tional QWERTY keyboard. Additional description of the virtual keyboard system embodied in the example of **FIG. 2** can be found in the related application for "Method and Apparatus for Entering Data Using a Virtual Input Device," referenced above.

[**0085**] As depicted in **FIG. 2**, microphone transducer **103** is positioned at the bottom of attachment **201** (such as a docking station or cradle). In the example of **FIG. 2**, attachment **201** also houses the virtual keyboard system, including camera **506**. The weight of PDA **106** and attachment **201** compresses a spring (not shown), which in turn pushes microphone transducer **103** against work surface **60**, thereby ensuring a good mechanical coupling. Alternatively, or in addition, a ring of rubber, foam, or soft plastic (not shown) may surround microphone transducer **103**, and isolate it from sound coming from the ambient air. With such an arrangement, microphone transducer **103** picks up mostly sounds that reach it through vibrations of work surface **60**.

[**0086**] Method of Operation

[**0087**] Referring now to **FIG. 3**, there is shown a flowchart depicting a method for practicing the present invention according to one embodiment. When the system in accordance with the present invention is turned on, a calibration operation **301** is initiated. Such a calibration operation **301** can be activated after each startup, or after an initial startup when the user first uses the device, or when the system detects a change in the environment or surface that warrants recalibration, or upon user request.

[**0088**] Referring momentarily to **FIG. 10**, there is shown an example of a calibration operation **301** according to one embodiment of the present invention. The system prompts **1002** the user to tap N keys for calibration purposes. The number of keys N may be predefined, or it may vary depending upon environmental conditions or other factors. The system then records **1003** the sound information as a set of N sound segments. In the course of a calibration operation, the sound-based detection system of the present invention learns properties of the sounds that characterize the user's taps. For instance, in one embodiment, the system measures **1004** the intensity of the weakest tap recorded during calibration, and stores it **1005** as a reference threshold level for determining whether or not a tap is intentional. In an alternative embodiment, the system stores (in memory **105**, for example) samples of sound waveforms generated by the taps during calibration, or computes and stores a statistical summary of such waveforms. For example, it may compute an average intensity and a standard deviation around this average. It may also compute percentiles of amplitudes, power, or energy contents of the sample waveforms. Calibration operation **301** enables the system to distinguish between an intentional tap and other sounds, such as light, inadvertent contacts between fingers and the typing surface, or interfering ambient noises, such as the background drone of the engines on an airplane.

[**0089**] Referring again to **FIG. 3**, after calibration **301** if any, the system is ready to begin detecting sounds in conjunction with operation of virtual keyboard **102**, using recognition function **903**. Based on visual input v from optical sensor **401** recognition function **903** detects **302** that a finger has come in contact with typing surface **50**. In general, however, visual input v only permits a determination of the time of contact to within the interval that separates two subsequent image frames collected by optical sensor **401**. In typical implementations, this interval may be between 0.01 s and 0.1 s. Acoustic input a from acoustic sensor **402** is used to determine **303** whether a concurrent audio event was detected, and if so confirms **304** that the visually detected contact is indeed an intended keystroke. The signal representing the keystroke is then transmitted **306** to PDA **106**. If in **303** acoustic sensor **402** does not detect a concurrent audio event, the visual event is deemed to not be a keystroke **305**. In this manner, processor **404** is able to combine events sensed in the video and audio domains so as to be able to make more accurate determinations of the time of contact and the force of the contact.

[**0090**] In one embodiment, recognition function **903** determines **303** whether an audio event has taken place by measuring the amplitude of any sounds detected by transducer **103** during the frame interval in which optical sensor **401** observed contact of a finger with typing surface **50**. If the measured amplitude exceeds that of the reference level, the keystroke is confirmed. The time of contact is reported as the time at which the reference level has been first exceeded within that frame interval. To inform optical sensor **401**, processor **404** may cause an interrupt to optical sensor **401**. The interrupt handling routine consults the internal clock of acoustic sensor **402**, and stores the time into a register or memory location, for example in memory **105**. In one embodiment, acoustic sensor **402** also reports the amount by which the measured waveform exceeded the threshold, and processor **404** may use this amount as an indication of the force of contact.

[**0091**] Referring momentarily to **FIG. 11**, there is shown an example of detected sound amplitude for two key taps. The graph depicts a representation of sound recorded by transducer **103**. Waveforms detected at time t**1** and t**2** are extracted as possible key taps **1101** and **1102** on projected keyboard **70**.

[**0092**] The above-described operation may be implemented as an analog sound-based detection system. In an alternative embodiment, acoustic sensor **402** is implemented using a digital sound-based detection system; such an implementation may be of particular value when a digital signal processing unit is available for other uses, such as for the optical sensor **401**. The use of a digital sound-based detection system allows more sophisticated calculations to be used in determining whether an audio event has taken place; for example, a digital system may be used to reject interference from ambient sounds, or when a digital system is preferable to an analog one because of cost, reliability, or other reasons.

[**0093**] In a digital sound-based detection system, the voltage amplitudes generated by the transducer are sampled by an analog-to-digital conversion system. In one embodiment, the sampling frequency is between 1 kHz and 10 kHz although one skilled in the art will recognize that any sampling frequency may be used. In general, the frequency used in a digital sound-based detection system is much higher than the frame rate of optical sensor **401**, which may be for example 10 to 100 frames per second. Incoming samples are either stored in memory **105**, or matched immediately with the reference levels or waveform characteristics. In one embodiment, such waveform characteristics are in the form of a single threshold, or of a number of