

The Seventeenth Annual Interactive Audio Conference

PROJECT BAR-B-Q 2012



Group Report: Form Factors and Connectivity for Wearable Audio Devices

Participants: *A.K.A. "Clothing Optional"*

Peter Drescher, Twittering Machine

Whit Hutson, IDT

Konstantin Merkher, CEVA

Alan Kraemer, DTS

Facilitator: Jim Rippie, Invisible Industries

David Tan, Beats Electronics

Dave Berol, Wolfson

Randy Granovetter, KNUVU

Karen Collins, University of Waterloo

Problem Statement

In the mobile social era, a wide variety of powerful communication devices will be developed for a myriad of purposes and use cases. What form factors might audio enabled wearable computing devices take? More importantly, how will they be connected to each other, and what kinds of data will they receive, record, and transmit?

A brief statement of the group's solutions to those problems

The group discussed various form factors wearable audio devices might take. Ranked from smallest to largest:

1. earbuds / earpods in-ear devices, basically as described @ [BBQ 2007](#)
2. Headphones portable "studio cans", as made popular by Beats Electronics
3. Helmets as worn by motorcyclists, with multiple built-in speakers and microphones, both inside and out
4. Clothing / Accessories speakers / mics built into jackets, earrings, sunglasses, augmented reality goggles, necklaces, pendants, lanyards, string ties, scarves, lapel pins, buttons, shirts/blouses, watches, wrist bands (ala Nike Fuel), belts (for power packs), et al.
5. Cars automobiles as the ultimate portable audio environment.

The group mainly focused on how audio data from various wearable devices would be recorded, transmitted, and rendered, using the multidimensional audio format from the 3D Audio Alliance (3daa.org) as a starting point. This initiative describes an open, royalty-free, specification whereby 3D audio is authored as an "object" containing content plus metadata, which can then be played back on any theater or television system regardless of speaker configuration. Audio is dynamically routed and mixed as necessary, based on the target platform's capabilities, as determined by the metadata.

The group considered turning that concept around, so that data from various wearable microphone and sensor inputs (location, ambient noise levels, et al) could be encoded into an audio object, to be streamed to the cloud. The object could then be streamed to any audio enabled device, and played back in an optimal manner, regardless of hardware configuration (mono cell phone, car stereo, gaming surround system, etc).

Thus, wearable audio devices would become nodes in a local mesh network, where some sensors (such as mics in hands-free systems in cars) would have known and stable locations, whereas others (such as mics in cell phones, worn accessories, and notebook computers) would need to be queried and discoverable. Each node becomes an access point for acquiring, encoding, and mapping the environment as audio objects, which are then rendered in the context of available wearable devices. The audio experience is thus dynamically optimized per use case, according to your personal profile, and profiles of other users in the network.

Expanded solution description

Use case scenarios:

a) In the car, with children riding in the back seat, the driver gets an invitation to join a conference call. Mics in multiple locations in the interior of the automobile encode the user's voice while suppressing noise from the environment (and the back seat). Metadata from the conference call audio object allows the various participants to be positioned left, right, and center, to increase intelligibility. Audio can even be routed so that the user hears the call, but the kids do not.

b) The World of Warcraft theme park: The user enters a walled garden, wearing augmented reality goggles, with connected earpods that comfortably fit over his ears (the fact that they look like pointy elven ears is merely an added bonus). Other players look like their avatars through the goggles, and Head Related Transfer Functions ensure that the roar of the dragon always comes from the same direction, no matter where the user looks. 3D directional audio also helps the player hear the heavy stomps of an ogre approaching from behind, and renders the sword swishes and magic spell whooshes of his teammates during battle.

c) The user is sitting outside a Starbucks, wearing Beats Headphones with built-in mics, listening to his favorite band, the Zombie Dinosaurs, on a **Cloud music service**. The external mics are used to perform active noise reduction on the (stationary) cafe background noise, while using a different algorithm to eliminate the (transient) sounds of a truck passing by, and a barking dog.

d) While walking to work from the train, the user receives a phone call from her boss. She flips up her earrings, which double as Bluetooth headset speakers, and puts them in her ears. Not only do the mics in her phone and necklace detect and characterize the ambient noise separate from her voice (allowing her boss to hear her clearly), the incoming audio environment is encoded so precisely, she can hear that the boss is in a large conference room, rather than in his office.

e) An Indy 500 race car driver is wearing a helmet festooned with 2 video cameras, multiple internal speakers, and 7 microphones, both inside and out. He talks easily with his pit crew while driving at high speed, as the sensor data from the cameras and mics is captured to be used as source material for the latest version of Need for Speed. Because all data is stored as interactive audio objects, the soundtrack can be rendered on any gaming system, from a PC with inexpensive little speakers, to an Xbox 5.1 system where you can really hear the roar of the engines, to the immersive race car model at the video arcade, which simulates the swerves, jolts, and vibrations of the real thing.

f) A small group of colleagues discuss a complicated problem in a noisy environment -- street traffic, metal carts rolling down the sidewalk, background Muzak leaking in from the hair salon next door. A person at one end of the table would not be able to hear someone speak at the other end in that environment, but fortunately, everyone is wearing network connected earpods. This not only provides noise cancellation, so they can all chat with each other in normal voices, but also lets a colleague in a distant city join the discussion as if he were there in person.

g) A large group of colleagues sit at a series of tables arranged in a horseshoe shape. They place their cell phones on the tables in front of them, and activate the internet connected mics, which collect audio and proximity data from around the room. When any one of the participants talks, the PA speakers in the corners automatically configure themselves and set volumes levels so that everyone in the room can clearly hear anybody speak.

h) A noisy political fund raiser in a large ballroom is captured as audio objects in three dimensions. A contributor accesses the event remotely, interacting dynamically with the audio in real time. He is able to converse interactively with various people at the event by isolating individuals from the general background noise floor (e.g. "the cocktail party effect").

i) An older guitar player is having issues in a concert setting; he can see the drummer playing a hi hat riff, but cannot clearly hear it. He adjusts his magical wearable audio thing, and now everything sounds frakkin' great!

Items from the brainstorming lists that the group thought were worth reporting

While adoption of various form factors may be heavily based on style and branding, product design is perhaps the most important aspect of creating a popular wearable audio device, as evidenced by the attention to mechanical detail, and materials used by Beats headphones. When fashion trends, lifestyles, and technology meet, new products can be successful that are both functionally superior and extremely cool, despite being somewhat unpredictable (who would have thought the "studio can" form factor would become in vogue outside the recording studio?)

"Big Brother" security concerns become an important aspect of mesh node sensor network technology. If location data from various sensors could be hacked, it would facilitate both criminal activities and government breeches of civil rights. Do we really want always-on mics installed in your wearables, or at your local Starbucks? How would military and anti-terrorist organizations deal with the possibility of this style of network hacking? In some cases, an individual might want to "opt out" of the environmental encoding system of the wearable audio network, by disabling upload of audio objects (e.g. "airplane mode"), while still allowing voice commands to be acknowledged and processed. On the other hand, young people growing up in an always-connected society might not be so concerned with that level of privacy.

How to power wearable audio devices? While we are overdue for a breakthrough in battery technology, the wearable sensors being discussed (microphones, location sensors, etc) actually require relatively little power. Nor do they have to transmit signals over large distances, as they will mostly communicate with a nearby cell phone, computer hub, or automobile that acts as a "mothership" to myriad peripherals. Thus, sensor systems could be partially powered by biometrics as users perform normal activity during the day. Shoes with built in kinetic generators could be used to charge cell phones ("My battery's running low; I'm going for a run"). Devices built into clothing could be charged by hanging them up in a wireless-powered closet. In any case, one might expect wearable audio devices to take the form of removable accessories (pins, lanyards, et al), rather than being woven directly into shirts and blouses (although these could also be considered a fashion statement).

How To Make Money With Wearable Audio Devices:

- Sales of branded accessories (e.g. Beats headphones, Nike Fuel band, etc)
- Cloud services for encoding data
- Aggregating and processing audio data
- Connectivity for transmitting audio objects
- Rendering engines and algorithms
- Related audio-enabled systems, such as personal assistants, virtual audio environments, adword marketing, voice recognition and search, enterprise and corporate services
- Other Internet audio services and subscriptions (music, gaming, etc)

People will need quality audio devices that are comfortable enough to wear for up to 8 hours a day, since the user will interact with multiple audio software applications during their day: phone calls, voice commands, voice recognition, video/web conferencing, transcription and translation services, as well as listening to music or playing games. Intelligent hands-free audio devices will serve not only as a universal interface to these applications, they will also become increasingly important to the automobile industry, due to anti-distraction vehicle operating laws in the United States and Canada. Additionally, there is a growing market for Bluetooth enabled smartphones (estimated to be 1.1 trillion by 2014).

Given multiple connected devices, and a plethora of audio enabled applications, we can expect users to wear a variety of audio devices, based on personal preference, and the environment they interact with throughout the day.

Other reference material

BBQ earpod report 2007 - <http://www.projectbarbq.com/bbq07/bbq07r5.htm>

BBQ Crossbar group 2005 - <http://www.projectbarbq.com/bbq05/bbq05r4.htm>

Annoying Audio blog: earpods - <http://blogs.oreilly.com/digitalmedia/2007/12/earpods-can-you-ihear-me-now.html>

MDA spec - <http://www.3daa.org/>