[0124] It should be borne in mind, however, that all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities. Unless specifically stated otherwise as apparent from the discussion, it is appreciated that throughout the description, discussions utilizing terms such as "processing" or "computing" or "calculating" or "determining" or "displaying" or the like, refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic) quantities within the computer system's registers and memories into other data similarly represented as physical quantities within the computer system memories or registers or other such information storage, transmission or display devices.

[0125] The present invention also relates to an apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, or it may comprise a general-purpose computer selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a computer readable storage medium, such as, but is not limited to, any type of disk including floppy disks, optical disks, CD-ROMs, and magnetic-optical disks, read-only memories (ROMs), random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, or any type of media suitable for storing electronic instructions, and each coupled to a computer system bus.

[0126] The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general-purpose systems may be used with programs in accordance with the teachings herein, or it may prove convenient to construct more specialized apparatuses to perform the required method steps. The required structure for a variety of these systems appears from the description. In addition, the present invention is not described with reference to any particular programming language. It will be appreciated that a variety of programming languages may be used to implement the teachings of the invention as described herein.

[0127] The present invention improves reliability and performance in detecting, classifying, and interpreting user actions, by combining detected stimuli in two domains, such as for example visual and auditory domains. One skilled in the art will recognize that the particular examples described herein are merely exemplary, and that other arrangements, methods, architectures, and configurations may be implemented without departing from the essential characteristics of the present invention. Accordingly, the disclosure of the present invention is intended to be illustrative, but not limiting, of the scope of the invention, which is set forth in the following claims.

What is claimed is:

1. A computer-implemented method for classifying an input event, the method comprising:

receiving, at a visual sensor, a first stimulus resulting from user action, in a visual domain;

receiving, at an auditory sensor, a second stimulus resulting from user action, in an auditory domain; and

responsive to the first and second stimuli indicating substantial simultaneity of the corresponding user action, classifying the stimuli as associated with a single user input event.

2. A computer-implemented method for classifying an input event, comprising:

receiving a first stimulus, resulting from user action, in a visual domain;

receiving a second stimulus, resulting from user action, in an auditory domain;

classifying the first stimulus according to at least a time of occurrence;

classifying the second stimulus according to at least a time of occurrence; and

responsive to the classifying steps indicating substantial simultaneity of the first and second stimuli, classifying the stimuli as associated with a single user input event.

3. The method of claim 2, wherein:

classifying the first stimulus comprises determining a time for the corresponding user action; and

classifying the second stimulus comprises determining a time for the corresponding user action.

4. The method of claim 3, wherein:

determining a time comprises reading a time stamp.

5. The method of claim 1 or **2**, further comprising:

generating a vector of visual features based on the first stimulus;

generating a vector of acoustic features based on the second stimulus;

comparing the generated vectors to user action descriptors for a plurality of user actions; and

responsive to the comparison indicating a match, outputting a signal indicating a recognized user action.

6. The method of claim 1 or **2**, wherein the single user input event comprises a keystroke.

7. The method of claim 1 or **2**, wherein each user action comprises a physical gesture.

8. The method of claim 1 or **2**, wherein each user action comprises at least one virtual key press.

9. The method of claim 1 or **2**, wherein receiving a first stimulus comprises receiving a stimulus at a camera.

10. The method of claim 1 or **2**, wherein receiving a second stimulus comprises receiving a stimulus at a microphone.

11. The method of claim 1 or **2**, further comprising:

determining a series of waveform signals from the received second stimulus; and

comparing the waveform signals to at least one predetermined waveform sample to determine occurrence and time of at least one auditory event.

12. The method of claim 1 or **2**, further comprising:

determining a series of sound intensity values from the received second stimulus; and

comparing the sound intensity values with at a threshold value to determine occurrence and time of at least one auditory event.