FingerPing: Recognizing Fine-grained Hand Poses using Active Acoustic On-body Sensing

Cheng Zhang [1], Qiuyue Xue [1], Anandghan Waghmare [1], Ruichen Meng [1], Sumeet Jain [1], Yizeng Han [2], Xinyu Li [1], Kenneth Cunefare[1], Thomas Ploetz [1], Thad Starner [1], Omer Inan [1], Gregory D. Abowd [1]
chengzhang, qiuyue, anandghan, rcmeng93, sumeet, xinyu.li, thomas.ploetz, thad, inan, abowd@gatech.edu; hanyz14@mails.tsinghua.edu.cn ; ken.cunefare@me.gatech.edu;
1.Georgia Institute of Technology 2. Tsinghua University

ABSTRACT
FingerPing is a novel sensing technique that can recognize various fine-grained hand poses by analyzing acoustic resonance features. A surface-transducer mounted on a thumb ring injects acoustic chirps (20Hz to 6,000Hz) to the body. Four receivers distributed on the wrist and thumb collect the chirps. Different hand poses of the hand create distinct paths for the acoustic chirps to travel, creating unique frequency responses at the four receivers. We demonstrate how FingerPing can differentiate up to 22 hand poses, including the thumb touching each of the 12 phalanges on the hand as well as 10 American sign language poses. A user study with 16 participants showed that our system can recognize these two sets of poses with an accuracy of 93.77% and 95.64%, respectively. We discuss the opportunities and remaining challenges for the widespread use of this input technique.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Gesture recognition; wearable input; acoustic sensing

INTRODUCTION
Despite years of research and development and substantial progress made, providing appropriate means for input to wearable devices remains a considerable challenge. The size and comfort required for continuous use of a wearable device, as well as the need to operate in mobile contexts with minimal difficulty and attention, make the options of keyboards and touchscreens less desirable. Voice input is one viable alternative, but it is not always the most socially appropriate solution. Another alternative utilizes a user's phone as input device for their wearable(s). However, such a proxy (or remote control) solution in many cases over-complicates the interaction as it requires the user to first reach for the phone, which is inconvenient in many scenarios such as when the hands are occupied with another task. Furthermore, the user may need to input in a more discreet fashion for privacy and social appropriateness in certain scenarios, such as during a meeting or interacting with smart home devices, such as Google Home or Amazon Alexa. Finally, heads-up displays for augmented and virtual reality present opportunities for non-voice, eyes-free input.

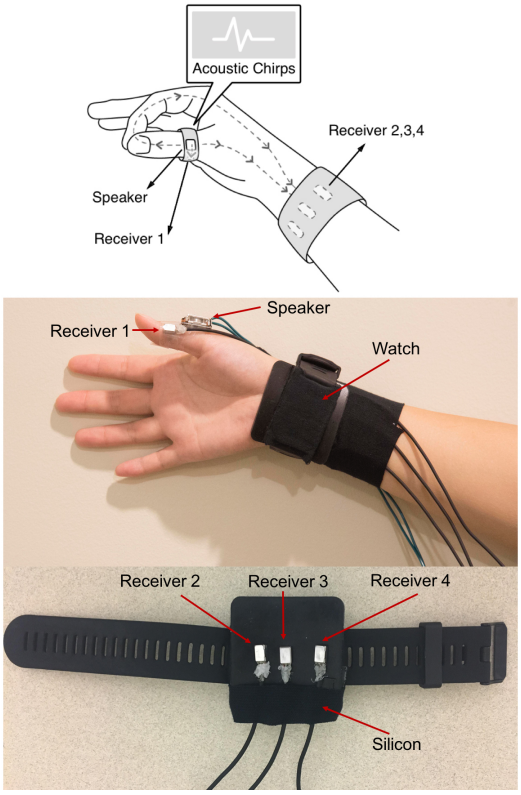

Figure 1. FingerPing

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada
© 2018 ACM. ISBN 978-1-4503-5620-6/18/04…15.00
DOI: https://doi.org/10.1145/3173574.3174011

Given this motivation for a convenient and socially appropriate wearable input solution, we introduce FingerPing, a novel wrist- and thumb-mounted sensing solution to enable one-handed input. FingerPing relies on detecting various hand configurations (e.g., the thumb touching the tip of a finger) based on how that configuration impacts the propagation of sound waves injected at the thumb and propagating around the hand. The human body is a good medium for sound propagation [20, 12], and its frequency response varies based on which path a sound wave travels through the body. A change in hand configuration, which results from forming the hand into a variety of different poses or gestures, creates sufficiently distinct propagation paths for sound waves. To utilize this phenomenon for recognizing different hand poses, FingerPing injects acoustical chirps (20Hz–6kHz) ten times per second from the base of the thumb. These chirps travel through the hand and are received by microphones present on the thumb and wrist. The received signals are then classified to match a set of known poses.

To demonstrate the capabilities of FingerPing, we designed and evaluated two sets of poses in this paper. The first pose set consists of thumb taps to the 12 phalanges across the four fingers, which can be used for number input and potentially text input with a T9 keyboard[1]. The other pose set consists of the ten number poses from American Sign Language, further demonstrating the flexibility in reliably distinguishing a large number of simple hand poses. Note that our system effectively detects endpoints of gesture input. However, the actual data analysis –after segmentation– is based on classifying *static* hand configurations (poses). Consequently, we denote our approach as pose or hand configuration recognition rather than gesture recognition – even though its purpose is clearly targeting the latter. Our technology may suffer more false-positive errors caused by the on-body acoustic noise during daily activities. However, false-positives from daily activities can be addressed with a reliable activation pose. The user could perform the activation pose to activate the system which would then enable the full set of poses for recognition. Our technology may also request additional calibration for different users. A user calibration procedure can be adopted to address this issue. More details is provided in the paper.

The contributions of this paper are:

- An active acoustical sensing system that recognizes various hand poses by retrieving and recognizing the acoustical signatures generated on the hand.
- The design of a one-handed hand poses set, which maps a standard 12-key number pad to the natural structure of the hand and can be used for number input and potentially text input. A second example pose set for input is the set of American Sign Language poses for the ten digits 1–10.
- An empirical evaluation of the two pose sets, with the results discussed in terms of viability for practical applications.

In the remainder of the paper, we will first discuss the previous work and highlight the innovation and contribution of FingerPing. We then present the underlying theory, design and implementation, and empirical evaluation of the system. In the last section, we discuss the challenges and opportunities for using this novel technique in everyday applications.

[1]https://en.wikipedia.org/wiki/T9_(predictive_text)

RELATED WORK

Providing appropriate input for wearable devices has been a research topic in the community for years. Various sensing modalities and form factors have been explored to improve the wearable input experience. Here we categorize these past research based on the form factors and placement of the various solutions.

Input with an armband

Performing hand gestures can introduce or alter various signals in the arm region, either the forearm or upper arm. To utilize this phenomenon for recognizing finger or hand gestures, these armbands detect different signals. [9, 14] recognized tapping positions on forearm using active and passive acoustic sensing, and [17] uses electromyography signals. Unfortunately, wearing an armband is rarely a convenient experience from the user's perspective.

Input with a wrist-mounted device

Compared to armbands, wristbands like wrist watches are more commonly worn and are also more convenient. Therefore, many researchers have developed wrist-mounted devices to recognize finger gestures by capturing different types of changes around the wrist area with respect to a number of different sensing modalities, from inertial movement [12, 23, 12], computer vision [13], forces [5], proximity [8], electrical signals [4], or acoustic signals[19, 15, 1].

Input with a ring

More recently, accompanied by the advancement of the wearable computing technology (e.g., battery, processor), researchers have been able to work on smaller wearable input devices such as a ring. Similar to the wrist-mounted device, wearing a ring is also socially appropriate in most scenarios. In addition, because the ring is worn on the fingers, it can gather more information and is at an even better position to capture the finger gestures. Thus, rings built with different sensing modalities have been explored to classify finger gestures, such as acoustic and motion sensing [21, 22], computer vision [2] , magnetic sensing [3, 10], or electrical sensing [24].

All of these previous solutions involve passive sensing, that is, the sensors placed on the body are detecting phenomenon produced by either the body itself or some agent external to the body. FingerPing uses an active acoustic sensing mechanism to recognize the hand poses, meaning that the sensor solution produces a powered signal that is then sensed as it propagates along the hand and wrist. Active acoustic sensing has also been applied to find content level in a container [6] and distinguish laptop and earbud configurations[11]. While this solution does require a power source for both transmission and reception (as opposed to just reception for a passive solution), the improvement in number and resolution of distinct poses and recognition accuracy justifies this added constraint.

THE DESIGN OF HAND POSES

In this paper, we demonstrate the capability of FingerPing by recognizing two sets of hand poses.

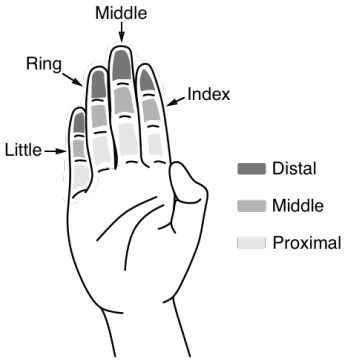

Figure 2. Tap on 12 Phalanges

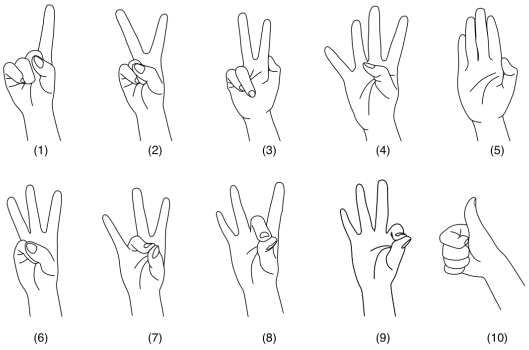

Figure 3. Digits '1' to '10' from American Sign Language

The first set of poses enables the user to input digits by touching any of the 12 phalanges of the index, middle, ring, and little fingers (see Figure 2) with the thumb. Because of the similarity between the layout of the 12 phalanges and a number-pad, this layout could be used to input digits or potentially input text using T9 keyboard without the heavy mental efforts to memorize different gestures.

The second set of poses are the numbers '1' to '10' from American Sign Language as shown in Figure 3. We use this set of poses to demonstrate the potential of FingerPing to recognize various other poses of the hand. In addition, recognizing this set of hand poses can potentially be used to translate the digits expression from American Sign Language(ASL) for people who do not understand ASL. For instance, the recognized results can be played in audio to assist the communication. Furthermore, these poses can also be used as shortcuts to access different functions in wearable computers.

THEORY OF OPERATION

FingerPing exploits the physical phenomenon that the spectral properties of sound waves change based on the paths they travel between sender and receiver. Performing different hand configurations has an impact on how the sound travels through the hand. For instance (Figure 1), when the thumb is open (not touching any fingers), there is only one major path the signal can travel from the speaker on the thumb to the receivers on the wrist, namely the direct path. Once the thumb touches one of the phalanges, a secondary path is created for sound propagation: starting from the speaker, via the touched phalanx and the corresponding finger, and finally to the receivers. In this case, the signals received would stem from at least two major paths: *i)* the direct path; and *ii)* the "detour" path, which goes through the phalanx and the finger. Depending on which phalanx is touched, the second path– the "detour"–changes, which also changes the energy of different frequencies. They can be either amplified or reduced depending on the amount of tissue/ bone in the path taken by the wave. Different components (e.g., tissue, bone) present different acoustic frequency response. This property of sound wave propagating through the human body constructs the unique fingerprints in frequency response for different hand configurations. The same phenomenon can also be observed for non-touch poses as shown in Figure 3, where the second path (detour path) varies depending on which hand pose is performed. Figure 4 shows the frequency response of chirps received from three receivers on the wrist area for the first pose set (digits; explanation below).

IMPLEMENTATION

Hardware Design

FingerPing consists of two parts of hardware: *i)* A surface transducer for emitting sound (sender); and *ii)* Four contact microphones –receivers– that capture sound signals after they have traveled through the user's body. In what follows we will describe both components in detail.

Surface Transducer

The surface transducer[2] is used to emit sound into the hand. The transducer is driven by a function generator (Agilent 33500B) and attached on the thumb using Kinesiology tape which is stretchy and elastic. We use a function generator to send ten chirps of 2 Vpp per second. For each chirp, we first linearly sweep the frequency range from 20Hz to 6,000Hz for 0.05 seconds, and then hold on at 6,000Hz for another 0.05 seconds as shown in figure 5. The range for the frequency sweep was optimized in an experimental evaluation (results not shown here) that unveiled that frequencies up to 6,000Hz retained maximum information while propagating in the body, which is also in line with previous findings in the literature (e.g., [20]).

Sound Receivers

The second part of the hardware includes four contact microphones (Knowles BU-21771) used to capture the signals from the body, each of which is 7.92 mm by 5.59 mm by 4.14 mm in size. These contact microphones provide a low noise floor and very flat frequency response. One of the microphones is attached on the thumb and the remaining three are aligned on the wrist as shown in Figure 1. We built a watch-like device to attach the microphones to the wrist of the wearer. A

[2]https://www.sparkfun.com/products/10917

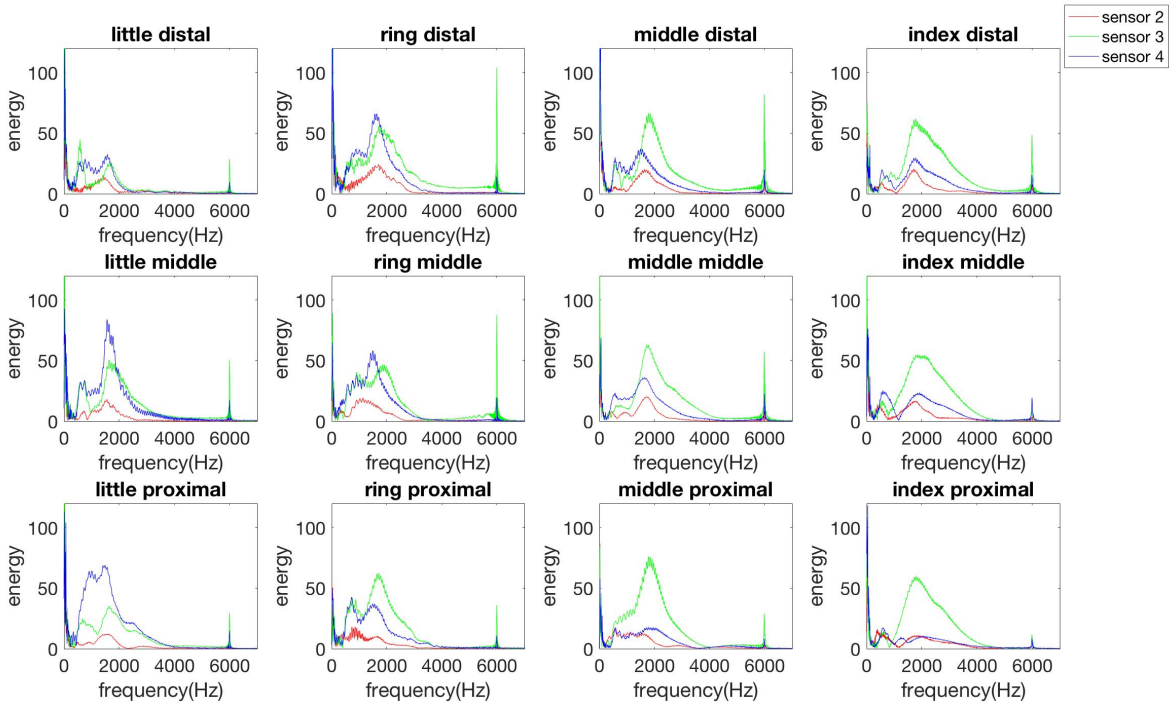

Figure 4. Frequency responses of taps on 12 phalanges

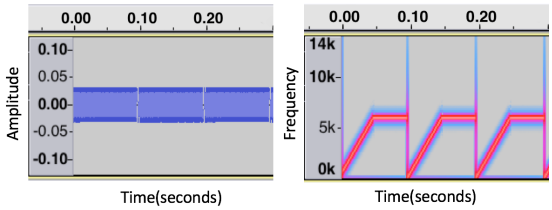

Figure 5. Sweep Signal

piece of silicone is placed between the watch case and the microphones to fix the positions of the sensors and match the acoustic impedance as demonstrated in [16], which contributes to a better quality of the captured acoustic signals.

All microphones are connected to pre-amplifiers to amplify the received signals (factor: 100). The amplified signals are then relayed to a Macbook Pro laptop computer for data processing via an audio interface (Fireface 800). The audio interface samples the audio at 44,100Hz. The laptop runs a software program written in C language using the PortAudio library to communicate with the audio interface. The data read from the audio interface is then sent to a Java program for real-time processing over network socket.

Data Processing Pipeline

The processing of the received data stream (four channels) can be divided into three steps as described below and shown in figure 6. Channels 1 to 4 of the audio interface are mapped to the signals from the four microphones, respectively.

Chirp Localization

In order to recognize a given pose, we first explicitly localize (temporally) each chirp within the continuous data stream – segmentation. Channel 1 is the signal from the microphone right next to the speaker. The amplitude of the chirp is highest here and least influenced by potential other noise. We perform peak localization on the audio signal of this channel by finding the maximum absolute amplitude, and then segment the chirps from all four –synchronized– channels using a window size of 0.046 seconds (2,048 data points to facilitate the use of Fast Fourier Transformation) centered at the peak position extracted from the first channel.

Pose Segmentation

Pose segmentation is based on comparison to a reference signal. This reference signal is recorded during system start when we ask a user to hold their hand still and open –that is to not perform any pose– and record the received signals from all four microphones as reference.

For each segmented chirp, we perform Faster Fourier Transform (FFT) to extract the energy distribution across frequency 0-10kHz. To detect whether a pose is performed, we calculate the Euclidean distance of the FFT results between the current chirp and the reference chirp recorded during system start as described above. If the distance is larger than an empirically determined threshold, we infer that a pose is being performed. We then use the subsequent 0.5 seconds of data from all four channels for the hand pose recognition. This data is also saved for post-analysis.

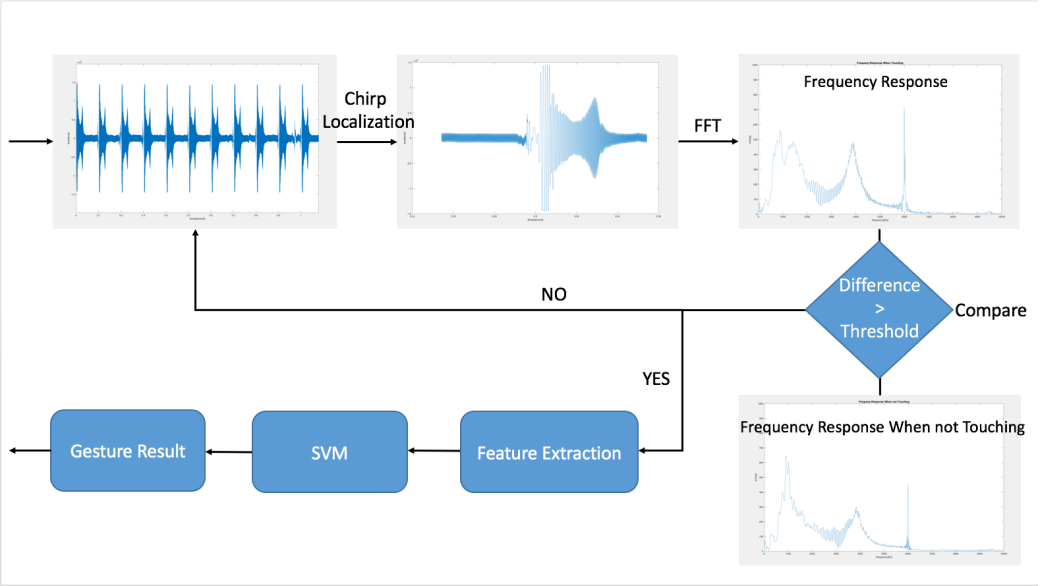

Figure 6. Data Processing Pipeline

Pose Recognition

In the second step, we extract features from each chirp collected from the four channels. For each chirp, we first pass it through a band-pass filter (100 Hz – 5,500 Hz), which is the most informative frequency range based on early exploration. On the filtered chirp. we then extract 35 features, namely: zero crossing rate, energy, entropy, spectral centroid, spectral flux, spectral chroma, spectral roll-off, and Mel-frequency cepstral coefficients [7]. Then we extend the feature vector to 294 by adding the dominant frequency and its energy, as well as spectral energy bins from 100 Hz to 5500 Hz as extracted through the FFT. Finally, we combine the feature vectors of all four channels resulting in a global descriptor of dimensionality $d = 1,176$, which is then fed into a support vector machine pose classification backend. We use the sequential minimal optimization (SMO) implementation of SVM provided by Weka [18].

Since the pose segmentation step actually sends data segments of a length of 0.5 seconds for pose recognition and each chirp takes 0.1 seconds, each channel may contain up to 5 chirps. The final recognition result of a particular pose is thus based on majority voting over the five individual chirp classifications.

USER STUDY

To evaluate the performance of FingerPing and to understand how users would use our system, we conducted a user study with 16 participants (10 male; average age of 26.6). 8 randomly selected participants were requested to test the first pose set (tapping on 12 phalanges) and the remaining 8 participants were instructed to evaluate the second set of poses (digits '1' to '10' from American sign language).

At the beginning of the study, a researcher first introduced and demonstrated the poses that the users were required to perform. Then the researcher helped the participant to put on the system hardware on their hand and wrist. Each participant was allowed to practice each pose before they proceeded to the actual test. Participants were sitting on a chair duing the study.

The study consists of seven separate sessions. During each session the participants were given visual prompt for the pose to be performed. The first session was a practice session, where each participant was instructed to perform all the poses in a random order with five instances per hand pose. Sessions 2 through 5 were used for collecting the training data, in each of which the participants were requested to perform each pose in a random order with five instances per pose. No feedback was given for these first 5 sessions. The last two sessions were used as testing sessions, where the participants were instructed to perform each pose five times in a random order. Unlike the practice and training sessions, the participants were given feedback for real-time classification results. If the classified pose was recognized as the one the participant was asked to perform then the interface showed an icon of green (red otherwise). Manual ground truth annotation was provided by observing researchers during the user study.

Results

We removed the mis-performed poses caused by the participants from the final analysis, where participants failed to perform the hand pose as the stimuli indicated. We dropped 87 out of 1760 instances (22 gestures * 80) from 16 participants. As a result, after removing mis-performed poses, there are 1.5 false positive errors in average in each testing session for both two pose sets. The average accuracy across all participants for the phalanges pose and American sign language are 93.77% and 95.64% respectively. The confusion matrix for the two poses sets is shown in Figures 10 and 11, respectively. For 12 phalange poses, recognition was most accurate for touch events involving the little finger, least accurate when the middle finger is targeted, which can be explained through the very similar path propagation compared to both index and

ring finger. The other common confusion exists between the similar positions in the neighbor fingers. For instance, 6.58% little-distal was misclassified as ring-distal and 6.67% middle-middle was misclassified as index-middle. Figure 9 shows the accuracy for 12 phalanges respectively.

For ASL poses, number '10' presents the highest accuracy of 100% and number '5' has the lowest accuracy 90.54%. We think the reason why '10' is the most accurate pose is that it is very different from any other poses as figure 3 shows. We also noticed there were much confusion between poses '7','8' and '9', which look similar in shapes. Figure 7 and figure 8 show the accuracy for each participant on phalanges and ASL poses. It shows that participants in general achieved a relatively high accuracy on recognizing ASL hand configurations. Also, we received some participants reporting tiredness at the end of the study. This was an issue of the study setup, which required the participants to keep focused on performing different hand poses for around 30 minutes, which is rarely the case in daily scenarios.

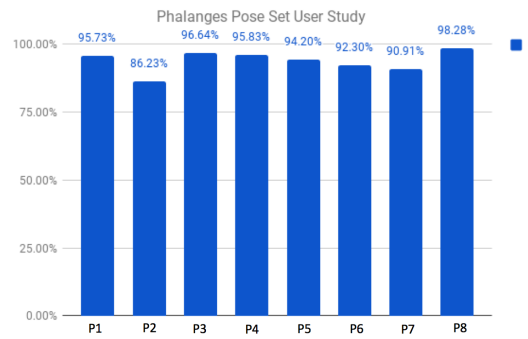

Figure 7. Accuracy for each participant on Phalanges poses

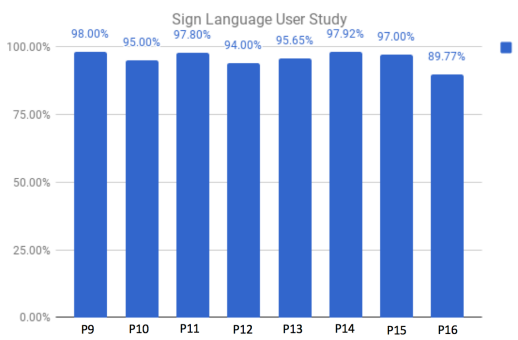

Figure 8. Accuracy for each participant on ASL poses

DISCUSSION

Applications

One obvious application for the phalange pose set is to use them directly for a number input system due to structural similarity and human intuition. If the layout is mapped to a T9 keyboard, it can be extended to be used as an input system for text as well.

Depending on the applications, these poses can be combined or selected to form new pose sets which may influence the recognition accuracy. Many applications do not need to have all 12 poses to be functional. For instance, to control a music player, generally only four buttons (next song, previous song, pause, play) are needed, which can be mapped with the distal of the four fingers in phalange pose set with an accuracy of 93.69% in post-analysis using the data collected in the user study. Similarly, we can also use a subset of the phalange poses to control a D-Pad , which is very helpful to navigate through menus with hierarchy on wearables (e.g., Google Glass, smart watches). The buttons of D-Pad can be naturally mapped to the index-middle, middle-distal, middle-proximal and ring-middle positions. Another interesting mapping can be to use the four corner phalanges (index-distal,index-proximal,little-distal,little-proximal), which presented 94.55% accuracy in our post-analysis. These four poses can be used to control music player or shortcuts to applications.

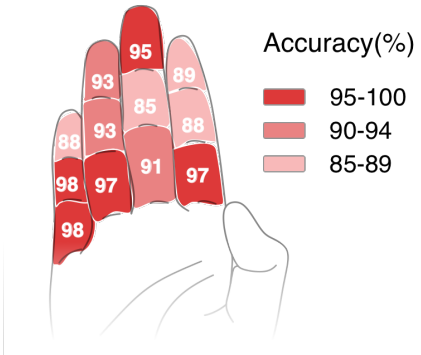

Figure 9. Accuracy for 12 Phalanges (Rounded to nearest integer)

Hardware Improvement

In the current hardware setting, the surface mounted speaker is directly taped on the thumb. However, we actually built different shapes of 3D-printed rings (hard plastic material) to attach the speaker to the thumb. Unfortunately, due to the relatively large size and the rigid shape of the speaker, this solution did not fit well with everyone's hand. One potential solution in the future is to 3D-print rings with flexible material (e.g., rubber), which would be adjustable for different thumb sizes. Another drawback of taping the speaker on the thumb is the airborne communication. In the study, the chirps are audible in the quiet study room. Future design of the ring should consider isolating the speaker to airborne communication by wrapping it up with sound-absorbing material, such that the speaker would not generate much noise to the environments surrounding the participant.

Addressing false-positives

The current implementation utilizes a threshold-based segmentation method to detect the start of the interaction, which can be prone to noise in tough scenarios, such as when the user's hands are touching different objects. One way to address this, is to apply more advanced machine learning pipeline (e.g., hidden Markov model) to automatically transit between different states with a much larger set of training data. The other possible solution is to introduce an activation pose. For instance, to

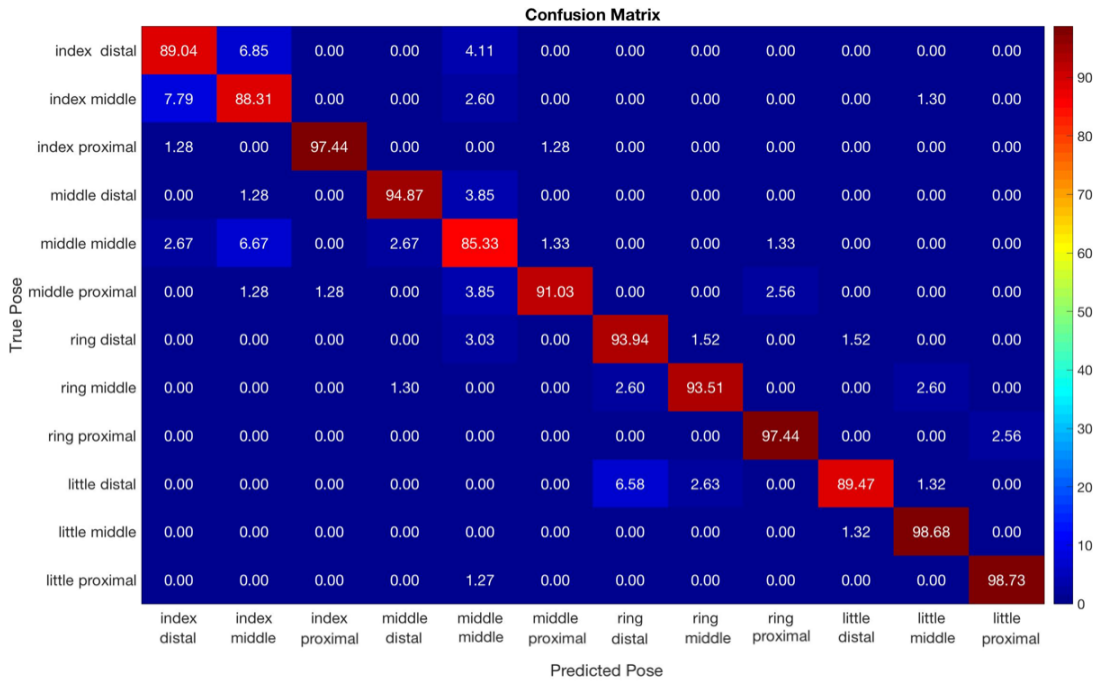

Figure 10. Confusion Matrix for recognizing touch events related to the 12 phalanges.

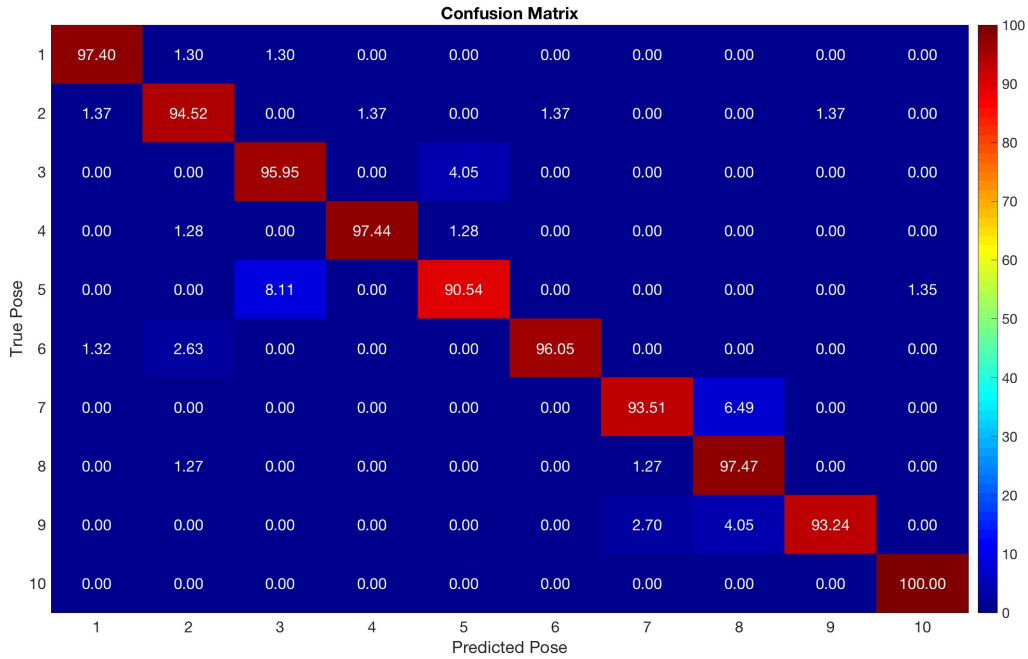

Figure 11. Confusion Matrix for recognizing 10 poses from American Sign Language.

activate the system, the user needs to tap on three phalanges on index finger in a certain, specific sequence, which is hard to trigger by accident. Therefore, the system stays inactive during regular activities and would only be active once the user chooses to perform the activation pose. The activation pose can be selected from the most distinguishable poses such as '10' in the American Sign Language.

Calibration

One limitation of the current system is its potential user or session dependency. A user may need to provide reference data during system start. The reason for this is that the composition (tissue, bones) of the human body is different for each person, which impacts acoustic frequency response. Certainly, with large amounts of training data from huge user cohorts one could create a generic, user-independent (baseline) model. Utilizing more training data, is to attempt a user adaptive system, where we match a few training gestures made by the user to similar-seeming past users and load in those users' training data (or have a user independent model where we re-train with the addition of the new training gestures, weighted highly). Going further, we may discover a calibration process can minimize the per session or per user differences. The calibration process could be incorporated into the use of device. For example, we might require the user to perform a few poses, explicitly chosen to represent the space of input, to "pair" the device to a phone - for example, 6, 9,1, and 10 (thumbs-up). Alternatively, instead of a single pose, we can require that valid input requires a sequence of poses (i.e., a gesture), selected so that using the change in features for recognition is more session/user independent. Also, sensor placement consistency may be improved by well-designed form factors in the future. More experimentation is needed to prove the effectiveness of these approaches. In this current research, we focused on the general proof-of-concept and left the refinement to a user-independent (baseline) model for future work.

Limitation

Another limitation of the current system is its user dependency. A user needs to provide reference data during system start. The reason for this is that the composition (tissue, bones) of the human body is different for each person, which impacts acoustic frequency response. Certainly, with large amounts of training data from huge user cohorts one could create a generic, user-independent (baseline) model. However, given that the initialization is very short and would only be required when a user first puts the sensing device on, in this current research we focused on the general proof-of-concept and left the refinement to a user-independent (baseline) model for future work.

Future Work

In this paper, we have provided preliminary evaluation results on recognizing 22 hand poses, including 10 from American Sign Language. Our natural next step is to explore the recognition of the whole 26 characters in the American Sign Language, which requires much longer practice and training sessions. Note that no structural changes would be required for scaling up the system from the current proof-of-concept to a full alphabet ASL recognizer. Also, while our current results focus on poses to show the concept, the same apparatus should work for gesture with the addition of Dynamic Time Warping or Hidden-Markov-Model for recognition.

The current system requires 0.5 seconds of touch data to recognize hand poses, which limits the theoretical maximum input speed to 2Hz. The length of chirps can be potentially shortened to enable faster input speed, which we plan to investigate in the next step.

The development of the current prototype was focused on the proof-of-concept of the general apparatus as well as the sensor data analysis pipeline. Wider application would require engineering a more compact and especially wireless system. Given the availability of miniaturized, and energy-efficient wireless components, such a development does not represent a substantial burden that would be impossible to overcome.

CONCLUSION

In this paper, we presents FingerPing, a active acoustic sensing technology that can recognize fine-grained hand poses by analyzing how the frequency response changes after traveling through different paths in hands. A user study with 16 participants shows that FingerPing can recognize the tap locations at 12 phalanges and 10 poses from ASL with an accuracy of 93.77% and 95.64% respectively.

ACKNOWLEDGMENTS

We would like to thank all the reviewers for their detailed feedback and the participants in the user study.

REFERENCES

1. Brian Amento, Will Hill, and Loren Terveen. 2002. The Sound of One Hand: A Wrist-mounted Bio-acoustic Fingertip Gesture Interface. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 724–725. DOI: http://dx.doi.org/10.1145/506443.506566
2. Liwei Chan, Yi-Ling Chen, Chi-Hao Hsieh, Rong-Hao Liang, and Bing-Yu Chen. 2015. CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions Through a Fisheye Ring. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 549–556. DOI: http://dx.doi.org/10.1145/2807442.2807450
3. Ke-Yu Chen, Kent Lyons, Sean White, and Shwetak Patel. 2013. uTrack: 3D Input Using Two Magnetic Sensors. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 237–244. DOI: http://dx.doi.org/10.1145/2501988.2502035
4. Gabe Cohn, Sidhant Gupta, Tien-Jui Lee, Dan Morris, Joshua R Smith, Matthew S Reynolds, Desney S Tan, and Shwetak N Patel. 2012. An ultra-low-power human body motion sensor using static electric field sensing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 99–102.
5. Artem Dementyev and Joseph A. Paradiso. 2014. WristFlex: Low-power Gesture Input with Wrist-worn Pressure Sensors. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*. Association for Computing Machinery (ACM). DOI:http://dx.doi.org/10.1145/2642918.2647396

6. Mingming Fan and Khai N. Truong. 2015. SoQr: Sonically Quantifying the Content Level Inside Containers. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 3–14. DOI: http://dx.doi.org/10.1145/2750858.2804264
7. Theodoros Giannakopoulos and Aggelos Pikrakis. 2014. *Introduction to Audio Analysis: A MATLAB Approach* (1st ed.). Academic Press.
8. Jun Gong, Xing-Dong Yang, and Pourang Irani. 2016. WristWhirl: One-handed Continuous Smartwatch Input Using Wrist Gestures. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 861–872. DOI: http://dx.doi.org/10.1145/2984511.2984563
9. Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: Appropriating the Body As an Input Surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 453–462. DOI: http://dx.doi.org/10.1145/1753326.1753394
10. Da-Yuan Huang, Liwei Chan, Shuo Yang, Fan Wang, Rong-Hao Liang, De-Nian Yang, Yi-Ping Hung, and Bing-Yu Chen. 2016. DigitSpace: Designing Thumb-to-Fingers Touch Interfaces for One-Handed and Eyes-Free Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1526–1537. DOI: http://dx.doi.org/10.1145/2858036.2858483
11. Gierad Laput, Xiang 'Anthony' Chen, and Chris Harrison. 2016a. SweepSense: Ad Hoc Configuration Sensing Using Reflected Swept-Frequency Ultrasonics. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. ACM, New York, NY, USA, 332–335. DOI: http://dx.doi.org/10.1145/2856767.2856812
12. Gierad Laput, Robert Xiao, and Chris Harrison. 2016b. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 321–333. DOI: http://dx.doi.org/10.1145/2984511.2984582
13. Christian Loclair, Sean Gustafson, and Patrick Baudisch. 2010. PinchWatch: a wearable device for one-handed microinteractions. In *Proc. MobileHCI*, Vol. 10.
14. Adiyan Mujibiya, Xiang Cao, Desney S. Tan, Dan Morris, Shwetak N. Patel, and Jun Rekimoto. 2013. The Sound of Touch: On-body Touch and Gesture Sensing Based on Transdermal Ultrasound Propagation. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces (ITS '13)*. ACM, New York, NY, USA, 189–198. DOI: http://dx.doi.org/10.1145/2512349.2512821
15. Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. FingerIO: Using Active Sonar for Fine-Grained Finger Tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1515–1525. DOI: http://dx.doi.org/10.1145/2858036.2858580
16. Tauhidur Rahman, Alexander T. Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury. 2014. BodyBeat: A Mobile System for Sensing Non-speech Body Sounds. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '14)*. ACM, New York, NY, USA, 2–13. DOI: http://dx.doi.org/10.1145/2594368.2594386
17. T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. 2009. Enabling Always-available Input with Muscle-computer Interfaces. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology (UIST '09)*. ACM, New York, NY, USA, 167–176. DOI: http://dx.doi.org/10.1145/1622176.1622208
18. Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. 2016. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
19. Cheng Zhang, AbdelKareem Bedri, Gabriel Reyes, Bailey Bercik, Omer T. Inan, Thad E. Starner, and Gregory D. Abowd. 2016. TapSkin: Recognizing On-Skin Input for Smartwatches. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces (ISS '16)*. ACM, New York, NY, USA, 13–22. DOI: http://dx.doi.org/10.1145/2992154.2992187
20. Cheng Zhang, Sinan Hersek, Yiming Pu, Danrui Sun, Qiuyue Xue, Thad E. Starner, Gregory D. Abowd, and Omer T. Inan. 2017a. Bioacoustics-Based Human-Body-Mediated Communication. *Computer* 50, 2 (Feb. 2017), 36–46. DOI: http://dx.doi.org/10.1109/MC.2017.43
21. Cheng Zhang, Anandghan Waghmare, Pranav Kundra, Yiming Pu, Scott Gilliland, Thomas Ploetz, Thad E. Starner, Omer T. Inan, and Gregory D. Abowd. 2017b. FingerSound: Recognizing Unistroke Thumb Gestures Using a Ring. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 120 (Sept. 2017), 19 pages. DOI:http://dx.doi.org/10.1145/3130985
22. Cheng Zhang, Xiaoxuan Wang, Anandghan Waghmare, Sumeet Jain, Thomas Ploetz, Omer T. Inan, Thad E. Starner, and Gregory D. Abowd. 2017c. FingOrbits: Interaction with Wearables Using Synchronized Thumb Movements. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers (ISWC '17)*. ACM, New York, NY, USA, 62–65. DOI: http://dx.doi.org/10.1145/3123021.3123041

23. Cheng Zhang, Junrui Yang, Caleb Southern, Thad E. Starner, and Gregory D. Abowd. 2016a. WatchOut: Extending Interactions on a Smartwatch with Inertial Sensing. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA, 136–143. DOI: http://dx.doi.org/10.1145/2971763.2971775

24. Yang Zhang, Junhan Zhou, Gierad Laput, and Chris Harrison. 2016b. SkinTrack: Using the Body As an Electrical Waveguide for Continuous Finger Tracking on the Skin. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1491–1503. DOI: http://dx.doi.org/10.1145/2858036.2858082