Ambient Addition:

How to Turn Urban Noise Into Music

Noah Vawter
B.S., Electrical Engineering
Worcester Polytechnic Institute, May 1997

Submitted to the Program in Media Arts and Sciences School of Architecture and Planning in partial fulfillment of the requirements of the degree of Master of Science in Media Arts and Sciences at the Massachusetts Institute of Technology May 2006
[June 2006]

Author: Noah Theodore Vawter
Program in Media Arts and Sciences
May 17, 2006

Certified by: Chris P. Csikszentmihályi
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by: Dr. Andew B. Lippman
Chair, Departmental Committee on GraduateStudies
Program in Media Arts and Sciences

Ambient Addition:

How to Turn Urban Noise Into Music

Noah Vawter

Submitted to the Program in Media Arts and Sciences School of Architecture and Planning in partial fulfillment of the requirements of the degree of Master of Science in Media Arts and Sciences at the Massachusetts Institute of Technology

Abstract

As human civilization devises ever more powerful machines, living among them may become more difficult. We may find ourselves surrounded by incidentally created sounds and noises which are out of synchronization with our momentary needs and discordant. Currently, legislating noise pollution is the only articulated solution and clearly it is not very effective. Our impression of sound, however, may be mediated and manipulated, transformed into something less jarring. So far, Walkmans and sound canceling headphones have done this, isolating us from noise but also from one another. In their place, a next generation headphone system is proposed which integrates environmental sound into a personal soundscape. It allows one to synthesize music from environmental sound using a number of digital signal processing (DSP) algorithms to create a sonic space in which the listener remains connected with his or her surroundings, is also cushioned from the most harsh and arrhythmic incursions and may also be drawn to appreciate the more subtle and elegant ones.

Ambient Addition:

How to Turn Urban Noise Into Music

Noah Vawter

Chris Csikszentmihályi
Associate Professor of Media Arts and Sciences
Fukutake Career Development Professor
Massachusetts Institute of Technology

Thesis Advsior

Ambient Addition:

How to Turn Urban Noise Into Music

Noah Vawter

Barry Vercoe
Professor of Media Arts & Sciences
Director, Music, Mind and Machine Group
MIT Media Laboratory

Thesis Reader

Ambient Addition

How to Turn Urban Noise Into Music

Noah Vawter

Douglas Repetto
Director of Research
Columbia University,
Computer Music Center

Thesis Reader

Acknowledgments

I would like to thank Jim Barabas for letting me use his Linux machine before I could get JTAG Tools working under Cygwin.

I would like to thank Nathan Wray for compiler assistance, among dozens of other things.

I would like to thank Ethan Bordeaux, Ben Recht and Brian Whitman, co-members of the DSP-Music Syndicate, for listening and believing. Some of the concepts for Ambient Addition were brainstormed by the four of us together.

I would like to thank Elisa Archer for her support and encouragement throughout this project.

I would like to thank every teacher I have ever had for preparing me for this world.

I would like to thank Ayah Bdeir, Kelley Dobson, Limor Fried, Tad Hirsh, jroth, Lira Nikolova, Gemma Shusterman, and Adam Whiton - my fellow Computing Culture research group members - for being part of it all...and especially to Chris Csikszentmihályi for directing it all.

I would like to thank my readers Doug Repetto and Barry Vercoe for contributing their time to this thesis.

I would like to thank my parents for teaching me to see the world in such an unusual way.

Figure 1. Bust of Roy Lamson enjoying the busy environment of the Media Lab transformed into music by Ambient Addition.

Table of Contents

“NOISE, n. A stench in the ear. Undomesticated music. The chief product and authenticating sign of civilization.” -Ambrose Bierce (1, 1911)

Introduction

Noise is a problem for people in urban environments. To shield themselves from it, many people listen to Walkmans* while walking in the city. Interestingly, they reserve the use of earplugs and noise-canceling headphones - devices which eliminate noise - for airplanes, industrial work, sleeping, etc. Although people block most city noise out, they want to hear part of it as well. As Jean-Paul Thibaud, a sociologist at Cresson Laboratories, Grenoble describes:

> The walking listener uses it [headphone listening] not only to protect himself from the sonic aggressions of the city but also to filter and enhance the events that
> the place its meaning. [...] Depending on the places and what is happening in them, the sound volume of the Walkman is used in order to be able to listen to or to mask conversations, bells ringing, children's screams, traffic noise, and so forth. [...] 'When there's a great deal of noise around, I make the Walkman louder.” [...] 'The other day, I lowered the volume because I wanted to hear the bells...” (5, 2003)

* In 1986, "Walkman" was entered into the Oxford English Dictionary. Its definition is broad enough to encompass all personal stereo systems, including the popular iPod. (2, 1999) Curiously enough, the 'W' at the beginning of the word is capitalized. The acceptable plural forms of Walkman are Walkmans, and Walkmen. (3, 2006) The Merriam-Webster Dictionary concurs, although it does not list the plural forms. (4, 2006)

In many cases though, people are acoustically isolated from the outside world. As activities like laughter and conversation take place around them outside their attention, Walkman listeners become socially isolated.

The following paragraphs offer evidence of the social isolation that Walkmans may produce. Following that discussion, some approaches artists have taken to turning noise into music are examined. Finally, in the conclusion to the introduction, the technology of Ambient Addition, which addresses these ideas, is introduced.

The Isolation That Walkmen Create

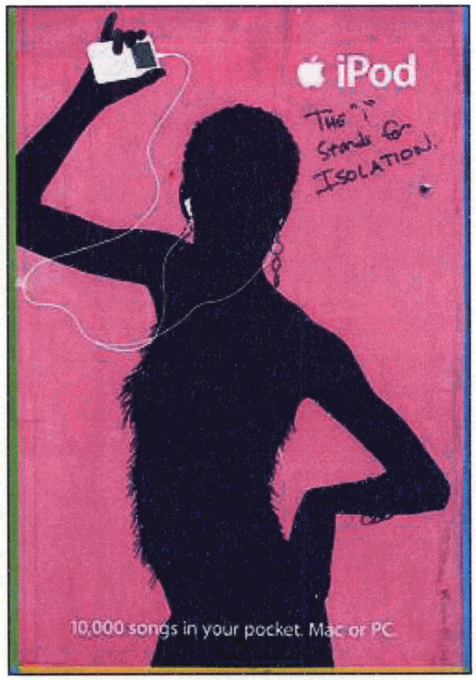

Figure 2. A defaced advertising poster in NOHO, New York City. (7, 2004)

Figure 2 depicts an anonymously defaced poster from the NOHO area of New York City in 2004. The unsolicited remark in the upper right hand corner states that the 'i' in Apple's "iPod"* stands for isolation. This grafito echoes complaints by principals and cultural critics about the isolation iPods create, which have rekindled in the last few years with the commercial success of the iPod. Some of the complaints mention concerns about isolation from when Sony first introduced the Walkman. For example, Nicholas Taylor, who

*a popular digital Walkman.

writes on music and culture in *Pop Matters* and *New York Spirit*, wrote:

> These fears concerning the social effect of the iPod are not new. They mirror very closely the iPod's analog progenitor, the Sony Walkman, which sparked the portable music revolution in the 1980s. Critics feared then just as they do now that the promise of portable music listened to through headphones would make people social zombies, unable to relate to the world around them, trapped between their left and right audio channels, in a world of their own. (8, 2005)

In 2005, iPods were banned in an Australian school. Taylor reported that "according to the school's principal, Ms. Kerrie Murphy, her decision to do away with a device taking the world by storm coincides with a wider discussion of how iPod listeners are 'not tuning into other people because they're tuned into themselves.'" (8, 2005) In the same article, Taylor defended the iPod, because he values the complex world of emotional experiences listeners are experiencing when they use them. While this may be valid, iPods create such a gap between the external and internal worlds of sound, that they sometimes lead to safety issues. Journalist Kerry Dougherty describes her iPod experience while jogging: "What I couldn't see, because it was right behind me, and couldn't hear, because 'Ring of Fire' was echoing through my eustachian tubes, was a big white sport utility vehicle on my heels. [...] She slammed on her brakes. Screamed to a stop. Glared at me. I deserved it. Lucky for me, she wasn't listening to her own iPod." (9, 2006)

When describing the Walkman's isolation effects, Phil Patton, a writer for *The New York Times* and contributing editor of *ID Magazine*, *Wired*, and *Esquire*, invoked anthropologist Edward Hall's model of human movement:

> Three decades ago the sociologist Edward Hall introduced the concept of the "space bubble," that culturally conditioned distance that dictates how close we stand to another person and how much space we need around us to feel comfortable. The Walkman might be said to have introduced another kind of bubble: a technological bubble of concentration and obliviousness to surroundings, a private space in public. Today, the streets are full of cellular telephone users enveloped in similar bubbles of communication and concentration. Palm pilots and other small digital devices have similar effects. (10, 1999)

Figure 3. Arthur Elsenaar and Taco Stolk's BuBL SPACE jams cellular teelephone signals within a 3m radius. (11, 2004)

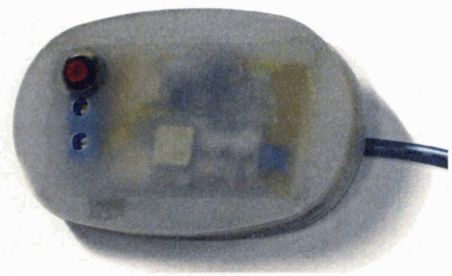

Figure 4. Limor Fried's "Wave Bubble" jams cellular telephone signals in a bubble around its user. Also, instructions to build it are available online. (12, 2005)

One response to the "technological bubbles of concentration and obliviousness to surroundings" has been the development of devices to "pop" the bubbles by jamming cellular telephone signals. In 2004, Arthur Elsenaar and Taco Stolk introduced BuBL SPACE as an art project which jams signals within a 3m radius. At the time of this writing, several cell phone jammers are available for purchase from the UK (their sale and possession inside the United States is a violation of FCC regulation) for around $300. Notably, Limor Fried published instructions for constructing her implementation of such a device, "Wave Bubble," in 2005 (see Figure 4). (12, 2005)

As stated earlier, complaints about social isolation began with the release of Sony's original Walkman* in 1979. Even its inventors were concerned about its possible solitary effects. On the 20th anniversary of the Walkman, Phil Patton published an article in which he explained:

> On the original model Sony Walkman, introduced twenty years ago this month under the trademark "Soundabout," was an orange button [...] it also had a second earphone jack and when you pushed the orange button, the sound emerged into two sets of phones and two listeners could talk to each other through a microphone. It was a revealing feature: even Sony was apparently worried about the solitary qualities of the Walkman. The orange button was like a panic button, an emergency "share" feature. The company was that hesitant, Sony cofounder Akio Morita wrote later, to release a product that was somehow so selfish.
>
> [...]
>
> "I noticed my experiment was annoying my wife, who felt shut out," he reported in his book, and so he ordered the addition of

* Recently Andreas Pavel has been recognized as the true inventor of the Walkman, so in this paragraph, 'Sony's original Walkman' actually refers to their model TPS-L2.

a second headset jack—and the orange button [...] Mr. Morita reported that he "thought it would be considered rude for one person to be listening to his music in isolation..." (14, 1999)

Figure 5. Sony's original TPS-L2 Walkman, with dual headphone jacks, microphone, and an orange button for allowing two listeners to communicate with each other. (13, 2004)

Summary of Walkman Isolation

Although Walkmans do not completely eliminate noise, they still inhibit social interaction to a degree. Some strategies exist to integrate external sound with headphone sound, such as permeable headphones, modulating volume levels (Thibaud), or temporarily mixing a microphone in with the music (Sony's TPS-L2). Perhaps though, something else can be done to cushion the blow of urban noise while keeping people involved with their acoustic environment. For ideas on how to do this, the next section discusses relevant examples from musicians who work with noise and music.

Artistic Strategies For Dealing With Noise

This section discusses works from five artists which influenced Ambient Addition. In particular, I'm analyzing works that were created to address our relationship to noise.

Noise Instruments

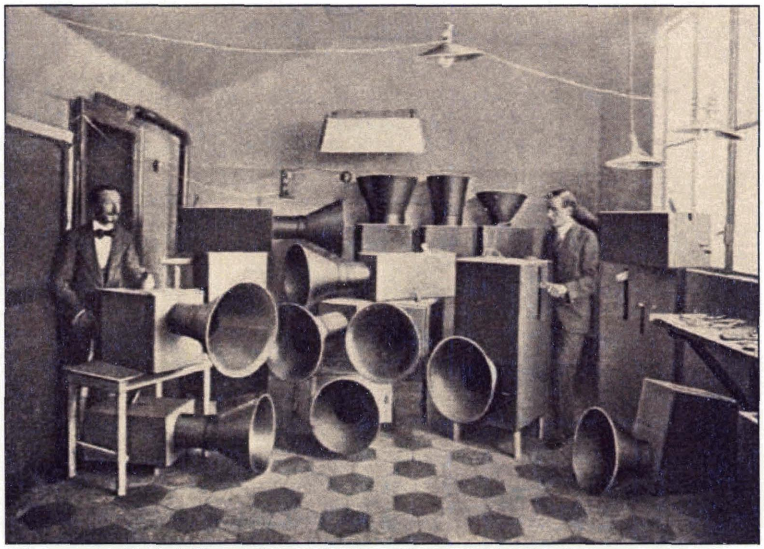

Figure 6. Luigi Russolo and Ugo Piatti with the Intonarumori, circa 1913 (15, 1987)

One of the early 20th century artists who described themselves as “Futurists,” Luigi Russolo, is a well-known sound artist for his manifesto "The Art of Noises" and for his *intonarumori* (noise-generators, depicted in Figure 6). He wanted to create music that would compete in power with machines of the modern world, declaring that traditional instruments paled in comparison to the sounds of industrialization. One of his famous quotes is "Do you know of any sight more ridiculous than that of twenty men furiously bent on redoubling the mewing of a violin?" (15, 1987) He performed concerts using his intonarumori, which were designed to emulate the machines of his age. Composer and sound poet Larry Wendt describes

Russolo's goals:

> He wanted to go back to the roots of sound in modern instruments, and also suggested using the sounds of nature (wind, water, and various animal sounds), the sounds of the modern urban environment, and 'all the noises which are made with the mouth without talking or singing.' He had no formal training in acoustics, instrument building, or musicianship, however as an enthusiastic amateur scientist, he proceeded to categorize the basic sounds in his environment and build instruments to mimic these sounds. (16, 1998)

Part of Russolo's goal was to move away from harmonic instruments.

Musique Concrète

Pierre Schaeffer's work from 1948 is interesting because he created music out of everyday items in his Études, such as "Study with Baking Pans" and "Study with Whirligigs." He was an early embracer of magnetic tape technology. The manipulations he performed with tape, like loops, speed-changing, phasing and reversing are still being used more than 50 years later. (17, 2001) Speed-changing is especially relevant, because it can be used to bring many incidental sounds into harmony.

It should also be acknowledged that Filippo Tommaso Marinetti was the first to create "noise scores" - notation to describe a composition in terms of what noises to play, when to play them, and for how long. (18, 1981)

Sampling: Acoustic

Figure 7. Xerophonics by Stefan Helmreich organizes the sounds of photocopiers into compelling songs. (19, 2003)

Musician Stefan Helmreich is a reference point because his technique involves editing environmental sound into something more listenable. His album *Xerophonics* extracts the rhythms of 13 different photocopiers and sequences them into instrumental songs. One reviewer wrote "The unending rhythms of the copiers relax and capture me, taking me back to the countless copier rooms and desks I've occupied as a worker drone in America. From now on, I hope to hear more of the music that always plays around me." (20, 2005) A second reviewer wrote: "The rhythmic clunks and motorized whirs can be sort of soothing..." (21, 2003) *Xerographics* stands out because of its excellent composition. These are not merely the sounds of photocopiers; they are formed into compelling songs.

Sampling: Social

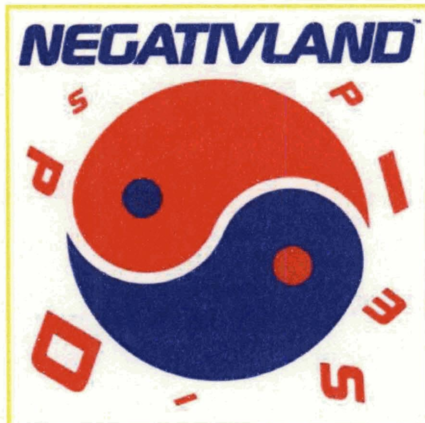

Figure 8. Negativland's album Dispepsi treats commercially produced sound as fodder for collage. (22, 1997)

Like Stefan Helmreich, collage artists Negativland consider the noise surrounding them fodder for collage. They create collages from sound which they consider themselves bombarded with, particularly advertisements. Instead of concentrating on musical features like melody, harmony, and rhythm, they focus on multi-track montages of sampled phrases. For instance, a highlight of their album "Dispepsi" (Figure 8) features advertising jingles encouraging people to drink Pepsi-Cola interspersed with celebrity Bill Cosby's voice from a public service announcement saying, "I am really tooth decay."

Field Recording

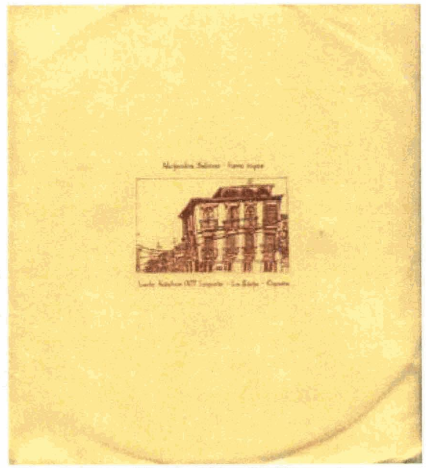

Figure 9. Alejandra Salinas' album "Home Tapes" demonstrates the fascination some people have with field recordings. (23, 2000)

One artist in particular, Alejandra Salinas, provides listeners a chance to perceive environmental sound from a completely different perspective. Her album "Home Tapes" which "consists of odd background sounds and indigenous music rising and receding between tape hiss and clips of family dialogue and children singing," (24, 2000) seems to be noisy, but at the same time, it stimulates the listener to recall intimate moments of his or her own family life. It shows that even a minimum of editing techniques combined with a shift in context can produce a strong effect.

Conclusion to Introduction

Having considered urban noise, the social isolation Walkmans create trying to tame noise and musical works created around noise, I submit Ambient Addition as a response. It is a portable Walkman that immerses the listener in an artificial sonic world incorporating the ambient sound around him or her. Consisting of a pair of headphones with small, embedded microphones, and a pocket-sized digital signal processing (DSP) system, it continuously records, analyzes, transforms and plays back environmental noise into more musical form. Through its outward appearance and sonic-bridging capabilities, it reduces isolation. The remainder of this thesis presents compositional techniques and details on design decisions, development, algorithms and results.

Technical Possibilities

Now that the motivation to transform ambient sound into music has been explained, this chapter will discuss techniques for doing it. A number of familiar musical characteristics such as rhythm and harmony will be identified, followed by descriptions of how Ambient Addition can synthesize the same effects out of environmental noise.

Before explaining how Ambient Addition can create music from environmental noise, it should be noted that the term "music" can be intractably broad. To assist future musicians in creating a broad range of musical styles, I will describe the operations Ambient Addition can perform in a very general way. Nevertheless, the reader should be warned that this model of music will inevitably reflect my own musical biases. As George Lewis states, "Musical computer programs, like any texts, are not 'objective' or 'universal,' but instead represent the particular ideas of their creators." (1, 2000)

Repetition and Rhythm

Music generally contains perceptible chunks which repeat at regular intervals. These repetitions with minimal variation act as a backdrop or skeleton over which other sonic activity takes place. In techno music, for example, bass thumps repeating from two to three times per second (120-180 BPM) form such a solid framework that a myriad of unusual sounds can be overlaid on top of them without losing coherency. In other forms of music, a repeating bassline performs a similar function, giving wandering and exploring melodic lines a place to return to. Repetition can also be nested to traverse larger time scales, such as when a pop song's repeated verses consist of repeated melodies. Repetition also forms the basis of rhythm.

In urban environments, repetition is fleeting. Since urban sound is generally the by-product of activity, a street scene will rarely sound musically repetitive unless someone or something is purposefully performing repetitive activity. Moreover, the repetition in street activity is often too irregular or disorganized to be perceived as rhythmic. For example, on garbage day, the clamor of steel cans may be heard every 45 seconds as a garbage truck makes its progress. However, the random sounds during those intervals make it difficult to perceive a regular rhythm. In this case, Ambient Addition can employ sampling to create a virtual sonic reality which resembles music more than the original environment.

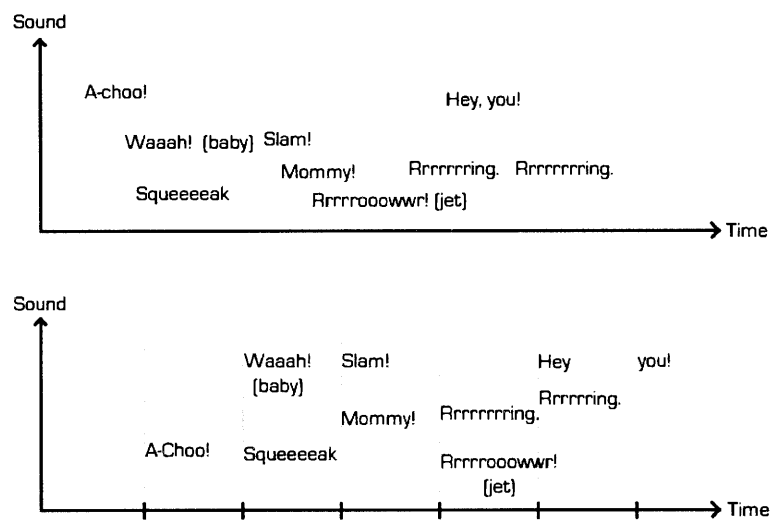

Figure 1. In the top graph, a typical sonic environment. In the bottom graph, the same sounds are played back at regular intervals to create a rhythm.

The fundamental technique, depicted in Figure 1, is to detect discrete sounds as they occur, record them, and play them back at regular intervals. In a simple scenario, a listener hears a door slamming shut. Since the door does not slam again for some time and no regular activity takes place in between, there is no perception of rhythm. However, if Ambient Addition records the slam and plays it back once every second, it creates the illusion of repetitiveness. An example of this compositional technique can be heard in the song "In the Army Now" by the band The Art of Noise (2, 1986). This basic technique can be varied and expanded in many ways to produce a range of effects.

For example, a composer may feel that listeners would prefer to hear a door slamming less frequently than a gentle sound. Therefore, Ambient Addition could analyze each sound it records, assessing its degree of harshness, and use that to inform the sequencer how often to play the sound back. This simple scheme has its limits though, because in most music, sounds do not simply repeat at regular intervals for very long. Instead, groups of sounds tend to repeat in patterns. Therefore, Ambient Addition could hold samples of several sounds at once in its memory and play them back like the rhyme scheme of a poem: A B A C A B A D. That is just one example of creating rhythm from a number of sounds. The complexities of rhythm are a tremendous area of research on their own, and are explored in many sources, such as Stephen Handel's Listening. (3, 1989)

Sounds which can not be fully eliminated from the headphones are a special consideration. For example, a car horn is almost always audible through a playing Walkman. Even though such a sound is random with respect to the internal sequencer of Ambient Addition, it can be integrated into the composition. One compositional suggestion is to simulate all possible ways in which a car horn may interrupt an internal rhythm. To do this, a computer sequencer or drum machine may be used to create 16 identical copies of a rhythm. Each copy should be slightly modified by triggering a car horn sample in it at a different place, such as at multiples of 1/16th notes. Then, for each rhythm, one should try to figure out the best place to trigger the car horn a second time to resolve that rhythm, or respond to the first sound. This information can be used to play the car horn sample back optimally in real-time.

Harmony

In addition to fleeting rhythm, another feature of urban noise is that the frequencies one hears are almost never in harmony with each other. This comes as little surprise because the audible emanations of sources like car engines, human voices and telephones are not produced according to a master specification. Although some sources like cellular telephones, motorcycle engines and Segways have a human composing or sculpturing their sound and are meant to sound good on their own, these devices do not consider the frequencies of the surrounding sound and can only create harmony with other sounds by accident. Finally, many sound sources generate constantly-varying pitches without adhering to any particular scale.

Figure 2. In "Sonic Authority," 2005 (Noah Vawter), official-looking tags identified tonal components in noisy spaces.

This concept was explored in a project I developed in 2005 titled "Sonic Authority." In that piece, I recorded noisy scenes, analyzed them to determine the dominant pitches amongst the noise, and posted official-looking metal tags to identify them. See Figure 2. The suggestion was that all noise sources, including car engines, cellular telephone ringtones, and bus brakes could be tuned to harmonious groups of notes.

Ambient Addition could be used to create the sensation of harmony at a personal level for a listener in a noisy environment. Since this requires invention, addition of information, or at least selective removal of information, there is no single, "correct" technique for doing it. All methods are based on a mixture of subjective thought and application of technology. I chose five techniques for synthesizing harmony. These five should not be considered exhaustive. Creative composers may invent many more.

<u>Techniques for Creating Harmony from Environmental Noise</u>

1. Resonance/Vocoding
2. Analysis and Attenuation
3. Analysis and Pitch-Scaling
4. Replacing Inharmonious Sounds
5. Selectively Modifying Overtones' Amplitudes

The following sections describe the above five techniques, with particular emphasis on taming the music pitches contained within environmental noise:

Harmony - Resonance/Phase Vocoding

A radio story inspired the first technique for creating harmony out of street noise. It was a very brief segment about how an artist whose name I never learned roamed New York City with a cardboard tube. He had determined that the resonant frequency of sound in the tube was about 233 Hz, a B-flat. His performance was simple: he walked up to people and invited them to listen to the sound of the city through the tube. The radio program was able to replay the sounds people heard. Laughter, cars and everything else could be recognized, but they had been "tonalized" as if everything was reflected through only one note. It did not seem to matter what sound went into the tube, it all came out with a similar "flavor."

I quickly wanted to hear the audio scenes reflected through various other notes. I imagined what it would be like to have an array of tubes with resonant tubes forming a chromatic scale. Clearly, creating a melody with them might not work well, because one would have to switch the tubes too quickly. Additionally, I noticed it took a few seconds for my ear to register the tone from the recording of the tube because of the irregular rhythm of the events happening in the street. On the other hand, I imagined the technique would work very well for creating a progression of harmonies. For instance, if three tubes at once were employed with resonance frequencies at E, G and B, an E minor chord could be effected. After 10 seconds or so, those three could be swapped out for D, F# and A tubes to create a dramatic II-I transition.

Simulating the resonance effects of an acoustic tube can be done in the digital domain with a device like Ambient Addition. With sufficient processing power, any number of acoustic tubes at once may be simulated, creating rich, complex harmony*. If realistic "acoustic tube sound," is important to a

* This is essentially how a pipe organ works.

composer, techniques are available to simulate them, such as Aren Jansen's Master's thesis "The Manifold Nature of Vowel Sounds," (4, 2005). In Ambient Addition, though, a different technique is used, because it is less complex and allows more control over the sound's timbre: Instead of simulating an acoustic tube, the outside noise stimulates a linear model constructed by inverting a frequency domain representation of the desired sound. Put simply, a sequencer supplies a list of "notes to be synthesized" to a function which produces a list of harmonics and amplitudes for each of those notes. When the outside noise and the desired sound are available in frequency domain representation, this amounts to simply multiplying the two together and inverting. I came up this technique myself, but eventually discovered it is almost exactly how phase vocoding works. The details of the implementation can be found in the "Algorithms" section of the Development chapter of this thesis.

Harmony - Analysis and Attenuation

After developing and implementing the Resonance/Vocoding algorithm, I began testing it and discovered it worked to my satisfaction except in the case of police sirens. It emphasizes the components of the siren's wailing which are supposed to be in the chord, but does not attenuate the off-spectrum parts sufficiently. A different technique can address the problem. The strategy is to continuously analyze the input signal in a manner similar to Sonic Authority. When the dominant pitch of the outside world is determined, it gets checked against the notes specified in the sequencer's current chord. If the dominant pitch matches one of the notes in the chord, the sound is allowed to enter the listener's headphones at full volume. If not, the outer sound can be attenuated proportionately to the level of mismatch producing a version of the sonic landscape which emphasizes the notes in the composed chord progression.

Of course, the intermittent amplitude of this technique may result in an undesirable, choppy effect. One way to work around this is to repeat the playback of the last audio segment which properly matched the chord progression until the outside world returns to compliance or the sequencer reaches a chord which matches the outside world. To maintain the impression of a rhythm, it would be best if the boundaries of the repeated segment were aligned to the playback sequencer.

Harmony - Analysis with Pitch-Scaling

Another technique to fill the inharmonious gaps between moments of natural harmonicity is to use pitch-scaling. First, the synthesizer should determine the shortest distance in semitones from the outside world's dominant pitch to a valid note in the current chord progression. Then, it should scale the pitch of the outside world until it is where it is supposed to be. For example, if the desired chord is F# minor (F# A C#) and the dominant pitch of the outside world is reported as E, the pitch could be scaled up two semitones to make everything match the desired F#. Also, multiple copies of the outside world can be scaled at once to create the sensation of a chord. There are a great number of pitch scaling algorithms which vary in quality and computational resources. A survey of them is outside the scope of this thesis*, but for the small intervals this technique is intended to fill, either resampling the frequency domain data of the environmental noise, or resampling the audio with a double-buffer to prevent clicks will suffice.

* See Stephan Bernsee's overview at http://www.dspdimension.com/data/html/timepitch.html

Harmony - Replacing Inharmonious Sounds

Since one goal of Ambient Addition is to create versions of the sonic landscape which are transformed as little as possible, it may be too heavy-handed to eliminate unwanted sounds completely. For example, if a driver unloads a metal handtruck from the back of a truck, he or she may cause a clanging sound containing unwanted frequencies. Such sounds, including cars honking, etc., are loud with good reason. Removing the sound altogether, like the attenuation technique, may be too significant an intrusion of reality, as well as fatal, yet such sounds can still be transformed to fit with the internal music of the sequencer inside Ambient Addition.

The previously mentioned pitch-scaling technique might not suffice for this if there are unwanted partials in the original sound. Therefore an alternate technique is suggested: use filtering to remove the offensive sound, then play a sample of something else that has the preferred pitch and similar acoustic characteristics as the original sound. For example, if the chord sequencer is attempting to make a G Major chord (using notes G, B and D), but the handtruck sounds like a C, the handtruck does not fit in well. However, it would only be a slight shift in perception if the listener were to hear a sample (or synthesized version) of a metal xylophone at note D with the same volume level and decay rate as the offensive sound.

Harmony - Selectively Modifying Overtones' Amplitudes

The previous technique of replacing inharmonious sounds may not be subtle enough in some cases. To interfere even less with the sound of the external world, while still removing inharmonious sounds, the following technique, depicted in Figure 3, is recommended. The sequencer should analyze the dominant pitch of the external world, creating a list of the top three dominant pitches. It should then compare the harmonics of the desired tone against those pitches, generating a list of unallowable overtones.

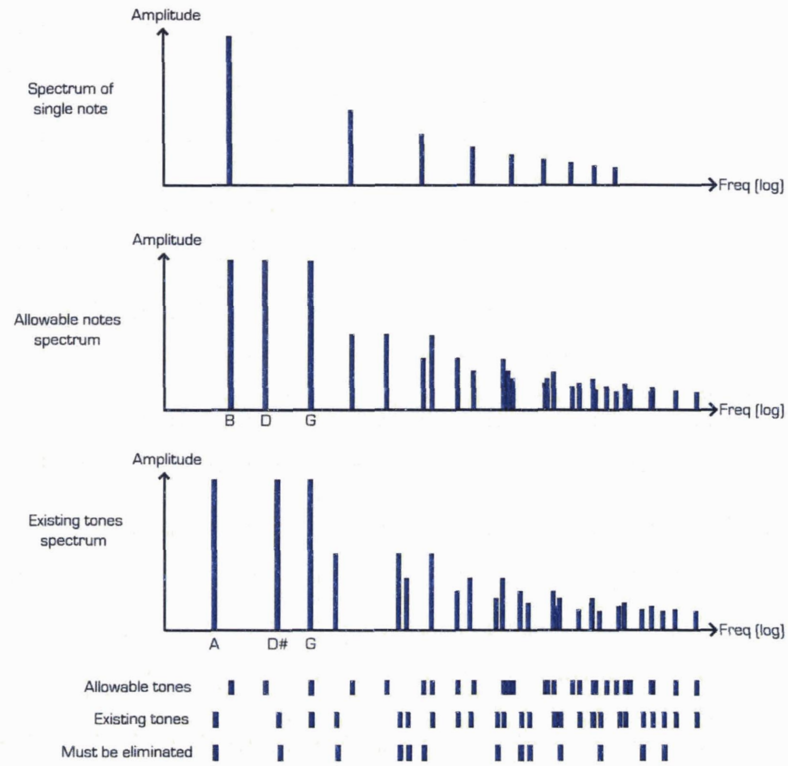

Figure 3. Choosing which overtones to mask out.

At the top of figure 3, the spectrum of a single, harmonic note is shown for reference. Below that, the figure shows the spectrum of a three-note chord the sequencer would *like* to play and at the bottom, the spectrum of the dominant notes in the environment. In this technique, partials which are in the environment, but not allowed by the sequencer, get eliminated through FFT filtering. Observe that this technique differs from the Resonance/Vocoding technique previously described in that it does not remove all out-of-chord tones. Instead, it only removes partials if they are both out of the current chord and detected as being part of a dominant pitch in the environment.

Melody

This section will address the problem of synthesizing melody from environmental noise. As fleeting to catch as a naturally occurring rhythm, melodies rarely, if ever, arise from the cacophony of a busy area, unless one has been intentionally created by a radio or instrument player. Occasionally, flecks of tones from car horns and laughter suggest one, but with no master composer, these fragments can never be expected to exhibit characteristics of a melody such as phrasing, resolution or "catchiness." Ambient Addition can intervene in several ways to create melodies in this context. Melody synthesis techniques may be organized into two general classes: In the first, pre-composed melodies are played using sonic characteristics of the surrounding world. In the second, pitches from the outside world are used as inspiration to generate a melody.

Melody - Pre-composed Melodies, External Sonic Characteristics

In this section, three audio synthesis methods suited for converting ambient sound into melodic music are discussed. They are granular synthesis, pitch-scaling, and overtone synthesis.

Granular Synthesis

Employing granular synthesis is one way to synthesize sound that reflects an environmental soundscape. This technique was used successfully by Barry Truax, the pioneer of granular synthesis, in his works *Pacific* and *Dominion* as described in his paper "Composing with Time-Shifted Environmental Sound" (5, 1992). For compositional reasons Truax changed the lengths of his samples instead of their pitches. Nevertheless, it is possible to change the pitches when using granular synthesis in order to play a pre-composed melody with a timbre derived from environmental sound. Although this technique can consume lots of processing power when many simultaneous grains are used, granular synthesis has the advantage of being relatively simple to implement.

Pitch Scaling

A second method is to record a sample in real-time which has a clearly identifiable dominant pitch, then scale it as needed to reflect the pre-composed melody. This can be done by analyzing all incoming transients and selecting one which is harmonically pure. The harmonicity check is recommended because it can sound especially harsh when inharmonic tones are scaled parallel to harmonic ones. A suggested way to compute the harmonicity from the dominant pitch data is to compare the ratio of the average probability of each dominant pitch to the maximum probability. Alternatively, the sensory dissonance level (6, 1969) may be computed. If the sample is too inharmonious due to background noise or inharmonic tones, it may be comb filtered to purify the harmonic structure. One disadvantage of the pitch-scaling method is that the sample's length must be at least as long as the durations of notes in the melody. If it is not, the sample playback will terminate prematurely. Another disadvantage is that the formants of the sound scale with the pitch, creating unnatural sounds for larger intervals. Two advantages of this technique are its simplicity of implementation and low processing requirements.

Overtone Synthesis

One last scheme to synthesize melody from arbitrary sound is to synthesize overtones of the desired tone. This method is complex and its description is somewhat longer than the other four techniques. It is based on throat-singing. A thorough description of throat singing explaining its ethnography and mathematics can be found in the September 20th, 1999 issue of *Scientific American.* (7, 1999)

In the previous section, under "Harmony," it was indicated that listening to street sounds through a simulation of a series of rapidly switching acoustic tubes would be impractical for creating the impression of a melody due to the random stimulation of the tubes. I now propose that this technique be used with some changes. Before explaining those changes, I will explain the technique briefly.

Sounds are usually described using their fundamental pitches, even though overtones, or sinusoidal components whose frequencies are integer multiples of the fundamental, may be mixed into them to change their timbre. If the timbre is carefully manipulated in time, it can reflect a melody. If, for example, the dominant pitch of an air conditioner from the outside world is calculated to be D-1 (meaning a D in the first octave), and the internal sequencer calls for an A-3 to be played, Ambient Addition can synthesize the 6th harmonic of the outside world to create the sensation of an A-3, while interfering as little as possible with the stream of incoming sound.

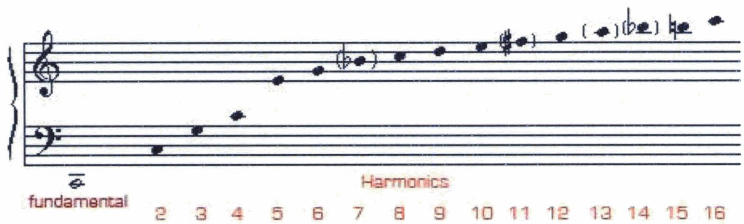

Figure 4. The overtone series illustrates how a musical note contains sinusoidal components of many frequencies. From (8, 2001))

Given that fundamental technique, I will now provide some additional details on the overtone synthesis method because many texts do not give enough information to implement it. Many texts depict a four octave staff such as figure 4 to explain the overtone series of instrument sounds. It is clear from such examples that playing only overtones of a fundamental note, one can create a melody based on a subset of the chromatic scale. Unfortunately though, the subset depicted does not cover the chromatic scale. In figure 4, the only valid notes in the highest octave are C, D, E, G. In order to create a full chromatic scale of at least one octave, the portion of the overtone series from the 16th through the 32nd overtones must be used. The following table shows which overtones are suggested to create a chromatic scale:

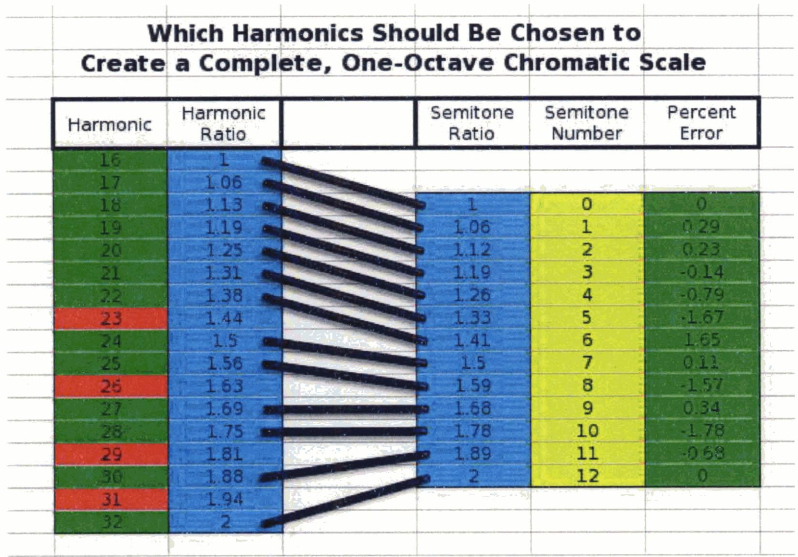

Figure 5. Harmonics 16 to 32 must be synthesized to create a complete chromatic scale.

This technique allows a sound of any fundamental to have an in-tune melody synthesized above it. It has several shortcomings, though. First, because the scale begins at the 16th harmonic, the synthesized tones are four octaves higher in pitch than the original sound. This means if the technique is used on sounds that are too high in pitch (above 750 Hz), the overtones will be inaudible (above 12 KHz). One may transpose the synthesized sounds down by one octave without ruining the harmonicity altogether. The second shortcoming is that it depends on steady stimulation of the fundamental overtone. If the fundamental used as a reference is varying in amplitude, such as a human voice, it may be necessary to artificially increase the amplitudes of the overtone frequencies. A third disadvantage is that the tone produced is merely sinusoidal. Such tones have less timbre than harmonically rich sounds, and are therefore limited in affect. As a workaround, overtones of the fundamental may be superimposed insofar as they do not ruin the harmonicity of the melody's fundamental. In the previous example, if the synthesizer were to generate an A-4 in addition to the A-3, the sound would be a little fuller without causing dissonance with the fundamental D-1. Generating an A-4 and an E-4, however, might push the dissonance too far.

Melody - External Melodies, Internal Sounds

With "internal melody, external sound" covered, this section will explore the opposite: creating a melody from fragments of found sound. For example, the pitches and timing of chirping birds can be tracked and used as the input to a process which generates a melody. Melody generation is an enduring area of research in computer music to which there have been many approaches. It is outside the scope of this thesis to investigate them all, but notable researchers in the field are David Temperley (see <u>The Cognition of Basic</u>

Musical Structures) (9, 2001) and Dirk-Jan Povel, who has embodied much of his research into his software "Melody Generator." (10, 2001)

Design

This chapter describes the design of Ambient Addition. It is divided into two sections, an audio section and one for hardware. The audio section describes what algorithms were chosen to create harmony and rhythm. It references the previous chapter Technical Possibilities. The hardware section describes the decisions made when designing the overall physical form, headphone/microphone combination, and the selection of the audio processing hardware. Details of construction are found in the next chapter "Development."

Design - Audio - How to Listen with Ambient Addition

Ambient Addition is intended to be listened to on a 30-minute walk. To maintain the listener's interest for that amount of time, it changes its style of processing regularly. Each style of processing is called a frame. Like most music, some frames repeat. With Ambient Addition, repeated frames have a special meaning because the source material - the outside world - changes, creating a unique part of a song even with the same processing.

Furthermore, when the listener recognizes a familiar frame, it influences his or her behavior. He or she is enticed to gravitate toward a new sonic source, or to create a sound, in order to influence what is coming in through the headphones. This type of anticipatory behavior is described in numerous books on musical theory, such as Robert Jourdain's book Music, the Brain and Ecstasy. (1, 1997) and John Ortiz's The Tao of Music. (2, 1997). This behavior is also *participatory*, in that it encourages the listener to take part in the music. Since this movement may divert the listener into exploring terrain in an atypical way, Ambient Addition is related to psychogeography. Psychogeography is defined as "the study of the exact laws and specific effects of geographical environments, whether consciously organized or not, on the emotions and behavior of individuals," (3, 1958) A deeper description of psychogeography is outside the scope of this thesis, but those interested can look into the the films of Stewart Home, or the Situationist International.

Design - Audio - Effects

Of the processing techniques described in the Technical Possibilities chapter, not all could be implemented in the timeframe of this project. Those selected are the ones which create harmony and rhythm. The next paragraphs explain which algorithms are implemented and why. Their mathematical details can be found in the Development chapter.

Analysis

On the analysis side, it is important that Ambient Addition can identify dominant pitches in the environment. This was determined based on the author's experience with the Sonic Authority project described in the Technical Possibilities chapter. Detecting external pitches enables Ambient Addition to synthesize music in harmony with the outside world. The method chosen is based on Wei Chai's integration of overtones described in her paper Automated Analysis of Musical Structure (4, 2005).

It is also important to be able to generate rhythms from transient sounds like shouts, bangs, and slams. To detect such sounds, an onset detector was implemented. Ambient Addition uses one based on Tristan Jehan's Ph. D. thesis Making Music by Listening. (5, 2005) Since accuracy was not as important as in Jehan's application, a number of optimizations were developed to reduce processing power and implementation time.

Synthesis

On the synthesis side, the effects implemented are largely based on a continuously running Fast Fourier Transform/Inverse Fast Fourier Transform (FFT/IFFT) loop. An FFT/IFFT loop means that all sound coming in through the microphones gets converted into the frequency domain, operated on, then converted back into the time domain. There are numerous reasons for basing synthesis on this method. First, it allows Ambient Addition to perform effects which preserve the original sonic landscape as much as possible. In theory, the FFT/IFFT loop is lossless and can recreate the input signal perfectly*. Second, since Ambient Addition is already performing an FFT continuously in order to compute the dominant pitch of the outside world, half of the work of the loop is already done. Third, the FFT/IFFT loop encompasses the synthesis method called additive synthesis, which is one of the most flexible audio synthesis methods. As Paul Wiffen, regular contributor to Sound on Sound magazine, writes:

> To continue in this vein, additive methods of synthesis are

* In practice, the quality of reconstruction of the FFT/IFFT loop on a fixed-point system like Ambient Addition hinges on the word size used in the calculation. Experiences related to this are described in the Results chapter.

> closest to the oldest of the visual arts, painting. The sound is built up from its constituent parts, just as a painter mixes together different hues to achieve the required colour, and then lots of different colours are used to create the final picture. Additive synthesis uses combinations of harmonics to create the basic tone colours or 'timbres' and on more sophisticated systems several of these timbres can be combined to make the overall sound. (6, 1997)

Finally, although additive synthesis can be difficult to control on common "panel of knobs" synthesizers, it is a natural fit for the demands of Ambient Addition, in which the parameters of the synthesis are controlled by other processes, such as the sequencer which imposes tonality on environmental noise.

One disadvantage of using the FFT/IFFT loop for the Resonance/Vocoding effect is the delay it necessitates for processing time. Larger FFT blocks are more efficient and offer more resolution and accuracy, but they also increase delay time. Large delay times destroy the illusion of real-time transformation that Ambient Addition seeks to create. An alternative would be to use William Gardner's Zero-Delay Convolution technique (7, 1995). However, Gardner's technique requires more complex programming, and more processing power. For simplicity, this prototype uses the FFT/IFFT loop with a reasonable delay time. Details are found in the Algorithms section of the Development chapter.

In addition to the harmony synthesis algorithms, techniques for replaying segments of the external sound were implemented. These were chosen because they can be used to create rhythm.

All algorithms which have been described in this section have numerous implementation details like buffer sizes, attack constants and harmonic rolloff rates. Such implementation details can be found in the Algorithms section of the Development chapter.

Design - Physical Form

As stated in the introduction, the goal of Ambient Addition is to portably transform environmental noise into something more musical. Furthermore, it should done in a way which reduces the social isolation of typical Walkmans. The physical design of Ambient Addition works along with its hardware and software to achieve that goal. This section describes the design decisions of Ambient Addition's physical form. This includes the overall form, the audio processing unit and the headphones/microphones combination.

Design - Physical Form - Overall Form

Since the goal of the thesis is to enhance the capabilities of Walkmans/iPods, I began the design of the physical form by inheriting the constraints to which those devices adhere. For example, the device must be portable and fit in a pocket. Its controls should be easily accessible, transforming sound with minimal operator intervention. That means it must run without requiring a connection to a desktop computer, laptop computer or other console. It should run on batteries and have no exposed integrated circuits and no excessive dangling cables. The device should also be easy to put on and take off.

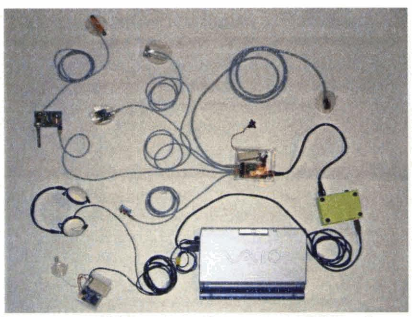

Figure 1. The complete set of hardware a user wears when experiencing Sonic City. (8, 2002)

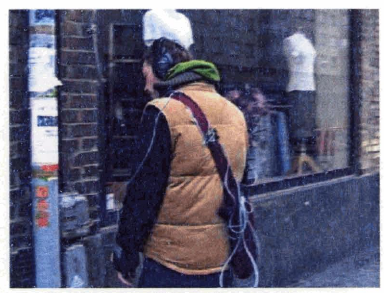

Figure 2. Sonic City: Sensors are mounted to the body with duct tape and a laptop is carried in a shoulder bag. (8, 2002)

Few previous attempts at headphone-based sonic mediation by other artists have adhered to such constraints. For example, "Sonic City," (8, 2002) a 4 year long project from the Future Applications Lab in Göteborg, Sweden, involves a large number of sensors fastened to the listener's clothing, each requiring a separate cable. It also requires a laptop computer and a shoulder bag to lug all the parts around. Admittedly, Sonic City is taking a different approach and operating under different constraints, but I want to make it clear that I do not intend to subject listeners to such a level of bondage.

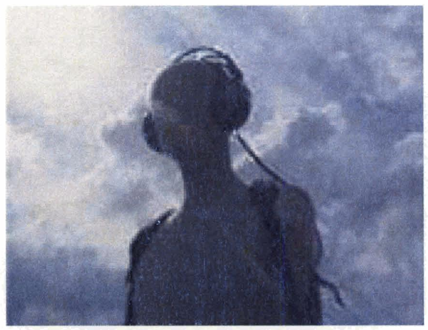

Figure 3. Maebayashi's "Sonic Interface" requires the user to wear a backpack containing a laptop. (9, 1999)

Also, Maebayashi's "Sonic Interface" includes a small backpack with a laptop computer inside (note the straps in figure 3). This makes it inconvenient for everyday use. For example, it becomes impossible to carry one's own backpack full of belongings. In prototyping my thesis hardware I also discovered the value in a having a small device which people could easily pass to one another.

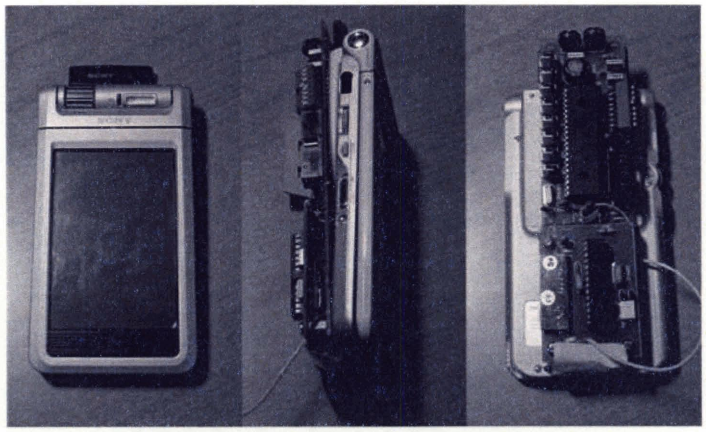

Figure 4. Atau Tanaka's "Mobile Music Making" (10, 2004)

A form like Atau Tanaka's "Mobile Music Making" (see figure 4) (10, 2004) is an improvement, but also presents difficulty in the form of a large amount of exposed hardware crudely mounted to a PDA. It is clearly not robust and appears to be too thick to keep in a pocket. While such bulky packages are acceptable in the context of demonstrating concepts, they also suggest the incompleteness or infeasibility of an idea, giving the impression that widespread acceptance is farther away than it may be.

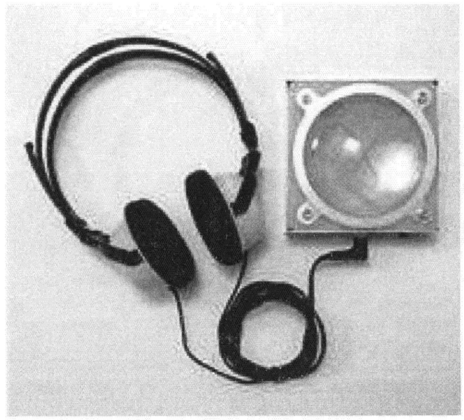
Figure 5. Toshio Iwai's "Sound Lens" is elegant. (11, 2001)

In my work, I would like to not only advance in the direction of musical and conceptual complexity, but also present a more complete design which may catch on more easily. The design at which I will arrive should be so refined that it may even be incorporated into an existing MP3 player with minimal changes. A positive example I would like to emulate is Toshio Iwai's "Sound Lens" (see figure 5) (11, 2001). It has only two parts: a light-to-sound transducer and a pair of headphones.

Returning to the original constraints, the "No laptop rule" is the one most tempting to violate. Laptop computers offer a number of convenient features needed to portably mediate audio in real time such as headphone jacks, microphone jacks, adequate processing power and plastic enclosures.

Additionally, they are readily available and to a large extent their operating systems and software applications are standardized. On the other hand, the size of the screen and keyboard interfaces limits their usefulness in this specific application. The next section discusses how to get the germane features of a laptop into a Walkman-sized package.

Audio Processing Hardware

Audio Processing Hardware is the general name given to the part of the device which will sample sound from a microphone, transform it, and play the transformed sound back through headphones to the user. Even considering the constraints in the previous section, there are many ways to do this. In a typical engineering application, the hardware selection decision is made based on specifications like processing power, cost, memory, peripherals, etc. Hardware is then sought which matches those specifications. In this case however, the final algorithms, and thus the processing power, were unknown. As explained in the previous chapter, a minimum number of CPU cycles were necessary for operations like continuous Fast Fourier Transform (FFT) and

Inverse FFT (IFFT), but beyond that, the device acts as a sandbox for algorithmic development, so there should be a good deal of processing power to expand into. Therefore, a variety of types of hardware were considered.

This section evaluates a number of hardware platforms, representing several strategies. I sincerely wanted to use an existing piece of hardware from the consumer market in order to make this project accessible quickly to a large number of people, but could not find anything suitable. In the end a DSP development board based on a the Analog Devices Blackfin processor was selected. The results of an exhaustive survey of consumer hardware are provided here for the benefit of future experimenters.

Leveraging Existing Consumer Hardware - iPod

The first strategy is to leverage an existing piece of sound hardware. For example, a common MP3 player like the iPod could be reprogrammed. This technique has several intriguing advantages. First, because the hardware is mass-manufactured, the price is often low and the availability is high. Second, once the project is developed, it can be published and used by many people - Apple's website claims over 15 million iPods have been sold as of March 31st, 2005 (12, 2005). Not every consumer product can easily run arbitrary code, but there is, in fact, an established "homebrew" movement for the iPod, enabling people to install the Linux operating system on their iPods (including dual booting capability, enabling people to run their original iPod software alongside user code). It is also possible to write applications that involve the audio hardware, but certain details preclude this as a feasible choice. In particular, it is impossible to use the microphone input and the headphone output simultaneously on the iPod.

Still, it might be possible to design external hardware to compensate for this shortcoming. A small circuit board with microphone inputs, a pre-amplifier, an analog to digital converter and a digital interface could deliver the input data to the processor, allowing the headphone jack to be used for audio output. The iPod has three digital interfaces which might be used for this purpose: a bidirectional serial port, a USB port and an IEEE1394 (usually referred to as Firewire) port. In the next paragraphs, the possibilities of each of these ports are analyzed.

First, it should be mentioned that a lack of documentation hinders development efforts. Since iPod user development is not Apple-supported, the only information available for programming these interfaces comes from reverse-engineering the iPod's firmware. This means interfaces may sometimes appear to be more limited than they are. For example, it appears that the speed limit of the iPod serial port is 115,000 baud. It may be capable of higher speeds, but it will have to be evaluated at that baud rate. Since the data require framing, e.g. one stop bit and one start bit, the throughput of audio data could only reach 8/10ths of that rate, or 92,000 bits per second (bps). Some UARTS have alternate framing options and longer word sizes, but it is not known whether the iPod serial port possesses those capabilities. Even if it did have ideal framing, the throughput could only approach 115 kbps, which is not fast enough to transmit high quality audio. High quality requires 705,600 bits per second per channel. (44.1KHz * 16 bits) It is worthwhile to calculate what quality level of audio could be transmitted with 92,000 bps. Quality level can be subjective since there is a tradeoff between word size (and therefore dynamic range) and sampling rate, but it is still useful to have a ballpark figure. E.g. at 92Kbps, a typical quality level would be 8 bits at 11.5KHz. This is not totally unacceptable as an input throughput, but is less desirable. It would have to be considered as a last resort.

One possible work around to increase the input quality would be to employ compression hardware in the input system. That would involve a circuit board outside the iPod that includes a microphone input, pre-amplifier, audio compression hardware such as the Micronas MAS3587F MP3 encoder chip, or the Analog Devices Melody chipset, possibly a microcontroller or FPGA to format the data and a UART to send it into the iPod. The iPod would then use its CPUs to decompress the input stream to the full data rate. This could work, because 92Kbps, the expected transfer rate of the serial port, is acceptable (at a quality level roughly equivalent to FM radio broadcasting) for transmitting compressed audio data. Despite the technical feasibility, such a solution tramples out the "everyone can use it" advantage of using consumer hardware. It is also quite complex and uses nearly all of the iPod CPU cycles to decompress. Such a solution will be considered, but it would be more preferable to use a digital interface that has a higher bandwidth, such as the iPod USB port.

Even at its slowest rate, USB can transfer data at 1.5Mbps. This is just about enough for stereo, 16-bit, 44.1KHz CD quality data. The major disadvantage is that USB is much more intricate than UART buses. This is probably why there was almost no information on the USB interface in the unofficial iPod development kit at the time of this project. To employ this bus would mean a major reverse-engineering effort. Although it would be very valuable in the long run for other projects, it is not within the scope of this thesis.

The iPod's IEEE 1394 interface remains a final possibility. It is more than fast enough for audio data (100Mbps) and simple enough that it has already been reverse-engineered to the point where the networking protocol TCP/IP has been implemented over it. IEEE 1394 is the best candidate so far, with only a few disadvantages. First, it would make the final product only usable to people with the special hardware. Second, due to the high data rate of the bus, the circuit board design would be challenging. Nevertheless, the task of designing the board relies on well known engineering theory, and is therefore much less risky than reverse-engineering the USB bus.

Perhaps the most important factor in considering the iPod is that its CPUs are not very fast. They are only fast enough to perform their main task of MP3 and AAC audio decompression. For example, the 3G models have a pair of 32-bit microprocessors running at 66 MHz each, for a rough speed of 133 million fixed-point operations per second (although they can be overclocked in software to 75 MHz). So far, this discussion has not revolved around computation speed, but as it continues, it will become clear that this is near the low end of the spectrum. The combination of tricky external hardware and low processing power makes the iPod platform less than optimal.

Since the hardware evaluation phase of this project completed, newer, video-playing iPods have been "opened," meaning user development for them can begin. Since they have new capabilities, there may be more processing power available to users. It is also likely, however, that their video decompression capabilities are implemented in Application-Specific Integrated Circuits (ASICs) which may be completely inaccessible to users without documentation or unsuited for general audio processing.

Leveraging Existing Consumer Hardware - Portable Video Games

Besides the iPod, few other consumer MP3 players have microphone capability, significant processing power and user development capability behind them. Another class of products, however, is interesting: portable video game playing devices. Two of the most popular video game devices released in the last year are the the Sony Playstation Portable (PSP) and the Nintendo DS, selling about 4 million units each in the US, as of March 2006. (13, 2006). The two devices cost about $200 each, are both intended for portable use and fit into pockets quite easily. They also have large user development efforts behind them. Their video capabilities are irrelevant to this project, but other specifications are relevant.

The biggest difference between them is in their processor speeds. The DS languishes below iPod speeds with two processors, one running at 67 MHz, the other at 33 MHz. The PSP on the other hand has a pair of processors, one running at 333 MHz (when overclocked from software) and one running at 200 MHz. Another major difference is that the DS includes a built-in microphone, while the PSP does not. The maximum sampling rate of the DS microphone is unknown, however its bit depth can have a maximum of 12 bits. (13, 2005) Upon further consideration, however, the DS's microphone is not useful because it is built in to the unit. Therefore, when carried in a pocket like a Walkman, the microphone is inaccessible.

Another possibility is that the PSP has a microphone input on the same jack as the headphone output. A commonly available headset is available for use with it. The exact specifications are unknown, but there is at least one piece software available (the game SOCOM) that makes use of it and which could be reverse-engineered. The difficulty of this reverse-engineering effort is considerably less than that of USB hardware. Furthermore, it would still function while the PSP sits in a pocket. Given its high processor speed, microphone inputs and development community, the PSP is an attractive platform. The main disadvantage is that the code to use the microphone must be reverse-engineered, which may be time consuming or otherwise prohibitive. The second disadvantage is that the microphone appears to be only a single channel one. Ultimately, I would like the project to be able to handle stereo input. For that to take place, the built-in microphone is inadequate. On the other hand, there are other digital interfaces to the PSP.

The first possibility is its USB port, which, quite similar to the iPod, is a complex device which has not been reverse-engineered and would be too time-consuming or otherwise prohibitive to involve in a project with a fixed deadline. The second possibility is the PSP's wireless ethernet interface. The bandwidth of its 802.11b is 2Mbps, so in theory a microphone, or pair of microphones, could transmit their data wirelessly to the PSP for processing. This would require external hardware, but should be considered. Returning to the DS for a moment, it, too, has a wireless interface, making it roughly equivalent to the PSP, but with significantly less processing power.

DSP Development Hardware

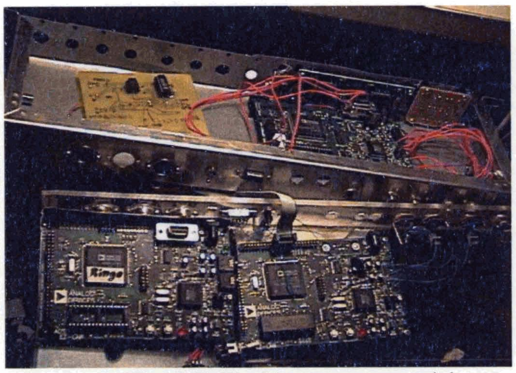

Figure 6. The EZ-Kit Lite was an exciting electronic music development tool. (15, 2001)

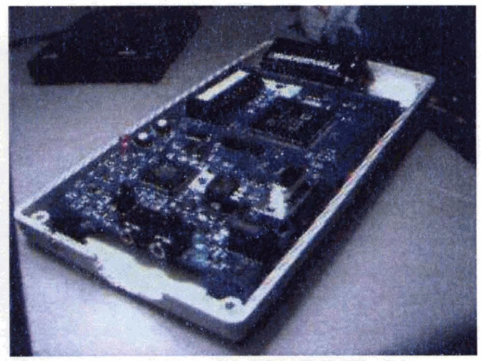

Figure 7. At 3" x 5", The EZ-Kit Lite was almost portable. (15, 2001)

In addition to considering consumer electronic devices for use in the project, Digital Signal Processing (DSP) development boards should also be considered. These are pieces of hardware freely available as commodity items, but not familiar to the public. They are mainly intended as evaluation boards for engineers to develop products on. For example, software for a processor can be developed on a "dev board" before final hardware is ready. In some cases though, the development boards suffice for a final application. From about 1998 to 2002, for example, I had great success developing electronic music projects on a dev board called the "EZ-Kit Lite" for the Analog Devices ADSP-2181 chip. Like most dev boards today, the EZ-Kit was small, had a DSP chip, stereo input and output and not much else (see figures 6 and 7).

Recalling this success, I made an investigation into the dev boards currently available. Dev boards produced by the manufacturers today are intended to demonstrate as many different possibilities of their DSP chip as possible, so they are physically large, with many peripherals and connectors on them. Looking further, I discovered another market of dev boards. Companies like DSP Research, Danville Signal Processing, etc, produce boards with DSP chips and application-specific peripherals on them. They are intended for sale to third-party application developers. Intended applications generally include video surveillance, video conferencing, and military signaling such as radar.

In this class of hardware, I discovered two promising boards, the Video Server HX (intended for video compression) from Mango DSP and the VicCORE-OEM53x from voiceINTERconnect (intended for voice control products). They were both small enough to fit in a pocket, had powerful processors (near 1 billion operations per second), and audio input and output. The HX board used a DSP from Texas Instruments (TI), while the VicCORE used one from Analog Devices (ADI). They were similar in price ($600).

Based on my prior experience with Analog Devices hardware, I chose the board from voiceINTERconnect board.

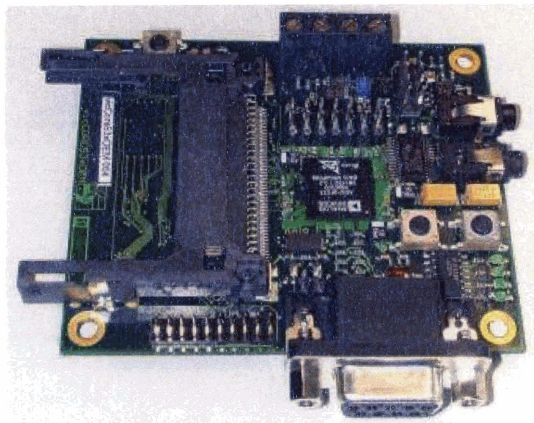

Figure 8. The VicCORE-OEM53x board is suitable for portable DSP experiments.

After an exhaustive search, these boards were the most suitable ones I could find, yet they still had some shortcomings. For example, the HX Board did not have microphone pre-amps. Similarly, the vicCORE only had one microphone pre-amp. Finally, it turned out that floating-point processors would have been a better choice than fixed-point, due to noise in the FFT/IFFT Loop. The details of this point are discussed in the algorithms section of the Development chapter.

Headphones/Microphones Combination

As the introduction discussed, Walkmans tend to create social isolation. Part of this effect is due to sonic dissociation, which the microphones of Ambient Addition address. The visual presentation of Walkmans is partially responsible, too, as evidenced by this anecdote:

> One of my employees let me know that he's going to take a week off to go back to Texas. I asked why and he stated 2 reasons - to see family and to meet some women. Huh? I asked what was wrong with the women in the Bay Area and to my (somewhat) surprise he said, "It's hard to meet them here because of the fuckin iPods. The iPods aren't as popular in Texas."(16, 2006)

Therefore, extra consideration for the outward appearance of most visible part of the system, the headphones, is warranted. It is necessary to create the impression that the person listening to Ambient Addition is not isolated in their own world. One strategy taken from Florian 'Floyd' Mueller and Matthew Karau's project "Transparent Headphones" (17, 2002) would be to make the microphones as visible as possible (see Figure 9).

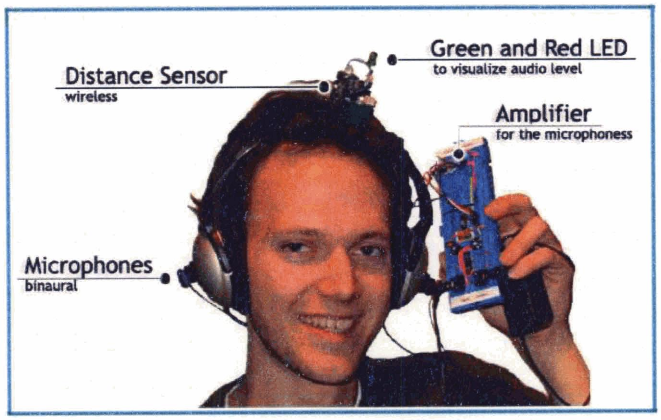

Figure 9. In the Transparent Headphones project, clearly visible microphones suggest the listener is in touch with what is around him.

I felt the Smart Headphones design worked but I wanted to try something different. Therefore, when my officemate Adam Whiton suggested using visually transparent* headphones during a brainstorming session, I chose that design. The listener's ears are visible and thus neither “stuffed” (as in earbuds) or obscured (as with over-the-ear headphones) and are accessible.

At the time this work was taking place, no headphones with transparent earcups could be located on the market. The most similar model was the Koss CL-20 (see Figure 10).

* The word 'transparent' is terminology in acoustic descriptions of headphones. It is a subjective term that indicates lack of audible distortion. This thesis only refers to the visual meaning of 'transparent.'

Figure 10. Koss CL-20 headphones are not transparent, but translucent.

The Koss CL-20 headphones actually had frosted plastic, making the visual effect more translucent than transparent. Literally, that metaphor is appropriate, since the project's sound is not an exact image of the surrounding world, but a slightly modified one. However, from the outside, the CL-20s simply look like a "cool pair of headphones," with nothing else unusual about them. Therefore, I decided to modify an existing pair of headphones to create my own transparent ones.

More than a year before the construction on my thesis began, I discovered a pair of distinctive headphones in the "give away" cabinet in the basement of the Media Lab. Large and comfortable, they seemed be about 20 years old and reflect the culture of audiophiles. They were made with soft leather on the headband and earpads to cushion the wearer's head and ears and certainly deserved to be recycled. I decided to replace their light brown earcups with transparent ones. It would have been possible to create a custom shape using computer design software and a 3-D printer. I pondered the specific shape of the earcups for several days, hoping to discover a shape which would convey the meaning of the project as well as the transparent headphones, but finally decided to simply clone the existing ones out of transparent plastic using a vacuum former. The details of that process are found in the Development chapter. See Figure 11 for a picture of the completed headphones.

Figure 11. The listener's ears are visible through the headphones, suggesting that he is not in his own world, but able to hear and respond to those around him.

Development

The following chapter describes the development of Ambient Addition. It is divided into two sections. The first section describes the physical construction of the hardware. The second describes the development of the software and digital signal processing (DSP) algorithms used in the project.

Hardware Development

This section describes the development of the hardware. It includes two essential parts: a pair of headphones with a pair of microphones built into them and a plastic case protecting a circuit board containing the electronics to process the audio. Additionally, it describes the power management.

Headphone Earcups

As discussed in the design section of this thesis, the headphones are a custom design, constructed at the Media Lab. This subsection describes that construction process. Since this is my first industrial design project, I include the details of construction here, but casual readers may safely skip this section.

I used the vacuum former to make the earcups because it is capable of making 3-D forms out of transparent plastic. To do this, it drapes a sheet of half-melted plastic over an object, then uses a vacuum to make the melted plastic conform to the object's surface. It can not make any completely arbitrary shape, but the cup shape needed for this project is within reach. To prepare my originals for vacuum forming, I separated the earcups from the headband, cut off all of the wires, and removed the earpads, speakers and volume control knobs from the earcups. I was left with two cup-shaped pieces of plastic. (see figure 1).

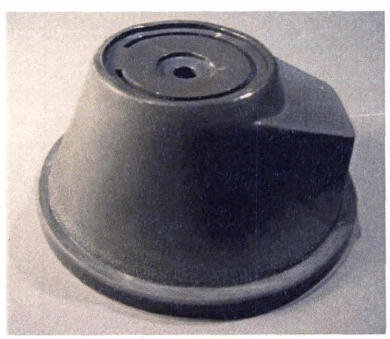

Figure 1. The original earcup used as a mold to form the earcups used in Ambient Addition.

When replicating an object for the first time with the vacuum former, the initial attempt can be somewhat experimental. The big questions are 1. Will the heat deform my original object, creating a poor form? and 2. Will the edges turn out well? On the first pass, the heat did not affect my original. However, I made about five copies in total, each with slightly different temperatures. By the end of the last forming, the original earcups had begun to show some signs of wear. Their texture became slightly smoothed out, and their flat parts had become somewhat concave. Nevertheless, their essential shape remained intact, allowing a good copy to be made. As for the edges, I was limited by the material I had available for vacuum forming. It did not respond as well as I would have liked. I found small bubbles were produced inside the plastic when it reached the forming temperature. Although the bubbles might be useful in another project (my advisor suggested illuminating them with LEDs), for my needs they were undesirable. Therefore, I reformed at lower temperatures which made the plastic somewhat less bubbly and less pliable. This resulted in rounder edges than the original, but was deemed workable.

The sight of sloppy edges and bubbles over the weeks following my first attempt eventually drove me to acquire some of the proper plastic material for the vacuum former and re-form. The second batch turned out much better, although I had to spend about an hour sanding the original earcups with four grades of increasingly fine sandpaper: 400, 600, 800, then 1200.

Once acceptable earcup forms were created on the vacuum former, more work was done to turn them into proper headphones. First, the new earcups had to be cut from the flat sheets of plastic in which they were formed. The bandsaw worked well here. One subtlety is that the replicated earcups were slightly larger than the originals. In this case, about 1/8" all around, due to the thickness of the forming material. Since the new earcups were larger than the original ones, the earpads did not fit over them until I performed trimming using a bandsaw. It would have been possible to trace the shape of the original earcups and cut to that specification, but it was simple enough to eyeball it. Also, the earpads always cover the cut edge, so an imperfect cut is acceptable. I made several passes with the bandsaw until the diameter of the earcups was small enough to fit the original earpads.

Finally, the earcups had to be mounted onto the headband. In the original headphones, the connection between the headband and the earcups was intricate because it enabled the earcups to swivel along two axes. The problem with that swivel hardware is it would obscure sight of the wearer's ears, so I made a simplified mechanism which did not get in the way. It was still able to swivel in one axis for an acceptable fit. It consisted of a single block of transparent acrylic with two holes tapped into either end.

Headphones - Speakers

One disadvantage of using the transparent headphones was the lowered acoustic isolation from the outside world. Simple tests with ear protection headphones demonstrated that the foam inside the earcups reduces a substantial amount of high frequency sound. Using such foam inside transparent headphones is impossible because it would block the sight of the wearer's ears.

Keeping the listener's ears visible also precluded traditional speaker placement within the earcups. Some headphones point the speakers down at the listener's ears, but that was impossible because the microphones were nearby. Therefore, the speakers were placed at the bottom of the earcups, facing up. Since the sound reaches the ears of the listener through an indirect path, the acoustic quality of the sound was affected negatively. First, the distance to the ears and increased acoustic volume reduced the overall sound level. To work around this, the original 32 Ohm transducers had to be replaced with 8 Ohm speakers. Second, in this configuration, each earcup acts as a resonating body, emphasizing a particular band of frequencies. A back of the envelope calculation of the resonant frequency of the earcups suggested 1.7KHz as the resonant point. This was not a serious concern though, because inverting the filter could correct the distortion in software.

Headphones - Microphones

In the initial development of this project, inexpensive microphone capsules from Radio Shack were used to pick up sound. In the final weeks of the project, they were replaced with "BMC" low end professional binaural microphone capsules from SoundProfessionals.com. The improvement in quality and input signal level was noticeable. These microphones were mounted on the outside of the headphones, above the level of the ears.

Microphone placement was one of the final design issues. They were mounted on the earcups so they would be mechanically isolated to prevent feedback. To do this, they were placed inside 1/2" thick black foam resting on the outside of the earcups. At first, the microphones' windscreens protruded offensively, creating a distracting, silly effect. To inhibit this while still defeating wind noise, the microphone capsules were buried in the 1/2" thick foam.

Hardware Development - Main Circuit Board Case

The VicCORE-OEM53x DSP board is expensive and fragile, so it needs protection from static electricity and physical damage. A laser cutter was used to make a basic shell for it consisting of a top and bottom layer of transparent acrylic. Between the two layers a ring of acrylic prevents damage from the side. "Sexy" curves were given to the edges to make it easier to hold on to and two versions were printed. A final version has no exposed parts and a development version has holes in it, allowing a JTAG cable to be plugged in and the reset button to be pushed. The case also has indents carved into it with the laser cutter which are used to hold a 9 volt battery to power Ambient Addition. In the future, it would be preferable to run Ambient Addition on a pair of 3.6 volt lithium batteries to give a longer running time. It would also make its shape physically shorter and slightly fatter, requiring a new case.

Software Development

This section describes the software development of Ambient Addition. It is divided into system software, such as booting and flash programming, operating system software, including the music sequencer, and the DSP algorithms section.

Software Development - System

The first real challenge in developing this project was getting the Blackfin to boot. It is a complex process summarized here for the benefit of future experimenters.

When the the Blackfin DSP boots, it reads the state of the BMODE pins to determine the memory device it is going to boot from and its word width. On the VicCORE board, these pins are hardwired to boot from 8-bit flash memory. After choosing the boot device, the DSP reads a stream of data from memory address 0x20000000. On the VicCORE board, that address is mapped to the socketed flash. To generate a suitable stream of data, the instructions in the help file of VisualDSP++ for "Create Loader File" must be followed to create a *.LDR file out of the compiled .DXE file. In particular, the options should be set for 8-bit data.

The socketed flash chip can be programmed using the IGLOO JTAG interface which is inexpensive relative to the JTAG interface available from Analog Devices. With the IGLOO JTAG interface, the socketed flash chip can be programmed via JTAG using the free software "JTAG tools." (1, 2005) JTAG Tools is intended to be run under Linux, but works under Cygwin in the Win32 environment as well. However, it is critical to install the "ioperm" package from Cygwin and turn it on to enable Cygwin to write to the parallel port. The interface plugs into the DB-25 parallel port and also requires power from a USB power port.

JTAG Tools is incapable of writing to the Samsung flash that comes with the VicCORE. It can write other flash chips, such as the common AMD29F040. There is one flaw in the software. It configures the asynchronous memory bus for 16-bit transfers. The result is one can not simply write the binary of one's data to the flash chip. Every byte must be duplicated, which can be done with a simple C program. Writing the flash chip can be regarded as very slow. It requires about 5 minutes to write 64KB, for a rate of about 200

Bytes per second. This is either a limitation of the parallel port, or the JTAG interface, not the flash memory. In the future, it would preferable to develop a serial port-based flash writer, which could run at e.g. 115,200 baud, for a theoretical write speed about of 6 seconds. This is technically feasible because the DSP executes its code from SRAM and SDRAM.

Development - Control Software

Given the long revision time of the DSP software, a means was necessary to modify variables without reprogramming. Therefore, the open-source command line code from AVRLib, the Atmega microcontroller library, was ported to the Blackfin.

The song sequencer was developed for maximum debugability. Therefore, any processing style or "frame" can be invoked from the command line. A song parser reads a list of these commands from the upper block of the socketed flash memory. The song is created by enabling various processing frames with delays in between.

Development - DSP Algorithms

This section describes the details of the implementation of the DSP algorithms.

FFT

The most critical algorithm in Ambient Addition is the continuously running Fast Fourier Transform (FFT). It is used as part of an FFT/IFFT loop as well as for dominant pitch calculation. For dominant pitch calculation, traditional offline FFT computation techniques are permissible, because gross pitch changes slowly enough that a few missed samples will not ruin the results. However, for inverse synthesis of the input sound, it must be artifact-free. This means the system must continue sampling into a second buffer while analyzing the first buffer. An additional concern when reconstructing the audio is to ensure that linear, as opposed to circular, convolution takes place. Circular convolution occurs when the samples at the end of a block interfere with those at the beginning of the block. To get around this, the sound buffers are zero-padded to twice their length before performing the FFT. This technique is called the Overlap-Add method and is described in Chapter 12 of The Scientist and Engineer's Guide to Signal Processing. (2, 1997)

One of the variables in the FFT system is the window length. As stated in the design chapter, there is a tradeoff between resolution and delay when choosing the window size. Larger FFT blocks are more efficient and offer more resolution and accuracy, but they also increase delay time. Large delay times destroy the illusion of real-time transformation that Ambient Addition seeks to create. For maximum accuracy in the Resonance/Vocoding effect, I began with a large window size - 4096 samples. As I developed the code, available high speed memory in the DSP became scarce, forcing the window size down to 1024 samples (after zero-padding, the FFT function operates on a 2048 sample window). The delay caused by this window size, including the double-buffer technique is approximately 1.5 buffer sizes. At the 32KHz sampling rate, the delay is about 1024*1.5/32000 = 48 milliseconds.

This delay is longer than what I normally expect for a realtime application* and was perceptible. However, it seems less critical because the important feature of the synthesized sound is that is related to the outside world, not that it occurs in exact synchronization with it. The frequency resolution with the 1024 sample window size is 32KHz/2/1024 = 15 Hz/bin. This means bins are off by more than 2% for any frequency lower than 15Hz * 1/(.02) = 750 Hz. This break point is just about one and a half octaves above middle C and

* My rule of thumb is usually 10-25 ms.

is well above the fundamental of human voices, a significant component of environmental sound. In the future, implementations of William Gardner's Zero-Delay Convolution should be implemented to achieve the vocoding effect with greater accuracy.

The word size of the Blackfin FFT routines was one final caveat with the FFT/IFFT loop. It is a very subtle fact which is not foregrounded in most DSP texts that the FFT creates something called *processing gain.* This means that at each stage of the FFT, the values are larger than the previous stage. This is particularly a problem with fixed-point processors like the Blackfin. Ideally, one wishes to accept 16-bit data from the A/D converters and take the FFT. However, each stage of the FFT increases the maximum possible value by 1 + sqrt(2). The greater the number of stages, the greater the processing gain. The resulting word size after N stages is (1+sqrt(2))^(N-1). This means that for a 10 stage FFT, data may grow as large as 2786 times its original size. One technique to work around this is to multiply every stage by the reciprocal of 1+sqrt(2), but that is very costly in processing cycles. The compromise which is implemented in the Blackfin development environment's FFT routine and the DSP hardware, is to simply shift the values each stage to the right by one bit. This effectively multiplies each value by 0.5, which is close to multiplying it by 0.414, the reciprocal of 1+sqrt

(2). Processing gain results in loss of dynamic range, and consequently a raised noise floor. The number of "wasted bits" is equal to the number of stages of the FFT minus one. Therefore, a 16-bit A/D converter effectively becomes a 7-bit one when processing gain for a 10 stage FFT is considered. In practice though, the loss is not quite that drastic because the scaling is configured to preserve a sinusoidal input at maximum amplitude. Especially when sampling environmental sound, the information is much more spread out across the frequency domain.

In the future, floating-point DSPs and zero delay convolution algorithms are recommended, even though floating-point DSPs are much more power-hungry.

Dominant Pitch Detection

The dominant pitch detection algorithm is copied directly from Sonic Authority which is a variation on Wei Chai's Ph. D. thesis (3, 2005). It simply works its way through a scale of 8 octaves semitone-by-semitone, accumulating the magnitudes of all FFT bins in the range of one quarter tone from the first 8 harmonics of each pitch. The bin with the highest amplitude is declared the dominant pitch.

Onset Detection

To detect discrete sounds like car honks and rubber tire squeals, an onset detector was implemented. Tristan Jehan's technique, from his Ph. D. thesis was particularly inspiring, although it contained a large sequence of processing (4, 2004). Since this application required less precision than his, I experimented by only implementing what I perceived to be the most salient steps.

The onset detector in Ambient Addition has two parts, a history buffer and an analyzer. The history buffer is a circular sample buffer of the last 65536 samples. When the analyzer detects a transient in the buffer, it copies the pertinent part of the buffer to the dedicated sample playback buffers. The analyzer works by performing FFTs every 128 samples. The loudness of each frame is then computed by summing the magnitude of each of the vectors in the frequency domain. The loudness of each frame is stored in a circular buffer which is smoothed with a one-pole low pass filter. In Jehan's original implementation, each time a local maximum is detected, a segment gets defined as the data between the two neighboring local minima. In Ambient Addition, this technique is simplified by declaring a segment as the data 128 samples before the local maximum until a variable length following it. Instead of relying on the local minima, then locating the nearest zero-crossing, Ambient Addition simply applies an exponential attack envelope to the onset of the sample. It also applies an exponential decay envelope to the end of the sample.

The create the illusion of rhythm, the onset detector periodically updates 4 dedicated sample buffers. A single bar sequencer triggers the playback of these sample buffers. The sequencer has an array of 32 scalar values from 0-4 which represent 0 for silence or 1-4 for a sample number. Ambient Addition uses an algorithm created for a previous algorithmic Walkman project, Chiclet (5, 2003) to create drum sequences. The sequencer is initially seeded with a simple rhythm that plays sample 1 every 8 steps. After playing the bar through twice, a mutation algorithm iterates the sequence. This is acceptable as an ambient rhythm generator, however it would be preferable in the future to be able to utilize pre-sequenced drum tracks and to analyze each sample to figure out if it sounds more like a snare drum, bass drum etc.

Conclusion

This section contains my thoughts after building Ambient Addition and experimenting with it. In the tradition of Albert Hoffman*, I begin this section with some observations of my own. After those, results of a user evaluation follow.

Self-Experiments

The most exciting result that came out of this project took place on a trip to New York City. When my Peter Pan bus pulled into the station after a 4.5 hour ride, I put on the headphones of an early prototype of Ambient Addition and switched it on in preparation for the sonic assault of the Times Square bus station. Upon hearing the familiar four chords of the beginning of the sequence play, I felt a goofy smile overcome my face. I listened to hear how the underlying rhythm of this new environment changed the song I had heard so many times in my home town of Cambridge. With so much commotion, the vocoding effect sounded as if it had been stimulated by white noise, but as I progressed through the station, a distinct rhythm emerged,

* Discoverer of LSD

with vocal-sounding features. With so many travelers swarming around, I could hear the rhythm, but not see its source. I followed the staccato utterances sputtering through the chords, until I found myself standing in the subway tunnels before a tall man, hollering and gesticulating.

The crowd had left a wide space around him, as he was hopping around rather energetically, but even though I could not tell if he was crazy, my usual tendency to stay away had been dissolved. I stood right under his nose, listening with a wide grin on my face. Although Ambient Addition rendered his words unintelligible, I realized he was a street preacher when he handed me a little colored card with religious propaganda printed on it. Then, my mind made an association between the harmony I was hearing and religion. This created a cognition that I was hearing church music, which was exciting, because I had never thought of Ambient Addition as church music before, despite its traditional four-part harmonic structure.

The tendency to behave less self-consciously repeated itself many more times while experimenting with Ambient Addition. For example, it enticed me sing along with it while walking down the street. This immediately raised the question, "Why sing along with this, but not another Walkman?" Without a full investigation, there can be no final answer, although it seems two things contribute to this tendency. The slowly changing chords (typically every 4- 20 seconds) allow one to get in tune with the device. Also, the interactive nature allows one to *play*, hearing the results immediately for comparison, as opposed to singing along with a tape, in which hearing one's own voice is difficult. Other spontaneous behavior observed while under the influence of the device included searching for things to bang on to make noise, bouncing a rubber ball, and blowing across bottlenecks.

Finally, I would like to mention that the sound of laughter is transformed in a curious and favorable way. I observed this while doing software development in the atrium of the Media Lab. While wearing the headphones for 10-20 minutes in between software revisions, I would simply listen to hear what the building, especially people interacting in the atrium sounded like. When an intriguing sound caught me by surprise, I would whip around to try to see what occurred, and many times, it was simply people laughing.

User Evaluation

A user evaluation was performed to get feedback on Ambient Addition. 20 strangers were asked to test the system on short walks in the city of Somerville, Massachusetts. After each user's walk, he or she was asked to fill out a survey with 23 short answer questions (see Appendix for survey). This section describes the results of that evaluation.

Social Isolation

One of the key questions Ambient Addition explores is whether headphone music produced by involving sounds around a listener can make him or her feel less socially isolated than common Walkmans. In this evaluation, very few people reported that Ambient Addition decreased their feelings of isolation. Of those who answered the question "How does Ambient Addition reflect your feelings toward social isolation?", 9 reported that it was an isolating experience, while 3 reported it was not. Another 4 users wrote that it had nothing to do with social isolation. Similarly, 6 users reported less interaction with those around them, while only 2 reported more. Another 2 users said their interaction level was about the same. These results suggest that, for most people, Ambient Addition created an isolating feeling.

On the other hand, all 17 users who responded to the question "Do you feel Ambient Addition made you do things on purpose to hear how they sounded?" reported "yes." Of the 17, 8 walked toward areas with varying levels of volume, such as near a "coffee machine," exhaust fans, traffic, a radio, behind a building, and people talking. The other 9 users spontaneously created sounds, such as vocalizations and tapped on objects around them or the headphones themselves. While these activities were not two-way, like conversation, they suggest that isolation was decreased from the user outward.

When asked "What do you think about the connection between the sound in the outside world and the sound Ambient Addition makes?", only 2 users responded that they did not see the connection, while 9 heard it clearly.

As to whether people were more likely to interact with someone wearing Ambient Addition than a typical Walkman, this was not the type of controlled experiment which could determine the answer. However, two questions attempted to gauge attitudes toward the transparent headphones. The first: "What do you think about the transparent headphones?" was too vague to get direct responses about interaction - 14 people responded with words like "nice," "cool," and "neat," etc. Still, three users gave responses which indicated they felt others were paying attention to them. One responded that the custom headphones attracted attention. Another wrote that he was nervous people were looking at his "dirty ears." A third said she "felt a little silly w/ the giant headphones.."

The next question: "What impression do you get seeing someone else with it on?" also resulted in answers indicating interest from non-listeners. Of the 10 non-blank responses, 5 people said they were curious about Ambient Addition or wondered what the listener was hearing. Two others mentioned the visibility of the ears explicitly. The remaining three made comments about the headphones' novelty.

Musical Questions

These questions addressed Ambient Addition as a musical work. The first one, "Did you think of this as music?" resulted in 15 out of 18 responders asserting that it is music. The other 3 gave answers like "a quirky experiment that could be music," and "musical, but not music. Musical sound?"

One of the questions which got the most variation in its answers was "How long and how often do you think you would listen to something like this?" Four people's responses could be paraphrased as "long periods of time," suggesting it as background music for working or teaching children. Five people suggested one hour. Six people said 15 minutes or less, including one user who said "probably not ever." Four others acknowledged the infancy of the project and said it would depend on the amount of variation in the song.

When asked if they would like Ambient Addition built into existing MP3 players/Walkmans, 13 users responded "yes," with four of those responses followed by exclamation signs. Four other users gave ambivalent answers such as "wouldn't go out of my way for it" and "I dunno." One user said he "would want to play with it more first."

Finally, two questions asked about the future development of Ambient Addition. To the first, 12 people requested things like "more complexity," "more settings," "more variations," "less predictability," and "more dynamic representation of the external landscape." One user responded "richer sound, less focused on percussive tones." When asked if they wanted any controls over the sound, 17 people responded. 8 of them said they wanted volume control. 9 of them had more complex requests like knobs to control the set of chords, timbre, randomization controls, tempo, and theme.

Future Work

With the completion of the Ambient Addition prototype, variations and enhancements can be explored as future projects. Here is a list of both major projects and incidental ones:

It would be interesting to attempt to convert MP3 music into something that Ambient Addition could play. Many people have suggested this after first hearing a description of what it does.

The onset detector could be much improved. Most importantly, it should record single audio events instead of fixed-size samples.

One of the problems of taking a completely intuitive approach to the tube resonance/vocoding technique is that some critical underlying theory was missed. In this case, I did not realize the importance of phase processing to this algorithm. The sound quality could be much improved in this area.

The overall volume level of environments varies greatly. Currently, Ambient Addition has a compile-time parameter to control the input gain/attenuation, but it would be better if an autogain module were implemented to compress the sound level into the useful range of the processing blocks.

Although the rhythm as implemented is amusing, it would be preferable to be able to import properly composed drum tracks.

From time to time, regular rhythms do appear in the environment. It would be intriguing to use the output of the onset detector to determine the tempo in such situations and synchronize playback to it.

Appendix - User Evaluation

Ambient Addition: User Evaluation

Date,Time__________ Location__________ Name or Alias____________

Approximate Age____ Gender_______

Did you think of this as music? Or as something else? What else?

How long and how often do you think you could listen to something like this?

Where do you imagine it would be interesting or fun to use Ambient Addition?

What, if anything, did you notice that you hadn't noticed before?

Would you like to see Ambient Addition built into MP3 players/Walkmans?

What music styles or artists does this make you think of?

Can you describe your state of mind while listening to Ambient Addition?

In what direction would you like to see this idea go?

Is there anything you wish Ambient Addition did beyond what you hear today?

Did you do anything differently while listening to Ambient Addition, compared with your average walk down the street? If so, what?

Did you interact with people more or less while listening to Ambient Addition than you normally do with headphones on? without headphones on?

What do you think about the transparent headphones?

What sort of controls (if any) would you like?

Do you feel Ambient Addition made you do things on purpose to hear how they sounded? If so, what?

What do you think about the audio quality?

What impression do you get seeing someone else with it on?

Does it seem more or less distracting than listening to music through headphones?

How does Ambient Addition reflect your feelings toward social isolation?

What kind of musical experience do you have? none, listener, player, composer, etc.

What do you think about the connection between the sound in the outside world and the sound Ambient Addition makes?

Did you like Ambient Addition? How much? (1-10) 1=hated, 10=loved it

Anything negative to report?

What else would you like to say?

Bibliography of Introduction

1 Bierce, Ambrose and R. Steadman (Illustrator), p. 85, Bloomsbury USA, January 17, 2004, original text 1911.

2 Sony Press Release, July 1, 1999.
http://www.sony.net/SonyInfo/News/Press_Archive/199907/99-059/

3 Compact Oxford English Dictionary online, 2006.
http://www.askoxford.com/concise_oed/walkman?view=uk

4 Merriam-Webster Online Dictionary, 2006.
http://webster.com/dictionary/walkman

5 Thibaud, Jean-Paul. "The Sonic Composition of the City," pp. 329-341. The Auditory Culture Reader, Eds. Michael Bull and Les Back, Berg, Oxford, 2003.

7 Schwartz, Staci, photograph, in the Village Voice, June 2 - 8, 2004.
As part of Laura Conaway's article "iPod, Annotated."
http://www.villagevoice.com/nyclife/0422,conaway,53965,15.html

8 Taylor, Nicholas. Give Chance a Chance, The Transformative Possibilities of the iPod, *New York Spirit*. Published October & November 2005.
http://www.nyspirit.com/Issue134/article5.html

9 Dougherty, Kerry, "With iPod, it's easy to run into isolation", The Virginian-Pilot April 1, 2006
http://home.hamptonroads.com/stories/story.cfm?story=102376&ran=125933

10 Patton, Phil. "Walkman Anniversary," 1999.
http://www.philpatton.com/walkman.html

11 Elsenaar, Arthur and Taco Stolk. Bubl Space, ISEA2004, 12th International Symposium on Electronic Art. 2004.
http://www.isea2004.net/mainframe.php?id=xh_bublspace

12 Fried, Limor. Social Defense Mechanisms: Tools for Reclaiming Our Personal Space. p. 32. January 2005.

13 Dahl, Axel, The Vintage Knob Online Audio Museum. May 2004.

http://www.thevintageknob.org/SONY/sonyvault/TPSL2/TPSL2.html
http://www.thevintageknob.org/index.html
14 Patton, Phil. "Walkman Anniversary," 1999.
http://www.philpatton.com/walkman.html

15 Russolo, Luigi. "The Art of Noises", Pendragon Press, 1987.
http://www.unknown.nu/futurism/noises.html

16 Wendt, Larry, "Narrative as Genealogy: Sound Sense in an Era of Hypertext", Switch, Vol. 2, No. 1. 1998.
http://cotati.sjsu.edu/spoetry/nghome.html#sp

17 White, Ray. "Musique Concrète," **Appendix** to *BBC Radiophonic Workshop, An Engineering Perspective,* 2001.
http://myweb.tiscali.co.uk/whitefiles/b2_h/4_radiophonic_workshop/rws/pgs/a_toc.htm

18 Lombardi, Daniele, "Futurism and Musical Notes," *Artforum*, January 1981.
http://www.ubu.com/papers/lombardi.html

19 Helmreich, Stefan. Cover art of Xerophonics. Published by Seeland, 2003.

20 Saperstein, Jeremy. Vital Source, January 2005.
http://www.vitalsourcemag.com/story/view/10080?PHPSESSID=763c5430407 3dea571d5619c4f21e283

21 Bennet, John. "The majority of music we never hear", The Campus Chronicle, Vol 3, No 22, May 2, Savannah, Georgia. 2003.
http://www.thecampuschronicle.com//archive/vol_3/05_02/inside_sound.html

22 Negativland, Cover art of "Dispepsi" Published by Seeland, July 1997.

23 Salinas, Alejandra, Cover art of "Home Tapes," published by Lucky Kitchen LK007, 2000.

24 Howell, David. The Wire (UK), 2000.

Bibliography of Technical Possibilities Chapter

1 Lewis, George E. pp. 33-39 "Ways & Means Too Many Notes: Computers, Complexity and Culture in Voyager" Leonardo Music Journal 10, 2000. http://muse.jhu.edu/demo/leonardo_music_journal/v010/10.1lewis.html

2 Dudley, Anne, Gary Langan, and Paul Morley, a.k.a. The Art of Noise, Daft,"In the Army Now", 1986.

3 Handle, Stephen. Listening. Chapter 11, "Rhythm" pp. 383-460. MIT Press, Cambridge, 1989.

4 Jansen, Aren, Master's Thesis *The Manifold Nature of Vowel Sounds. The Physics of Acoustic* Tubes July 2005. http://people.cs.uchicago.edu/~aren/mpaper/mpaper.html

5 Truax, B. "Composing with Time-Shifted Environmental Sound," Leonardo Music Journal, 2(1), 1992, 37-40.

6 A. Kameoka and M. Kuriyagawa, (1969) "Consonance Theory Part II: ``Consonance of Complex Tones and Its Calculation Method", The Journal of the Acoustical Society of America, 1969b, Vol. 45(6), 1460-1469, 1969.

7 Levin, Theodore C., and Michael E. Edgerton. "The Throat Singers of Tuva," Scientific American, September 20, 1999.

8 Frank, Robert J. Theory on the Web, presented at the 2001 TMEA conference, 2001. http://www.smu.edu/totw/overtone.gif

9 Temperly, D. The Cognition of Basic Musical Structures. MIT Press, 2002.

10 Povel, Dirk-Jan. http://www.socsci.kun.nl/~povel/
http://www.socsci.kun.nl/%7Epovel/Melody/index.html
http://www.socsci.kun.nl/%7Epovel/Research.html

Bibliography of Design Chapter

1 Robert Jourdain's book Music, the Brain and Ecstasy: How Music Captures Our Imagination. William Morrow Press, 1997.

2 John Ortiz's The Tao of Music. Weiser Books, 1997.

3 Debord, Guy. Report on the Construction of Situations and on the International Situationist Tendency's Conditions of Organization and Action. Revised translation by Ken Knabb in 1981, 1957.

4 Chai, Wei. (2005) "Automated Analysis of Musical Structure."

5 Jehan, Tristan. Creating Music by Listening, 2005.
http://web.media.mit.edu/~tristan/phd

6 Wiffen, Paul. SYNTH SCHOOL : Sound on Sound Magazine, October 1997
http://www.soundonsound.com/sos/1997_articles/oct97/synthschool4.html

7 Gardner, W. G. (1995). Efficient convolution without input-output delay. Journal of the Audio Engineering Society, 43 (3), 127-136.

8 Holmquist, L., Gaye L., Mazé, R., et al. "Sonic City: The Urban Environment as a Musical Interface" Proceedings of the 2003 Conference on New Interfaces for Musical Expression.
http://www.viktoria.se/fal/projects/soniccity/pdf/NIME03_Gaye.pdf

9 Maebayashi, Akitsugu, "Sonic Interface", exhibited at Mediaselect, Nagoya, Japan, 1999.

10 Tanaka, Atau. Mobile Music Making, Proceedings of the 2004 Conference on New Interfaces for Musical Expression, Hamamatsu, Japan.
http://hct.ece.ubc.ca/nime/2004/NIME04/paper/NIME04_3B02.pdf

11 Iwai, Toshio. "Sound Lens." 2001.

12 Apple Computer press release. June 2005.
http://www.apple.com/pr/library/2005/jun/28ipod.html

13 S., Rushan, "PSP vs. Nintendo DS - Closest Race Since SNES vs. Genesis". PSPUpdates - 24/7 Coverage of the Latest PSP News
http://pspupdates.qj.net/index.php?pg=49&aid=25898

14 Korth, Martin. "DS Tek - Nintendo DS Technical Information." 2005. http://neimod.com/dstek/

15 Vawter, Noah. "The Home-Built Effect Processor Express." 2001. http://gweep.net/~shifty/death/

16 Father Dan. "Ipod Isolation and the Single Guy," anecdote, 2006. http://www.danielcurran.com/2006/04/ipod-isolation-and-single-guy.php

17 Mueller, Florian 'Floyd', Karau, Matthew. Transparent Hearing, CHI 2002 Extended Abstracts on Human Factors in Computing Systems, Minneapolis, 20 - 25 April 2002. http://web.media.mit.edu/~stefan/hc/projects/transparentheadphones/

Bibliography of Development Chapter

1 Zhang, Jie. Developer. 2005. http://blackfin.uclinux.org/projects/jtagtools/

2 Smith, Steven. The Scientist and Engineer's Guide to Digital Signal Processing. California Technical Publishing, 1997. http://www.dspguide.com/

3 Chai, Wei. (2005) "Automated Analysis of Musical Structure."

4 Jehan, Tristan. Creating Music by Listening, 2005. http://web.media.mit.edu/~tristan/phd

5 The DSP-Music Syndicate. *Chiclet*, 2003.