Hand Input: High-Resolution Skin Localization

Murium Iqbal, Robert Xiao, Benjamin Shih, Chris Harrison, Bhiksha Raj
18797 Machine Learning with Signal Processing Course Project
Human-Computer Interaction Institute, Carnegie Mellon University
{muriumi, brx, bshih1}@andrew.cmu.edu,

Abstract: The majority of on-body interfaces are limited by their ability to resolve inputs at high-resolution proximities. We address this problem for the hand and arm region by using innovative approaches to generate feature sets and classify the input signals. Incorporating this design into modern electronic devices may expedite the transition to, and increase the use of, on-body interfaces. This paper discusses the progress in prototyping and classifying the hand input signals.

Keywords
Bio-acoustics, hand input, on-body interaction.

I. INTRODUCTION

As interactive electronic devices become increasingly common in our everyday lives, we seek a more elegant, intuitive way to use them. One option is to increase the available surface area with which we can interact, which is undesirable because of the need for lightweight, portable equipment. Alternatively, we can use something that we always carry with us – our hands. We present Hand Input, a hands-on method for interacting with our devices.

Current technology is restricted by the need for touch interfaces. Our project proposes the use of a person's own hands as a surface with which to interact with his/her electronic devices, thus removing the need for bulky touch screens. This approach provides a portable and accessible on-body input system. Specifically, we used a wristband outfitted with two vibrational sensors in order to resolve taps and features on the hand. Various spectral features such as power spectral density, spectral centroid, and band energy ratio, and log spectral band ratio helped us distinguish between taps on various regions of the hand and arm.

Our initial goal of extremely high resolution on the palm was not achievable with the hardware we had because the propagation of vibrational waves throughout the fleshy areas of the hands were not sufficiently distinguishable for classification with our feature set. Nevertheless, by choosing different areas of the hand and arm, we were able to replicate playlist functionality with start, stop, play, pause, and playlist switching.

The contributions of this project are:

1) A wearable wristband device, containing two vibrational microphones placed near the wrist.

2) A novel analysis of the wristband sensor data, using signal processing algorithms.

To improve this project in the future, we would like to increase both the quality and number of sensors on the wrist band, as well as the quality of the data that we collect for training the support vector machine (SVM).

II. PREVIOUS WORK

Skinput

Hand Input hopes to improve the work of Skinput by using more sensors to provide significantly higher resolution to the different areas of the hand. As shown below in figure 1, Skinput has 5 locations on the hand, one for each of the fingers.

Fig. 1. Skinput resolution on the hand.

Hambone

Hambone approaches the hand interface from a similar perspective. A problem that Deyle et al. encountered was the low signal resolution, which they addressed by sacrificing comfort, subtlety, and intuition for easily distinguishable movements such as moving fingers in the air and rubbing or tapping fingers together. In addition, they chose to include a separate device for foot gestures.

The Sound of One Hand

Like Hand Input and Hambone, this project aims to characterize the differences between certain gestures and flicks using one hand. However, Amento et al. focus only on classifying fingertip gestures. In addition, their device uses a simple classifier which we would like to improve on.

III. MATERIALS AND METHODS

In order to record the gesture data, we constructed a wristband embedded with two vibrational microphones. The sensors are sewn to an elastic wristband and placed directly on the skin surface at the wrist. Through experimentation, we discovered that vibrational waves travel much faster through bone than through flesh. As a result, one microphone is located on top of the ulna, while the other is placed on the fleshy part of the wrist. This placement allows us to localize and distinguish between different inputs on the hand.

Using two piezoelectric-contact microphones, which are stitched into a wristband, we can listen to sounds traveling through a person's arm in response to taps or touches. Localizing these taps allows us to assign a function associated with a tap on a specific part of the hand or arm and, thus, ultimately create a touch interface on a person's skin.

Data from the microphones is taken as stereo input into an amplifier and then fed into a computer. If the signal on either channel exceeds a preset threshold, this is considered a "tap" and a sample of the signal at this instance is stored as data for the tap. A set of 100 samples from each of these regions is taken and features are extracted to train a classifier to recognize taps from the recorded regions.

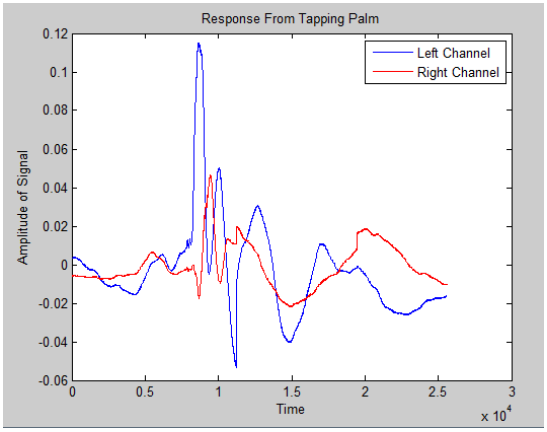

Fig 2. Sample stereo recording of a tap on the palm.

Next, we localize and classify the inputs using the responses obtained from the two microphones. Due to some troubles with stereo signal accuracy for locations that are really close in similar areas of the hand, such as adjacent areas on the palm, we selected the four tap locations to be the finger tip, elbow, wrist, and palm. Prior to classification, training data is collected. Given this training set, we start our analysis by combinatorially computing on the left, right, and cross-correlation channels a list of various spectral features.

Note that F_n is the nth spectral coefficient, and P_n is the nth normalized power spectrum coefficient:

$$P_n = \frac{|F_n|^2}{\sum_n |F_n|^2}$$

Average Spectral Density

Average spectral density is a single value that represents the total average power carried by the signal, or in other words the average energy of the impulse. The physical significance of this attribute is how loud the tap was.

$$\frac{\sqrt{\sum_n |F_n|^2}}{N}$$

Spectral Centroid

The spectral centroid is a scalar that is equivalent to the center of mass of a given spectrum. It represents the average frequency, and physically is the average pitch of the tap.

$$C = \sum_n n \cdot P_n$$

Band Energy Ratio

Band Energy Ratio (BER) is the ratio of energy across various bands of frequencies, or how much energy is concentrated in a specific band. It shows variations of energy in the signal across differently sized bins, where each band is some range of hertz. BER physically represents the amount of a type of frequency (such as bass or treble) exists in the signal.

Band Energy Ratio, $BER_i = \sum_{n \in S_i} P_n$ for every octave ($S_i = [2^i, 2^{i+1})$ for each i), every third ($S_i = [2^{i/3}, 2^{(i+1)/3})$ for each i), and for 20 evenly-spaced "linear" bins ($S_i = [iN/20, (i+1)N/20)$).

Log Spectral Band Ratio

Log Spectral Band Ratio (LSBR) emphasizes the differences between bands and helps reveal relationships in spectrums between signals in response to identical stimuli. For example, if two of the bands in the signal are bass and treble, LSBR tells us how the energy is balanced and distributed across the two different classes. In other words, it represents the balance of bass and treble in that signal.

Log Spectral Band Ratio, $SBR_{ij} = \log \frac{BER_i}{BER_j} = \log BER_i - \log BER_j$ for every pair of octaves, and every pair of thirds

The logarithm is for linearizing differences in orders of magnitude because SVM is a linear classifier. It partitions the data points by drawing straight lines, which for some arbitrary data point represents either yes or no for whether the data point is a member of that particular class. Thus, the classification requires linearity.

We computed these features in Python and compared each tap location using Weka, a machine learning toolbox. Then, we used a SVM to classify the collected data based on the initial training data.

IV. RESULTS AND ANALYSIS

Weka

Using a training set of approximately ~200 taps per location, we obtained ~95% accuracy on these four locations of the hand by using the above features with Weka and training an SVM. The data was classified according to the features described in the previous section. Figure 3 is a sample of the classified data. The features displayed below represented examples of features that were easily separable, and thus were good for classification.

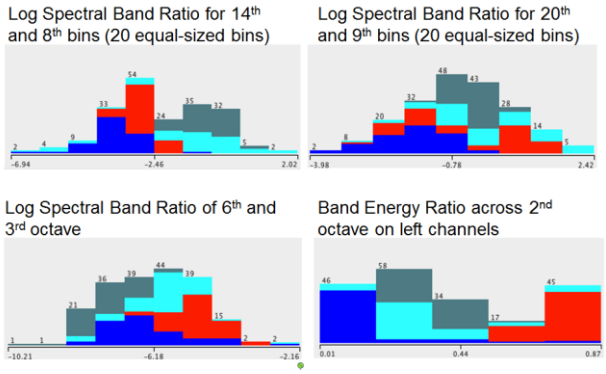

Fig 3. Sample classification of data using SVM.

Initially, we tried to record training data that was performed mechanically and precisely, but found that our classification accuracy for testing data was only ~80%. However, we realized that this procedure could be improved. Modeling this problem graphically, precisely-measured data is easy to classify because each class' data points are bunched closely together, but problems occur when trying to fit data that is not recorded in the exact same way. By recording data with some flexibility in the way each part of the hand was tapped, the point clouds for each class become looser and larger, while ideally maintaining easy-to-partition clusters. Thus, introducing variations in our training data counter-intuitively improved our results.

Hidden Markov Model

Murium's section

V. DEMO AND OTHER PROOFS-OF-CONCEPT

For our demo, we made a playlist shuffler that mimicked the functionality of iTunes as an audio player software. Using python libraries, we assigned each tap location a command: "pause/play", "select playlist", "next song", and "return to menu". In real-time, actions performed while wearing the device were processed in software and converted to the appropriate action and executed on the computer.

Below is a list of other potential applications that this type of interface would be useful for:

A. Number Pad

In this concept, Hand Input will be able to distinguish between the ten keys (0 through 9) that make up a standard number pad, similar to the ones found on phones and computers. This idea represents the pinnacle of our project – it involves extremely high resolution on the palm of the user, and a low rate of false-positives so that the device can ensure the right numbers are detected.

B. MP3 Click Wheel

The user will be able to mimic the click wheel found on many mp3 devices. This includes the commands for next song, previous song, pause/play, and a scrolling wheel for volume. The wheel would be the hardest to detect in terms of signal amplitude, but because it requires a different gesture pattern, we still expect to be able to recognize it.

C. Mouse Track Pad

By tapping on the respective cardinal directions on the hand, the user's mouse will follow in the same direction. This idea is the simplest among the three because it only involves four distinct regions of detection.

VI. FUTURE OBJECTIVES

There are many aspects of our device that can be improved upon. In addition, much functionality remains to be added before the device would be available as a commercial product.

Conceptual Improvements

The device resolution would drastically increase with better microphones. The microphones used in the prototype had difficulty distinguishing between different areas on the palm because the palm was too homogeneously fleshy. It is also very difficult to distinguish between very specific regions on the arm/hand, because it is hard for an individual to tap consistently in adjacent skin areas. However, with more sensitive microphones, we believe that we can detect taps on the different regions of the palm.

Extra microphones would be useful for assisting in the tap localization. The additional microphones would also improve the accuracy of our device, while trading off for computational power.

A next step would be to make the system robust to changes in the positioning of the wristband in order to de-sensitize the system to the shifting of the microphones. Currently, once the user has trained the device, even slight movements can have a significant negative effect on the accuracy of the system. We seek a way to mitigate the problems with positioning.

As with machine learning projects in general, larger sets of training data, more advanced features, and better classifiers may help localize taps on more specific locations, such as specific fingers or specific locations of the palm.

Commercial Improvements

As a supplement to vibrational microphones, the device might incorporate accelerometers into the wristband in order to expand the range and types of motions that can be detected. This property would allow for new gestures such as wiggling the hand back and forth.

In addition, the on-body interface would be significantly more practical if it was not tethered. The device would also be more useful if it were able to operate wirelessly, and were able to ubiquitously interface with any electronic device. Therefore, a future objective is to enable the device to run wirelessly. In parallel to this concept, the device would be easier to use if all

the processing was done on-board with a digital signal processing (DSP) chip.

References

[1] C. Harrison, D. Tan, D. Morris, “Skinput: Appropriating the Body as an Input Surface”, Microsoft Research.
[2] Deyle, T., Palinko, S., Poole, E.S., and Starner, T. Hambone: A Bio-Acoustic Gesture Interface. In Proc. ISWC '07. 1-8.
[3] Amento, B., Hill, W., and Terveen, L. The Sound of One Hand: A Wrist-mounted Bio-acoustic Fingertip Gesture Interface. In CHI ‘02 Ext. Abstracts, 724-725.