Things that talk: Using sound for device-to-device and device-to-human communication

by V. Gerasimov
W. Bender

Nonlexical sound is explored as both a device-to-device and device-to-human communication medium. Considerations for device-to-device communication include robustness in various environments, potential interference, frequency limitations of conventional and piezoelectric devices, computational complexity, and strategies for ultrasonic and human-audible frequencies. Algorithms include modem protocols, information-hiding techniques, impulse coding, and dual-tone modulation. Considerations for device-to-human communication include the use of sounds that are unobtrusive in public settings and sounds that enable attention to be divided between the performance of complex tasks and real-time feedback.

Hearing is fundamental to our perception of the world, and sound can considerably enrich interaction with computers. Research in data-auralization techniques, for instance, demonstrates that humans easily learn to distinguish objects by associated sounds.[1,2] The use of sound does not have to be limited to accenting what the user can see on the computer monitor; it can also provide information from objects and events beyond the screen.[3] Yet computers and hand-held devices rarely use sound to tell us more than "look at me."

Sound is a reliable way to alert someone about an event. "We need not turn to hear something; in fact we cannot turn away or close our ears."[3] For this reason many of our communication devices "ring" to cause us to be attentive to a telephone call or a new message. Unfortunately, we cannot be unattentive if someone else's device is ringing.

The primary purpose of these audio alarms is to attract attention, and most of the research in this area is focused on how to do that quickly and with persistence. For example, Patterson et al.[4] explored how to alert medical personnel in critical situations. In the hospital, airplane cockpit, or nuclear power station, alarms help to avoid life-threatening situations. When it comes to saving lives, an alarm cannot be too disturbing or annoying.

"Classic" audio alarms may vary to distinguish among different alert situations, but they always attract the attention of everyone around them. These same alarm sounds are used in personal communication devices. As more people carry their pagers and cellular telephones in public places, there will be an imperative to design personalized alarms that attract the attention of a specific individual while not disturbing everyone else.

Looking beyond the task of alerting the user, one could ask: Why does a telephone ring not tell you or your computer who is calling? Why does your wristwatch alarm not signal your door to unlock or your television to turn on? Many commonplace devices and appliances contain information that can

0018-8670/00/$5.00 © 2000 IBM

be shared with other devices. Although many of these devices have mechanisms, e.g., speakers and displays, to transfer information to humans, they are deficient in their ability to communicate information to other devices. Almost all modern electronic devices have internal clocks, but almost none are capable of accomplishing such mundane tasks as automatically transferring the time and date settings from one device to another.

Of course, one solution to device-to-device communication is to add additional I/O infrastructure, such as the capability to transmit or receive infrared light or radio frequencies, or do both. In this paper, we explore the possibility of using the existing audio capability found in many commonplace devices.

About sound. Sound is radiant energy transmitted in longitudinal waves that consist of alternating compressions and rarefactions in a medium. The maximum possible frequency of sound in any medium is approximately 1.25×10^{13} Hertz (Hz). The speed of sound depends on properties of the material through which it travels (see Table 1). Sound has propagation features quite different from that of light, which makes it an excellent alternative in situations where light does not work well.

Specialized imaging applications, e.g., ultrasonic microscopy and echolocation, make sophisticated use of sound. Sound may be a good alternative for optical computer vision; for example, rather than trying to recognize people by imaging their faces or fingers in visible light, sonar technology can be used to measure unique body parameters, such as volumes, shapes, motion, and locations of different organs. But for the most part, the use of sound in computers and communications applications has been limited to music, simple sound effects, and speech synthesis and recognition. (Notable exceptions include dual-tone modulation frequency [DTMF] used in telephony and various modulations used in modems and facsimile machines. However, modems and faxes are pseudoacoustic devices and cannot effectively communicate through air. They work well only with electric signals and require a wire to isolate the system from external disturbances.)

Many electronic devices in and around the human-made environment are designed to speak or listen. Many devices have a speaker or beeper to relay status to a human. These same speakers can be used for device-to-device communication. For example, the existing speakers in even the most low-end television could be used to communicate with toys and appliances if only they had the capability to listen. (At one time, some televisions could listen. The first television remote controls were mechanical devices that generated different sounds.) Many devices, such as telephones or faxes, have microphones. Piezoelectric beepers in wristwatches can also be used as microphones[6] (albeit with limited fidelity).

Table 1 Speed of sound in various media[5]

Medium	Speed (m/second)
Air at 0° C	331 (increases with temperature and pressure)
Water	1400–1600 (increases with temperature, salinity, and depth)
Lead	1210
Rolled aluminum	5000

Medical concerns. Concern over noise includes annoyance, communication interference, work interference, hearing loss, and other health issues, such as interference with sleep, impact on vasoconstriction, vertigo, and change in galvanic skin response.

Annoyance has been roughly correlated to a decibel (dB) scale. White noise at less than 55 dB is not annoying, whereas noise at 93 dB or greater is extremely annoying.[7,8] In isolation, speech sounds require 35 dB over background noise in order to achieve 50 percent intelligibility and 80 dB over background noise in order to achieve 90 percent intelligibility. However, there is little evidence of communication interference in office situations, where background noise is intermittent.[9] Studies of noise and efficiency are somewhat inconclusive, with varying impact on perceptual-motor performance, selected and sustained attention tasks, verbal learning and memory, and intellectual tasks. In some contexts, noise was a cause of facilitation of performance, in others, deterioration.[10]

In general, audible sound can be used safely either in short data exchanges when people control the transmission (e.g., phone ringers, acoustic remote controls, data exchange between wristwatches and computers) or for long data exchanges when humans are not present. However, there is evidence that noise at levels greater than 80 dB can be of concern. The U.S. Occupational Safety and Health Administration (OSHA) regulates noise exposure in the workplace as a function of decibels and exposure time.[11]

For example, there is a restriction of a maximum of eight hours per day exposure to noise at 90 dB and a maximum of one hour per day at 105 dB.

Landström reports that exposure to infrasound (6–16 Hz) has an impact on fatigue and wakefulness even with exposures as short as 20 minutes.[12] However, Slarve and Johnson report no long-term impact from exposure to infrasound at levels up to 144 dB.[13]

Ultrasound in the range of 20–500 kHz (and probably higher) is harmless if it is used intermittently and is not very loud. (It is safely used in medical applications such as fetus imaging.) If ultrasound is extremely intense, it may cause heating and scattering of hard objects. Noise at frequencies lower than 100 kHz may be unpleasant for pets. There are many natural sources of ultrasound—bats and some insects produce very loud ultrasonic noise. However, these sounds do not seem to affect humans or animals. Almost all televisions and computer monitors constantly squeal at line-scanning frequency. This sound has not been recognized as a source of medical problems.

Things that talk. The remainder of this paper explores "things that talk"—the use of sound as a device-to-device and device-to-human communication medium. In the discussion of device-to-device communication, considerations are made for robustness in various environments, potential interference, frequency limitations of conventional and piezoelectric devices, computational complexity, and strategies for ultrasonic and human audible frequencies. Also, several algorithms are described. In the device-to-human communication section, the problem of generating audio alerts is discussed. Two applications, a device that uses real-time audio feedback to coach athletes and an audio computer game, are also described.

Device-to-device communication

Sound has many positive features that make it an attractive alternative to infrared (IR) light and radio frequencies (RFs) for device-to-device communication. Humans (and most other animals) use sound to communicate. Therefore, sound is a likely choice for communication wherever people live or work. The human environment is usually designed to localize and preserve sound, and in such environments, sound is more predictable than IR and RF. Sound, unlike IR and RF, does not have unexpected foes such as sunlight, rain, metal objects, etc. Sound can be hidden from people if its frequency is higher than 20 kHz. IR usually requires direct visibility and is hampered by sunlight or bright interior light. Also, IR has some unexpected features: It propagates through materials that are not transparent in visible light and reflects from different materials than does visible light. RF suffers from interference problems and can be blocked by metallic objects. Sound does not have these problems. It travels around corners and can still be localized within a room.

But one feature of sound makes it difficult to use—sound travels slowly (about 300 meters per second in air). This creates several problems: (1) a signal reaches its target with a noticeable delay; and (2) reflections of the signal from walls or other objects can be comparable in intensity to the original signal and can reach the receiver with significant delays, interfering with the main signal. (Because of the slow speed of sound, an echo may create long disturbances of the signal, making it necessary to use relatively long coding packets.) Other problems include ambient noise, which tends to be a combination of low-amplitude broadband and moderate-amplitude narrowband components, and the nonlinear characteristics of most audio equipment and environments that have very different and practically unpredictable acoustic characteristics.

Sound is a likely choice for communication wherever people live and work.

Sound signals and the environment. Audible background noise generally drops off exponentially with frequency. Different environments have intermittent peaks that can climb as high as 90 dB. In an office environment, monitor line frequencies vary from approximately 15 kHz to greater than 100 kHz. Extreme noise levels are found in industrial settings, where some noise levels reach as high as 132 dB.[14] In the home, the kitchen tends to be the noisiest room. Kitchen appliances are the source of intermittent noise with peaks up to 80 dB. Even pouring water for a bath causes noise with peaks of 73 dB.

Male human speech at a distance of one meter is measured at 75 dB.[15]

Since the environment has little effect upon the relative amplitude of a sound within narrow bands of frequency, spectral analysis is a reliable method of acoustic signal analysis in unknown environments. The relative amplitude changes of the spectral components of a transmitted signal can be detected by a receiver. In general, if it is possible to measure ambient acoustic characteristics, choice of protocol or the parameters of a given protocol can be adjusted to maximize signal to noise.

Many small speakers and microphones used in electronic devices perform reasonably well at frequencies up to 30 kHz. The frequency limit of piezoelectric devices is much higher. The higher limit means that the same devices can use ultrasound whenever interdevice communication might be distracting or not be relevant to human listeners.

Sound signals have several convenient mathematical representations. These representations split the received sound signal into a function of the original signal plus noise:

$$f_{recv}(t) = f_{noise}(t) + \int_0^{\infty} f_{sent}(t - x)\, g(x)dx \qquad (1)$$

A received signal can be represented as a sum of the convolution of the transmitted signal with an impulse response of the environment and noise[16] (Equation 1). Unfortunately, whenever air or any objects in a given environment move, the parameters of the system change, and as a result, the impulse response may change in time. Therefore, $g(x)$ depends on t. $g(x)$ cannot be calculated at the source and then used to predict and correct the received signal. In any environment that can be used for acoustic communication, $g(x)$ has to decrease exponentially as $x \to \infty$. Also the integral of $g(x)$ over the whole axis must be less or equal to 1.

The noise component in Equation 1 represents both the deviation of the system behavior from the regular convolution formula caused by nonlinearities and all other sources of sound in the environment. Noise is uncontrollable and largely unpredictable.

An impulse function gives the most complete description of an environment. Unfortunately, calculating the impulse response of the system in real time is a complicated task even if ample computational resources are available. Therefore, working with that signal representation on small (4–8 MHz) microcontroller-based devices is not feasible.

It is possible to use a simpler representation of the received signal. The original signal can reach the receiver either directly or after reflection from one or more surfaces. Therefore, the received signal can be represented as a sum of the directly received signals, all reflections (echoes) of the original signal that reach the receiver, and noise:

$$f_{recv}(t) = f_{noise}(t) + \sum_i \alpha_i f_{sent}(t - \tau_i) \qquad (2)$$

where $0 < \alpha_i \leq 1$ is the energy reduction coefficient, and τ_i is the corresponding delay of a reflected or directly received signal.

Equation 2 has some useful features that help to explain why spectral analysis works in these situations. If the original signal is a sine wave, then the sum of all reflections of the original signal is also a sine wave of the same frequency[17] (Equation 3). Thus, the amplitude of the reflected signal is proportional to the amplitude of the original signal.

$$\sum_i \alpha_i A_{orig} \sin(\omega t - \tau_i) = A_{orig} \sum_i \alpha_i \sin(\omega t - \tau_i)$$

$$= A_{orig} A_{sum} \sin(\omega t - \tau_i) \qquad (3)$$

Since sound reflects differently from different objects and propagates rather slowly, a sharp change in the original signal spectrum will cause several changes in the received signal spectrum. A computationally simple receiver can correctly analyze the signal only after a *settle* time. Therefore, a system can work only if either the coding frames are longer than the average settle time or it has an adequate echo-tolerance reserve. Although the detectable echo tail in a medium-sized room can be as much as several seconds long, the strongest echo components are the first reflections from large surfaces, e.g., walls, floor, ceiling, furniture, etc. These strongest echo components in a room usually have delays only up to several hundredths of a second.

Communication protocols. Acoustic computer communication requires special protocols to be developed. The protocols used in modems, faxes, tele-

Table 2 Standard Touch-Tone codes

	1209 Hz	1336 Hz	1477 Hz	1633 Hz
697 Hz	1	2	3	A
770 Hz	4	5	6	B
852 Hz	7	8	9	C
941 Hz	*	0	#	D

phones, and remote controls are designed for mediums other than sound through the air. However, those protocols can be modified to work reliably with acoustic signals.

Modem protocols. An obvious first step in device-to-device communication research is to adopt existing fax and modem protocols for acoustic transmission. The fastest modem protocols exploit almost all of the properties of existing telephone lines. They rely on phase information, and they do not tolerate multiple echoes. Less sophisticated modem protocols (up to 2.4 Kbps, or bits per second) are less dependent upon specific telephone-line characteristics and can be modified for use with through-the-air communication.

The audio-signal stream in low-speed modem protocols consists of a sequence of short coding frames. The transmitter generates several predefined frames that must be recognized by the receiver. All of the coding frames have the same length. Different coding frames must have different spectral characteristics to be distinguished by the receiver.[16,18] Coding frames can contain either a single frequency or a mixture of several frequencies. A simple technique used in modems is frequency shift key (FSK), which uses single-frequency frames. The FSK frequencies are chosen so that there are a whole number of waves per frame and so that the frames can be easily distinguished by the receiver.

Usually, after traveling through the air, the signal is blurred because of echoes. A technique for circumventing this problem is to increase the frame length beyond the average echo settle time. Echo tolerance can also be increased by shifting the coding frequencies used in adjacent frames.

Information-hiding protocols. Bender et al.[19] suggest several possible techniques for adding recoverable inaudible data to a host acoustic signal. Spread-spectrum and phase encoding are useful for digital or high-quality analog transfers. One of their suggested techniques, echo coding, preserves hidden data after the sound has traveled through the air. An additional advantage of this technique is that it can be used in conjunction with a host signal that is meaningful to a human listener, i.e., it can be used in situations where it is desirable for the device-to-device communication to be monitored.

Impulse protocols. To send data, IR remote controls use a sequence of light bursts with different delays between them. This method can be used with sound. An acoustic remote control can send short impulses with pauses between them. (As previously mentioned, Zenith Electronics Corporation manufactured a mechanical-acoustic television remote control in the 1950s. Each button produced a single tone.) This protocol is easy to implement and requires minimal processing to recover the data. But it cannot deliver high, continuous data-transfer rates.

Touch tones. Touch tones used to dial telephone numbers and send information among telephone nodes work fine in air, but are slow and subject to noise. However, it is a ready-to-use method for computers to communicate with telephones and faxes. Recognition and generation algorithms are easy to implement. Each Touch-Tone** code is a mixture of two tones chosen from a set of eight basic tones. Table 2 lists the standard telephone codes.

A personal-computer sound board can be programmed to communicate with touch tones. Patterns generated for each Touch-Tone code can be sent to the sound driver when the user presses the corresponding buttons. The recognition algorithm uses eight digital filters to detect the basic tones. Touch tones work well if they are at least 20 ms (millisecond) long. Thus, the maximum reliable data transmission speed that can be achieved using touch tones is about 200 baud. Standard touch tones are 70 ms or longer, which is above the typical echo settle time. Therefore, computers, telephones, and faxes can communicate reliably using touch tones. The set of tones can be expanded to achieve higher data-transfer rates.

Other protocols. Almost any modulation scheme of amplitudes, frequency, or phase can be used for acoustic transmission, including frequency hopping or spread-spectrum techniques. As with the modem protocols described above, compensation for the blurring of the signal caused by echoes is a major design issue.

Explorations. A study of different data encoding schemes for acoustic data transmission was conducted.[20] In order to verify robustness of the protocols, several working prototype systems were built that use sound to communicate. The study was conducted in an office environment. The transmitter was approximately two meters from the receiver. Both the speakers and the electret microphone were consumer-grade, typical of those shipped with a personal computer.

Echo coding. In the echo-coding studies, a one-bit modulation was used. An echo added to a signal corresponds to one and subtracted from the signal corresponds to zero:

$$f_{out}(b, t) = f_{in}(t) + (-1)^b \alpha f_{in}(t - \Delta t) \tag{4}$$

where Δt is the fixed echo delay, $(0 < \alpha < 1)$ is the echo-to-signal ratio and b is the encoded bit.

If the synthetic echo should change its sign abruptly, then any change from zero to one or from one to zero will result in a "click." A way to avoid these clicks is to increase and decrease the echo component linearly at either side of each frame.

There are several methods of detecting an echo with a given delay. The most robust methods require a Fourier transform of the signal and an analysis of its spectrum. Consequently, these methods require more processing power than is available in most "ordinary" devices. For the purpose of this research, it is desirable to find less computationally demanding methods. A simplified echo-coding method that requires only one multiplication per sample was developed:

$$F_{out}(t) = \sum_{i=0}^{N-1} F_{in}(t + i\tau) \cdot F_{in}(t - \Delta t + i\tau) \tag{5}$$

where τ is the sampling period, Δt is the echo delay, and N is the number of samples used by the filter.

The filter value of a modulated signal can be presented as a sum of two components as follows:

$$\begin{aligned} F_{out}(t) &= \sum_{i=0}^{N-1} F_{in}(x, t + i\tau) \cdot F_{in}(x, t - \Delta t + i\tau) \\ &= \sum_{i=0}^{N-1} [f_{in}(t + i\tau) + (-1)^b f_{in}(t + i\tau - \Delta t)] \cdot \\ &\quad [f_{in}(t + i\tau - \Delta t) + (-1)^b f_{in}(t + i\tau - 2\Delta t)] \\ &= (-1) \sum_{i=0}^{N-1} f_{in}^2(t + i\tau - \Delta t) \; + \\ &\quad \sum_{i=0}^{N-1} (f_{in}(t + i\tau) \cdot f_{in}(t + i\tau - \Delta t) \; + \\ &\quad (-1)^b f_{in}(t + i\tau) \cdot f_{in}(t + i\tau - 2\Delta t) \; + \\ &\quad f_{in}(t + i\tau - \Delta t) \cdot f_{in}(t + i\tau - 2\Delta t)) \end{aligned} \tag{6}$$

The first component is a sum-of-squares of the original function, and the second component is a sum-of-products of the original function at different points. The detection of echo is based on the fact that the absolute value of the first component is usually greater than the absolute value of the second component. In most cases, the sign of the filter output is defined by the encoded bit.

The program does not analyze the original signal and, consequently, may have poor results if the signal contains pauses or resonance frequencies. It is necessary to use an error-correction code and to verify check-sums to support reliable data transfers.

Experience with echo coding suggests that this method can give positive results at speeds of up to 100 baud, but it requires too much computational power to implement a robust real-time communication protocol. Although the simplified echo-detection algorithm is fast enough to run in real time on 8-MHz computers, its noise tolerance and signal independence leave something to be desired. The algorithm can reliably recover only about 10 bits per second. However, since the coding frame size is longer than the typical echo tail, echoes in the environment do not degrade this method.

Table 3 Sonicom two-frequency protocol

Transmitter output	1 channel, 44 100 8-bit samples/sec
Receiver input	1 channel, 44 100 16-bit samples/sec
Frame size	120 samples (~2.72 ms)
Coding frequencies (Hz)	735, 1470, 2205, 2940, 3675, 4410
Hail frequency (Hz)	5512.5
Encoding	4 bits/frame
Transmission speed	1470 bps

Table 4 Sonicom ultrasonic protocol

Transmitter output	1 channel, 44 100 8-bit samples/sec
Receiver input	1 channel, 44 100 16-bit samples/sec
Frame size	30 samples (~0.68 ms)
Coding frequency (Hz)	18 375
Hail frequency (Hz)	5512.5
Encoding	1 bit/frame
Transmission speed	1470 bps

The principal advantage of the echo-coding technique is that it adds information to any acoustic signal in an audible spectrum (speech, music, noise), so that human listeners do not clearly distinguish the communication between devices. One possible use of echo coding is to intermix a message for humans with a message for devices.

Frequency shift key and amplitude modulation. Three protocols based on variants of FSK and amplitude modulation were developed. The first protocol uses frames of two frequencies to transmit four bits simultaneously. Packets of 64 bytes are sent as a synchronous sequence of four-bit frames demarcated by a hail frame at the start of each sequence. The receiver, upon recognizing the hail signal, attempts to synchronize with the incoming data. Once synchronization is achieved, 64 bytes of information are received and decoded. The system uses a partial Fourier transform to identify frequency patterns in the received signal. The second protocol uses an amplitude modulated 18.4-kHz sound signal. Again, a hail signal is used for synchronization. Data are sent as a set of one-bit frames. A third protocol uses 40 groups of four frequencies to encode 80 bits of information in each coding frame. The 80 bits include a Reed-Solomon error-correction code. This "spread-spectrum" protocol achieves a data rate of up to 3.4 Kbps. All three protocols are implemented as part of Sonicom, a device-to-device communications prototype that uses an ordinary sound board to send and receive signals. Sonicom transfers data at speeds of up to 3.4 Kbps (see Tables 3, 4, and 5). A block diagram of Sonicom is shown in Figure 1.

Table 5 Sonicom spread-spectrum protocol

Transmitter output	1 channel, 44 100 16-bit samples/sec
Receiver input	1 channel, 44 100 16-bit samples/sec
Frame size	1024 samples (~23 ms)
Coding frequency (Hz)	4306–18 087 in ~86 Hz increments
Encoding	80 bits/frame
Transmission speed	3.4 Kbps

The Sonicom transmitter program generates and holds samples for each coding frame. The two-frequency protocol uses two tones to encode four bits. Each frame is comprised of either two different tones mixed together or silence. See Figures 2A and 2B.

In order to define 16 different frames (four bits), the system uses six different frequencies. There are 15 composite frames from pairs of frequencies (C_6^2). The sixteenth frame is silence.

A "hail" frame designates the beginning of each data packet (128 frames). Hail frames use a different frequency than data frames. This difference ensures that the receiver is able to readily identify the synchronization frames.

All of the frames are multiplied by the Blackman function[7] in Equation 7:

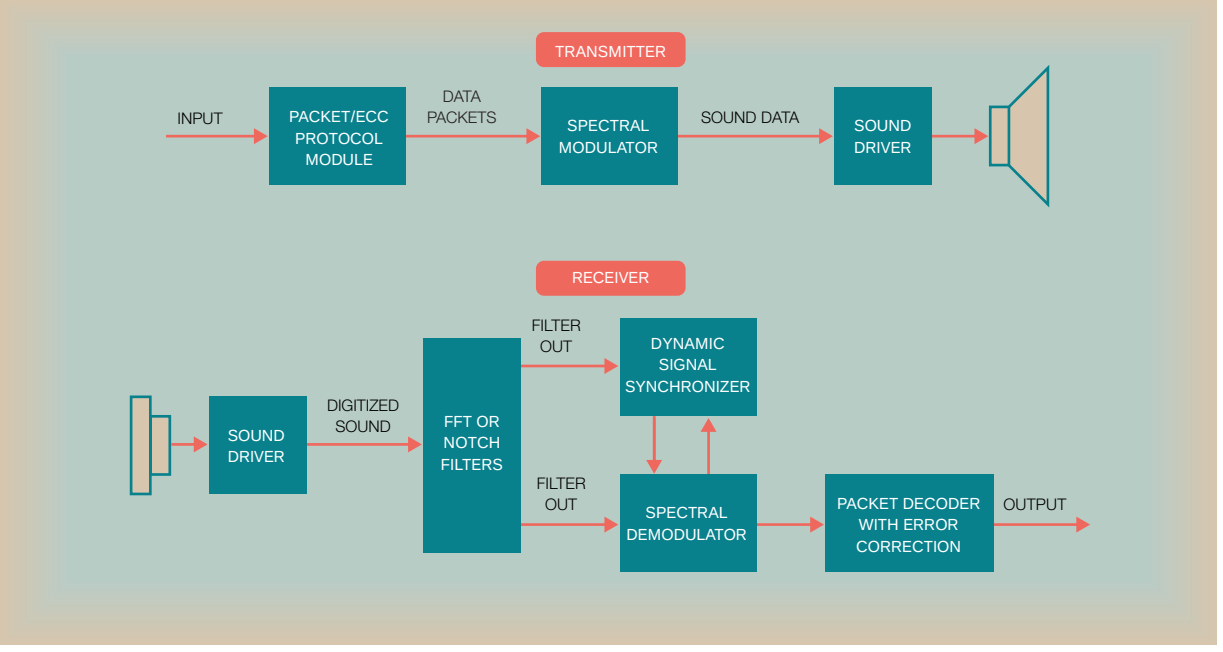
$$f(x) = 0.42 - 0.5 \cdot \cos\frac{2\pi x}{N} + 0.08 \cdot \cos\frac{4\pi x}{N} \qquad (7)$$

where N is the frame width.

Performance is improved when the Blackman function is used by the transmitter to form frames. Use of the Blackman function also reduces the number of calculations required in the receiver.

In general, it is always better to encode data using as few discrete frequencies as possible. Fewer numbers of frequencies require fewer numbers of filters, thereby reducing the amount of computation necessary to decode a signal. Also, using fewer frequencies improves the overall frequency separation and noise tolerance of the system. (The spread-spectrum

Figure 1 A block diagram of the Sonicom system



protocol is relatively computationally expensive since it requires a 1024-sample FFT [Fast Fourier Transform] for decoding.)

The amplitude protocol uses a high-frequency (ultrasonic) tone to encode data (see Table 4). The data frames are 30 samples wide at 44.1 kHz. With this protocol, each frame represents only one bit. Typical computer microphones and speakers can handle frequencies up to about 30 kHz, but amplifiers and digital-to-analog converters and analog-to-digital converters have a limit of approximately 22 kHz. Conventional sound boards have maximum sampling rates of 44.1 kHz. However, they do not work well at this upper limit. Consequently, a 18375-Hz tone was chosen to correspond to one. Silence was used to correspond to zero. The same hailing frequency was used as in the two-frequency protocol—at 5.5 kHz it was audible. This frequency was adequate for a demonstration system but must be increased if the protocol is to be ultrasonic.

The Sonicom receiver program conveys data from the sound-board driver to a set of digital filters.[21] The system uses the filter outputs to detect and decode incoming data. The receiver program maintains an idle state when no incoming packets are detected. When in this idle state, the system uses only the one filter that is needed to detect the hailing frequency. As a hail frame is received, the filter output first increases and then decreases. This output results in a peak that is one frame wide. By estimating the midpoint of the peak, the receiver is able to synchronize with the incoming data. It is only necessary to activate multiple filters in order to decode incoming data packets. Thus, computational load is reduced between data transmissions.

The digital filters are written in assembler using the Intel 80386 command set. Fixed-point arithmetic is used, enabling the program to run in real time on Intel 80486DX 33-MHz computers. The output of each filter is calculated by the following formula:[17]

$$F_{out} = \left(\sum_{k=0}^{N-1} x_k \sin \frac{2\pi i f_{filter}}{f_{sample}} \right)^2 + \left(\sum_{k=0}^{N-1} x_k \cos \frac{2\pi i f_{filter}}{f_{sample}} \right) \quad (8)$$

where N is the length of the filter (in samples), f_{filter} is the filtering frequency, f_{sample} is the sampling frequency, and x_k are the sample values. The value of each filter is a squared amplitude of the Fourier coefficient corresponding to the filtered frequency.

Figure 2 (A) A packet comprised of the hail signal and 16 different coding frames; (B) the same packet, after traveling approximately one meter through air, includes noise and echo added by the environment

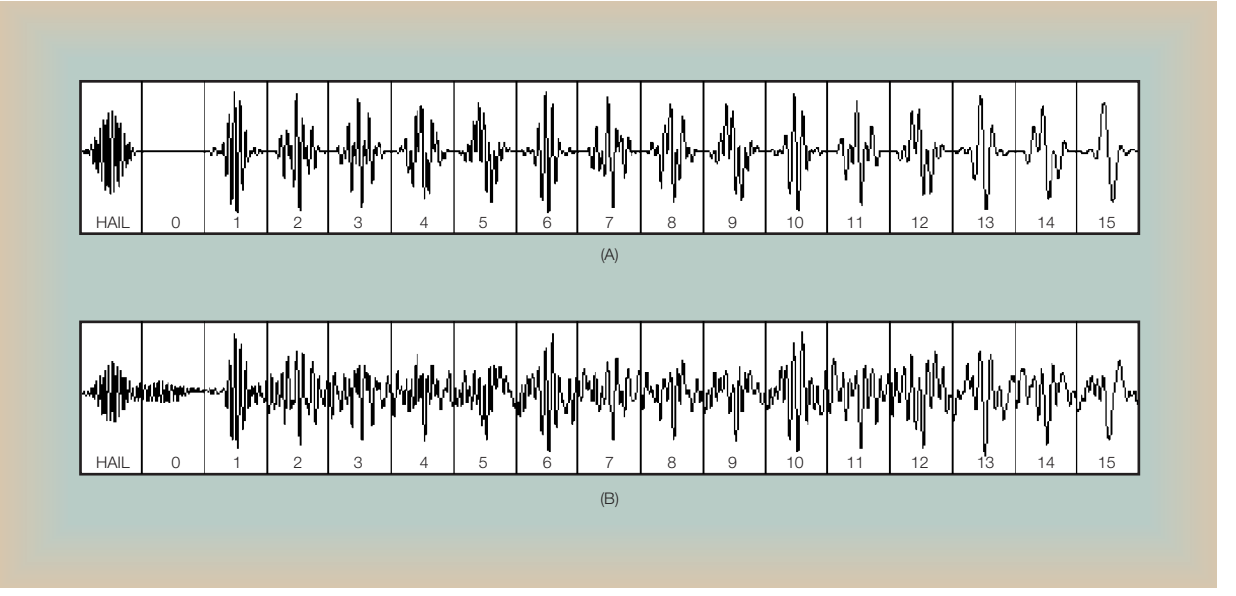


Table 6 Impulse coding technical characteristics

Transmitter output	1 channel, 1-bit, timer-modulated
Receiver input	1 channel, 44 100 16-bit samples/sec
Coding impulse width	~5.7 ms (50 semiwaves)
Coding impulse frequency (Hz)	4405
Bit 0 pause	~5.7 ms (50 semiwaves)
Bit 1 pause	~17 ms (150 semiwaves)
Average transmission speed	59 bps (7–8 bytes/sec)

Filters are the most computationally expensive part of the system. The filter blocks are optimized so that they perform only three multiplication operations per filter per sample. The program calculates the values of sine and cosine during initialization and stores them in memory arrays for subsequent low-cost filter state updates.

The Sonicom has good echo tolerance, making it possible to use a short frame size. The protocol works well with distances between transmitter and receiver of up to two meters (without obstacles). The reception improves if the receiver uses two microphones and sums the inputs.

Impulse coding. Another protocol, impulse coding, is similar to Morse code, but instead of using impulses of different length, it sends impulses of constant length with variable length pauses between them. The protocol encodes one bit at a time. The value of an encoded bit corresponds to the distance between two consecutive sound impulses. A short pause equals zero, and a long pause equals one (see Table 6). A similar protocol is implemented in IR remote controls. The protocol is easy to implement, requiring little computational overhead by either the sender or the receiver.

The receiver uses one digital filter to detect impulses. It dynamically adjusts its response level depending on the maximum filter output level from the last received impulse. The program measures the distance between the rising edges of two consecutive impulses. If the distance approximates the coding length of either a zero or a one, the receiver adds the decoded bit to a buffer. Otherwise, it resets the buffer.

The implementation has two parts: (1) a program for sending data using sound; and (2) a receiver in the form of software programs that decode the sound signals.

The data-transfer rate of impulse coding is slower than protocols that use multiple-frequency encoding. The impulse protocol also suffers somewhat from echo but otherwise proves to be reliable. Finally, if audible impulses are used, it creates a disturbing noise.

Noise tolerance. Each of the protocols used in the study has a noise tolerance of ~20 dB. However, each behaves somewhat differently to intermittent sounds typical of an office environment. The DTMF protocol is quite sensitive to human speech. The ultrasound encoding is somewhat sensitive to broadband noise (e.g., a clap). The impulse protocol is sensitive to human speech and to midrange frequencies. The noise tolerance of the echo protocol is dependent upon the original sound source. The spread-spectrum protocol is not sensitive to single-frequency noise sources and is quite tolerant of human speech.

A change in the path between the transmitter and the receiver may affect the signal delay. The resultant distance between coding frames at the receiver, as measured in samples, may vary significantly. A synchronization mechanism is necessary to overcome this problem. The receiver has to dynamically estimate the coding frame boundaries to correctly decode bits from the data stream. For example, when the FFT is used to analyze the spectrum of the transmission, the receiver has to find a set of consecutive samples that belongs to the same coding frame and is least affected by the noise and echo. In the case of a single-frequency amplitude-modulated signal, the receiver has to determine when the filter output is most likely to have the proper modulation level.

Dynamic synchronization is achieved in the beginning of the reception by calculating FFT or filter values several times within a single frame. The error-correction code, which indicates both whether the data are recoverable and how many errors have been corrected, is used to choose the best time to decode the frame. Once a satisfactory synchronization is achieved, the receiver assumes that the subsequent frames arrive with the same delay, until it detects an unrecoverable error.

More sophisticated multifrequency protocols can be implemented if the transmitter and the receiver are powerful enough to calculate the FFT in real time. The FFT is an efficient way to modulate and demodulate sound signals simultaneously carrying many amplitude or frequency-modulated bits. This type of modulation puts more information in the frequency domain and relaxes the time-domain constraints, increasing the data-transmission rates and decreasing the impact of the echo and signal delay.

Observations. Five subjects participated in a study conducted to determine which encoding techniques produced the least disturbing noise. Subjects listened to examples of DTMF, ultrasonic coding, FSK, impulse coding, and echo coding. As expected, since ultrasonic and echo-coding techniques operate near the threshold of the human auditory system, all of the subjects found these the least disruptive.

Four subjects preferred touch tones to the other audible techniques. They explained that either they were accustomed to touch tones or that the data transmission sounded melodic to them. One subject pointed out that since FSK has a much faster data transfer rate than touch tones, it might be preferable to listen to an FSK squeal for a short period of time rather than to listen to touch tones or impulse coding for a longer period.

Discussion. One could conclude that ultrasound should be used for device-to-device communication and that audible signals be sent only when devices communicate to human listeners. FSK modulation can be used ultrasonically to achieve even higher data transfer rates than reported here. However, it will require special hardware that can operate at higher sampling rates than conventional sound boards or that can process analog sound signals. Touch tones are attractive, not only because they are commonplace, but also because the combination of two frequencies produces a beat (characteristic pulsation) that makes sound alternately soft and loud. This beat is similar to the sounds produced by musical instruments. In moderation, the tones seem pleasant. Touch tones are typically much longer than frames in FSK modulation. This characteristic explains why these data coding techniques sound so different to human listeners. Perhaps double-frequency tones can be used to make an encoding that sounds like background music. The only advantage of impulse modulation is its simplicity. It requires minimal acoustic hardware quality and processing power and can be used in small microcontroller-based devices. (See Table 7 for an overview of all of the protocols used in the study. The data and software used in the study can be found at http://vadim.www.media.mit.edu/ttt.html.)

Applications. A wristwatch utilizing the sound from its alarm has been programmed to transmit data to

Table 7 Summary of the protocols used in the Sonicom study

	Data Rate	Frequencies	Computational Overhead	Noise Tolerance	Disruption Level
DTMF	38 bps	697–941 Hz & 1209–1633 Hz	8 filters	Overlaps human speech	Melodic
Ultrasound	183 bps	18.375 kHz	1 filter	Only sensitive at one relatively rare frequency and to broadband noise	Not disruptive
Spread Spectrum	3.4 Kbps	4306–18087 Hz	FFT	Not sensitive to single-frequency noises and tolerant to human speech	Relatively quiet high-frequency noise
Impulse	59 bps	4405	1 filter	Sensitive to human speech and mid-range frequencies	Relatively loud mid-frequency noise (most disruptive)
Echo	10 bps	1 ms delay at −10 dB	1 filter	Depends upon original sound signal	Can spoil experience of original sound signal

a computer by using an impulse-coding technique. These data serve as a return channel for address book and telephone directory data stored on the watch. Using either the ultrasound or dual-tone technique, an airport public-address announcement could transmit data to personal data assistants (PDAs) or laptop computers. A computer might have an easier time than a human in decoding the contents of those routinely garbled announcements and would be more tolerant and exploitative of repetitious messages.

Several computers close to the same wall, floor, ceiling, or table can use an acoustic transceiver to transfer data through a solid object. It would work as a "wireless" network. The sound would travel much faster and at higher frequencies than it does through air.

Three acoustic responders placed on a ceiling can serve as reference points to track the position of other objects in a room. A device that wants to know its position can send a "ping" to the responders, receive responses, measure time delays, and calculate the distances to each of the responders. This method can provide relative coordinates of the device with very high precision. The theoretical precision limit for digital processing with a typical sound board (44.1-K samples per second) is about 15 mm, but analog ultrasonic devices can achieve much better results. The advantage of this method over electromagnetic sensors is that all of the coordinate calculations are linear. (Ishii et al. used a similar technique in designing a "reactive" Ping-Pong** table.[22])

Device-to-human communication

Acoustic human-to-device communication might be in the form of a clap or a whistle. It is rumored that some people can simultaneously whistle the multiple tones necessary to dial a touch-tone telephone. Presumably this skill is not easily acquired and hence not a candidate for a general-purpose protocol. Voice is the obvious candidate for such a protocol and has been the subject of a great deal of research. The use of voice generation and recognition is beyond the scope of this paper. The following discussion is restricted to nonlexical device-to-human communication.

A principal consideration in designing a device-to-human communication protocol is the balance between the human listener's attention to the task at hand and attention to the communication itself. A secondary concern is robustness. A tertiary concern is bandwidth.

In this section, device-to-human communication is examined from the perspective of three applications: audio that only changes the focus of attention of an intended human recipient (Personal Alarm), audio that divides attention between a complex task and performance feedback (Batting Belt), and audio that

Table 8 Spectral characteristics of the sounds. The spectral features are indicated by: sharp spectral peaks (P); significant temporal variations in the spectrum (T); repetition (R); and sharp onset (O).

Spectrum Features	Sounds
PORT	Old phone ring, modern phone ring (2 variations), dog bark, Turkish march, classic music (2 variations), jazz, rock (2 variations), MS Windows™ start sound, laughter (several people), baby cry, "psycho" sound effect
POR	"Wah-wah" (trombone)
PO T	"Hi!" (cartoonish voice), "uh-huh" (male voice), "phaser" sound effect, bird chirp, "welcome" (female voice), "message" (male voice), "message" whisper, "koshmar" (male voice), "yozhiki" (male voice), "toska" (male voice), "Walter Bender" (male voice), "mommie" (young girl's voice), monkey scream, horse neigh, sonar ping
ORT	Cough (variation #1)
PO	Glass break, gong, small bell, door squeak, large bell, frog, elephant, mechanical alarm clock
P T	Horses galloping by, "UFO" sound effect
OR	Keyboard clicks, clock ticking, cough (variation #2)
O T	Pneumatic door
P	Train going by, large bubbles in water
O	Rattlesnake, finger pop, pneumatic gun, cough (variation #3)
R	Wind in trees (variation #1)
T	Paper rustle
None	Surf, waterfall, car going by, wind in trees (variation #2)

is used as the primary focus of attention in a traditionally nonaudio interface (Audio Search).

Personal alarms. Pagers and cellular telephones have become a source of annoyance in public places. They often ring and attract attention at unacceptable moments. They are barred from many establishments for that reason. Furthermore, a single cellular-telephone "ring" in a meeting room may cause most of those present to retrieve whatever devices they have from their pockets or bags because all the alarms are so similar. Meanwhile, the use of alarms is rapidly expanding beyond traditional telephone applications. Researchers such as Mynatt[23] and Schmandt[24] are developing systems that utilize audio alerts for a wide variety of messaging applications.

Alarms have two problems: (1) they are too obtrusive; and (2) they are not directed. The objective of the Personal Alarm project is to find audio alarms for communication devices that can both attract the attention of the user and not disturb others. (Note that the use of vibration as an alarm is limited in its applicability. It requires that the device be worn by the user rather than in a purse, bag, or briefcase. Vibration also has a limited range of expressiveness.)

An informal study was conducted to test the level of obtrusiveness of sounds in a group-meeting environment. A wide variety of sounds was prerecorded on an audiotape. The tape was then played back by a microcassette recorder with an internal speaker placed on a desk at weekly student and faculty meetings. The reactions of five groups of 15–20 subjects to various sounds were observed and recorded. The meeting room had no noticeable external noise. Sound-damping panels on the walls minimized the reverberation.

The prerecorded sounds were played at approximately three-minute intervals in order to prevent the subjects from anticipating them. (One to two minutes of concentrated group discussion was sufficiently immersive to cause the subjects to forget about the ongoing study.) Twenty sounds were tested in each session.

Sounds were taken from various sources, including the Internet, sound sampler CDs, and sound schemes for Microsoft Windows**. Some speech and noise sounds were recorded by the researchers. (A complete list of the sounds is shown in Table 8.) The sound waveforms were scaled so that they had approximately equivalent perceived loudness. Spectral diagrams were generated of all the test sounds. These diagrams were visually analyzed for temporal variation, rhythmic regularity, energy distribution, and onset time.

One of the main obstacles in picking sounds for this study and grouping the results is the absence of a perceptual sound classification scheme.[25] The human auditory system analyzes sounds along many different dimensions, and these dimensions have little direct correlation with the physical characteristics of

the sound. The spectral analysis used in the human auditory system is fundamentally different from that used on the computer.[26,27]

During each meeting of the group, the sounds were subjectively classified by their *obtrusiveness*, depending on how many subjects were distracted, and their *directness*, depending on which subjects were distracted; i.e., were only the subjects proximal to the sound distracted or were all subjects distracted? The subjects' reactions were scaled from "not noticeable" to "very distracting."

Observations. Nonspeech sounds tended to distract and annoy. These sounds shared common characteristics such as attack time, spectral type, and rhythmic regularity. Sounds with sharp spectral features, steep onset, and large temporal variations were more noticeable than sounds with flat spectrum, gentle onset, and little temporal variations. These observations confirm results by Patterson et al.,[4] Edworthy et al.,[28] and Swift et al.[29] who classified alarm sounds and environmental noise based on their physical characteristics in terms of perceived urgency, annoyance, and impulsivity.

The results further suggest that for the purpose of the calm-alarm research, sounds can be separated into the following groups:

- Sounds that conclusively attracted much attention: regular phone alarms (repeated sound with sharp onset); any kind of music (sharp onset, repetition), any short sounds repeated two or more times such as ticking, clicking, etc. (repetition); animal sounds (sharp spectral characteristics, instinctive response); child's voice or cry (sharp spectral characteristics, instinctive response); laughter (social response); unusual sound effects such as "UFO," "psycho," or "phaser."

- Sounds that conclusively attracted little attention: moderate cough (socially acceptable); surf (broad spectrum); wind (broad spectrum); phrases said by an adult in a calm, even tone (socially acceptable).

- Conclusively direct sounds (attract attention of only one particular person and only a few people closest to the source): slight cough and rustling paper were heard only by people close to the sound source; adult voice or whisper alarms, especially calling a person by his or her name.

A somewhat surprising result was that music (classic, rock, jazz, slow, fast, loud, soft) was immediately noticed by everybody, making it unacceptable as a nonobtrusive alarm. The probable reasons are large temporal variations and abundance of sharp onsets typical of all genre of music. Even background music seems socially unacceptable in the meeting environment. (Ironically, many of the alarm options that come preprogrammed into cellular phones are musical.) Surf, wind, and other broad-spectrum soft sounds are less noticeable.

The candidates for nonobtrusive alarms were the socially excusable sounds such as moderate cough, paper rustling, or a whisper. Subjects near these sound sources were distracted, but subjects far from the source did not notice or did not pay attention to these sounds.

One way to attract the attention of a particular person in a crowd is to call that person by name.[30,31] This feature of human auditory perception (a selective but heightened sensitivity to certain sounds) can be exploited in order to attract the attention of the user of a device without distracting others. The user, for example, may record his or her name for the device to play back as an alternative to a ring.

Discussion. Sound aside, there are design issues that may improve alarm usability. When it comes to alarms, communication devices are at best clumsy. Unless settings are explicitly changed, devices "speak" with the same tone and loudness in a quiet auditorium or on a busy street, whether they are in a bag or in hand. Measurement of ambient noise before contributing to it is straightforward and useful—telephones already have microphones with which to listen. A device may estimate how loud it must be to be heard above background noise. Determining whether or not the device is in a bag or a pocket is also useful. In these situations an alarm must be louder.

Even without detecting where they are, communication devices can take a better approach to sounding alarms. For example, they may use the distinction between important and not-so-important messages to control ring volume and persistence. Only when the user does not respond to an important message may it be necessary to increase volume.

Sounds produced by personal devices have to be loud enough to be heard. As the devices become smaller and lighter, it is reasonable to mount them on the

shoulder[32] instead of on the belt, in a pocket, or in an attaché case. If a shoulder-mounted phone or pager rings, the sound has to travel only a short distance to the ear, and thus can be at a lower volume. It is not necessary to shoulder-mount multiple devices. Having a single "parrot" device that can speak and listen from the shoulder is sufficient if it can also talk to other electronic devices.

Finally, people and even some pets understand not only bells and whistles but also speech. It may not yet be possible to implement a complete and robust speech interface for communication devices, especially small and inexpensive ones, but it is possible to substantially improve the usability of existing devices with elements of synthesized or stored speech.

Real-time audio feedback. In many situations one's attention must be divided between a complex task and performance feedback. The limits of our ability to divide our attention between tasks or stimuli has been studied extensively.[33–38] Although agreement is not universal as to how the mechanisms of cognitive resource-sharing work, researchers have demonstrated that subjects can avoid distraction when listening to one audio stream while immersed in distractor streams when these distractor streams differ in voice or pitch from the target.[37]

The ability to manage tasks that require attending to multiple stimuli involves a trade-off between focused and divided attention. Wickens[35] has proposed a "proximity compatibility principle," which states that it is easier to switch focus when there is mental proximity (similarity) and where information that is related can be treated as a unit. Miyata and Norman argue that the ability to keep all of the tasks in short-term memory (without "swapping") is a key factor in best utilizing the limited human processing and memory capacity.[38] This suggests that real-time feedback, tightly coupled to the performance of a task, might be optimal in the context of divided attention.

The Swings That Think project[39] was directed toward developing a family of devices that provide real-time motion analysis and audio, tactile, or visual feedback to users engaged in tasks that require coordination of body movements and, possibly, some extra-body affordance (e.g., a golf club, tennis racket, fishing pole, or baseball bat). The devices perform three functions: sensing, analyzing, and providing feedback to the user. Each device consists of a collection of "wearable" sensors such as ankle and wrist straps, belts, and hats that sense characteristics of posture and motion while engaged in various activities.

One of the systems built for the Swings That Think project was the Batting Belt. The sensor part of the system includes a set of accelerometers and gyroscopes placed inside a baseball bat and on a player's body. An RF transmitter inside the bat and a belt-pack sends digitized data to a personal computer that analyzes the motion of both body and bat during swings. The system provides the player with audio feedback aimed at improving his or her batting technique.

From the player's perspective the system works as a coach. From the human-computer interaction perspective the Batting Belt system is an unusual computer peripheral. The physical batting system includes a human or machine pitcher that throws a ball toward a player who has a bat. In order to hit the ball with the bat, the player has to anticipate the trajectory of the ball. The high speed of the ball does not allow an untrained player to visually track both the ball and bat. The ball usually flies between the pitcher and player in less than 500 ms. A complete loop of perception to action feedback in people is so long (200–500 ms)[40] that the player does not have enough time to perceive and process the information in a consecutive way. The batter cannot rely on sight alone to coordinate motion but has to anticipate the ball and bat trajectories and act upon indirect (in this case, audio) clues.

The only way to gain batting skills is practice. The first important skill is the ability to hit the ball with the bat. The player has to learn when to start the swing, depending on various visual cues, and how to correct the path of the bat according to the trajectory of the ball (a timing uncertainty of ± 0.01 seconds makes the difference between a hit and a foul ball[41]). After the basic skill is learned, the player can work on improving technique.

The sound generated at the impact between the bat and the ball is one source of real-time audio feedback. Two important measures of successful batting are power and precision. When the bat hits the ball, the kinetic energy of the bat transfers to the ball. Therefore, the player has to make the bat move as fast as possible just before it hits the ball. The ball has to touch the bat in the right place to transfer as much energy as possible. If the ball touches the bat too close to the handle or to the tip, a substantial part of the energy may be wasted in the vibration of the

bat. (For the typical bat, the frequency of oscillation is approximately 260 cycles per second [middle C]. Minimizing this oscillation maximizes the energy imparted to the ball. "The highest-frequency sound produced in the ball-bat contact is roughly one-half the inverse of the impact time [so that a collision that lasts less than 1/1000 of a second generates sound frequencies greater than 500 cycles per second, which is one octave above middle C]. One hears the characteristic high frequencies in the 'crack' of the bat.") The "crack" becomes a "thunk" when the ball is hit off-center, since the collision time is longer.[42]

The Batting Belt project involved both hardware and software design. One of the main objectives of the hardware design was to make the system nonobtrusive. The bat sensors had to alter the bat balance as little as possible. The bat sensors in the later prototypes were embedded inside an aluminum bat.

Since batting requires that players rely on their visual system to start swinging and to correct the bat path, visual feedback would be distractive and even unsafe. Under normal circumstances, the auditory system provides little or no information before and during the swing. Hence, sound may be used to deliver real-time feedback and coaching information. In this situation audio is the preferred channel to provide a robust user interface for the computer system.

The Batting Belt system estimates the speed of the bat using an embedded accelerometer and gyroscope that measured translational and rotational velocity components of the bat and in real time generated a tone with a pitch proportional to the compound velocity value. The tone helps the players to objectively judge how fast they can swing the bat. Sound makes it possible to deliver real-time performance information, avoiding any visual distraction. After each swing, the player may also visually check the maximum speed of the bat by the number of LEDs (light-emitting diodes) lit on the bat.

After each swing, the system analyzes the parameters received from all the sensors to verify how closely the player followed basic batting guidelines. An ideal model of batting for the system was taken from Adair's *The Physics of Baseball*.[42] The model describes a sequence of events, including a forward step, rotation of the hips, the torso, the shoulders, and the arms. A foot-mounted accelerometer helps to detect whether and when the player made a step forward. Body-, head-, and shoulder-mounted gyroscopes track the phases of rotation of the player's body. The accelerometer and gyroscope inside the bat help the system to detect the moment of the bat-to-ball collision and estimate the amount of energy transferred from the bat to the ball.

In addition to real-time audio feedback, the system prepares a report on player performance during the swing and generates an audio stream that includes a set of phrases pointing out the achievements and mistakes during the swing. An example of such an audio stream is "Try to hit the ball. Turn your hips later. Make a step forward." The system is able to detect errors in body coordination, too early or too late bat-to-ball collision, and absence of the transfer of weight by a step.

Batting practice is an example of a complex performance-oriented situation in which the user has to train both visual and motor systems. Unlike most computer systems, the Batting Belt has to rely primarily on sound to convey information to the user. The visual information containing the recorded summary of the training session was complimentary and delivered to the user only after batting practice.

Audio computer games. Even though modern computer games use a rich spectrum of sound effects, the use of sound is almost always limited to the emphasis of action on the screen. Almost all computer games can be played with sound turned off—the terms video game and computer game are almost synonymous. An example of a computer game that is based on sound—although it has enough visual clues to be played without sound—is Loom** by LucasArts. The player has to memorize and play magic melodies to go through the game. Sounds in some simulator games, such as SimCity**, may help the player identify an event even when it is outside the current field of view.

To break the tradition, a simple nonvideo computer game, called Audio Search, was written. The game provides all the information necessary to play the game by using sound. (The game can be played with eyes shut.) The objective of the game is to capture creatures that make noises. The player moves around the game space using the keyboard and tries to approach the targets that can be heard but cannot be seen. In the prototype implementation, the computer generated stereo sound that was based on the position of the player and targets. As the player moves around the space, changes in direction and proxim-

ity of the targets are heard. To capture a creature, the player has to come close to it.

The human auditory system has an excellent facility for resolving direction. The human ear can detect the difference in both intensity and phase of the sound coming from the left and right directions. Although, in a static position, it is impossible to clearly distinguish between sounds coming from the front and back, humans can easily compensate for that by turning their heads or moving laterally. When a computer accurately generates signals in the game for two or more speakers in the room, a player can readily find the position of virtual sound objects.

People who played the game found it engaging and relaxing—unlike many computer games. The game is an extreme case of complete visual interface denial and points to a way of enriching video games in general. Future computer games may use sound to add information that cannot be seen on the display.

Conclusion

Audio has been largely overlooked as a device-to-device communications medium. Although it will not generally replace IR or RF, the use of human-audible frequencies is either a replacement or auxiliary channel of communication. Also, creative use of sound in device-to-human communication may substantially improve usability, especially in public settings. Alarms can be both efficient in alerting the owner and less obtrusive and irritating. Embedding environmental sensors into communication devices will enable the development of more efficient alert strategies.

Many extreme high-performance activities such as batting a baseball require complete, undivided visual attention. Sound can be used as a communication medium in these situations.[43] Hearing may deliver nearly as much information as vision. An audio user interface implemented in learning tools[44,45] or computer games may improve understanding of the subject and augment and expand the perceptual experience.

Acknowledgments

The authors would like to acknowledge IBM and the News in the Future and Things That Think research consortia at MIT for sponsoring this work in part.

**Trademark or registered trademark of Fred D. Hinger, Parker Brothers, Inc., Microsoft Corporation, LucasArts, or Electronic Arts.

Cited references

1. S. Bly, *Sound and Computer Information Presentation*, unpublished doctoral thesis (UCRL-53282), Lawrence Livermore National Laboratory and University of California, Davis, CA (1982).
2. W. Buxton, S. A. Bly, S. P. Frysinger, D. L. Lunney, D. L. Mansur, J. J. Mezrich, and R. C. Morrison, "Communicating with Sound," panel session, *CHI'85 Conference on Human Factors in Computing Systems*, San Francisco (April 14–18, 1982), pp. 115–119.
3. W. Gaver, "Auditory Interfaces." *Handbook of Human-Computer Interaction*, *2nd Edition*, M. G. Helander, T. K. Landauer, and P. Prabhu, Editors, Elsevier Science, Amsterdam (1997).
4. R. Patterson, J. Edworthy, M. Shailer, M. Lower, and P. Wheeler, "Alarm Sounds for Medical Equipment in Intensive Care Areas and Operating Theatres," *Report AC598*, University of Southampton Auditory Communication and Hearing Unit, Southampton, UK (1986).
5. *Encyclopedia Britannica*, http://www.eb.com.
6. J. Paradiso, "The Interactive Balloon: Sensing, Actuation, and Behavior in a Common Object," *IBM Systems Journal* **35**, Nos. 3&4, 473–487 (1996).
7. F. J. Langdon, "Noise and Annoyance," *The Noise Handbook*, W. Tempest, Editor, Academic Press, New York (1985), pp. 143–176.
8. U. Landström, P. Löfstedt, E. Åkerlund, A. Kjellberg, and P. Wide, "Noise and Annoyance in Working Environments," *Environment International* **16**, 555–559 (1990).
9. D. R. Davies and D. M. W. A. Ainsworth, "Noise and Communication," *The Noise Handbook*, W. Tempest, Editor, Academic Press, New York (1985), pp. 69–86.
10. Jones, "Noise and Efficiency," *The Noise Handbook*, W. Tempest, Editor, Academic Press, New York (1985), pp. 87–141.
11. *Occupational Safety and Health Studies* **39**, No. 125, Part II, U.S. Department of Labor, Washington, DC (1974).
12. U. Landström, "Noise and Fatigue in Working Environments," *Environment International* **16**, 471–476 (1990).
13. R. N. Slarve and D. L. Johnson, "Human Whole-Body Exposure to Infrasound," *Aviation, Space, and Environmental Medicine*, 428–431 (April 1975).
14. W. Tempest, "Noise in Industry," *The Noise Handbook*, W. Tempest, Editor, Academic Press, New York (1985), pp. 179–194.
15. G. M. Jackson and H. G. Leventhall, "Noise in the Home," *The Noise Handbook*, W. Tempest, Editor, Academic Press, New York (1985), pp. 237–277.
16. M. E. Frerking, *Digital Signal Processing in Communication Systems* (1994).
17. I. N. Bronstein and K. A. Semendiav, *Spravochnik po Matematike*, Moscow (1962).
18. *Signal Processing Methods for Audio, Images and Telecommunications*, P. M. Clarkson and H. Stark, Editors (1995).
19. W. Bender, D. Gruhl, N. Morimoto, "Techniques for Data Hiding," *IBM Systems Journal* **35**, Nos. 3&4, 313–336 (1996).
20. V. Gerasimov, *Things That Talk*, master's thesis, MIT, Media Arts and Sciences, Cambridge, MA (1997).
21. N. R. Shanbhag, *Pipelined Adaptive Digital Filters* (1994).
22. H. Ishii, C. Wisneski, J. Orbanes, B. Chun, and J. Paradiso, "PingPongPlus: Design of an Athletic-Tangible Interface for

Computer-Supported Cooperative Play," *CHI'99 Conference on Human Factors in Computing Systems*, Pittsburgh, PA (May 15–20, 1999), pp. 394–401.
23. E. Mynatt, M. Back, and R. Want, "Designing Audio Aura," *Proceedings of CHI'98* (1998), pp. 566–573.
24. C. Schmandt, N. Marmasse, S. Marti, N. Shawhney, and S. Wheeler, "Everywhere Messaging," *IBM Systems Journal* **39**, Nos. 3&4, 660–677 (2000, this issue).
25. T. L. Bonebright, N. E. Miner, T. E. Goldsmith, and T. P. Caudell, "Data Collection and Analysis Techniques for Evaluating the Perceptual Qualities of Auditory Stimuli," *ICAD'98 5th International Conference on Auditory Display*, British Computer Society, Glasgow, UK (November 1–4, 1998).
26. B. L. Vercoe, "Computational Auditory Pathways to Music Understanding," *Perception and Cognition of Music*, I. Deliège and J. Slobodo, Editors, Psychology Press, London (1997), pp. 307–326.
27. B. C. J. Moore, *Psychology of Hearing*, *4th Edition*, Academic Press, San Diego, CA (1997).
28. J. Edworthy, S. Loxley, and I. Dennis, "Improving Auditory Warning Design: Relationship Between Warning Sound Parameters and Perceived Urgency," *Human Factors* **33**, No. 2, 205–231 (1991).
29. C. Swift, I. Flindell, and C. Rice, "Annoyance and Impulsivity Judgments of Environmental Noises," *Proceedings of the Institute of Acoustics 1989 Spring Conference* **11**, Part 5, 551–555 (1989).
30. E. C. Cherry, *Journal of the Acoustical Society of America* **25**, No. 5, 975–979 (1953).
31. N. Moray, "Attention and Dichotic Listening: Affective Cues and the Influence of Instructions," *Quarterly Journal of Experimental Psychology* **11**, 56–60 (1959).
32. N. Sawhney and C. Schmandt, "Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging," *CHI'99 Conference on Human Factors in Computing Systems*, Pittsburgh (May 15–20, 1999), pp. 96–103.
33. J. Duncan, "Divided Attention: The Whole Is More than the Sum of the Parts," *Journal of Experimental Psychology: Human Perception and Performance* **5**, No. 2, 216–228 (1979).
34. L. H. Schaffer, "Multiple Attention in Continuous Verbal Tasks," *Attention and Performance V*, P. M. A. Rabbit and S. Dornic, Editors, Academic Press, New York (1975).
35. C. D. Wickens, *Engineering Psychology and Human Performance*, Harper Collins, New York (1992).
36. A. Allport, "Visual Attention," *Foundations of Cognitive Science*, M. Posner, Editor, MIT Press, Cambridge, MA (1989).
37. R. A. Barr, "How Do We Face Our Attention?" *American Journal of Psychology* **94**, No. 4, 591–603 (1981).
38. Y. Miyata and D. Norman, "Psychological Issues in Support of Multiple Activities," *User Centered System Design*, Norman and Draper, Editors, Lawrence Erlbaum Associates, Hillsdale, NJ (1986), pp. 268–270.
39. V. Gerasimov and W. Bender, *Swings That Think*, http://vadim.www.media.mit.edu/stt/bat.html (1998).
40. S. K. Card, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ (1983).
41. P. Kirkpatrick, "Batting the Ball," *American Journal of Physics* **31**, 601–613 (1963).
42. R. Adair, *The Physics of Baseball*, Harper & Row Publishers, New York (1990).
43. E. Mynatt, M. Back, and R. Want, "Designing Audio Aura," *ACMCHI 98 Proceedings* (1998), p. 566.
44. M. Cooley, "Sound + Image in Computer-Based Design: Learning from Sound in the Arts," *ICAD'98 5th International Conference on Auditory Display*, British Computer Society, Glasgow, UK (November 1–4, 1998).
45. B. Shneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (1997).

Accepted for publication May 5, 2000.

Vadim Gerasimov *MIT Media Laboratory, 20 Ames Street, Cambridge, Massachusetts 02139-4307 (electronic mail: vadim@media.mit.edu).* Mr. Gerasimov is a Ph.D. candidate at the MIT Media Laboratory. He codeveloped the game Tetris at age 16, then received his undergraduate degree in applied mathematics from the Moscow State University and his M.S. degree in Media Arts and Sciences from the Massachusetts Institute of Technology. He has been studying and working at the Media Laboratory since 1994.

Walter Bender *MIT Media Laboratory, 20 Ames Street, Cambridge, Massachusetts 02139-4307 (electronic mail: walter@media.mit.edu).* Mr. Bender is a senior scientist at the MIT Media Laboratory and principal investigator of the laboratory's News in the Future consortium. He received the B.A. degree from Harvard University in 1977 and joined the Architecture Machine Group at MIT in 1978. He received the M.S. degree from MIT in 1980. Mr. Bender is a founding member of the Media Laboratory.