amount of force to be exceeded before the device is turned on or off can prevent inadvertent switching of the device in question.

[0033] The present invention is able to classify and interpret detected input events according to the time and force of contact with the typing surface. In addition, the techniques of the present invention can be combined with other techniques for determining the location of an input event, so as to more effectively interpret location-sensitive input events, such as virtual keyboard presses. For example, location can be determined based on sound delays, as described in related U.S. patent application Ser. No. 10/115,357 for "Method and Apparatus for Approximating a Source Position of a Sound-Causing Event for Determining an Input Used in Operating an Electronic Device," filed Apr. 2, 2002, the disclosure of which is incorporated herein by reference. In such a system, a number of microphones are used to determine both the location and exact time of contact on the typing surface that is hit by the finger.

[0034] The present invention can be applied in any context where user action is to be interpreted and can be sensed in two or more domains. For instance, the driver of a car may gesture with her right hand in an appropriate volume within the vehicle in order to turn on and off the radio, adjust its volume, change the temperature of the air conditioner, and the like. A surgeon in an operating room may command an x-ray emitter by tapping on a blank, sterile surface on which a keyboard pad is projected. A television viewer may snap his fingers to alert that a remote-control command is ensuing, and then sign with his fingers in the air the number of the desired channel, thereby commanding the television set to switch channels. A popup menu system or other virtual control may be activated only upon the concurrent visual and auditory detection of a gesture that generates a sound, thereby decreasing the likelihood that the virtual controller is activated inadvertently. For instance, the user could snap her fingers, or clap her hands once or a pre-specified number of times. In addition, the gesture, being interpreted through both sound and vision, can signal to the system which of the people in the room currently desires to "own" the virtual control, and is about to issue commands.

[0035] In general, the present invention determines the synchronization of stimuli in two or more domains, such as images and sounds, in order to detect, classify, and interpret gestures or actions made by users for the purpose of communication with electronic devices.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0036] **FIG. 1** depicts a system of detecting, classifying, and interpreting input events according to one embodiment of the present invention.

[0037] **FIG. 2** depicts a physical embodiment of the present invention, wherein the microphone transducer is located at the bottom of the case of a PDA.

[0038] **FIG. 3** is a flowchart depicting a method for practicing the present invention according to one embodiment.

[0039] **FIG. 4** depicts an overall architecture of the present invention according to one embodiment.

[0040] **FIG. 5** depicts an optical sensor according to one embodiment of the present invention.

[0041] **FIG. 6** depicts an acoustic sensor according to one embodiment of the present invention.

[0042] **FIG. 7** depicts sensor locations for an embodiment of the present invention.

[0043] **FIG. 8** depicts a synchronizer according to one embodiment of the present invention.

[0044] **FIG. 9** depicts a processor according to one embodiment of the present invention.

[0045] **FIG. 10** depicts a calibration method according to one embodiment of the present invention.

[0046] **FIG. 11** depicts an example of detecting sound amplitude for two key taps, according to one embodiment of the present invention.

[0047] **FIG. 12** depicts an example of an apparatus for remotely controlling an appliance such as a television set.

[0048] The figures depict a preferred embodiment of the present invention for purposes of illustration only. One skilled in the art will readily recognize from the following discussion that alternative embodiments of the structures and methods illustrated herein may be employed without departing from the principles of the invention described herein.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0049] For illustrative purposes, in the following description the invention is set forth as a scheme for combining visual and auditory stimuli in order to improve the reliability and accuracy of detected input events. However, one skilled in the art will recognize that the present invention can be used in connection with any two (or more) sensory domains, including but not limited to visual detection, auditory detection, touch sensing, mechanical manipulation, heat detection, capacitance detection, motion detection, beam interruption, and the like.

[0050] In addition, the implementations set forth herein describe the invention in the context of an input scheme for a personal digital assistant (PDA). However, one skilled in the art will recognize that the techniques of the present invention can be used in conjunction with any electronic device, including for example a cell phone, pager, laptop computer, electronic musical instrument, television set, any device in a vehicle, and the like. Furthermore, in the following descriptions, "fingers" and "styluses" are referred to interchangeably.

[0051] Architecture

[0052] Referring now to **FIG. 4**, there is shown a block diagram depicting an overall architecture of the present invention according to one embodiment. The invention according to this architecture includes optical sensor **401**, acoustic sensor **402**, synchronizer **403**, and processor **404**. Optical sensor **401** collects visual information from the scene of interest, while acoustic sensor **402** records sounds carried through air or through another medium, such as a desktop, a whiteboard, or the like. Both sensors **401** and **402** convert their inputs to analog or digital electrical signals. Synchronizer **403** takes these signals and determines the time relationship between them, represented for example as the differences between the times at which optical and