrace{\texttt{mute}}^{\text{source}} \overset{\text{or}}{||} \overbrace{\texttt{deafen}}^{\text{sink}}$*), or, b) peers are explicitly included (with* $\overbrace{\texttt{select}\ [\texttt{solo}]}^{\text{sources}} \overset{\text{or}}{||} \overbrace{\texttt{attend}}^{\text{sinks}}$*) when the respective object is not. Narrowcasting attributes are not mutually exclusive, and the dimensions are orthogonal. Because a source or sink is active by default, invoking* `exclude` *and* `include` *operations simultaneously on an object results in its being disabled.*

or perhaps nonlinearly (as in μ- or a-law representations). This audio signal is filtered by the computer's DSP hardware and software, including amplification (increased envelope size) and attenuation (decreased scale). The amplification or attenuation of a signal can be accomplished by adjusting its gain, a scalar coefficient which multiplies the raw signal and controls the dynamic range. (Balancing or panning a stereo signal involves coupled gain adjustments to a left–right signal pair.) Frequency-based adjustments— such as equalization, "sweetening," aural enhancement, etc.—are also possible, typically by specifying frequency-band-specific amplifications or attenuations.

4.3 Binaural Parallax: Localizing Sound

Binaural sound enables an immersive listening experience, giving one a sense of presence and space (Bormann, 2005). Wonderland native audio spatialization supports stereo (two channels) and only horizontally flattened positioning, including delay effects (ITD, interaural time difference) and panning (IID, interaural intensity difference). Two interleaved channels are sent in each packet, one delayed by 0–0.63 ms (Equations 1 and 2), depending on the location of the source relative to the respective sink.

$$\mathrm{ITD} = \frac{a}{c}(\theta + \sin\theta), -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}. \tag{1}$$

Approximating the radius of a head a as 8 cm and the speed of sound c as $300\,\mathrm{m{\cdot}s^{-1}}$, maximum ITD can be estimated as

$$\mathrm{ITD_{max}} \approx \frac{0.08}{300}\left(\frac{\pi}{2} + 1\right) \approx 0.6\ \mathrm{ms}. \tag{2}$$

An RTP packet contains 20 ms (50 packets/s) of audio data. For μ-law encoding as used in jVoiceBridge, a packet comprises 160 8-bit samples (8 kHz × 1 sample/s/Hz × 0.020 s) with a 12-byte RTP header, totally 172 bytes of data (with $20 \log 2^8 \approx 48$ dB dynamic range). For PCM stereo streams sampled at 44100 Hz, there are 44100/50 × 2 16-bit samples/packet, which is 3528 bytes of data payload (with $20 \log 2^{16} \approx 96$ dB dynamic range), plus the RTP header. Overhead for each audio packet, besides the 12-byte RTP header, includes 8 bytes for the UDP header and 20 for the IP header (Seo, Htoon, Zimmermann, & Wang, 2010). After whatever local processing, the audio signal is packetized and sent over the internet to one or more distal nodes, corresponding to other participants in a session. These packets are resequenced by receiving nodes, allowing the reconstructed audio signals to stream through their own respective filters. The user interface or distributed processing might allow articulated control, featuring, for simple example, both source-wise and local computer audio amplification adjustment (colloquially called "dynamics," "loudness," or "volume" [not to be confused with 3-dimensional extent]). One could change, for instance, the loudness of each of the conferees individually, as well as amplification of the entire mix.

Finally, at the audio interface of each attending computer, the digital signal is pumped through a DAC (digital–analog converter), presented to an amplifier as a continuous (in time and amplitude) voltage, which power signal in turn drives speakers, such as stereo loudspeakers or headphones. The value of the amplified signal controls the displacement of speaker cones, and the gain of the signal controls the amplitude of its envelope, corresponding (assuming linearity of the filters) to maximum excursion of the speaker cones. The electrical signal is thereby transduced back into the physical domain of moving air molecules, that is, sound, which acoustic signal propagates through air to the ears of the listener. The ears and the listener's brain have their own complicated physiological (biological reception) and psychoacoustic (psychophysical stimulus–sensation behavior) response to the sound, which is perceived, interpreted, and "heard." Qualities such as amplitude of an auditioned signal are therefore a combination of source-side processing (including muting and softer gain adjustment such as "muzzle") and sink-side processing (including deafening and softer gain adjustment such as "muffle"). The intensity of a signal corresponds to its instantaneous power, or energy/time, the time-averaged square of its running value. For a digital signal, intensity is the (windowed and averaged) sum of the squares of the samples. The "level" is proportional to the logarithm of the intensity or RMS (root mean square, the standard deviation of a centered [zero-mean] signal), associated with a channel's subjective loudness or volume, since human perception of loudness is approximately logarithmic with power. Distance attenuation in the voice bridge is shown in Figure 7 and expressed by Equations 3 and 4.

$$d' = \frac{\mathrm{d} - d_{\mathrm{fvs}}}{d_{\mathrm{e}} - d_{\mathrm{fvs}}}, \tag{3}$$

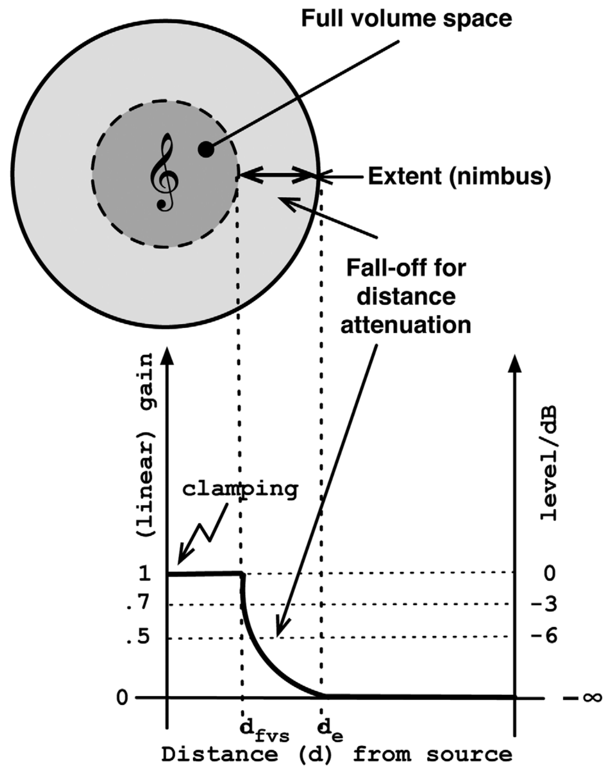


Figure 7. *The "full volume space" (d_{fvs}) of a source is represented by the dotted circle in the diagram, within which audio is heard at full volume. The "extent" or nimbus (d_e) of the displayed source refers to the space in which sinks can receive the signal.*

$$\mathrm{gain(d)} = \begin{cases} 1 & d \leq d_{\mathrm{fvs}} \\ \mathrm{falloff}^{d'} & d_{\mathrm{fvs}} < d < d_{\mathrm{e}}. \\ 0 & d_{\mathrm{e}} \leq d \end{cases} \tag{4}$$

In contrast, the ordinary free-field inverse-square intensity attenuation for a nominal point source (representing a constant power source radiating acoustic power P through a notional sphere with area proportional to the square of its radius, which corresponds to source-to-sink distance d) is $I = P/4\pi d^2 \propto d^{-2}$. This is a power law relation, in contrast to the exponential function implemented by the voice bridge model and expressed above.

Each avatar is both a source (projecting a nimbus) and a sink (sensing within a focus), but music tracks are only sources, not sinks. This means that avatars can be `deafened` or `attended`, but tracks cannot. However,

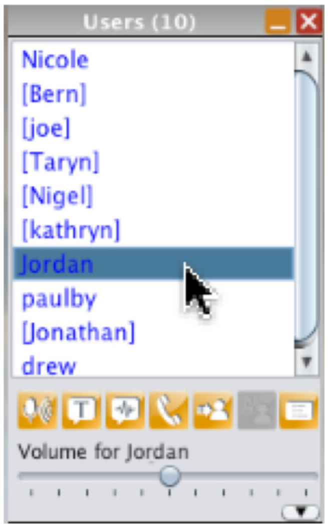


Figure 8. *Original Wonderland voice-chat* HUD *(Head Up Display) panel with basic audio functions, including mute and individually adjustable volume for each user and self.*

Figure 9. *Extended FiW voice-chat panel indicating current status and matrix of narrowcasting operations {disable (exclude), enable (exclusively select, selectively include)} × {production (sources), reception (sinks)}.*

any source can be `muted` or `selected`. Furthermore, only avatars can issue such narrowcasting commands, as tracks lack an associated user with agency or volition. The narrowcasting commands are considered separately in the following paragraphs, {`source` ⇒ `nimbus`, `sink` ⇒ `focus`} × {`disable`, `enable`}.

4.3.1 Mute (Nimbus Disable): $A \rightarrow \bar{B}$**.** Mute is a common media control function, available in most conferencing systems (Alam et al., 2009). Narrowcasting `mute` is a source-related command that blocks media coming from a source. Wonderland allows muting one's self as well as blocking audio from other participants in a conference by adjusting a sensitivity controller (as shown in Figure 8), as well as muting by an administrator (like public branch exchange [PBX] mute; Alam et al., 2009). But such functionalities are not available for individual music audition. Users can share audio by uploading audio files to a Wonderland server or accessing audio sources on the internet via URLs, dynamically adding new tracks to a mix. One might turn such shared samples on and off individually, but that would result in a kind of audio chaos, someone turning off what another had just turned on. To overcome such complications, a new capability was developed for the FiW music browser to allow a user to turn individual tracks off and on without affecting others' soundscapes. Narrowcasting `mute` can be compactly expressed as in $A \rightarrow \bar{B}$ (with a minus sign stacked above a symbol for a source) or $A \rightarrow [B]$ (with surrounding square brackets), for A muting B.

4.3.2 Select or Solo (Nimbus Enable): $A \rightarrow \overset{+}{B}$**.** While exploring a collaborative musical space, users might hear a mix of musical tracks when multiple tracks are nearby. By `muting` tracks individually, one can focus on a particular track. Alternatively, narrowcasting `select` (a.k.a. `solo`), a source-related command which limits projected sound (the nimbus) to particular sources, can be invoked to avoid unwanted cacophony. By extending the `mute` function, the `select` function resets all projection amplifications (as described in Greenhalgh & Benford, 1995) to zero, excepting a selected track. The `select` or `solo` operation can be abbreviated as in $A \rightarrow \overset{+}{B}$ (with a plus sign above a symbol for a source) or $A \rightarrow]B[$ (with square brackets facing outwards), for A soloing B.

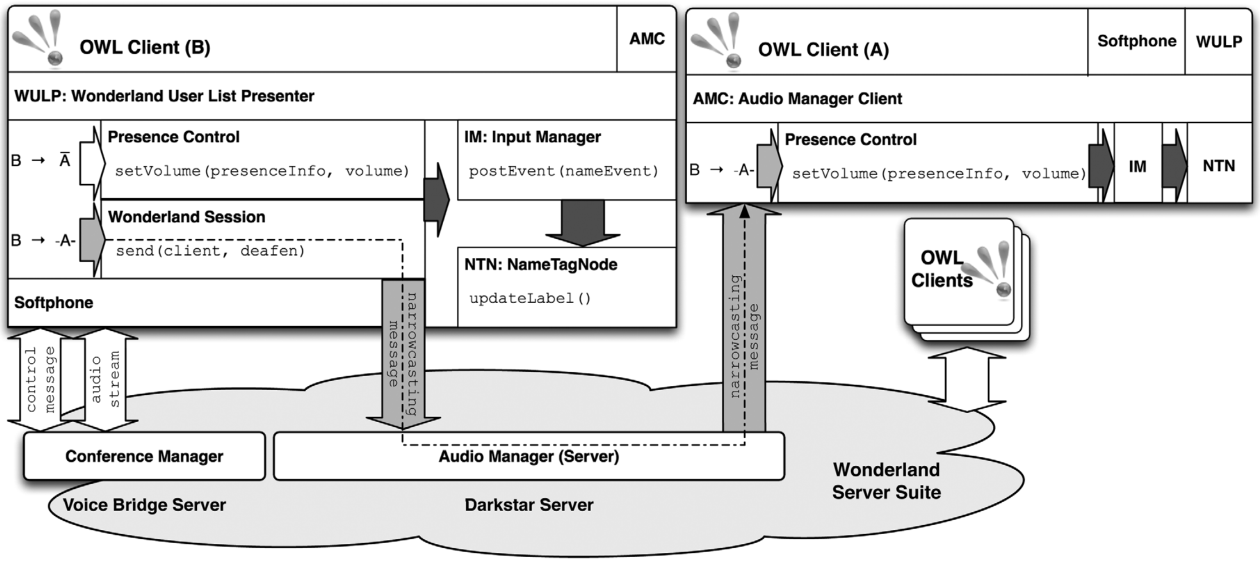


Figure 10. *Extended voice-chat architecture: A Wonderland client consists of a* `Softphone` *(middle left) which communicates with the* `Conference Manager` *in the Voice Bridge server; the* `Wonderland User List Presenter` *(**WULP**, top left), which allows one to adjust other participants' apparent sensitivity locally; and the* `AudioManagerClient` *(**AMC**, top right), which is the message receiver for control signals sent from the* `AudioManager` *in the Darkstar server.* `Mute` *and* `select(solo)` *are implemented by adjusting sensitivity in* `PresenceControl` *in the **WULP**. By sending a* `Narrowcast` *message via* `AudioManager`*, a receiver (**AMC** at a different client) can* `mute` *a sender, realizing* `deafen` *and* `attend`*.*

4.3.3 Deafen (Focus Disable): $A \rightarrow -B-$.

`Deafen` is a sink-related command which blocks media going to a sink. Such a function can be useful to a virtual ethnomusicological field worker (Cooley, Meizel, & Syed, 2008) for private discussions with colleagues, excluding, for instance, casual tourists. `Deafen` can be compactly represented as in $A \rightarrow -B-$ (with minus signs straddling a symbol for a sink), for A deafening B. A `deafen` operation can be realized in different ways in a client-server architecture:

- A source can block outgoing media to particular sinks.
- A crossbar matrix mixer at the soundscape or media server can block streams to a particular sink.
- A sink can block incoming media from particular sources.

We have selected the last method, exploiting the duality relationship between `mute` and `deafen` ($A \rightarrow -B- \equiv B \rightarrow \bar{A}$). This simplicity can be useful when implementing narrowcasting functions in any conferencing system, because all the functions described here can be elaborated just from basic `mute`. For example, if user A wants to `deafen` B, a `Narrowcast (Deafen)` message is sent from A to B, and B sets A's apparent intensity to zero (as traced by the stippled line in Figure 10's component diagram).

4.3.4 Attend (Focus Enable): $A \rightarrow +B+$.

`Attend` is a sink command which limits received media streams to only those explicitly apprehended by particular sinks. When user A wants to `attend` B, so that only B can hear A, all except B are `deafen`ed to A. `Attend` can be expressed as in $A \rightarrow +B+$ (with plus signs straddling a symbol for a sink), for A attends B.

4.4 Textual Narrowcasting

As previously mentioned, besides voice-based communication, Wonderland provides public and private text-based chatspaces. Voice and text-chats have respective merits: text-chat has advantages regarding simplicity

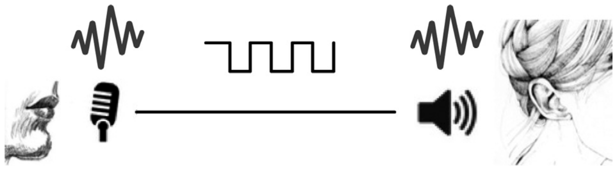


Figure 11. *Voice transmission from human source to human sink over network: source (mouth) → sink (microphone) → source (speaker) → sink (ear).*

and bandwidth considerations, whereas voice-chat is richer and more natural (Geerts, 2006; Yankelovich et al., 2005). In order to provide a more articulated experience for textual conferencing in FiW, narrowcasting features were added to the native chat system. When a source is `muted`, text typed by the user associated with that source's avatar is blocked. When a `selection` exists, all text is blocked except that from the selected source(s). Text is not delivered to `deafened` sinks and is delivered only to particular sinks when any are `attended`.

4.5 Narrowcasting State Representations

When the audition interface is extended using narrowcasting features, complementary visual cues (Lee & Kim, 2008) should support it without many distractions or complications. The symbology for narrowcasting operations is complicated by the conflation of associations of source and sink icons. As illustrated by Figure 11, the chain of simplex transmission for voice-chat starts with a source, a mouth, emitting sound which is sensed by a sink, a microphone, conveying the signal through the network for reproduction by another source, a loudspeaker, which signal is captured by another sink, an ear. Historically, a slashed speaker icon is used for mute and a headphone icon (connoting private listening) is used for `select` (`solo`), but such associations don't generalize to soundscape models with sinks. In contemporary conferencing equipment, "mute" usually means local reticence, like a video "sneeze button": distal parties' voices are still audible, but the muting side has "put its hand over the mouthpiece." Table 1 shows how narrowcasting operations are represented across different situations in our cyberworld.

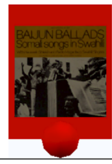
Narrowcasting is multimodal, applicable to music, voice, and text streams. Our system supports two different kinds of sound sources: musical selections and avatar conversation ("voice-chat"). Narrowcasting for music enables aesthetic focus; narrowcasting for talk, like that for text-chat, enables cognitive focus. The former is required for dense presentation of musical sound, the latter for virtual worlds in which many avatars are expected to be able to interact. Music tracks and voice-chat are controllable by narrowcasting functions. In the FiW user interface, narrowcasting widgets (controls and displays) are distributed across two panels— solo and mute buttons in the metadata pop-up window, and {source, sink} × {include, exclude} (as seen in Figure 9) in the user head-up display, a floating panel inside the main view window. Narrowcasting state is also displayed with the respective source or sink, as shown in Figure 9.

Narrowcasting controls are grayed-out when narrowcasting is disabled. Such inhibition is currently established locally, by each user (rather than a super-user administrator or server custodian). This narrowcasting enable/disable switch (as seen in the last row of Table 1) was deployed in order to test the usefulness of the narrowcasting privacy suite, so that we could instruct experimental test subjects to disable it before performing some task, and then re-enable it before performing similar tasks. That is, the narrowcasting absence or presence is an optional experimental condition.

Nimbus, or auditory extent of sources, is visualizable as a translucent boundary (as seen in Figure 12), enabled via an extensibility module capability of Wonderland. We generalized its heretofore rigid parameters, and it is now adjustable on both the server and the client sides. (However, the focus, or sensitivity of the sinks, is just a degenerate notional point inside the respective avatars' heads.)

The audio server has a virtual crossbar matrix for a session's sources and sinks. Distributed narrowcasting state, controlled by the respective clients, is compiled into a Boolean audibility attribute for each source–sink combination. Each client streams its voice channel to the

Table 1. *Different Methods for Representation of Narrowcasting Operations*

Representation	Exclude (Disable)		Include (Enable)	
	`Mute`	`Deafen`	`Select`	`Attend`
Figurative icons or symbols are used in metadata window (as shown in Figure 1) and voice-chat panel (as shown in Figure 9).				
Distinctive border colors are used when a track (as described in § 3) is `muted` or `selected`.				
Combinations of distinctive colors and glyphs decorate monikers in the voice-chat panel (as shown in Figure 9).	[Alice] **Alice (Mute)**	–Bob– **Bob (Deafen)**	]Carol[**Carol (Select)**	+David+ **David (Attend)**
Figurative avatars with decorated names floating over head indicate applied narrowcasting operations.				

FiW Control Panel: Users can toggle preferred narrowcasting decorations (or even disable such display completely). Asserting "Arrange" resizes and places popped-up windows automatically (as in Figure 1). Text search and tour windows can also be opened from here.

server, which otherwise manages all the Folkways musical sources ("treatments" in Wonderland jargon). The respective source–sink spatial arrangements, including direction and range, extended by multipresence and disambiguated by autofocus (described in section 5.2), is used to parameterize sound spatialization, as the server renders 2-channel (binaural) contributions for each musical source for each client, which composite stereo pair is streamed back to the terminals, the respective clients. At the client side, local volume is effectively multiplied by the distributed audibility flags (bits), and the respective source stream mixed or not into the locally displayed composite soundscape.

It would have been more elegant to inhibit transmission of inaudible streams at the server side, rather than depending upon each respective client to ignore

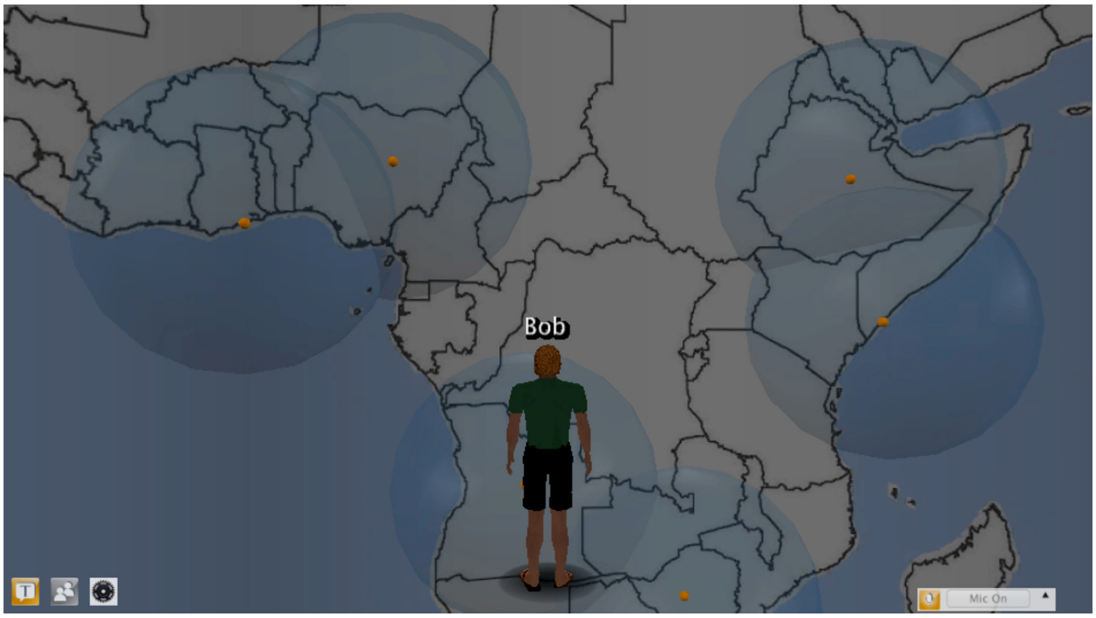


Figure 12. *Translucent spheres represent nimbi of sources. Overlapped sections visualize overlapped sonic spheres where cacophony is present.*

them, since doing so would conserve network bandwidth and also prevent eavesdropping. However, the audibility attribute is separate from volume control, so there is no way a client could override such discretion (without meta-techniques such as editing and recompiling client software), and engineering considerations (details of Wonderland architecture) trumped logical economy.

4.5.1 Figurative Icons. Icons and symbols as part of GUIs play a significant role in user experience. The extended FiW voice-chat and music browser functions use carefully designed icons for narrowcasting operations. Mute is a complex operation to represent, as a speaker icon is used in operating systems and media players, but a microphone icon is used in contemporary messenger systems such as Skype,[16] LINE,[17] and Google Talk.[18] Three types of mute icons are used in our cyberworld:

- A slashed microphone iconifies Wonderland voice-chat mute.
- Narrowcasting `mute`, which was previously figuratively represented as hand clapped over a mouth (Cohen, 2000), is now iconified by a slashed mouth.
- As in many media players, muting a track is represented as a slashed speaker icon.

Previous research (Cohen, 2000; Fernando et al., 2006) figuratively used a megaphone to represent `select` (`solo`) and ear trumpets to represent `attend`. In our cyberworld, a headphone icon is used to represent `solo` for music auditioning, as donning a headphone naturally implies auditioning a particular sound source privately. Narrowcasting `deafen`, formerly represented as hands clasped over ears, is now iconified by a slashed ear.

5 Multipresence in Wonderland

Multipresence allows each user in a virtual environment to have presence in several places or spaces at once by designating multiple representatives as “self,” effectively increasing one’s attendance in groupware activities (Fernando et al., 2006; Taylor, 1999). Multiple sources are useful for broadening one’s exposure within or across virtual spaces. Multiple sinks are useful for

16. `www.skype.com`
17. `line.naver.jp/en`
18. `www.google.com/talk`

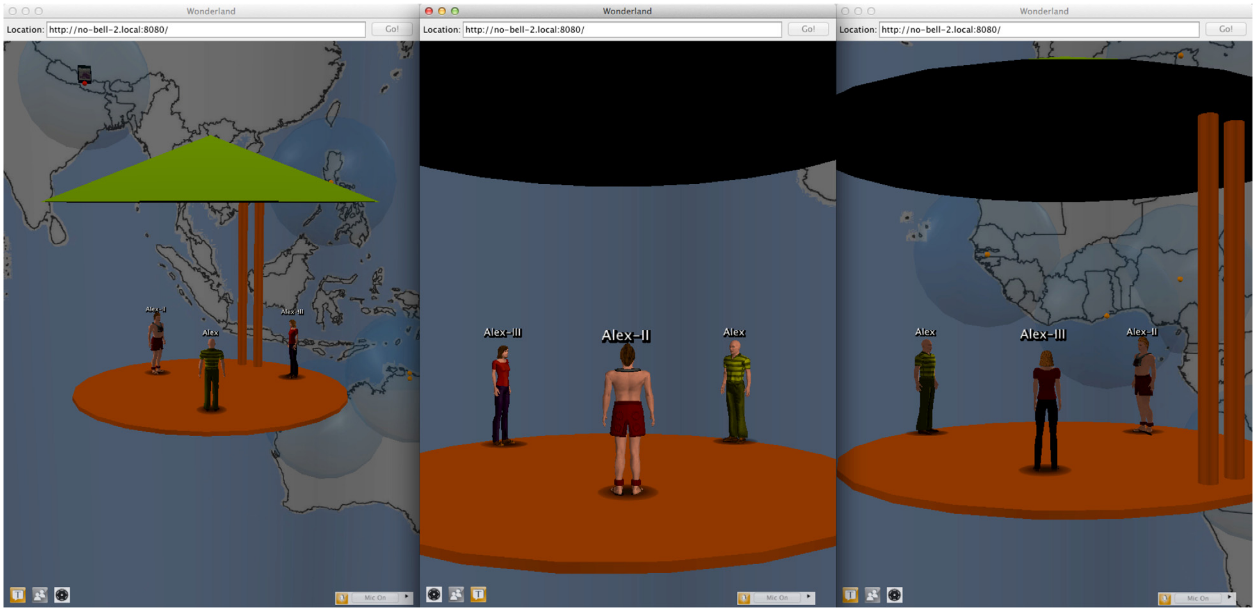


Figure 13. *Disambiguating multipresence: To acquire control of a self-identified avatar, a user should select one of the Wonderland windows. Confusion about "Who am I now?" can be intuitively resolved, since an avatar in Wonderland always stands in the middle of its client window, even across different perspectives. Users can manifest different identities (Alex-II: the Mohawk-coiffed avatar in the center window) or even change apparent gender (Alex-III: in the right window).*

monitoring distributed sources. Wonderland ordinarily allows running a single instance of a client per computer. It caches assets (local settings, 3D objects, and avatar models) into a local directory when a client connects to a server for the first time. This cache is used to populate a client scene in subsequent sessions. By allowing a client to have multiple caches, we are able to execute multiple clients acting as clones of a local user (as shown in Figures 1 and 13). When an avatar is "forked," the Wonderland server can recognize if multiple avatars represent the same user (host), but there is no way others can recognize such difference. While exploring world music, one might want to simultaneously pay close attention to tracks associated with different countries or genres, maintaing consistency with other participants. An active listener can spawn "doppelgänger" presence and locate representatives at each location of interest, the clones capturing each respective avatar's soundscapes, individually controlled using narrowcasting functions such as self-`deafen`. Similarly one can participate in a conference and at the same time join a world tour of music.

5.1 Autoventriloquism

An interesting example of a configuration enabled by multipresence and narrowcasting is *autoventriloquism*. One's own voice can be heard from another place by forking one's avatar, deafening one (the source) and muting the other (the sink) to avoid the usual "sidetone" feedback (as shown in Figure 14).

5.2 Autofocus

Because of multipresence, multiple selves might be within hearing range of a particular track, separately spatialized for each. Using an "autofocus" technique to avoid the paradoxes of such seemingly conflicted soundscapes (Fernando et al., 2006), the nearest self-identified sink (one of the clones) is automatically, implicitly `attended`, and the others are implicitly `deafened` to that source.

As multipresence can be thought to effectively enable virtual cloning, which is metaphorically a "fork" operation that splits a process in an operating system, so can autofocus be considered analogous to "join" that

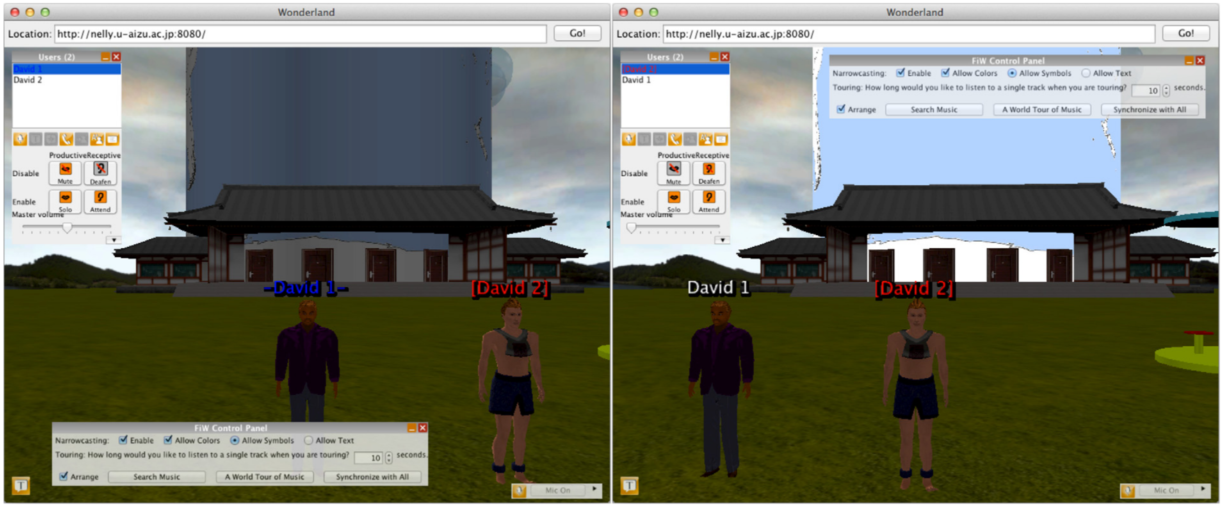


Figure 14. *Autoventriloquism: Generally one hears one's own voice in both ears, but with this multipresence configuration, one can hear one's voice coming from the left. Image on the left, "David 1," is self-`deafen`ed and "David 2" is self-`mute`d on the right image.*

coalesces processes: multipresence articulates distributed self-identified avatar existence (violating the usual singleton cardinality of existence), and autofocus resolves the ambiguity of non-unique perspective.

5.3 Coexistence: Not Only One→Many but Many→One

Not only may a single user have multiple avatars, but also multiple users can share a location as, for instance, when joining a tour. A docent can initiate a tour, sharing a track with others in a session. Users may join the tour, allowing their avatars to be automatically guided through a sequence of track placemarks. An avatar automatically guided in a tour could passively relax while one in parallel conference might actively react. Even though this technique does not support pure implementation of coexistence in which multiple users share a single avatar, it provides a satisfying user experience when one joins a tour of world music.

6 Evaluation and Assessment

6.1 Experimental Design

Using Folkways in Wonderland as a virtual laboratory, we pose the following question: How do social actors, represented by avatars, interact in such an immersive cyberworld, when presented with a specific collaborative task? A laboratory environment enables us to control variables and thus answer—at least within this restricted environment—questions about such dependencies with rigor that cannot be achieved in the real world. In particular, we are concerned with two primary clusters of independent variables known by ethnomusicologists to shape the emergence of musical community: the *social* and the *musical*. Here, social variables include the number and demographic profiles of participants populating the cyberworld, while musical variables include the number and kinds of music tracks populating the cyberworld. Variables within either cluster can be manipulated: the former through participant selection, the latter by loading different collections of music tracks into FiW.

6.1.1 Procedure. The experiment, which conformed to institutional ethics guidelines for subjective experiments (and required no special dispensation), was conducted in two sessions (a week apart) in three stages: preparatory formalisation, individual and group tasks, and a music scavenger-hunt game. Prior to entering the space, each subject was profiled, recording age, gender, and ethnic background, all considered as variables within the social cluster. During the second and third stages participants were organized into teams of two or three members and asked to perform group activities including narrowcasting.

6.1.2 Musical Scavenger Hunt.

Organized into teams, participants were encouraged to work together by interacting in the space. Understandably and predictably, participants tended to team-up according to nationality and native language. A "target track" (hint only, such as "find a track with no instruments") was announced by the referee, and the goal of the game was to locate this track on the map. Once each target track was located, the referee announced to all participants that it was found, and the winners were asked to explain their strategy before everyone was prompted with the next target track.

6.1.3 Subjects.

Participants were computer science undergraduate students, age 20–24: 6 females and 17 males; 12 Chinese, 10 Japanese, and 1 American. After 30 minutes of orientation, a feature-spanning exercise was conducted to train the participants. Instantiation ("spawning") jitters creation locations so everyone doesn't "beam down" together. Avatars can be initialized with randomly varied appearance, but students were encouraged to personally customize their puppets, which have costume and features that persist across subsequent sessions (as start-up preferences), while exploring FiW.

6.2 Experimental Results

Even though audio narrowcasting features and multipresence have been previously described, they had never been evaluated before. Further, narrowcasting has been applied here not only to musical audition and voice-chat, but also to text-chat.

Full narrowcasting is more powerful than conferencing systems with room models for privacy, since narrowcasting explicitly includes sink operations (`deafen` and `attend`) as well as source functions (`mute` and `select[solo]`). In combination with multipresence, the granularity of narrowcasting control is finer than that allowed by systems that use a conference room as the metaphor of isolation. For instance, Dolby Axon requires making a private chatroom to adjust bidirectional privacy. Therefore the experimental protocol was specifically designed to have participants use and evaluate multimodal narrowcasting features, i.e. repeat some tasks enabling and disabling narrowcasting while evaluating the FiW music browser. Salient characteristics of narrowcasting include: when such features are used in a cyberworld, how useful they are, and how easily they can be used. During the training session participants were taught how to use the system, including its narrowcasting commands. Participants were then asked to compare and contrast songs from different parts of the world, recall the techniques or features they used, and answer a questionnaire regarding their experience. A summary of the results of the post-exercise questionnaire is shown in Table 2.

Thirteen out of 18 participants answered the technique used to locate tracks as "Search" facility and two stated "Placemark" feature. Thirteen out of 19 participants rated "Tours" as their favourite feature of FiW. Six liked the diverse music itself, four liked the "Search" function, and others cited other features. All participants who evaluated the usefulness of "`select`" and "`mute`" for multiple track audition found it useful. Out of 19 participants, 9 cited "chats" as helpful for private communication where narrowcasting is built-in and 7 specifically mentioned "narrowcasting." When participants were asked whether narrowcasting decorating glyphs (++, −−,][, []) were useful to understand, 17 out of 21 answered positively. The favorite display combination was the combination of colors and symbols (e.g., "**]Carol[**"). Four preferred the combination of colors and text (e.g., "**Carol (Select)**"). All except one agreed that the figurative icons based on mouth and ear representations in the HUD are intuitive.

7 Discussion

Social networks such as Facebook, Twitter, and LINE have gained global popularity during recent years. Users exchange text messages and multimedia (graphics, video, and audio) or conduct conferences among private

Table 2. *Results of Questionnaire Based on 6-Point Score (0: Worst, 1: Poor, 2: Basic, 3: Normal, 4: Good, 5: Excellent)*

	Rating			
Questionnaire Item	Responses	2	4 or 5	Quality
Overall experience	22	1	16	77%
Helpfulness of narrowcasting features	20	0	15	75%
Helpfulness of multipresence	20	1	15	80%
Intuitiveness of visual cues (icons and colors)	21	2	15	81%
Ease of finding track compared to navigating around Smithsonian Folkways web site	20	1	12	65%
Helpfulness of placemarks and history mechanism	21	0	18	86%
	Responses	No	Yes	Agreement
Usefulness of colors to distinguish narrowcasting operations	23	2	21	91%
Usefulness in sharing a secret	21	4	17	81%
Intuitiveness of using mouth and ear as base for icons	22	1	21	95%

Two users rated a few of the items as "basic," but no "worst" or "poor" judgments were recorded. Agreement is expressed as a fraction of total ratings (where negative is [0–2] and positive is [3–5]). Quality is expressed "good" or "excellent" as a fraction of positive ratings.

or public groups. Even though there are limited privacy settings—such as blocking users (related to "`mute`"), following a certain topic (analogous to "`select`"), and sharing information only among a set of friends—narrowcasting could provide more articulated control over shared media. For example, if one wants to share an exclusive secret, narrowcasting provides a way to form a coterie without having to make a new group or private chat room with selected confidants. The problem with making a new group or chat room is that generally recipients could accept such group request (in Facebook or LINE, but not Dolby Axon) and they might send a subsequent group invitations to others, in which case the original user's goal would be compromised. Using "`deafen`," one can simply exclude some members and share a secret with the rest of a group without making a new clique. The same approach can be used against "cyber bullying" (Campbell, 2005): one may just exclude (`mute` and/or `deafen`) a boor without leaving a group of colleagues. The narrowcasting `solo` function can be effective in situations such as virtual classrooms (Gardner et al., 2011; Ibanez et al., 2010; Peña-Ríos, Callaghan, Gardner, & Alhaddad, 2012), allowing a student to `select` a teacher or mentor to closely monitor and thereby avoid distractions. Since our experiment concluded that most users have positive feelings about such user interface conventions regarding narrowcasting, we hope to evangelize such functions with social media applications developers.

8 Conclusion and Future Research

We have presented a novel application for listening to world music inside a virtual space. Rather than finding tracks using traditional interfaces, an avatar- or avatars-represented user can explore music immersively while adjusting their soundscape with narrowcasting. Users can invoke `mute` or `select` functions to listen only to particular songs when cacophony might distract. The same functions can be used to exclude distracting members when focusing on particular speaker or writer in voice- or text-chats. To prevent one's voice or words from being delivered to other members, users can use `deafen` or `attend` functions. By cloning, one can be at multiple spaces at the same time.

This is the first enterprise-quality, distributed narrowcasting and multipresence system, which flatters and is showcased by the unique Smithsonian Folkways world music collection. Conforming to MX: IEEE 1599, we embrace standard encoding format, encouraging possible future extensions. For instance, the authors are preparing a space that emphasizes shared musical heritage between Cape Breton (Nova Scotia) and
Fn 19 Ukraine.[19]

Research will at the outset be exploratory, but we anticipate that this preliminary phase will quickly lead to the formulation of hypotheses and, subsequently, more focused experimentation designed to test them. We believe that this process will produce results suggesting better ways of designing musical cyberworlds for research, discovery, learning, entertainment, and e-commerce, as well as indicating broader principles underlying the role of music in human interaction and community–formation in general. In this way, controlled research in and about a custom-built musical cyberworld can usefully supplement, without supplanting, traditional real-world fieldwork in ethnomusicology.

[19] `http://diversitycapebreton.ca/ir/visit-virtual-cape-breton`

References

Ackerman, M. S., Starr, B., Hindus, D., & Mainwaring, S. D. (1997, March). Hanging on the 'wire: A field study of an audio-only media space. ACM *Transactions on Computer-Human Interaction*, *4*(1), 39–66.

Alam, S., Cohen, M., Villegas, J., & Ahmed, A. (2009, March). Narrowcasting for articulated privacy and attention in SIP audio conferencing. *Journal of Mobile Multimedia*, *5*(1), 12–28.

Baggi, D., & Haus, G. (2009, March). IEEE 1599: Music encoding and interaction. IEEE *Computer*, *42*(3), 84–87.

Baggi, D. L., & Haus, G. M. (2013). *Music navigation with symbols and layers: Toward content browsing with* IEEE *1599* XML *encoding*. New York: John Wiley & Sons, Inc.

Barz, G. F., & Cooley, T. J. (Eds.) (2008). *Shadows in the Field: New perspectives for fieldwork in ethnomusicology* (2nd ed.). Oxford: Oxford University Press.

Bormann, K. (2005, June). Presence and the utility of audio spatialization. *Presence: Teleoperators and Virtual Environments*, *14*(3), 278–297.

Boustead, P., & Safaei, F. (2004). Comparison of delivery architectures for immersive audio in crowded networked games. *Proceedings 14th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, 22–27.

Boustead, P., Safaei, F., & Dowlatshahi, M. (2005). IEEE: Internet delivery of immersive voice communication for crowded virtual spaces. *Proceedings* IEEE *Conference 2005 on Virtual Reality*, 35–41.

Campbell, M. A. (2005, July). Cyber bullying: An old problem in a new guise? *Australian Journal of Guidance and Counselling*, *15*(1), 68–76.

Chen, Y.-X., & Butz, A. (2009). Musicsim: Integrating audio analysis and user feedback in an interactive music browsing UI. *Proceedings International Conference on Intelligent User Interfaces*, 429–434.

Cohen, M. (2000, February). Exclude and include for audio sources and sinks: Analogs of mute & solo are deafen & attend. *Presence: Teleoperators and Virtual Environments*, *9*(1), 84–96.

Cohen, M., & Győrbiró, N. (2009, November). Mobile narrowcasting spatial sound. *Proceedings of* IWPASH*: International Workshop on the Principles and Applications of Spatial Hearing*, Zao, Miyagi, Japan.

Cohen, M., & Kawaguchi, M. (2003, July). Narrowcasting operations for mobile phone CVE chatspace avatars. *Proceedings of* ICAD*: International Conference on Auditory Display*, 6–9.

Cooley, T. J., Meizel, K., & Syed, N. (2008). Virtual fieldwork: Three case studies. In *Shadows in the Field: New perspectives for fieldwork in ethnomusicology* (2nd Ed., pp. 90–107). Oxford: Oxford University Press.

Crisostomo, S., Safaei, F., & Platt, D. (2004, June). On the provision of immersive audio communication to massively multi-player online games. ISCC*: Proceedings Ninth International Symposium on Computers and Communications, 2*, 1000–1005.

Damm, D., Fremerey, C., Kurth, F., Müller, M., & Clausen, M. (2008). Multimodal presentation and browsing of music. ICMI*: Proceedings 10th International Conference on Multimodal Interfaces*, 205–208.

Downie, J. S. (2002, October). Toward a theory of music information retrieval queries: System design implications. *3rd International Conference on Music Information Retrieval*, 299–300.

Fernando, O. N. N., Adachi, K., Duminduwardena, U., Kawaguchi, M., & Cohen, M. (2006, January). Audio narrowcasting and privacy for multipresent avatars on workstations and mobile phones. IEICE *Transactions on Information and Systems, E89-D*(1), 73–87.

Fine, G. A. (2001). Participant observation. In *International Encyclopedia of the Social and Behavioral Sciences* (pp. 11073–11078). Amsterdam: Elsevier Science.

Frank, J., Lidy, T., Peiszer, E., Genswaider, R., & Rauber, A. (2008). Ambient music experience in real and virtual worlds using audio similarity. *Proceedings of the International Workshop on Semantic Ambient Media Experiences*, 9–16.

Funkhouser, T., Min, P., & Carlbom, I. (1999). Real-time acoustic modeling for distributed virtual environments. SIGGRAPH*: Proceedings 26th Annual Conference on Computer Graphics and Interactive Techniques*, 365–374.

Gardner, M., Gnem-Gutirrez, A., Scott, J., Horan, B., & Callaghan, V. (2011). Immersive education spaces using Open Wonderland from pedagogy through to practice. In *Multi-User virtual environments for the classroom: Practical approaches to teaching in virtual worlds* (pp. 190–205). Hershey, PA: Information Science Reference.

Geerts, D. (2006). Comparing voice chat and text chat in a communication tool for interactive television. *Nordi*CHI*: Proceedings Nordic Conference on Human-Computer Interaction: Changing Roles*, 461–464.

Greenhalgh, C., & Benford, S. (1995). MASSIVE: A distributed virtual reality system incorporating spatial trading. *Proceedings 15th International Conference on Distributed Computing Systems*, 27–34.

Hughes, B., & Kamat, A. (2005, Feb.). A metadata search engine for digital language archives. *D-Lib Magazine*, *11*(2), 891–908.

Ibanez, M. B., García, J. J., Galán, S., Maroto, D., Morillo, D., & Kloos, C. D. (2010). Multi-user 3D virtual environment for Spanish learning: A Wonderland experience. ICALT*: Proceedings International Conference on Advanced Learning Technologies*, 455–457.

Kaplan, J., & Yankelovich, N. (2011, September). Open Wonderland: An extensible virtual world architecture. IEEE *Internet Computing*, *15*(5), 38–45.

Kong, L. (2001, December). Religion and technology: Refiguring place, space, identity and community. *Area*, *33*(4), 404–413.

Kuhn, M., Wattenhofer, R., & Welten, S. (2010). Social audio features for advanced music retrieval interfaces. *Proceedings of the International Conference on Multimedia*, 411–420.

Lee, S., & Kim, G. J. (2008, September). Effects of visual cues and sustained attention on spatial presence in virtual environments based on spatial and object distinction. *Interacting with Computers*, *20*(4–5), 491–502.

Lomax, A. (1997). *Saga of a folksong hunter: A twenty-year odyssey with cylinder, disc and tape*. The Alan Lomax Collection Sampler—Rounder CD 1700, 1997.

Ludovico, L. A. (2008). Key concepts of the IEEE 1599 standard. *Proceedings of the Conference on the Use of Symbols to Represent Music and Multimedia Objects*, 15–26.

Ludovico, L. A. (2009). IEEE 1599: A multi-layer approach to music description. *Journal of Multimedia*, *4*(1), 9–14.

Naef, M., Staadt, O., & Gross, M. (2002). Spatialized audio rendering for immersive virtual environments. VRST*: Proceedings* ACM *Symposium on Virtual Reality Software and Technology*, 65–72.

Nettl, B. (2005). *The study of ethnomusicology: Thirty-one issues and concepts*. Chicago: University of Illinois Press.

Peña-Ríos, A., Callaghan, V., Gardner, M., & Alhaddad, M. J. (2012). Towards the next generation of learning environments: An InterReality learning portal and model. *Intelligent Environments*, 267–274.

Ranaweera, R., Cohen, M., & Frishkopf, M. (2013, May). Narrowcasting enabled immersive music browser for Folkways world music collection. CASA*: Proceedings of the International Conference on Computer Animation and Social Agents*, 420.

Ranaweera, R., Frishkopf, M., & Cohen, M. (2011). Folkways in Wonderland: A cyberworld laboratory for ethnomusicology. CW*: Proceedings of the International Conference on Cyberworlds*, 106–112.

Seo, B., Htoon, M. M., Zimmermann, R., & Wang, C.-D. (2010). Spatializer: A web-based positional audio toolkit. *Advances in Computer Entertainment Technology*, 57–61.

Taylor, J. (1997, April). The emerging geographies of virtual worlds. *Geographical Review*, *87*(2), 172–192.

Taylor, T. L. (1999). Life in virtual worlds: Plural existence, multimodalities, and other online research challenges. *American Behavioral Scientist*, *43*(3), 436–439.

Waldo, J. (2008, August). Scaling in games and virtual worlds. *Communications* ACM, *51*(8), 38–44.

Yankelovich, N., Kaplan, J., Provino, J., Wessler, M., & Di-Micco, J. M. (2006). Improving audio conferencing: Are two ears better than one? CSCW*: Proceedings of the Conference on Computer Supported Cooperative Work*, 333–342.

Yankelovich, N., McGinn, J., Wessler, M., Kaplan, J., Provino, J., & Fox, H. (2005). Private communications in public meetings. CHI*: Extended Abstracts on Human Factors in Computing Systems*, 1873–1876.

Yankelovich, N., Walker, W., Roberts, P., Wessler, M., Kaplan, J., & Provino, J. (2004). Meeting central: Making distributed meetings more effective. CSCW*: Proceedings of the Conference on Computer Supported Cooperative Work*, 419–428.