thresholds associated with different locations on typing surface **50**. Processing then continues as described above for the analog sound-based detection system. Alternatively, the sound-based detection system may determine and store a time stamp with the newly recorded sound. In the latter case, processor **404** conveys time-stamp information to optical sensor **401** in response to a request by the latter.

[**0094**] In yet another embodiment, processor **404** compares an incoming waveform sample in detail with waveform samples recorded during calibration **301**. Such comparison may be performed using correlation or convolution, in which the recorded waveform is used as a matched filter, according to techniques that are well known in the art. In such a method, if s_n are the samples of the currently measured sound wave, and r_n are those of a recorded wave, the convolution of s_n and r_n is defined as the following sequence of samples:

$$c_n = \sum_{k=-\infty}^{\infty} s_{n-k} r_k.$$

[**0095**] A match between the two waveforms s_n and r_n is then declared when the convolution c_n reaches a predefined threshold. Other measures of correlation are possible, and well known in the art. The sum of squared differences is another example:

$$d_n = \sum_{k=n-K}^{n} (s_k - r_k)^2,$$

[**0096**] where the two waveforms are compared over the last K samples. In this case, a match is declared if d_n goes below a predefined threshold. In one embodiment, K is given a value between 10 and 1000.

[**0097**] The exact time of a keystroke is determined by the time at which the absolute value of the convolution c_n reaches its maximum, or the time at which the sum of squared differences d_n reaches its minimum.

[**0098**] Finally, the force of contact can be determined as

$$\max_n \frac{\sum_{k=-\infty}^{\infty} s_{n-k} r_k}{\sum_{k=-\infty}^{\infty} r_k^2},$$

[**0099**] or as any other (possibly normalized) measure of energy of the measured waveform, such as, for instance,

$$\frac{\sum_{k=-\infty}^{\infty} s_k^2}{\sum_{k=-\infty}^{\infty} r_k^2}.$$

[**0100**] Of course, in all of these formulas, the limits of summation are in practice restricted to finite values.

[**0101**] In one embodiment, sample values for the current sample are stored and retrieved from a digital signal processor or general processor RAM.

[**0102**] In some cases, if the virtual keyboard **102** is to be used on a restricted set of typing surfaces **60**, it may be possible to determine an approximation to the expected values of the reference samples r_n ahead of time, so that calibration **301** at usage time may not be necessary.

[**0103**] Gesture Recognition and Interpretation

[**0104**] For implementations involving virtual controls, such as a gesture-based remote control system, the low-level aspects of recognition function **903** are similar to those discussed above for a virtual keyboard. In particular, intensity thresholds can be used as an initial filter for sounds, matched filters and correlation measures can be used for the recognition of particular types of sounds, and synchronizer **403** determines the temporal correspondence between sound samples and images.

[**0105**] Processing of the images in a virtual control system may be more complex than for a virtual keyboard, since it is no longer sufficient to detect the presence of a finger in the vicinity of a surface. Here, the visual component of recognition function **903** provides the ability to interpret a sequence of images as a finger snap or a clap of hands.

[**0106**] Referring now to **FIG. 12**, there is shown an example of an apparatus for remotely controlling an appliance such as a television set **1201**. Audiovisual control unit **1202**, located for example on top of television set **1201**, includes camera **1203** (which could possibly also be a three-dimensional sensor) and microphone **1204**. Inside unit **1202**, a processor (not shown) analyzes images and sounds according to the diagram shown in **FIG. 9**. Visual feature computation module **902** detects the presence of one or two hands in the field of view of camera **1203** by, for example, searching for an image region whose color, size, and shape are consistent with those of one or two hands. In addition, the search for hand regions can be aided by initially storing images of the background into the memory of module **902**, and looking for image pixels whose values differ from the stored values by more than a predetermined threshold. These pixels are likely to belong to regions where a new object has appeared, or in which an object is moving.

[**0107**] Once the hand region is found, a visual feature vector v is computed that encodes the shape of the hand's image. In one embodiment, v represents a histogram of the distances between random pairs of point in the contour of the hand region. In one embodiment, 100 to 500 point pairs are used to build a histogram with 10 to 30 bins.