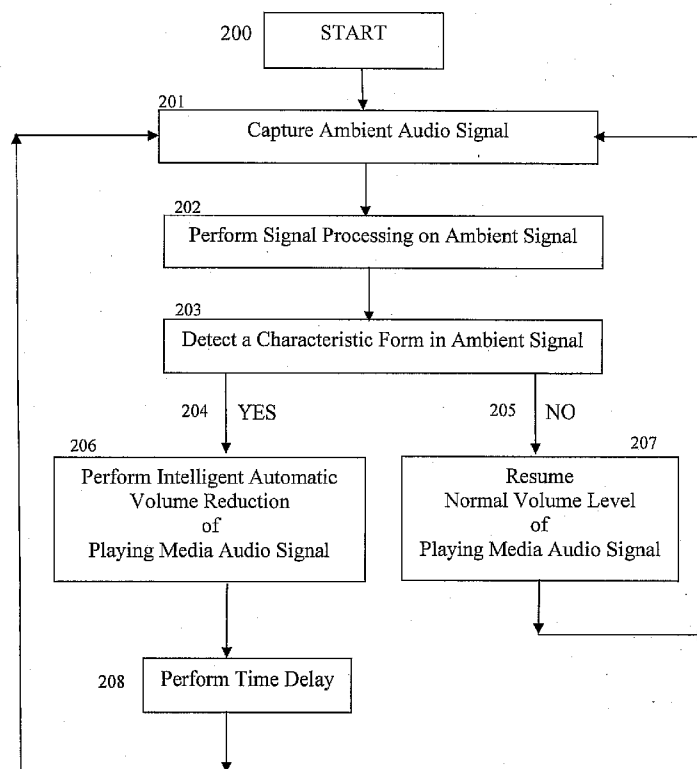US 20070189544A1

(19) **United States**

(12) **Patent Application Publication** (10) Pub. No.: US 2007/0189544 A1

Rosenberg (43) **Pub. Date:** **Aug. 16, 2007**

(54) **AMBIENT SOUND RESPONSIVE MEDIA PLAYER**

(75) Inventor: **Louis B. Rosenberg**, Pismo Beach, CA (US)

Correspondence Address:
SINSHEIMER JUHNKE LEBENS & MCIVOR, LLP
1010 PEACH STREET
P.O. BOX 31
SAN LUIS OBISPO, CA 93406 (US)

(73) Assignee: **Outland Research, LLC**, Pismo Beach, CA (US)

(21) Appl. No.: **11/697,706**

(22) Filed: **Apr. 7, 2007**

Related U.S. Application Data

(63) Continuation-in-part of application No. 11/267,079, filed on Nov. 3, 2005.
Continuation-in-part of application No. 11/223,368, filed on Sep. 9, 2005.
Continuation-in-part of application No. 11/610,615, filed on Dec. 14, 2006.

(60) Provisional application No. 60/665,291, filed on Mar. 26, 2005. Provisional application No. 60/648,197, filed on Jan. 27, 2005. Provisional application No. 60/644,417, filed on Jan. 15, 2005. Provisional application No. 60/793,214, filed on Apr. 19, 2006. Provisional application No. 60/841,990, filed on Aug. 31, 2006.

Publication Classification

(51) **Int. Cl.**
H03G 3/20 (2006.01)
H03G 3/00 (2006.01)

(52) **U.S. Cl.** **381/57**; 381/104

(57) **ABSTRACT**

Some embodiments of the present invention provide a method of adjusting an output of a media player comprising capturing an ambient audio signal; processing the ambient audio signal to determine whether one or more characteristic forms are present within the ambient audio signal; and reducing an output of a media player from a first volume to a second volume if the one or more characteristic forms are present within the ambient audio signal. The characteristic forms may be, for example, a name or personal identifier of a user of the media player, the voice of a user of the media player, or an alarm or siren.
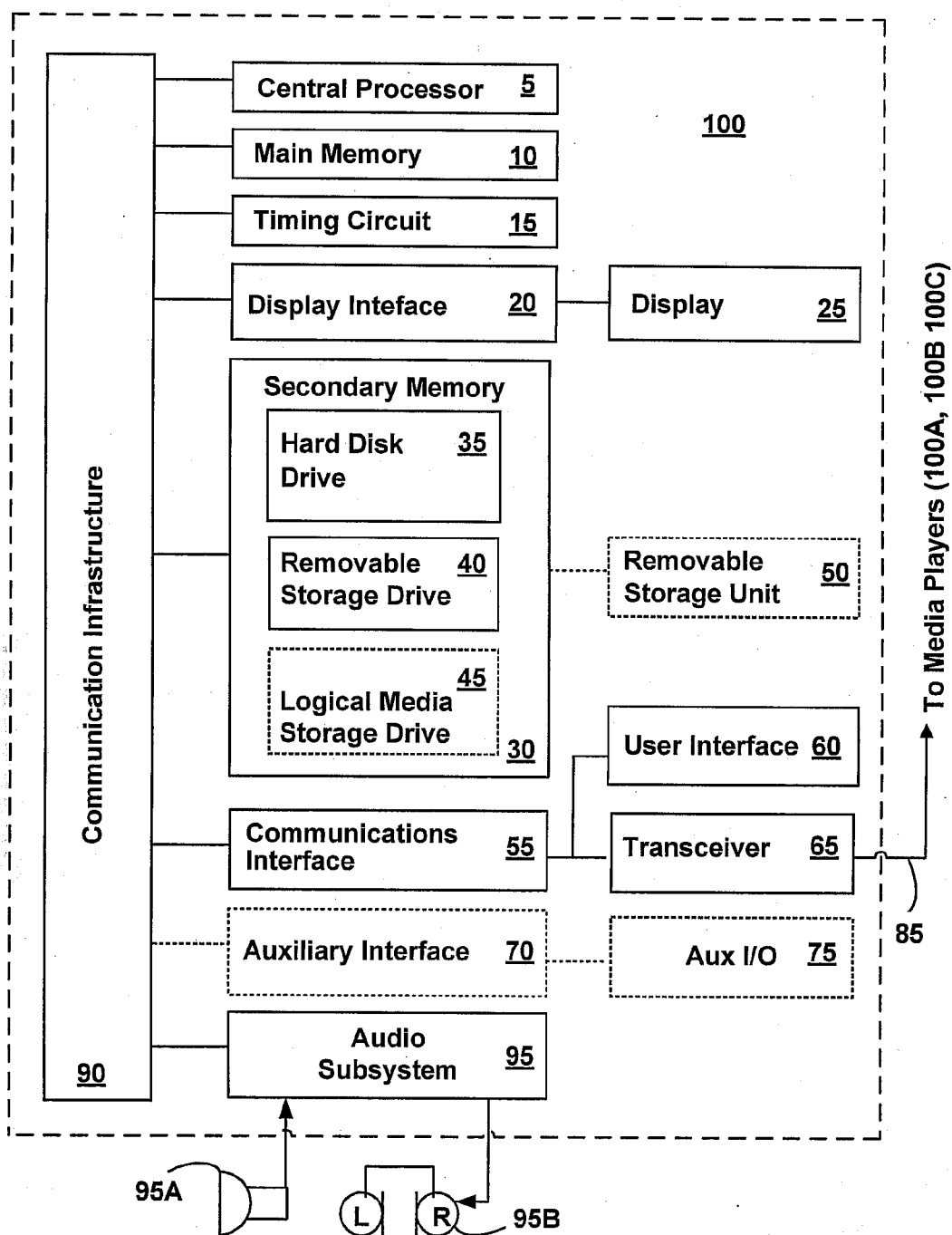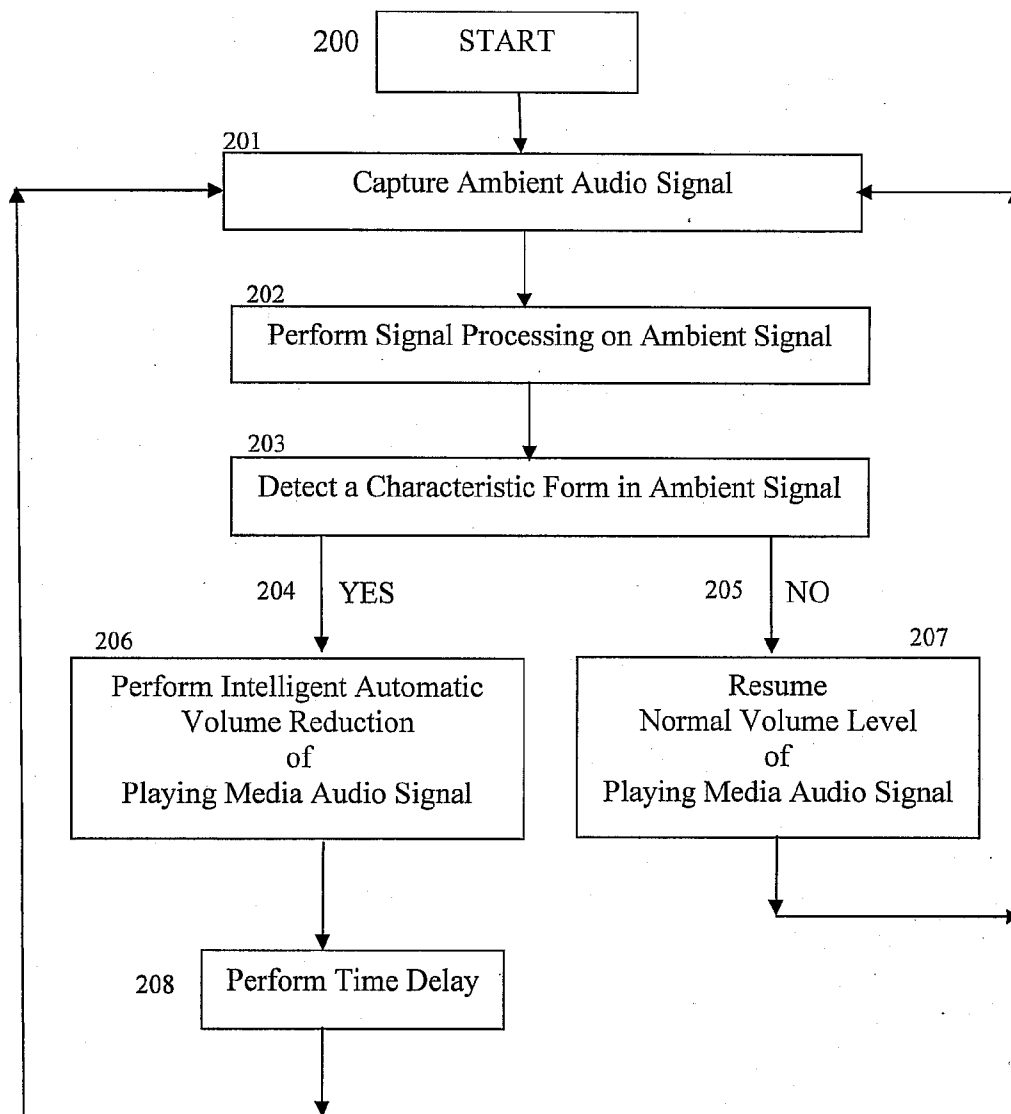
Flow Chart of Program Flow

Communication Infrastructure

90

Central Processor   **5**

Main Memory   **10**

Timing Circuit   **15**

Display Inteface   **20**   —   Display   **25**

Secondary Memory

Hard Disk Drive   **35**

Removable Storage Drive   **40**   ······   Removable Storage Unit   **50**

Logical Media Storage Drive   **45**

30

User Interface   **60**

Communications Interface   **55**   Transceiver   **65**

Auxiliary Interface   **70**   ······   Aux I/O   **75**

Audio Subsystem   **95**

100

To Media Players (100A, 100B 100C)

85

95A

L   R   95B

FIG. 1

FIGURE 2: Flow Chart of Program Flow

200 | START

201
Capture Ambient Audio Signal

202
Perform Signal Processing on Ambient Signal

203
Detect a Characteristic Form in Ambient Signal

204 | YES

205 | NO

206
Perform Intelligent Automatic
Volume Reduction
of
Playing Media Audio Signal

207
Resume
Normal Volume Level
of
Playing Media Audio Signal

208 | Perform Time Delay

AMBIENT SOUND RESPONSIVE MEDIA PLAYER

[0001] This application is a continuation-in-part of U.S. patent application Ser. No. 11/267,079 filed Nov. 3, 2005, which claims the benefit of U.S. Provisional Patent Application No. 60/665,291 filed Mar. 26, 2005 and U.S. Provisional Application No. 60/648,197 filed Jan. 27, 2005, all of which are incorporated in their entirety herein by reference.

[0002] This application is also a continuation-in-part of U.S. patent application Ser. No. 11/223,368 filed Sep. 9, 2005, which claims the benefit of U.S. Provisional Patent Application No. 60/644,417 filed Jan. 15, 2005, both of which are incorporated in their entirety herein by reference.

[0003] This application is also a continuation-in-part of U.S. patent application Ser. No. 11/610,615 filed Dec. 14, 2006, which claims the benefit of U.S. Provisional Patent Application No. 60/793,214 filed Apr. 19, 2006, both of which are incorporated in their entirety herein by reference.

[0004] This application also claims the benefit of U.S. Provisional Patent Application No. 60/841,990 filed Aug. 31, 2006, which is incorporated in its entirety herein by reference.

BACKGROUND OF THE INVENTION

[0005] 1. Field of the Invention

[0006] The present invention relates generally to media players, and more specifically to responsive media players.

[0007] 2. Discussion of the Related Art

[0008] Portable media players have become popular personal entertainment devices due to their highly portable nature, their ability to provide accessibility to a large library of stored media files, and interconnectivity with existing computer networks, for example the Internet. The accessibility and simplicity in downloading music and other electronic media continues to fuel the popularity of these devices as is exemplified by Apple Computer, Inc.'s highly successful iPod™ portable media player. Other manufacturers have competing Media Players offering various functionalities and file playing compatibilities in an effort to differentiate their products in the marketplace.

[0009] As discussed in U.S. Patent Application No. 2004/0224638 A1, which is herein incorporated by reference in its entirety, an increasing number of consumer products are incorporating circuitry to play musical media files and other electronic media. For example, many portable electronic devices such as cellular telephones and personal digital assistants (PDAs) include the ability to play electronic musical media in many of the most commonly available file formats including MP3, AVI, WAV, MPG, QT, WMA, AIFF, AU, RAM, RA, MOV, MIDI, etc. With a wide variety of devices and file formats emerging, it is expected that in the near future a large segment of the population will have upon their person an electronic device with the ability to access music files from a library of media files in local memory and/or over a computer network, and play those music files at will. Such users generally wear headphones to experience music content in a personalized high fidelity manner.

[0010] Because most users of portable media players generally wear headphones to play music directly into their ears, users experience the beneficial effect of separating themselves from the noises of daily life, providing a serene audio environment of personally played music. Unfortunately, users often miss important sound events within the real world when listening to music through headphones of a portable media player. For example, another person might be talking to the media player user but because of the music playing through their headphones, the user is unable to hear the fact that they have been verbally addressed. Similarly, a siren or alarm may sound in the environment of a headphone-wearing media player user, but they may not hear the warning sound effectively, thus creating a dangerous situation for the user. Finally, a headphone-wearing media player user may try to talk to someone else within their immediate environment, but because they cannot hear their own voice, they may find themselves talking substantially too loud for the current situation. This may create an embarrassing situation for the user.

SUMMARY OF THE INVENTION

[0011] Several embodiments of the invention advantageously address the needs above as well as other needs by providing a media player that is responsive to ambient sound.

[0012] In some embodiments, the invention can be characterized as a method of adjusting an output of a media player comprising capturing an ambient audio signal; processing the ambient audio signal to determine whether one or more characteristic forms are present within the ambient audio signal; and reducing an output of a media player from a first volume to a second volume if the one or more characteristic forms are present within the ambient audio signal.

[0013] In some embodiments, the invention can be characterized as a method of adjusting an output of a media player comprising capturing an ambient audio signal; processing the ambient audio signal to determine whether one or more characteristic forms are present within the ambient audio signal; and mixing at least a portion of the ambient audio signal with a first output of a media player to generate a second output of the media player if the one or more characteristic forms are present within the ambient audio signal.

[0014] In some embodiments, the invention can be characterized as a An apparatus for use in a media player comprising a microphone; and one or more processors adapted to: process an ambient audio signal received by the microphone to determine whether one or more characteristic forms are present within the ambient audio signal, and adjust an output of a media player if the one or more characteristic forms are present within the ambient audio signal.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The above and other aspects, features and advantages of several embodiments of the present invention will be more apparent from the following more particular description thereof, presented in conjunction with the following drawings.

[0016] FIG. 1 depicts a generalized block diagram of a media player in accordance with some embodiments of the present invention;

[0017] FIG. 2 depicts a flow chart of a process of an ambient sound responsive media player unit in accordance with some embodiments of the present invention.

[0018] Corresponding reference characters indicate corresponding components throughout the several views of the drawings. Skilled artisans will appreciate that elements in the figures are illustrated for simplicity and clarity and have not necessarily been drawn to scale. For example, the dimensions of some of the elements in the figures may be exaggerated relative to other elements to help to improve understanding of various embodiments of the present invention. Also, common but well-understood elements that are useful or necessary in a commercially feasible embodiment are often not depicted in order to facilitate a less obstructed view of these various embodiments of the present invention.

DETAILED DESCRIPTION

[0019] The following description is not to be taken in a limiting sense, but is made merely for the purpose of describing the general principles of exemplary embodiments. The scope of the invention should be determined with reference to the claims.

[0020] There currently exists a need to provide intelligent volume control of media content play through headphones (or other similar headsets and ear pieces), such that a headphone wearing media player user may more easily hear when he or she is verbally addressed, when an alarm or siren sounds within his or her environment, and/or when he or she is speaking aloud.

[0021] This disclosure addresses the deficiencies of the relevant art and provides exemplary systematic, methodic and computer program product embodiments which provides an ambient sound responsive portable media player that enables a media player to intelligently adjust and/or vary the playing volume of a musical media file to a user based at least in part upon detected sounds from the ambient environment of the user. More specifically, the present invention provides an ambient sound responsive media player in which the musical sounds played to a user through the headphones of a media player are moderated based at least in part upon detected ambient sounds from within the user's local environment. The system works by incorporating a microphone in the media player system, the microphone configured to detect sounds from the ambient environment of the media player user as the user listens to music through headphones. The system further includes a processor for making volume adjustments to playing media content based at least in part upon detected ambient audio signals from said microphone. The processor of the present invention may be configured through hardware and software components to perform one or more of the following functions:

[0022] (A) Name responsive volume reduction. This is a function in which the playing volume of currently playing media file is automatically reduced by the processor for a period of time in response to the media players user's name being detected as verbal content within the audio signal captured from the ambient environment. In this way if another person calls the user's name, presumably to talk to that user, the media player is responsive to automatically reduce the playing volume of media content to that user.

[0023] (B) User voice responsive volume reduction. This is a function in which the playing volume of a currently playing media file is automatically reduced by the processor for a period of time in response to the media players user's

own voice being detected within the audio signal captured from the ambient environment. In this way if the media player user begins speaking aloud into the ambient environment, the media player is automatically responsive by reducing the playing volume of media content to that user so the user can more easily hear himself talk. This prevents the user from speaking too loudly into the ambient environment and embarrassing himself.

[0024] (C) Alarm sound volume reduction. This is a function in which the playing volume of currently playing media file is automatically reduced by the processor for a period of time in response to an alarm sound or siren sound being detected as within the audio signal captured from the ambient environment. In this way if an alarm or siren sounds within the user's local environment, presumably because there is a danger to be alerted to, the media player is responsive to automatically reduce the playing volume of media content to that user. In this way the user will more easily hear the alarm sound.

[0025] In some embodiments of the present invention, the media player is operative to mix musical audio content derived from a stored media file with ambient audio content captured from a microphone local to the user. In this way the user can listen to musical media content in audio combination with ambient audio signals from the local environment. While such a function may enable a user to more easily hear sounds such as other speaking users, the user's own voice, and/or alarms and sirens, such a mixed audio signal may be unpleasant during times when such events are not occurring. Thus some embodiments of the present invention include an inventive method in which the relative volume balance of the mixed signal (i.e. the relative volume of the musical media content and the ambient microphone content) are selectively adjusted in response to detected ambient audio events. More specifically, the relative volume of the microphone content is automatically increased with respect to the musical media content within the mixed audio signal in response to detected ambient audio events such as (A) detection of the media player's name being uttered within the ambient audio signal, (B) detection of the media player's own voice within the ambient audio signal, and/or (C) detection of an alarm or siren sound present within the ambient audio signal.

[0026] The present invention provides a system, method and computer program product which enables a media player to intelligently adjust and/or vary the playing volume of a musical media file to a user based at least in part upon detected sounds from the ambient environment of the user. More specifically, the present invention provides an ambient sound responsive media player in which the musical sounds played to a user through the headphones of a media player are moderated based at least in part upon detected ambient sounds from within the user's local environment. In some embodiments ambient sounds from the local environment are selectively mixed with digital media sounds such that their relative volumes are adjusted based at least in part upon detected ambient sound events within the user's local environment. Where necessary, computer programs, routines and algorithms are envisioned to be programmed in a high level language, for example Java™ C++, C, C#, or Visual Basic™.

[0027] Referring to FIG. 1, a generalized block diagram of a media player 100 is depicted. The media player 100

includes a communications infrastructure **90** used to transfer data, memory addresses where data items are to be found and control signals among the various components and subsystems of the media player **100**.

[0028] A central processor **5** is provided to interpret and execute logical instructions stored in the main memory **10**. The main memory **10** is the primary general purpose storage area for instructions and data to be processed by the central processor **5**. The main memory **10** is used in its broadest sense and includes RAM, EEPROM and ROM. A timing circuit **15** is provided to coordinate activities within the media player **100**. The central processor **5**, main memory **10** and timing circuit **15** are directly coupled to the communications infrastructure **90**.

[0029] A display interface **20** is provided to drive a display **25** associated with the media player **100**. The display interface **20** is electrically coupled to the communications infrastructure **90** and provides signals to the display **25** for visually outputting both graphics and alphanumeric characters. The display interface **20** may include a dedicated graphics processor and memory to support the displaying of graphics intensive media. The display **25** may be of any type (e.g., cathode ray tube, gas plasma) but in most circumstances will usually be a solid state device such as liquid crystal display.

[0030] A secondary memory subsystem **30** is provided which houses retrievable storage units such as a hard disk drive **35**, a removable storage drive **40**, an optional a logical media storage drive **45** and an optional removal storage unit **50**.

[0031] The removable storage drive **40** may be a replaceable hard drive, optical media storage drive or a solid state flash RAM device. The logical media storage drive **45** may be flash RAM device, EEPROM encoded with playable media, or optical storage media (CD, DVD). The removable storage unit **50** may be logical, optical or of an electromechanical (hard disk) design.

[0032] A communications interface **55** subsystem is provided which allows for standardized electrical connection of peripheral devices to the communications infrastructure **90** including, serial, parallel, USB, and Firewire connectivity. For example, a user interface **60** and a transceiver **65** are electrically coupled to the communications infrastructure **90** via the communications interface **55**. For purposes of this disclosure, the term user interface **60** includes the hardware and operating software by which a user executes procedures on the media player **100** and the means by which the media player conveys information to the user.

[0033] The user interface **60** employed on the media play **100** includes a pointing device (not shown) such as a mouse, thumbwheel or track ball, an optional touch screen (not shown); one or more pushbuttons (not shown); one or more sliding or circular rheostat controls (not shown), one or more switches (not shown), and one or more tactile feedback units (not shown); One skilled in the relevant art will appreciate that the user interface devices which are not shown are well known and understood.

[0034] To accommodate non-standardized communications interfaces (i.e., proprietary), an optional separate auxiliary interface **70** and auxiliary I/O port **75** are provided to couple proprietary peripheral devices to the communications infrastructure **90**.

[0035] The transceiver **65** facilitates the remote exchange of data and synchronizing signals between and among the various media players **100A**, **100B**, **100C** in processing communications with **85** with this media player **100**.

[0036] The transceiver **65** is envisioned to be of a radio frequency type normally associated with computer networks for example, wireless computer networks based on Blue-Tooth™ or the various IEEE standards 802.11.sub.x., where x denotes the various present and evolving wireless computing standards.

[0037] Alternately, digital cellular communications formats compatible with for example GSM, 3G and evolving cellular communications standards. Both peer-to-peer (PPP) and client-server models are envisioned for implementation of the invention. In a third alternative embodiment, the transceiver **65** may include hybrids of computer communications standards, cellular standards and evolving satellite radio standards.

[0038] Lastly, an audio subsystem **95** is provided and electrically coupled to the communications infrastructure **90**. The audio subsystem is configured for the playback and recording of digital media, for example, multi or multimedia encoded in any of the exemplary formats MP3, AVI, WAV, MPG, QT, WMA, AIFF, AU, RAM, RA, MOV, MIDI, etc.

[0039] The audio subsystem includes a microphone **95A** which is used for the detection of sound signals from the user's local ambient environment. The microphone **95A** may be incorporated within the casing of the portable media player or may be remotely located elsewhere upon the body of the user and is connected to the media player by a wired or wireless link. Ambient sound signals from microphone **95A** are generally captured as analog audio signals and converted to digital form by an analog to digital converter or other similar component and/or process. A digital signal is thereby provided to the processor of the media player, the digital signal representing the ambient audio content captured by microphone **95A**. In some embodiments the microphone **95A** is local to the headphones or other head-worn component of the user. In some embodiments the microphone is interfaced to the media player by a Bluetooth communication link. In some embodiments the microphone comprises a plurality of microphone elements.

[0040] The audio subsystem also includes headphones (or other similar personalized audio presentation units that display audio content to the ears of a user) **95B**. The headphones may be connected by wired or wireless connections. In some embodiments the headphones are interfaced to the media player by a Bluetooth communication link.

[0041] As referred to in this specification, "media items" refers to video, audio, streaming and any combination thereof. In addition, the audio subsystem is envisioned to optionally include features such as graphic equalization, volume, balance, fading, base and treble controls, surround sound emulation, and noise reduction. One skilled in the relevant art will appreciate that the above cited list of file formats is not intended to be all inclusive.

[0042] The media player **100** includes an operating system, the necessary hardware and software drivers necessary to fully utilize the devices coupled to the communications infrastructure **90**, media playback and recording applications and at least one ambient sound responsive volume adjust-

ment program operatively loaded into main memory **10**. Optionally, the media player **100** is envisioned to include at least one remote authentication application, one or more cryptography applications capable of performing symmetric and asymmetric cryptographic functions, and secure messaging software. Optionally, the media player **100** may be disposed in a portable form factor to be carried by a user.

[0043] Referring to FIG. **2**, shown is a flow chart of a process of an ambient sound responsive media player unit in accordance with some embodiments of the present invention. The program flow shown would generally be performed in parallel with other processes performed by the media player, including processes that select and/or play media items by accessing media content from memory and outputting an audio representation of such media content through headphones and/or other similar audio presentation hardware. The program flow shown would generally be performed, at least in part, by routines running upon a processor of the portable media player. The program flow shown is generally performed, at least in part, by at least a portion of at least one ambient sound responsive volume adjustment program operatively loaded into main memory **10**. In the particular embodiment shown herein, the entire program flow shown is performed by the at least one ambient sound responsive volume adjustment program operatively loaded into main memory **10**. At the time in which the program flow begins, the media player has already selected and begun to play a media file through a separate process (not shown).

[0044] The program flow of FIG. **2** begins at step **200**, generally in response to a function call or other programming flow construct. Once started, the program flow performs a continuous loop until terminated. The continuous loop includes a number of steps which may be performed in a variety of orders. In the particular flow shown in FIG. **2**, the first step in the continuous loop is step **201** wherein ambient audio signals are captured through microphone **95A**. This ambient audio signals are generally captured as analog signals from the microphone element and then are digitized through an analog to digital conversion process. In addition, noise reduction, filtering, and/or other commonly known signal processing steps may be performed upon the ambient signal. The ambient audio signals, once converted to a final digital form, are generally stored in a temporary local memory of the portable media player. It should be noted that this ambient audio signal capture step **201** may be performed by a separate process that runs in parallel with the program flow of FIG. **2**. This separate process may, for example, store digitized ambient audio signal into a shared memory space that is accessible by the steps of this program flow.

[0045] The process then proceeds to step **202** wherein additional signal processing is performed on the captured ambient signal. This signal processing may include sound recognition processing, speech recognition processing, and/ or vocal identity recognition processing steps and/or substeps. Because sound recognition, speech recognition, and/ or vocal identity recognition processes are known to the prior art the specifics of such processes will not be described in detail herein. For example, U.S. Pat. No. 4,054,749 and U.S. Pat. No. 6,298,323, each of which are hereby incorporated by reference, both disclose methods and apparatus for voice identity recognition wherein a particular user's voice

may be identified as being present within an audio signal within certain accuracy limits. Similarly, U.S. Pat. No. 6,804,643, which is hereby incorporated by reference, discloses a speech recognition system in which particular verbal utterances may be identified from within an audio signal, the particular verbal utterances including particular words, phrases, names, and other verbal constructs. Similarly, other pieces of art disclose methods and systems by which particular non-verbal sounds may be identified within an ambient sound signal. One example of such sound recognition methods is disclosed in HABITAT TELEMONI-TORING SYSTEM BASED ON THE SOUND SURVEIL-LANCE by Castelli, Vacher, Istrate, Besacier, and Sérignat which is hereby incorporated by reference. Another example of such sound recognition methods is disclosed in a 1999 doctoral dissertation from MIT by Keith Dana Martin entitled *Sound-Source Recognition: A Theory and Computational Model* which is hereby incorporated by reference. Another example of such sound recognition methods is disclosed by Michael Casey in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOL-OGY, VOL. 11, NO. 6, JUNE 2001 in a paper entitled, *MPEG-7 Sound-Recognition Tools* which is hereby incorporated by reference. In such papers it is explained that recent advances in pattern recognition methodologies make the automatic identification of characteristic environmental sounds, animal sounds, non-verbal human utterances, and other non-verbal environmental sounds possible. Using such techniques, for example, alarm sound and/or siren sounds may be identified from within an ambient audio signal.

[0046] Thus by using the prior art methods of speech recognition, voice identity recognition, and environmental sound identification, the ambient sound signal captured by microphone **95A** and stored in local memory, may be processed such that (A) the utterance of the media player user's name may be identified if substantially present within the captured ambient audio signal, (B) the unique voice of the media player user may be identified if substantially present within the captured ambient audio signal, and/or (C) the sound of an alarm and/or siren and/or other similar emergency related alert sound may be identified if substantially present within the captured ambient audio signal. To perform such identifications, processing is performed in step **202**. Note—in general this step is performed upon a certain time-sample's worth of ambient audio signal during each loop of the program flow. Also in general the time-samples generally proceed as overlapping time windows with each loop of the program flow.

[0047] The process then proceeds to step **203** wherein a set of conditional routines are performed based upon whether or not a characteristic form (e.g. a signal conforming to A, B, or C above) is identified as present within the ambient signal. A characteristic form is a sound or signal that when detected by the media player will cause an audible adjustment to the output of the media player such that the user will be enabled to better hear ambient sounds. Thus in step **203**, conditional routines are performed based upon whether or not the ambient signal has been identified to contain one or more of (A) an verbal utterance of the media player user's name by another user, (B) a verbal utterance of any kind from the media player user himself or herself, or (C) the non-verbal sound of an alarm and/or siren and/or other similar emergency related alert. If one or more of such characteristic forms are present within the ambient audio

signal, the process proceeds along arrow **204** to step **206**. If not, the process proceeds along arrow **205** to step **207**. These two alternate paths are described as follows:

[0048] In the "yes" branch, the process proceeds along arrow **204** to step **206**. At step **206**, the routines of the present invention perform an Intelligent Automatic Volume Reduction routine in which the currently playing media audio signal is automatically reduced in volume so that the user can better hear the ambient sounds around him or her. This reduction in playing volume of the currently playing media audio signal may be performed abruptly. Alternately, the volume reduction may be performed gradually over a period of time. In general the period of time is short, for example 1500 milliseconds. The volume reduction may be reduced by a fixed amount, for example to 65% of the nominal volume level set by the user, or may be reduced by an amount that is dependent upon the volume level of the identified characteristic ambient sound that triggered the reduction. In some embodiments the user may set a configuration parameter that indicates the desired volume reduction level upon the identification of a characteristic ambient sound event. The volume reduction level may be set as a percentage of the nominal volume level at which the user is currently listening. Alternately the volume reduction level may be set to a defined low value on the absolute volume scale of the unit (for example to a value of 2 out of a scale of 10). Once this automatic volume reduction step is complete, the process flows to step **208** which will be described further down.

[0049] In the "no" branch, the process proceeds along arrow **205** to step **207**. At step **207**, the routines of the present invention will resume the playing volume of the currently playing media content to (or approximately to) the normal (nominal) playing volume. By nominal playing volume, it means the volume it would be playing as if it had not been reduced previously by the Intelligent Automatic Volume Reduction routines. Thus if the volume had been reduced previously by the Intelligent Automatic Volume Reduction routines of step **206**, then step **207** will return the volume substantially to its normal volume level. This may happen abruptly. Alternately the return of the volume to the nominal level may be performed using gradual volume adjustment routine that gradually resumes the volume over a period of time. In some embodiments the period of time is on the order of 1500 to 3000 milliseconds. Such a time period is short enough that the event seems quick to the user, but long enough that it is not jarring. Note, if the volume was already at the nominal level when step **207** is performed, then step **207** does not perform any substantial change in volume level. Once step **207** completes, the process loops back to the beginning, returning to step **201**. In this way the routine continues to capture and process a steady stream of ambient audio signals and responds accordingly with volume reduction and/or resumption.

[0050] If a characteristic form was identified within the ambient signal in step **203** and the playing volume of the media content was reduced at step **206**, the process then proceeds to step **208** wherein a time delay may be optionally performed. The time delay is performed to ensure that the volume reduction lasts for at least some amount of time beyond the identification of the characteristic form within the ambient signal. In general, this amount of time may be set by the user through a configuration process. This amount

of time may be, for example, 3 to 6 seconds. In this way if the routines of the present invention, for example, identify that somebody called the name of the media player user, the volume reduction does not just occur for a split second upon the identification, but lasts for a number of seconds thereafter. In this way the user may hear what is being said to him immediately after his or her name was called. In some embodiments the volume reduction lasts indefinitely, or until the user explicitly resumes normal volume by pressing a button or otherwise engaging the user interface upon his or her media player. The process then loops back to step **201**. In this way the routine of the present invention are configured to continually capture and process a steady stream of ambient audio signals and responds accordingly with volume reduction and/or resumption. In general the volume reductions linger for some time delay period after each identified characteristic form within the ambient signal. In some embodiments the duration of the time delay is dependent upon the type of characteristic form identified. For example, if the characteristic form is an alarm sound, the time delay may not last long beyond the cessation of the alarm sound, presumably because the emergency alert is over. Alternately, if the characteristic form is a vocal call of the user's name by another user, the time delay is set generally long enough to allow the user to hear what else the other user says after the name call.

[0051] In a unique embodiment, the time delay is set to last for as long as the user who called the media player user's name continues to speak. This is performed based upon the detected vocal identity of this other user. Thus if a first user calls the name of the media player user and then continues to speak, the routines of the present invention may be configured to perform an automatic volume reduction upon the detection of the name call as uttered by the first user and will maintain the volume reduction for at least as long as the first user's voice continues to be identified without a timegap of more than some threshold amount of time. The threshold is generally set such that if the first user speaks at a typical speaking pace, the volume reduction will be maintained until the first user finishes talking.

[0052] Additional Non-Verbal Ambient Sound Triggers: As described previously, the routines of the present invention may be configured to trigger the automatic volume reduction of playing media content on a media player in response to the detection of a characteristic non-verbal sound within the local environment such as the sound of an alarm and/or siren and/or other similar emergency alert captured by microphone **95A** of the system. In some embodiments of the present invention, the automatic volume reduction routines may be configured such that additional and/or alternate characteristic non-verbal sounds within the ambient environment may be detected and trigger the volume reduction. For example, common household sounds that a user may desire to attend to such as the sound of a doorbell ringing, a telephone ringing, or a baby crying may be employed as characteristic ambient sounds that trigger the automatic volume reduction routines and methods disclosed herein. In this way a user may be wearing a media player within his or her house and if the microphone on the media player captures a characteristic sound that is substantially similar to a doorbell ringing, a phone ringing, or a baby crying, the volume of the playing media content is automatically reduced for a period of time following the detected characteristic ambient sound event.

[0053] System Configuration: For embodiments of the present invention that trigger a volume reduction period based upon the detection of an utterance of the media player user's name within the ambient environment, the system is generally configured to identify one or more proper nouns that are relationally associated with the user and stored in memory as a digitized sample, an audio template, or some other stored representation that may be used for pattern matching or other speech recognition methods. For example, the user's name was Theodore, he may configure his media player to be responsive to utterances that are substantially similar to the verbal utterance "Theodore" or the verbal utterance "Theo" or the verbal utterance "Teddy" or the verbal utterance "Ted". In this way a single user may configure his or her media player to be volume-reduction responsive to verbal utterances of a plurality of proper nouns, i.e. personal identifiers, that are set in memory to be relationally associated with an automatic volume reduction process of the media player. The user may also configure the unit to be responsive to a first name, last name, and middle name, and/or any combination thereof. The user may also configure the unit to be responsive only to name utterances that exceed a certain volume threshold. In this way the unit may be less likely to get falsely triggered by name calls that may not be meant for the user even if they conform with a characteristic utterance associated with that user. In addition, the user may set his or her unit to be responsive to utterances that are nick-names or pen-names or user-names or even other words that are not necessarily names. This Theodore in the example above may set his unit to be responsive to the utterance "dog-boy". So long as his friends know to use the utterance "dog-boy" to get his attention, the configuration will work well for this user. In this way a user may set a particular word or phrase to be effectively a volume reduction password that his or her friends can use to get his or her attention. In general, setting a particular verbal utterance to be an identified volume reduction trigger utterance within the ambient environment, involves the user uttering the word or phrase to the media player during a configuration process. Alternate methods of configuring speech recognition systems known to the art may be used as well. In addition, one or more generic words commonly used to summon attention, such as, for example, "sir" or "help" or "excuse me," may be additionally optionally configured to also trigger the automated volume reduction methods if such words are captured in the ambient audio signal at a volume that exceeds a certain threshold.

[0054] Audio Mixing Embodiments: In some embodiments of the present invention, the media player is operative to mix musical audio content derived from a stored media file with ambient audio content captured from a microphone local to the user. The methods and apparatus used to mix two separate audio signals into a single audio stream that may be listened to by a user is well known in the art and will not be described in detail herein. Regardless of the method used, a single audio signal is presented to the user through the headphones or other similar sound display hardware, the signal audio signal including an audio combination of a musical media file accessed from a memory of the media player and an ambient audio signal derived from the signal captured by Microphone 95A. The relative volume of the two component audio signals as represented in the combined mix audio signal may be dependent at least in part upon a mixing balance setting supplied by the user through a user interface of the media player. In this way the user can listen to musical media content in audio combination with ambient audio signals from the local environment. It should be noted that the ambient audio signal content may be filtered or otherwise processed to extract extraneous noise and/or sound content that is outside certain magnitude and/or frequency limits or thresholds.

[0055] While such an inventive audio mixing function may enable a user to more easily hear sounds from within his or her natural surroundings in a controlled and settable audio combination with music that he or she is listening to (including ambient sounds such as other speaking users, the user's own voice, and/or alarms and sirens), such a mixed audio signal may be unpleasant during times when such events are not occurring. For example, the user may be constantly distracted by ambient environment sounds in the mixed audio signal that are not important, relevant, or valuable for him or her to attend to. Thus some embodiments of the present invention include a further inventive method in which the relative volume balance of the mixed signal (i.e. the relative volume of the musical media content and the ambient microphone content) are selectively adjusted in response to detected ambient audio events. More specifically, the relative volume of the microphone content is automatically increased with respect to the musical media content, for a period of time, in response to detected characteristic ambient audio events within the ambient audio signal stream. The detected characteristic ambient audio events may include, but are not limited to: (A) the detection of the media player's name being uttered within the ambient audio signal, (B) the detection of the media player's own voice within the ambient audio signal, and/or (C) detection of an alarm or siren sound present within the ambient audio signal.

[0056] In this way, a user may be listening to an audio signal that is a mixed audio combination of a musical media file and an ambient microphone signal, the relative volumes being such that the musical media file is substantially louder than the ambient microphone signal as presented within the mixed audio content. In response to the detection of a characteristic ambient audio event such as A, B, or C, above, the routines of the present invention are configured to adjust the relative volumes in the mixed audio signal for a period of time, the adjustment such that the representation of the ambient audio signal is made substantially louder relative to the musical media content. Thus if a third party calls the name of the user of the media player, upon detection of that name being uttered, the user is presented with an audio mix of musical media and microphone data such that the user can easily hear the ambient environment as mixed with the musical media. When the period of time is over the relative volume levels are automatically returned to their nominal relative volume (i.e. a nominal relative volume such that the musical media content is substantially louder than the microphone content).

[0057] It should also be noted that in some embodiments the nominal relative volume levels of the two signals may be set such that the volume of the ambient microphone content is substantially zero at times when an ambient audio event has not been detected. In this way the user only hears the musical content until and unless an ambient audio event is detected. In response to such a detected ambient audio event (for example an event such as A, B, or C above), the

automatic routines of the present invention adjust the relative volumes of the two signals such that the ambient environment microphone signal is no longer zero, instead being substantial with respect to the musical media content. In this way the pure musical content is played to the user until an ambient audio event is detected, then in response to the detected event a mixed audio signal is presented with both musical content and ambient audio content such that ambient audio content is clearly audible at a substantial relative volume. This change in mix volumes may be abruptly enacted or gradually enacted. This mixed audio signal with new volume relative volume levels lasts for a period of time. Then after the period of time the routines of the present invention automatically resume the audio to the nominal volume levels (in this case the ambient audio content going to zero volume). The resumption of nominal values may be abrupt or gradual.

[0058] Note—in some embodiments the mixed volume level is such that the musical audio content is gradually decreased down to substantially zero while the ambient audio musical content is gradually increased up to the prior music volume level. Such a cross-fade enables the music to fade out while the ambient audio content fades in. This lasts for a period of time. After the period of time, the process reverses, the ambient audio content fading out to zero volume and the musical content fading back to its pre-event nominal volume.

[0059] User Interface: In some embodiments of the present invention the media player includes dedicated user interface elements such as buttons, touch screen elements, and/or other manual or vocal commands that enable a user to override the automatic volume adjustment methods disclosed herein. For example, a button may be provided upon the portable media player that causes the volume levels to return to nominal values upon it being pressed. In this way the automatic ambient sound responsive volume adjustment routines of the present invention may cause the musical media content to automatically drop in volume during an event, such as a user speaking or an alarm sounding, and the user may override the automatic volume reduction by pressing the dedicated button or engaging the other dedicated user interface element. In this way the user can quickly resume the volume back to nominal levels, if for example, the user realizes that the alarm is not relevant to him and/or the other detected ambient audio event is not important.

[0060] External Electronic Alert Signal Employed for Automatic Volume Reduction In some embodiments of the present invention, the automatic volume reduction routines of the present invention that are active to attenuate the volume of a playing media file to a user for a period of time and then resume volume to nominal levels thereafter, may be triggered by an external electronic signal alert detected by a wireless transceiver of the media player. In this way an external electronic device in the user's local environment, such as a home automation system, a home security system, a personal computer, or some other separate electronic device, can send a specific electronic alert signal to the portable media player of the user. In response to receiving the specific electronic alert signal from the separate device within the user's local environment, the media player may automatically reduce the playing volume of the media content to the user for a period of time. This feature is useful in a ubiquitous computing environment in which a plurality

of intelligent devices may coexist within a local environment of the user as he or she listens to music through the portable media player. A separate device, such as a home security system, may wish to gain the user's attention and thus can issue an electronic alert to the media player which causes the volume to be reduced for a period of time. In some embodiments, the electronic alert signal system is used in combination with the features of the ambient sound responsive media player disclosed herein.

[0061] Thus as disclosed on the pages herein, an ambient sound responsive media player is operative to alert a media player user to ambient audio events within his or her local environment that he or she may not be able to easily hear while listening to the currently playing media content. Furthermore the present invention enables the user to attend to the ambient audio event for a period of time following the detected ambient audio event by lowering the music volume during that period of time. The present invention may support one or more of a variety of ambient audio events, including the verbal call of the user's name by another party in the local environment, an siren or alarm or other emergency sound audible within the local environment, the utterance of a password phrase by another party within the local environment, a verbal utterance identified to be from a user with a particular verbal identity within the local environment, and/or a verbal utterance identified to be from the media player user himself. In these ways the present invention is operative to enable a user to listen to music without cutting himself off from important audio events within his or her local environment. In these ways some embodiments of the present invention also are operative to allow third party users to gain the verbal attention of a media player user who may be listening to loud music through headphones. In these ways some embodiments of the present invention are also operative to enable a media player user to hear emergency sounds that may be important within his or her local environment. And finally, in these ways some embodiments of the present invention are also operative to enable a media player user to spontaneously begin engage in a conversation and not talk too loud, because he or she can more easily hear himself or herself while talking.

[0062] The foregoing described embodiments of the invention are provided as illustrations and descriptions. They are not intended to limit the invention to the precise forms described. In particular, it is contemplated that functional implementation of the invention described herein may be implemented equivalently in hardware, software, firmware, and/or other available functional components or building blocks. While the invention herein disclosed has been described by means of specific embodiments, examples and applications thereof, numerous modifications and variations could be made thereto by those skilled in the art without departing from the scope of the invention set forth in the claims.

What is claimed is:

  1. A method of adjusting an output of a media player comprising:

    capturing an ambient audio signal;

    processing the ambient audio signal to determine whether one or more characteristic forms are present within the ambient audio signal; and

reducing an output of a media player from a first volume to a second volume if the one or more characteristic forms are present within the ambient audio signal.

2. The method of claim 1 wherein the one or more characteristic forms comprise a name or personal identifier of a user of the media player.

3. The method of claim 1 wherein the one or more characteristic forms comprise the voice of a user of the media player.

4. The method of claim 1 wherein the one or more characteristic forms comprise an alarm or siren.

5. The method of claim 1 wherein the one or more characteristic forms comprise one or more generic words commonly used to summon attention.

6. The method of claim 1 wherein the one or more characteristic forms comprise a common household sound selected from the group consisting of a doorbell ring, a telephone ring and a baby's cry.

7. The method of claim 1 wherein the one or more characteristic forms comprise the voice of a person other than a user of the media player.

8. The method of claim 1 wherein the reducing step further comprises reducing the output of the media player if a volume of the one or more characteristic forms exceeds a volume threshold.

9. The method of claim 1 wherein the reducing step further comprises reducing the output of the media player in a manner that is performed gradually over a period of time.

10. The method of claim 1 wherein the second volume is a fixed percentage of the first volume.

11. The method of claim 1 wherein the second volume is based at least in part upon a volume level of the one or more characteristic forms.

12. The method of claim 1 wherein the second volume is a fixed volume on an absolute volume scale of the media player.

13. The method of claim 1 further comprising:

resuming the output of the media player to the first volume in a manner that is performed gradually over a period of time.

14. The method of claim 1 wherein the media player maintains the output at the second volume for a fixed duration.

15. The method of claim 1 wherein the media player maintains the output at the second volume until the media player is manually reset to the first volume.

16. The method of claim 1 wherein the media player maintains the output at the second volume for a duration dependent upon the one or more characteristic forms.

17. The method of claim 1 wherein the media player is manually reset to the first volume by actuating a button on the media player.

18. The method of claim 1 wherein the reducing step further comprises reducing the output of the media player upon receiving an electronic alert signal.

19. A method of adjusting an output of a media player comprising:

capturing an ambient audio signal;

processing the ambient audio signal to determine whether one or more characteristic forms are present within the ambient audio signal; and

mixing at least a portion of the ambient audio signal with a first output of a media player to generate a second output of the media player if the one or more characteristic forms are present within the ambient audio signal.

20. The method of claim 19 wherein the one or more characteristic forms comprise a name or personal identifier of a user of the media player.

21. The method of claim 19 wherein the one or more characteristic forms comprise the voice of a user of the media player.

22. The method of claim 19 wherein the one or more characteristic forms comprise an alarm or siren.

23. The method of claim 19 wherein the one or more characteristic forms comprise one or more generic words commonly used to summon attention.

24. The method of claim 19 wherein the one or more characteristic forms comprise a common household sound selected from the group consisting of a doorbell ring, a telephone ring and a baby's cry.

25. The method of claim 19 wherein the one or more characteristic forms comprise the voice of a person other than a user of the media player.

26. The method of claim 19 wherein the mixing step further comprises mixing the at least a portion of the ambient audio signal with the first output of the media player to generate the second output of the media player if a volume of the one or more characteristic forms exceeds a volume threshold.

27. The method of claim 19 wherein the mixing step further comprises mixing the at least a portion of the ambient audio signal with the first output of the media player to generate the second output of the media player in a manner that is performed gradually over a period of time.

28. The method of claim 19 wherein a first volume of the at least a portion of the ambient audio signal is gradually increased and a second volume of the first output of the media player is gradually decreased.

29. The method of claim 19 wherein a first volume of the at least a portion of the ambient audio signal is substantial relative to a second volume of the first output of the media player, such that the at least a portion of the ambient audio signal is clearly audible.

30. The method of claim 19 further comprising:

resuming the first output of the media player in a manner that is performed gradually over a period of time.

31. The method of claim 19 wherein the media player maintains the second output for a fixed duration following the determination of the one or more characteristic forms.

32. The method of claim 19 wherein the media player maintains the second output until the media player is manually reset to the first output.

33. The method of claim 19 wherein the media player maintains the second output dependent upon a duration of the detected one or more characteristic forms.

34. The method of claim 19 wherein the media player is manually reset to the first output by actuating a button on the media player.

35. The method of claim 19 wherein the mixing step further comprises mixing the at least a portion of the ambient audio signal with the first output of the media player to generate the second output of the media player upon receiving an electronic alert signal over a wireless link.

36. An apparatus for use in a media player comprising:

a microphone; and

one or more processors adapted to:

  process an ambient audio signal received by the microphone to determine whether one or more characteristic forms are present within the ambient audio signal, and

  adjust an output of a media player if the one or more characteristic forms are present within the ambient audio signal.

37. The apparatus of claim 36 wherein the one or more processors are adapted to reduce the output of the media player from a first volume to a second volume.

38. The apparatus of claim 37 wherein the one or more characteristic forms are selected from the group consisting of a name or personal identifier of a user of the media player, the voice of a user of the media player, and an alarm or siren.

39. The apparatus of claim 36 wherein the one or more processors are adapted to mix at least a portion of the ambient audio signal with a first output of the media player to generate a second output of the media player.

40. The apparatus of claim 39 wherein the one or more characteristic forms are selected from the group consisting of a name or personal identifier of a user of the media player, the voice of a user of the media player, and an alarm or siren.

* * * * *