석 사 학 위 논 문
Master's Thesis

생체음향을 활용한 맥락 기반 미세 손 동작 감지 기법

Enabling context-aware micro hand gesture recognition using bio-acoustic sensing technique

2024

김 진 아 (金 辰 婀 Kim, Jina)

한 국 과 학 기 술 원

Korea Advanced Institute of Science and Technology

석 사 학 위 논 문

생체음향을 활용한 맥락 기반 미세 손 동작 감지 기법

2024

김 진 아

한 국 과 학 기 술 원

문화기술대학원

생체음향을 활용한 맥락 기반 미세 손 동작 감지 기법

김 진 아

위 논문은 한국과학기술원 석사학위논문으로
학위논문 심사위원회의 심사를 통과하였음

2023년 12월 5일

심사위원장 윤 상 호 (인)

심 사 위 원 우 운 택 (인)

심 사 위 원 이 기 혁 (인)

Enabling context-aware micro hand gesture recognition using bio-acoustic sensing technique

Jina Kim

Advisor: Sang Ho Yoon

A dissertation submitted to the faculty of
Korea Advanced Institute of Science and Technology in
partial fulfillment of the requirements for the degree of
Master of Science in Culture Technology

Daejeon, Korea
December 5, 2023

Approved by

Sang Ho Yoon
Professor of Graduate School of Culture Technology

The study was conducted in accordance with Code of Research Ethics[1].

[1] Declaration of Ethical Conduct in Research: I, as a graduate student of Korea Advanced Institute of Science and Technology, hereby declare that I have not committed any act that may damage the credibility of my research. This includes, but is not limited to, falsification, thesis written by someone else, distortion of research findings, and plagiarism. I confirm that my thesis contains honest conclusions based on my own careful research under the guidance of my advisor.

MGCT 김진아. 생체음향을 활용한 맥락 기반 미세 손 동작 감지 기법. 문화기술 대학원 . 2024년. 28+iii 쪽. 지도교수: 윤상호. (영문 논문)
Jina Kim. Enabling context-aware micro hand gesture recognition using bio-acoustic sensing technique. Graduate School of Culture Technology . 2024. 28+iii pages. Advisor: Sang Ho Yoon. (Text in English)

초 록

손동작 기반의 조작방법은 가상, 증강현실(VR, AR)에서 주된 입력방식으로 사용되어지고 있으며, 특히 작은 손의 움직임을 의미하는 마이크로 제스쳐는 미세한 손의 상호작용을 강건하고 자연스럽게 만든다. 하지만 기존의 제스쳐 인식 방식은 사용자가 제스처를 수행할 때 존재하는 맥락과 제약상황에 대한 고려가 부족하다. 따라서, 본 논문에서는 손가락 기반 마이크로 제스처를 인식하는 동시에 상호작용 맥락을 인식하기 위해 생체음향을 활용한 능동 및 수동 센싱 접근 방식을 채택한다. 이를 위해, 손에서 발생할 수 있는 맥락 정보를 분류하고 상호작용 기법의 동작 방식을 디자인하였다. 또한, 생체음향 시스템의 하드웨어 구성 및 신호처리방식을 조사하고 최종 파이프라인 알고리즘을 제안하였다. 나아가, 다수의 사용자 데이터 취득 후 시스템의 성능을 평가하였으며 이를 적용할 수 있는 사용자 시나리오를 구성하여 추후 증강현실에서의 적응형 입력방식의 가능성을 제시했다.

핵 심 낱 말 인간-컴퓨터 상호작용, 증강현실, 웨어러블 제스쳐 인식

Abstract

The use of microgestures has improved the robustness and naturalness of subtle hand interactions. However, the conventional approach often neglects the context in which users perform microgestures. Therefore, this thesis proposes context-aware tap and swipe gesture recognition using bio-acoustic sensing. The recognition system in this study adopts both active and passive sensing approaches to recognize finger-based microgestures while also recognizing the associated interaction contexts, including within-hand, surface, and hand grasp. In this study, we investigate the hardware configuration and sensing approach to form a bio-acoustic system with multiple band-pass filter processing. We also validate the accuracy of context-aware tap and swipe gesture recognition and propose a context-aware microgesture recognition pipeline to enable adaptive input controls for rich and affordable hand interactions.

Keywords Human-computer interaction, Augmented reality, Wearable gesture recognition

Contents

List of Figures

Chapter 1. Introduction

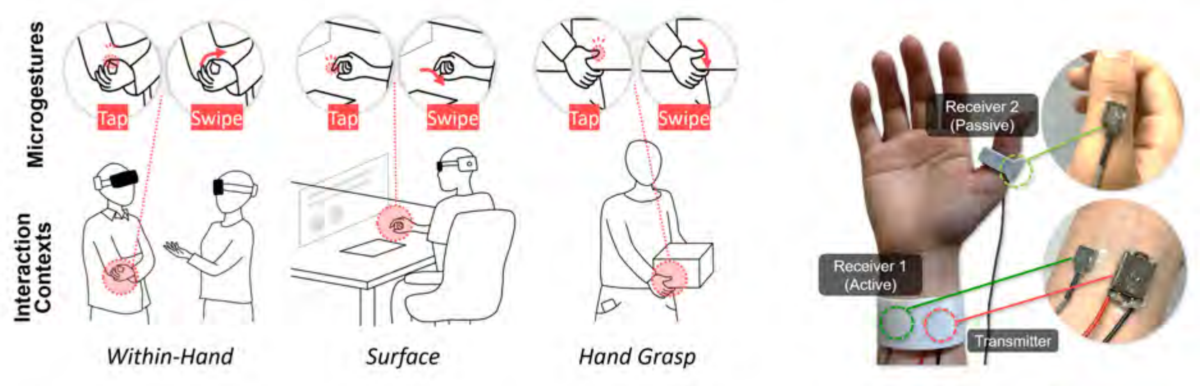

Figure 1.1: Our proposed system understands the contexts associated with tap and swipe gestures. We aim to support three interaction contexts including *Within-Hand*, *Surface*, and *Hand Grasp*. We employ sensing nodes at the wrist and finger which capture active and passive signals accordingly. The components are small enough to be placed on the finger and the wrist.

With the advancement in wearable sensing techniques, researchers have explored a broad category of hand interactions, including hand activity recognition [14], coarse-grained hand gesture [30, 21, 22], microgestures [31], and subtle finger interactions [2, 3, 8]. Particularly, micro hand gestures like tap and swipe using a single hand have been adopted for contemporary augmented and virtual reality (AR/VR) interfaces to provide less distraction, reduced fatigue, and improved privacy [42]. To this extent, on-body sensing approaches promoting user comfort in a socially acceptable form factor have been suggested for microgesture recognition [9, 49, 50].

Previously, various sensing approaches have shown the potential of microgesture recognition with robust performance [15, 23]. However, these works mainly focused on recognizing basic gestures or discrete hand postures. Still, extra care must be taken to recognize micro hand gestures under various contexts including surfaces and hand grasps. This is because the hand gesture itself cannot fully express the user's intent, as the same gesture may have different meanings depending on the situation. To further expand the scope of interaction, Grasping Microgestures [38] introduced subtle and rapid microgestures with busy hands. In addition, researchers explored tapping and swiping gestures on different interaction contexts, including rigid surfaces [48, 51] and within the hand [52, 31]. To this end, recent work utilized a high-frequency AC circuit with two inertial measurement units (IMUs) to support within-hand/hand-to-surface/and hand-to-object interactions [7]. In our work, we mainly focus on recognizing microgestures robustly under different contexts. We further advance the sensing capability by understanding the surrounding environment where tap and swipe gestures occur (*e.g.*, tap while grasping a pen, swipe on a rigid desk).

To enable microgesture recognition, a wide variety of sensing modalities has been considered in the form of wrist-worn and finger-worn devices. The wrist and fingers have been of particular interest since they convey user intent precisely with high flexibility and social acceptability [61]. Previously explored sensing approaches would include utilizing muscle activation [54], motion [15], computer vision [13], and RF waveguide [53]. These sensing mechanisms focus on either passive or active sensing methods for capturing coarse-grained hand gestures or microgestures accordingly. Instead, we take advantage of bio-acoustic, which can acquire both active and passive acoustic signals. Active sensing is to use actively emitting sound signals to analyze the responses & signal change and passive sensing means analyzing signals transmitted through bone conduction or skin propagation. For our proposed work, we capture motion-induced low bandwidth (passive) and transducer-generated broad bandwidth (active) bio-acoustic signals. It allows us to acquire rich information from microgestures to recognize discrete gestures and contexts associated with the gestures.

To this end, We propose bio-acoustic sensing that enables context-aware microgesture recognition through the use of both active and passive bio-acoustic sensing. Our approach focused on recognizing tap and swipe gestures while understanding their interaction contexts of *Within-Hand*, *Surface*, and *Hand Grasp*. This allows us to cover broader interaction scenarios even with the same gestures. Moreover, our system reduces instrumentation within the hand region by placing a single receiver on the thumb while locating a receiver and a transmitter on the wrist (Figure 1.1). We further quantify the accuracy of performing tap and swipe gestures in different contexts and demonstrate exemplary cross-context interactions for AR and grasp interfaces. Our contributions are as follows:

- A bio-acoustic sensing technique utilizing active and passive signals to recognize robust tap and swipe gestures;
- A context-aware gesture recognition pipeline for classification of associated contexts (*Within-Hand*, *Surface*, and *Hand Grasp*) with microgestures;
- Exploration of bio-acoustic hardware configuration to understand deeper interaction contexts with minimal instrumentation within a hand and wrist;

Chapter 2. Related Works

2.1 Hand Interaction with Wearables

Wearable sensing techniques have been employed to support natural hand interactions by recognizing hand activities and gestures [46]. The emergence of AR/VR devices facilitates the use of wearable sensing techniques for input control [47]. Among various body parts, the wrist and fingers have been preferred to employ wearable sensors with less obtrusiveness and availability of rich and direct sensing signals from hand gestures.

Wrist-worn Wearables Wrist-worn sensing approaches have been explored to recognize hand gestures and reconstruct hand posture by utilizing Surface Electromyography (sEMG) [19], electrical impedance tomography (EIT) [40], computer vision [13], pressure [41], capacitive [43], IMU [15, 12], and acoustic sensing [21, 48]. To further augment the performance of the recognition task, previous works combined distinctive sensing modalities. For instance, EmPress [20] combined sEMG and pressure sensing modalities to improve the accuracy of hand gesture recognition. On the other hand, TapID [15] added additional accelerometers to the existing smartwatch form factor to provide reliable and quick touch detection for the VR input. Recent works utilized off-the-shelf smartwatches to robustly recognize finger gesture [12], hand activity [14], and customized hand gestures with a few-shot learning [18].

Finger-worn Wearables Finger-worn wearable systems have been populated for detecting fine-grained hand gestures. These include fine-grained finger tracking with microphone [55], micro-finger poses with proximity sensors [3], hand poses with acoustic sensing [23], and subtle pinch and touch detection by coupling AC signal to the body [2, 7]. Additionally, Magnetic sensing [17, 16] and computer vision [45] approaches have been investigated. Furthermore, the finger-worn wearable devices robustly supported tap interactions with various surfaces [1, 8]. For robust and effective hand interaction, previous works focused on achieving highly accurate recognition for coarse- and fine-grained hand gestures. Our approach extends the capabilities of hand interactions by developing both wrist- & finger-worn system, as both locations contain rich implications of the environment where tap and swipe occur.

2.2 Bio-Acoustic Sensing in HCI

Bio-acoustic sensing technique has been employed in HCI to recognize various hand-related interactions including hand gesture [32], contact detection [27], tracking [51], and identification [28].

Passive sensing For hand gesture recognition, researchers utilized the passive sound signal transmitted through bone conduction propagation from the skin vibration [32, 56]. However, these works inherited an armband form factor that limits capturing fine-grained hand gestures. To this end, recent works adopt a smartwatch to acquire bio-acoustic signals to recognize various hand activities including grasp object sensing and gestures [30, 14]. Recent works placed sensors on the wrist and fingers to acquire robust signals [26, 25, 29, 35]. However, the passive sound signals generated by microgestures tend to degrade quickly over distance.

Active sensing To this end, researchers employed active bio-acoustic approaches using audible (∼18 kHz), ultrasonic (20 kHz∼), and wide (0∼48 kHz) [23, 31, 24] frequencies. By adding an acoustic transmitter (*e. g.*, surface transducer or speaker), these works robustly recognized microgestures including thumb-to-finger tap and fine-grained hand gestures. Still, these works require both receiver and transmitter to be equipped within the hand (*e. g.*, finger or back of the hand). However, removing the surface transducer within the hand would be desirable since it requires a large footprint due to the associated hardware and battery. Meanwhile, Touch&Active [6] implemented a grasp interface using active acoustic sensing applied to objects and Interferi [50] developed on-body gesture recognition using acoustic interferometry on arm and face. Still, previous works do not fully support various interaction contexts like different hand grasps gestures.

Based on the feasibility of acoustic sensing shown in the previous studies, we propose a bio-acoustic method using both passive and active sensing to achieve comparable microgesture recognition performance in various interaction contexts. We exploit rich features using numerous bandwidth covering from passive (10∼500 Hz) to active (10∼6,000 Hz) bio-acoustic signals.

2.3 Microgestures in HCI

Microgestures are defined as small movements of the digits that do not require moving the whole hand commonly performed but rarely noticeable [33]. This allows users to perform the gestures anytime and anywhere [36]. The main application of microgestures was eyes-free interaction during everyday activities [10, 11]. Nowadays, the scope of microgestures broadly covers from a full set of thumb-to-finger gestures to 3D microgestures for providing expressive and precise interactions in AR/VR [9, 42]. In particular, tap and swipe gestures using the thumb, index, and middle fingers take a large portion of microgestures based on the elicitation study [36].

To advance microgestures as input, various attempts have been explored. First, the grasping microgesture concept has been introduced with its superior performance when on the move and hands are busy [38]. Researchers showed that finger movement in grasping microgestures was rapid, easy, and elegant to perform [39]. Using hand grasp information [37], the same microgestures could be used to interact with different applications [33]. Moreover, hand grasp itself could be used as a user interface [34]. Other attempts were to recognize surface or object materials that users interact with to provide distinctive control based on detected materials [63, 62, 60]. Thus, recognizing deeper interaction contexts like *Surface* and *Hand Grasp* along with microgestures has a high potential to provide rich interactions. We aim to support surface- and grasp-aware tap and swipe gestures through a bio-acoustic sensing technique.

Chapter 3. System Design Rationales

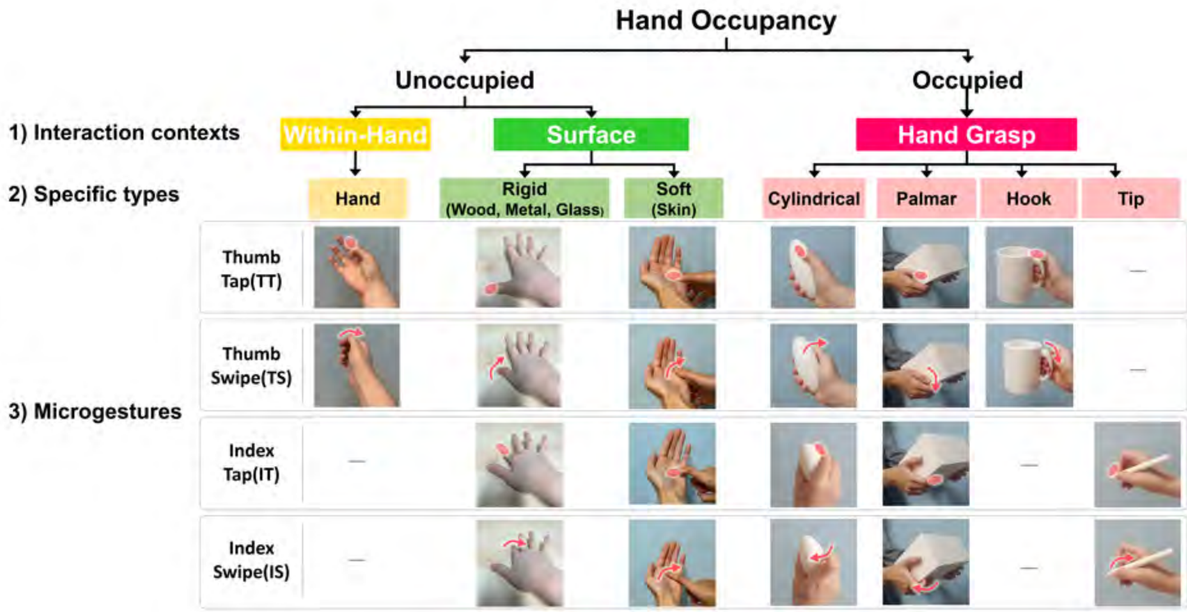

Figure 3.1: Our system supports a set of gestures categorized in *Within-hand, Surface* and *Hand Grasp*. Each context has a more specific type with a tap and swipe gesture.

3.1 Context-Aware Design Parameter

Interaction contexts To be more specific, we categorize the interaction context into *Within-Hand*, *Surface*, and *Hand Grasp* similar to previous work [7]. We take one step further by presenting each classifier of interaction context that instantly differentiates both deeper contexts and gestures, as well as the whole pipeline of how they distinguish context-aware microgestures, which was underinvestigated by the previous work. From previous studies, we learned that the interacting surface's material could provide context cues for interactions [4, 5]. Classifying the surface's rigidity makes it possible to change the interaction context without extra gestures or input commands. For example, users tap and swipe on the desk to perform work-related interface whereas on the skin to incur personal interface. Additionally, Grasping microgestures have also been highlighted to enable interactions under physical, temporal, and socially constrained hand-busy situations [38, 39]. In this work, we would like to explore the potential of using surface materials and hand grasps as an interaction medium.

Specific types We also consider the interacting surface material stiffness as another context cue for interactions based on several studies detecting the material of an object using acoustic signal. Classifying the surface by soft or rigid material makes it possible to incur mode changes in the interacting environment without requiring extra gestures or input commands.

To recognize grasping microgestures, the system needs to differentiate basic hand grasps [37] including *Power*, *Intermediate*, and *Precision* grips. We defined 4 representative hand grasps consisting of 3 palm-grasps (Cylindrical, Palmar, and Hook) and 1 side-grasp (Tip). Rather than recognizing the specific object users hold, we focus on detecting hand grasp type to infer interaction context. This approach would be suitable for context-aware interactions since utilizing hand configuration is more accessible than requiring actual artifacts for initiating interactions [59].

Microgestures Previously, researchers emphasized the importance of microgestures including *Tap, Press, Stretch, Swipe, and Draw* which have the potential to provide direct and subtle interactions [38]. Also, microgestures have been preferred for AR/VR interfaces among other available input modalities like voice or keyboard/mouse. Here, tap and swipe gestures are preferred with their ease of use, conceptual simplicity, and resemblance to interactions in touchscreen-based devices among various microgestures [42]. Furthermore, the thumb and index finger have shown high flexibility and comfort [58, 36]. To this end, we select tap and swipe performed by thumb and index finger as our representative microgestures (Figure 3.1).

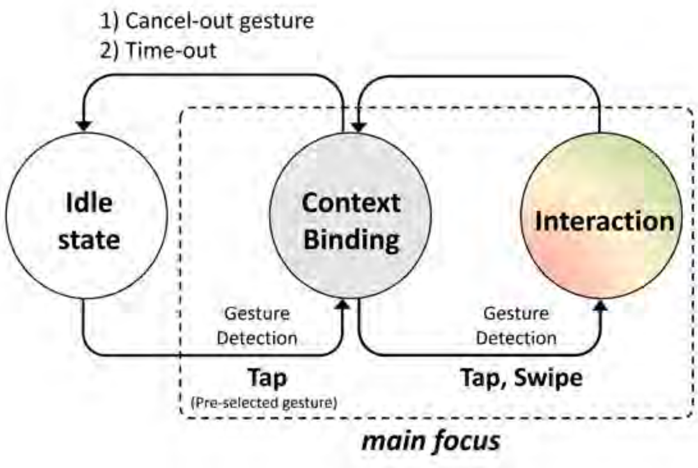

Figure 3.2: Our proposed interaction workflow focuses on the context binding and the execution of a microgesture after the binding is initialized.

3.2 Towards Robust Context-Aware Interaction

We propose a microgesture recognition with context binding approach to provide a robust interaction workflow. Figure 3.2 illustrates the workflow consisting of 3 states, including *Idle*, *Context Binding*, and *Interaction*. The system resides in *Idle* state if no pre-selected gesture is detected to prevent false triggers during daily activities. When users perform a pre-selected gesture (tap in our case) under defined contexts, the system enters *Context Binding* state where the system is ready to recognize tap and swipe gestures for the interaction. We intentionally add *Context Binding* state before *Interaction* state so that users could still cancel out executed interaction in case of 1) unintended trigger or 2) misclassification of interaction context. Here, users could either perform cancel-out gesture (*e.g.*, double-tap) or do nothing to go back to *Idle* state. The time-out duration remains 5 seconds. Otherwise, the system goes into the *Interaction* state, where users' subsequent tap or swipe gestures occur in the bound context that they intended. After the gesture is performed, it goes back to context binding state.

Although vision-based controls support intuitive and direct interaction with hand-tracking capabilities, it is not possible to support fast and subtle hand interactions while understanding the external field of view (FOV). To this end, we designed the system to enable robust and real-time interaction while understanding the surrounding interaction contexts by tapping or swiping with no FOV limitation. With the proposed system, we aim to integrate context awareness into microgesture-based interactions for a seamless cross-spatial interaction experience.

Chapter 4. Sensing Principle

4.1 Bio-Acoustic Sensing Principle

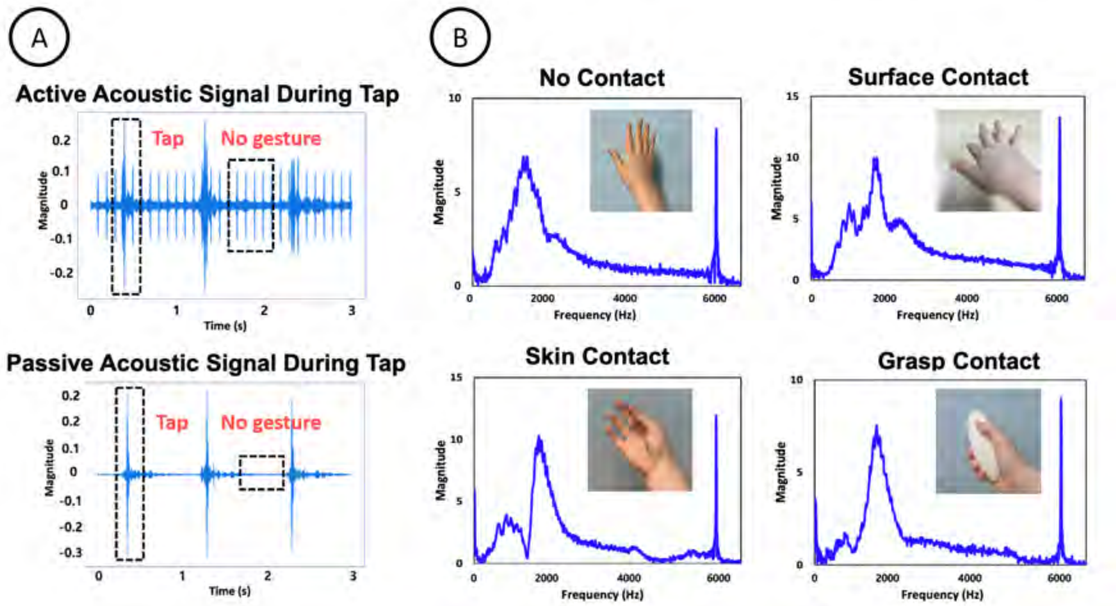

Figure 4.1: (A) Raw acoustic signals during tap gesture showed large amplitude change. (B) We measured frequency responses to different contact conditions that exhibited distinctive active acoustic signal patterns.

The proposed tap and swipe recognition method is based on on-body bio-acoustic sensing that analyzes and compares acoustic signals' temporal and spectral properties change. The acoustic signal contains anatomical information about body structures such as bone, cartilage, tendon, and muscle tissues [65]. The acoustic signal could also capture distinctive features directly from the objects with various physical properties [5].

First, we obtained a change in acoustic signal from various hand states. Based on the previous work, the shape of the hand with a varied configuration of bones and muscles affects properties of acoustic waves [31]. Varying hand configuration affects how sound travels on the hand. The intensity of the signal is either increased or decreased depending on the amount of tissue/bone that was in the path of the wave [23]. Second, we also obtained bone conduction sound propagation from the fingers to the wrist. The acoustic measurements at the wrist reflect the tendon movements related to the finger [25]. When users perform a tap gesture, the signal transfers through the bones of the hand to the wristband microphone [64].

When users perform microgestures, the sound waves generated and affected by finger movements and hand configurations are transmitted via bone conduction. This provides a better signal-to-noise ratio (SNR) than airborne sound [26]. Thus, we utilized on-body passive and active acoustic sensing approaches.

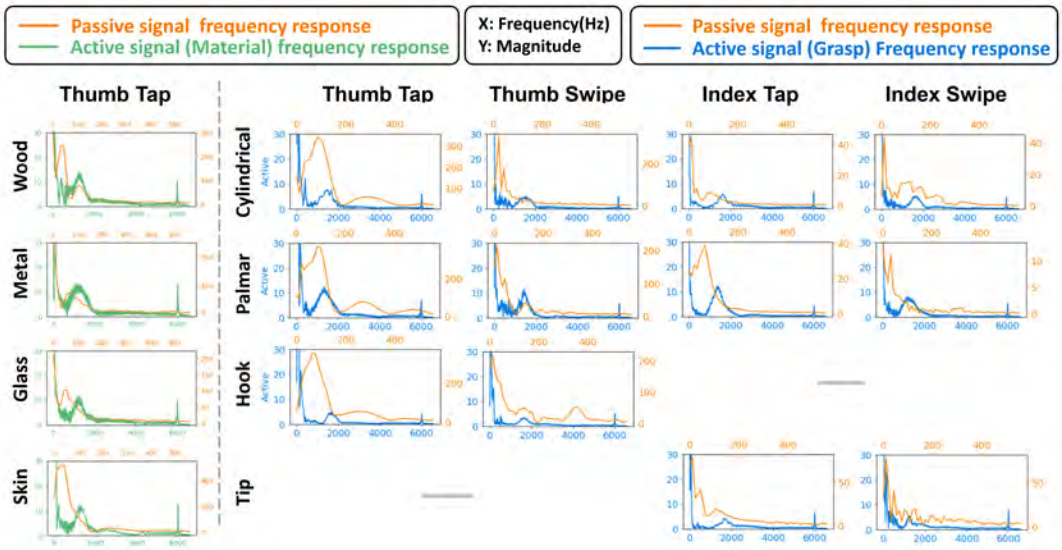

Figure 4.2: Observation of frequency response for various surfaces and hand grasps.

4.2 Bio-Acoustic Sensing Technique

Our sensing technique utilizes active and passive acoustic signal propagation around the hand. Rather than relying on either passive or active method [23, 7], we aim to capture both types to understand the further context from microgestures. We placed a transmitter and a receiver on the wrist (active setup), and another receiver on the finger (passive setup). This setup allows us to acquire the signal from the wrist and the finger separately.

Figure 4.1A shows the raw signal when the gestures are made for a certain period of time. We observe that vibration generated by gestures could affect both active and passive acoustic signals. Compared to the passive acoustic signal, the active acoustic signal shows the periodical amplitude change from the sweep signal emitted by the transmitter. In addition, we confirmed that active bio-acoustic sensing contributes to the unique behavior according to the contact state of the hand as shown in Figure 4.1B.

For active acoustic sources, we applied frequency sweep up to 6,000 Hz which retained maximum information during body propagation [57]. The frequency responses of different contact states calculated by the Fast Fourier transform are distinguishable where magnitude attenuates or increases compared to the non-contact state. The active bio-acoustic method supports microgesture recognition under different contact conditions of the hand, which was not possible with passive acoustic signals only.

We also examined acoustic signal behaviors when carrying out gestures under different conditions (Figure 4.2). The frequency response of active signals under 6,000 Hz and passive signals under 500 Hz exhibit distinctive behaviors upon gestures. These sensor behaviors demonstrate the potential of providing context-aware gesture recognition with discernible and rich sensor signals.

4.3 Apparatus

Figure 4.3A illustrates the overall hardware configuration of our work. The transmitter signal was generated by a surface transducer driven by a function generator. The 1-axis accelerometers were connected to an audio interface (US-4X4HR, TASCAM) to amplify and digitize the analog signal. The audio interface was connected to a 13” 2019 MacBook Pro with a 2.4 GHz Intel Core i5 processor at a

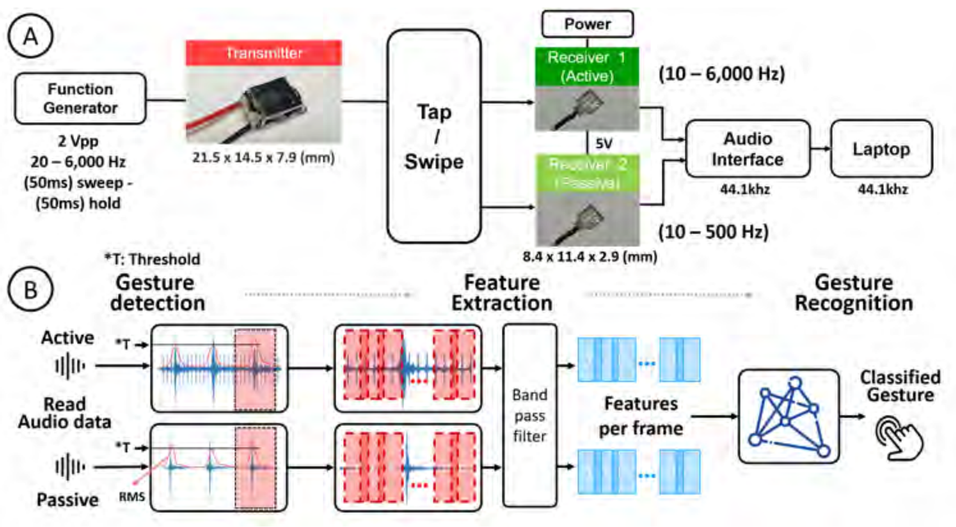

Figure 4.3: (A) Proposed system hardware configuration and (B) gesture recognition processing workflow.

sampling rate of 44.1 kHz.

Transmitter As an active acoustic source, we used the surface transducer (COM-10917, Sparkfun). We attached the transducer to the wrist with pressure-sensitive adhesive (PSA, 468MP, 3M) along with a silicone wristband to support firm attachment. The transducer was driven by a function generator (DG1022, 2CH, 100 MSa/s, RIGOL) that emitted sinusoidal sweep signals of 20∼6,000 Hz at 2 Vpp. The sweep signal increased linearly from 20 to 6,000 Hz for 50 ms and then held at 6,000 Hz for another 50 ms. The duration of one chirp was 100 ms. We specify the frequency range with reference to a previous work [23].

Receiver A two 1-axis accelerometers (VS-BV203-B, KEMET) were chosen as our receivers because of their high sensitivity, built-in amplifier, and wide working frequency bandwidth (10∼15,000 Hz). Using this receiver, we were able to robustly capture both passive (∼500 Hz) and active (∼6,000 Hz) acoustic signals to support the 10 to 6,000 Hz frequency range. The size of the accelerometer is 8.4 mm × 11.4 mm × 2.9 mm, which is applicable for wearables. The same adhesive and silicone wristband was applied to fix the receiver firmly on the user's wrist. We also attached the receiver to the back of the thumb where we used PSA.

4.4 Context-Aware Tap & Swipe Recognition

The processing pipeline is divided into 3 steps including *Gesture Detection*, *Feature Extraction*, and *Gesture Recognition* (Figure 4.3B). We expect that the change in the frequency response from microgestures would create unique features for the recognition tasks. With the proposed system, we further explore how different surfaces and hand grasp affect the frequency response.

Gesture Detection It is essential to detect the occurrence of the gesture for processing gesture recognition. Figure 4.3B illustrates the use of RMS (Root Mean Square) which is the average loudness in the waveform as a cue for detecting gestures. When a finger touches any surface, it produces distinguishable acoustic signals which also increases the RMS. We employed RMS over conventional signal processing like Short-time Fourier transform (STFT) since the RMS supports a real-time system with low computation requirements. We used both active and passive acoustic receivers for gesture detection and they compensate each other for the occurrence of false triggers. When both receivers' RMS values exceed thresholds, a fixed-length segment of data before and after the peak of RMS is extracted for machine learning purposes.

Feature Extraction For feature extraction, we applied multiple bandpass filters to obtain more unique features from raw acoustic signals. Inspired by previous work on applying diverse frequency bandwidths on wearables to improve the gesture recognition performance [24], we carefully selected bandpass filters that best reflect the characteristics of microgestures along with associated contexts. Here, we used spectral features including Linear Frequency Cepstral Coefficients (LFCC), centroid, roll-off, flatness, bandwidth, flux, entropy, mean, standard deviation, sum, maximum, and minimum. LFCC is suitable for equally extracting features over the sensing range compared to Mel-Frequency Cepstral Coefficients (MFCC) [24]. We also utilized waveform features including RMS, variance, entropy, and zero crossing rate. All features were extracted on sliding windows.

Gesture Recognition We used a Support Vector Machine (SVM) provided by the scikit-learn library as a classification algorithm. The extracted data sample used to train the model is applied min-max feature scaling normalization to ensure that all features have a similar range. We chose a polynomial kernel because it gave the best result. And finally, it classifies tap and swipe gestures.

Chapter 5. Pilot Study

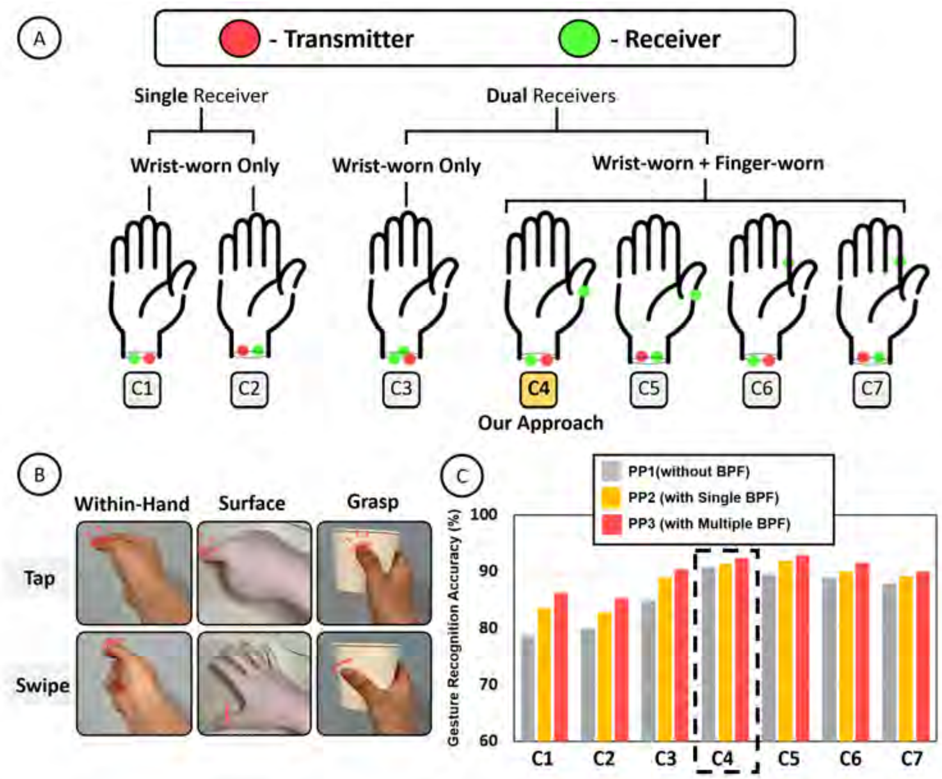

Figure 5.1: (A) Sensor configurations candidates, (B) gesture set for a pilot study, (C) Classification accuracy comparison for representative configurations and pre-processing methods.

5.1 Study Design

We conducted a pilot study to verify the basic performance of tap and swipe gesture recognition on a variety of hardware configurations (5 participants, 2 males, 3 females, mean age 24) to select the appropriate sensor placement. As shown in Figure 5.1A, we chose the wrist, thumb, and index finger as candidate locations for the receiver. Due to the relatively large size and the potential signal interference to the receiver, we only considered the wrist for placing the transmitter. Here, we compared the gesture recognition performance among various pre-processing methods to identify bandpass filters and features that work well with acoustic sensing.

Setup To investigate sensor configuration, we came up with 7 configurations (**C1**$\sim$**C7**) as shown in Figure 5.1A. We limit the number of receivers to 2 or below to minimize the sensor requirement. We placed the receiver and the transmitter on the anterior and posterior wrists while keeping them around the radius and the ulnar on each side. Regardless of the user's hand size, the landmarks for sensors of the wrist (radius, ulnar) and back of the finger (thumb, index) were marked and the sensor was attached to them. The selected locations on the wrist support efficient transmission of the bone-conducted vibration throughout the hand [44]. For **C4** to **C7** in Figure 5.1A, we added another receiver on either the thumb or index finger to capture passive acoustic sensing directly from gestures.

Procedure To acquire data, we recorded the amplified data from an audio interface using a 1-axis accelerometer (VS–BV203–B, KEMET). We provided visualized instructions for users to perform a set of gestures as shown in Figure 5.1B. Before the study, each participant had a practice session. The study contained 5 sessions, each consisting of 10 trials of 6 gestures in random order. Between each session, participants took a 30-second break and researchers checked and adjusted the location of the receiver and the transmitter if needed. A total of 18,000 samples were acquired: 3000 samples for **C1** & **C2** (single receiver, 5 participants × 2 configurations × 5 sessions × 6 gestures × 10 trials) and 15,000 samples for ***C3∼7*** (5 participants × 5 configurations × 5 sessions × 6 gestures × 10 trials × 2 receivers). We segment the data to contain at least 0.5 s long information for each sample (0.25 s before and after the onset of peak point). Then, we applied 3 different pre-processing methods to compare the gesture recognition performance as below.

Result We trained the SVM model (per-user) using 1 to 4 sessions and tested it on the 5th session. Figure 5.1C illustrates the gesture recognition accuracy on different configurations (**C1∼C7**) and pre-processing methods (**PP1∼PP3**). Regarding the pre-processing method, we observed an improvement in recognition accuracy from **PP1** to **PP3**. The performance was improved when employing multiple bandpass filters. This tells us that it is crucial to focus on the effective range of bandwidth to extract meaningful acoustic features. For all pre-processing methods, **C4** achieved the highest average accuracy of 91.5%. As expected, a single receiver shows worse gesture recognition accuracy (83%) compared to dual receivers (88%). We also noticed that the passive acoustic receiver worked best while attached to the thumb compared to the index finger and the wrist. To this end, we picked **C4** (1 transmitter & 1 receiver on the anterior wrist along with another receiver on the thumb) as the main hardware configuration.

5.2 Insight For Context-Aware Sensing

In this research, we adopted the **C4** (thumb & wrist anterior side) due to the overall better performance and future hardware design, even though the final result of **C5** (thumb & wrist posterior side) shows the highest accuracy with **PP3**. Regarding future hardware design, it is common for the MCU and other sensors of typical smartwatches to be located on the wrist's outer side. Due to the active setup, we found it more appropriate to avoid interfering with existing areas and add elements inner side of the wrist where there is room for integration.

Previously, we observed that using multiple bandpass filters influenced the performance of the ML model. As confirmed in the pilot study, we chose 4 bandpass filters for the active signal and a single bandwidth for the passive signal as shown in Figure 6.1. We decided on the final bandwidths based on the following reasons.

- *10∼100 Hz*: Bandwidth including coarse human activity
- *100∼3,000 Hz*: The most changeable bandwidth where the active signal can be affected
- *3,000∼6,000 Hz*: Rest bandwidth of whole bandwidth excluding 10∼3,000 Hz
- *10∼6,000 Hz*: Full bandwidth covering active acoustic source

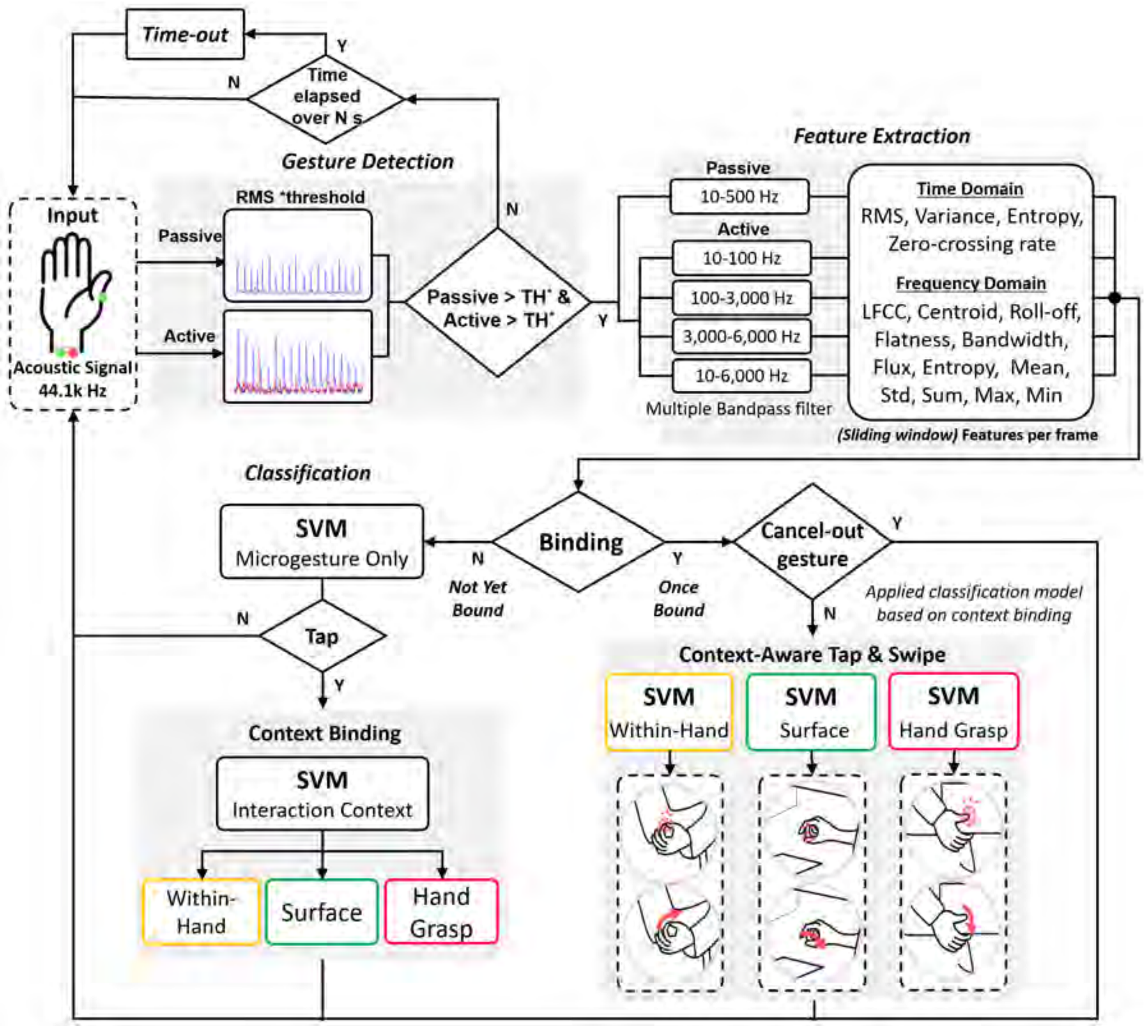

Figure 6.1: Context-aware microgesture recognition pipeline using bio-acoustic sensing.

Our system employed a context binding to initialize the context-aware gesture recognition workflow as shown in Figure 6.1. We designed multiple SVM classifiers to support robust interaction. Initially, we used the tap gesture as the pre-selected gesture to establish context binding. Once bound, the subsequent gesture recognition operates under the same bound context until cancel-out gesture or time-out events occur.

Feature Extraction Procedure We store acoustic signals from active and passive receivers in 2 buffer queues using *PyAudio*. Each queue contains 12,288 data points (6 chunks × 2048 data points) and computes each RMS value of the last chunk. If they both exceed each of their thresholds, the system stores and combines 5 additional chunks to form 11 chunks. We form 0.5 s amount of data (22,050 data points) by trimming 478 data points from the beginning. Then, we used bandpass filters and extracted the features from the filtered data with a 4,096-point Hamming window. Here, the window shifted with a length of 1,024 points. We selected features from time and frequency domains (Figure 6.1) which contain 3,535 feature dimensions from 707 features $\times$ 5 bandpass filters. Lastly, we concatenated features into a single list and applied normalization.

Overall Classification Process After feature extraction, the input features are fed into the classifier depending on several conditions. If the context has not been bound, the input features are first passed to *Microgesture Only Classifier* to detect the initialization gesture, and if it is a tap gesture, it goes to *Interaction Context Classifier* for context binding initialization. Meanwhile, if the context has already been bound, they are fed into the *Context-Aware Tap & Swipe Classifier*. Here, we applied the classification model based on the type of bound context, and it will run unless it is a cancel-out gesture. In our work, we recognize, down to the specific gesture, which is different from previous works [7, 60] where a series of operations were required to perform gesture recognition under various contexts. The overall processing of feature extraction and prediction took 78 ms and 15 ms, respectively. The total latency took up to 300 ms which reflects the time to capture 5 additional chunks of data (232 ms) upon gesture detection. A commonly permitted delay to work in real-time of hand gesture recognition is less than 300ms. Even though it is a little bit over, it could reduce the latency with advanced processing methods.

Chapter 7. System Evaluation

7.1 Study Setup

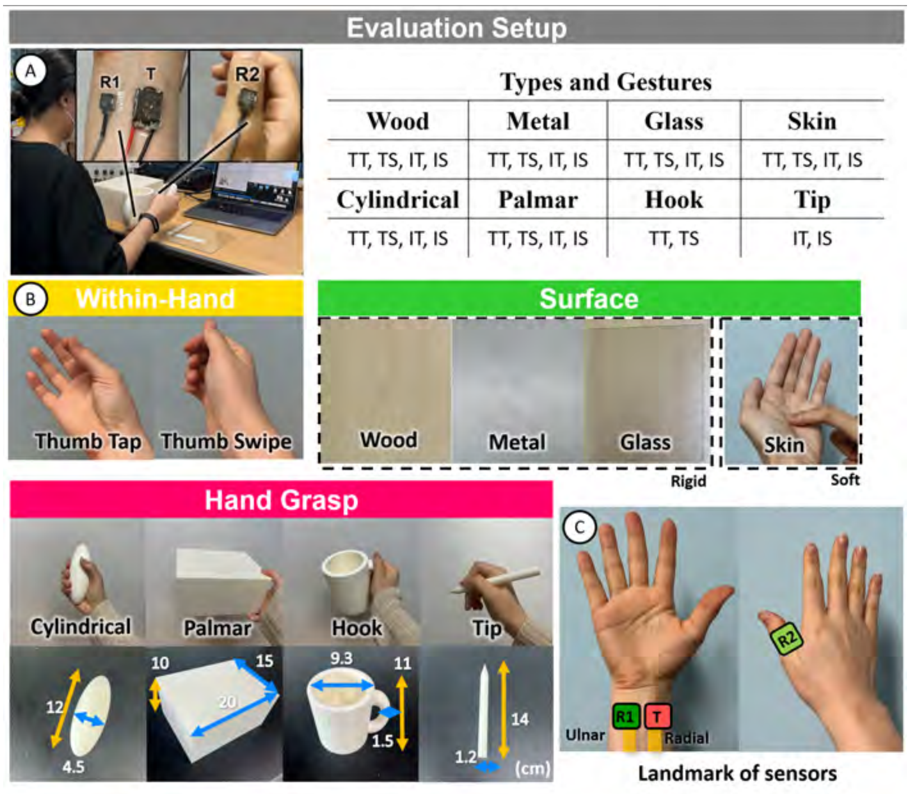

Figure 7.1: (A) System evaluation setup. The table shows a gesture set used in the user study (TT: Thumb Tap, TS: Thumb Swipe, IT: Index Tap, and IS: Index Swipe). (B) We evaluated *Within-Hand*, *Surface*, and *Hand Grasp*. For *Hand Grasp*, we provided 3D printed objects (dimensions: cm). (C) A landmark of sensor locations.

As shown in Figure 7.1A, we employed **C4** configuration (single transmitter & receiver on the anterior wrist and another receiver on the thumb). Participants were given a visual prompt to perform randomly ordered gestures for data acquisition. We provided a practice session before the study. Each study consisted of 5 sessions with 10 trials of all gestures in random order. Between each session, participants took a 1-minute break and researchers corrected the location of the receiver and the transmitter. Adjusting the hardware is to ensure that we collected data from the same locations on the hand. During data collection, we did not strictly constrain the participants' hand posture (*e.g.*, elbow position, wrist rotation) in a sitting state in order to evaluate our system in a wild setting.

We applied a per-user SVM classifier (C=5.0, γ=0.001, and polynomial kernel) for all evaluations since the different body composition requires a per-user classifier when using a bio-acoustic sensing [64]. We also conducted a *leave-one-session-out cross-validation* where we trained 4 sessions and tested the model on 1 session (not included in the training session) for each participant on all sections.

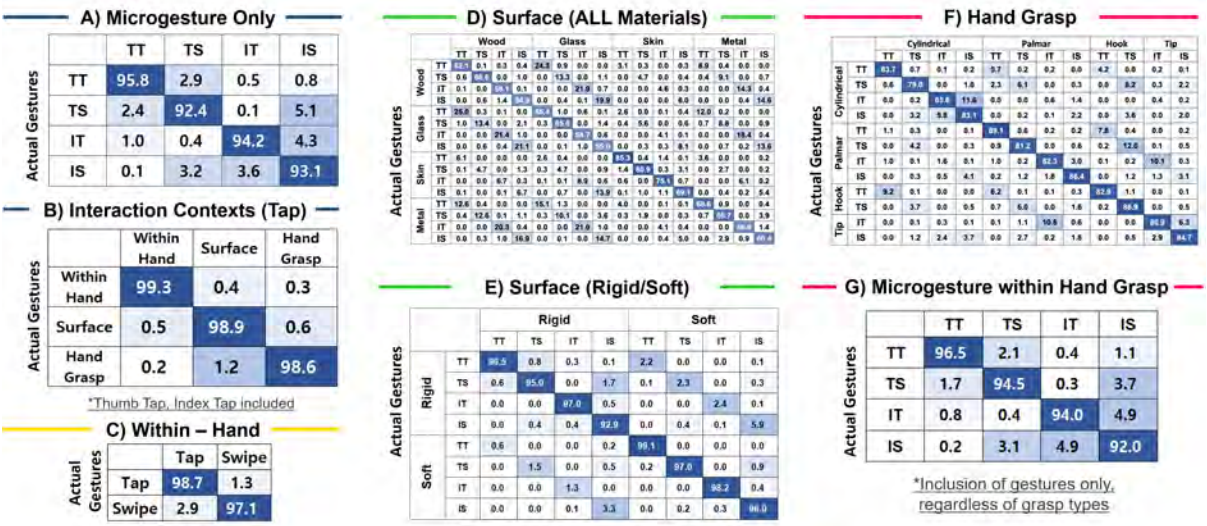

Figure 7.2: Confusion matrix across the users for classifying (A) microgestures regardless of interaction contexts, (B) interaction contexts, (C) within-hand tap & swipe, (D) microgestures on all surfaces, (E) microgestures on grouped surface types (rigid & soft), (F) microgestures with hand grasps, (G) microgestures within hand grasps.

7.2 Context Binding Classification

7.2.1 Microgesture Classification

We carried out the classification of four microgestures regardless of interaction contexts. We considered the same gesture from different interaction contexts as the same class. We also balanced the number of data for each gesture class. Figure 7.2A showed that the average accuracy of microgesture classification was 93.9% (SD=2.6).

7.2.2 Interaction Context Classification

We examined accuracy on interaction context classification using tap gestures among *Within-Hand*, *Surface*, and *Hand Grasp*. We used data from studies of system evaluation where all data for each interaction context were regarded as a single class. We acquired a total of 42,000 samples from 2 receivers. Since the number of data in *Within-Hand* case is smaller, we reduced the number of data from *Surface* and *Hand Grasp* when training the dataset. This prevents us from producing a biased model for interaction context classification. Using tap gestures, the average accuracy of interaction context classification was 98.9% (SD=0.7). Figure 7.2B shows that the proposed system robustly classifies interaction contexts.

7.3 Context-Aware Tap & Swipe Classification

7.3.1 Within-Hand Classification

We asked participants to perform tap and swipe gestures. A total of 2,800 samples were acquired (14 participants × 5 sessions × 2 gestures × 10 trials × 2 receivers). The *leave-one-session-out cross-validation* accuracy across all participants showed 97.9% (SD=2.0). Figure 7.2C indicates that our system supports robust *Within-Hand* tap and swipe gesture recognition.

7.3.2 Surface Type Classification

In this study, we confirmed the capability of microgesture recognition on various surfaces. We chose wood, metal, glass, and skin as representative surface materials. As shown in Figure 7.1A, we asked the participants to perform the tap and swipe gestures. In each session, the participants performed 16 gestures (4 material types × 4 gestures) 10 times in a random order (*e.g.*, *Glass TT-Skin IS-Metal TS-* etc). For skin surface data collection, we asked participants to perform the gestures on the participant's other palm. We acquired 22,400 samples (14 participants × 5 sessions × 16 gestures × 10 trials × 2 receivers).

We averaged the accuracies across participants for all 16 classes using *leave-one-session-out cross-validation*. As shown in Figure 7.2D, we observed low performance (64.0%, SD=9.0). We observed frequent confusion between the wood, glass, and metal surfaces. To explore the potential of classifying the surface type based on material properties, we categorize surfaces based on stiffness. Here, we consider wood, glass, and metal as "Rigid" and skin as "Soft" material, reducing the total number of classes to 8. We adjusted a number of data to keep a 1-to-1 ratio between "Rigid" and "Soft" materials for training the model. As shown in Figure 7.2E, the overall accuracy (94.4%, SD=2.4) improved when we grouped surface types by stiffness.

7.3.3 Hand Grasp Type Classification

In this evaluation, we asked participants to perform the gesture set defined in Figure 3.1 including *Cylindrical*, *Palmar*, and *Hook*, and *Tip*. We guided participants to use either the thumb or index finger with *Hook* and *Tip* grasps which better represent the natural hand behaviors on given grasps. We used 3D-printed objects made with PLA to induce representative hand grasps we selected. Although their properties may not be the same as real objects, we focus our investigation on the microgesture recognition under various hand configurations.

In each session, participants carried out 12 gestures (2 grasp types × 4 gestures and 2 grasp types × 2 gestures) 10 times in random order (*e.g.*, *Palmar IT-Hook TT-Cylindrical TS*). Each participant took 5 sessions. Since participants performed many gestures in a random order, we constantly provided visual prompts and verbal reminders. We collected a total of 16,800 samples (14 participants × 5 sessions × 12 gestures × 10 trials × 2 receivers).

The average accuracy across all participants came out to be 85.7% (SD=5.2). As shown in Figure 7.2F, the same gesture on different grasps was a main source of error. We also observed larger errors in **IT** and **IS** of the *Cylindrical* grasp due to the similar physical finger movements caused by the grasping posture. To further confirm the performance of microgesture recognition after context binding, we also trained a model for distinguishing 4 microgestures within all hand grasps. The overall accuracy was 94.1% (SD=2.6) which guarantees robust microgesture recognition after the context binding stage (Figure 7.2G).

7.4 Ablation Analysis on Bio-Acoustic Sensing

As reported in the pilot study, the performance of **C4**∼**C7** using both the passive and active sensing approaches was higher than that of **C1** and **C2** using only the active sensing approach. In this section, we examined the recognition accuracy among *active-only*, *passive-only*, and *active+passive* acoustic sensing approaches.

We used the collected data from the study to analyze the classification performances of all cases shown in Figure 3.1 for each sensing approach. As expected, the *active+passive* approach showed a higher accuracy (83.8%) than *active-only* (74.5%) and *passive-only* approaches (66.8%). We also analyzed performance only on 4 microgestures. Again, the *active+passive* showed the higher accuracy (93.9%) than *passive-only* (89.5%) and *active-only* (77.2%). With the results, we validated the superior performance of using both active and passive acoustic signals for context-aware microgesture recognition.

7.5 Robustness against False Positive

We validated the robustness of our system against false triggers during daily activities. We asked participants to carry out daily natural behaviors related to representative grasps (*e.g.*, Drinking water while holding a cup, lifting a box, etc.). It was specified in the grasp types, but the actions were to include cases that might occur in everyday life. We used 3 out of 5 sessions for the training set, and the rest 2 sessions as the test set. A total of 4200 s data (14 participants $\times$ 4 grasps related daily behaviors $\times$ 5 sessions $\times$ 15 s trial). We observed the false trigger error of 1.68%. Given the low error rate using a small set of training dataset, we expect to further reduce the error rate with more daily activity data collection.

Chapter 8. Example Applications

We present several example applications to showcase the benefits and usability of the proposed system. Through example applications, we confirmed the potential of context-aware gesture recognition for interactions with various spatial and environmental contexts.

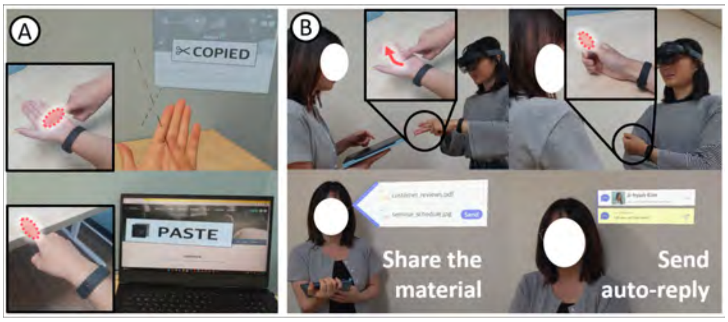

Figure 8.1: Example applications. (A), (B) The user can hover between various physical spaces for cross-spatial interactions.

Cross-Spatial Interaction Our system enables carrying out cross-spatial interactions. The first application utilizes various surfaces as a cue for different interactions. For example, users can perform a tap gesture on the skin to capture the scene using AR glasses. Then, users can display it on the monitor by simply tapping the desk (Figure 8.1A). Figure 8.1B illustrates potential applications as a communication medium. The users perform *Within-Hand* swipe to share information. Moreover, we could utilize *Within-Hand* tap & swipe to support subtle and private interactions.

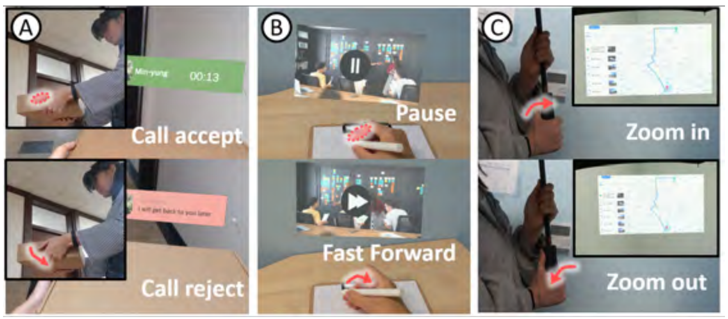

Figure 8.2: (A)∼(C) carry out quick & subtle finger interactions even if hands are occupied.

Quick & Subtle Interaction with Busy Hands Our system benefits users when hands are occupied with various *Hand Grasp*. For example, it infers that a user is carrying a heavy object if *Palmar* grasp is detected (Figure 8.2A). Since it is cumbersome for users to perform any interactions while carrying an object, the system only allows tap & swipe gestures. It can be employed for writing with *Tip* grasp (Figure 8.2B). In this case, users tap a pen body to play/pause a video and swipe the pen to skip forward. In addition, our system could distinguish between thumb and index finger actions as discrete selection commands. When users hold an umbrella using a *Cylindrical* grasp, they can utilize a thumb/index finger swipe gesture to zoom in or out in an AR map application (Figure 8.2C).

Chapter 9. Discussion

Our work focused on enabling microgesture recognition while understanding the surrounding context. Our demonstrated sensing method and pipeline allow quick and robust cross-spatial interactions with no FOV limitation like the vision-based approach.

By further analyzing the performances, we found that passive acoustic sensing contributes to recognizing gestures, while active acoustic sensing contributes to understanding interaction contexts. The low performance on the surface type (16 classes) is due to the similar hardness and a non-direct way of measurement. Thus, additional cues such as microphones could allow more detailed material-based interactions. Although it cannot be directly compared to previous studies [7], they showed 82.35% of discriminating 4 gripping objects with a double IMU on the thumb and index finger. We demonstrated a comparable accuracy of 85.7% from the hand grasp type, even with more cases. Based on the results and example scenarios, context-aware gesture recognition is helpful for spatial interaction in hand-busy situations.

About the scalability, the range of contexts in this study includes *Within-Hand*, *Surface*, and *Hand Grasp*. We confirmed the potential of utilizing bio-acoustic sensing to support context-aware microgestures. We plan to expand the scope of "Interaction Context" in terms of adding different properties (*e.g.*, texture, weight, or size) or increasing the number of surfaces and hand grasps. Meanwhile, since our primary research goal is to bind a context and execute microgesture recognition after context binding, we did not specify the implementation of the cancel-out gesture. We will consider simple and conventional gestures while different from tap&swipe gestures like flick or double-tap. For the time-out duration, we set 5s. However, a user study may be needed to explore the appropriate duration time.

The interaction phase consists of 2 steps to recognize the desired context and gesture. It could be a burden for users to require an additional step. However, the binding gesture is not significantly different from the gesture set, and it can be performed in a short time. In addition, the interaction scenario we are aiming for is not a continuous input command, but rather a short input command that can be performed on demand in the physical environment around us. It doesn't seem to have a serious impact on the user experience, but it will be necessary to get feedback on the user experience.

Our current system relies on wearing the device on the finger and wrist to support robust performance. Although our approach is aligned with upcoming wearables [67], wearing a ring could still cause inconvenience to users not familiar with wearing accessories. In future work, we will investigate sensing techniques that do not require a device worn on the finger. The "wrist-only" condition would be a viable option. Also, our hardware configuration does not support the wireless setup at the current stage. In this study, we employed a wired setup to obtain high-quality ground truth data. In future device configurations, we will utilize a wireless audio IC like AD5930 [66] to support wireless setup and robust performance. In addition, the audible frequency range was used for the driving signal. No participant problems were reported, although detectable. In a lab, ambient noise was 46 dB, increasing to 47 dB when the transmitter was on. Using a 3D-printed Thermoplastic Polyurethane cover brought levels back to 46 dB similar to home/office noise. The future design could include absorbent materials for the active source which will reduce the noise.

The cross-user performance was lower due to the potential user dependency on the bio-acoustic sensing approach. To this end, we plan to explore the newly introduced calibration approach to reduce user dependency on recognition, such as collecting a set of calibration data during the initialization phase. For example, we would ask the user to provide a set of interaction data for a few-shot learning to adjust the overall model for each user similar to [18].

Chapter 10. Conclusion

We present bio-acoustic sensing that enables context-aware tap and swipe gesture recognition. Utilizing an acoustic transmitter and accelerometer, we support active and passive acoustic sensing through wrist- and thumb-mounted approaches. Our system recognizes in which contexts the microgestures occur including *Within-Hand*, *Surface*, and *Hand Grasp*. This enables a broad interaction scenario even with the same set of gestures. We confirmed the hardware configuration and employed multiple bandpass filters to characterize frequency response. Evaluation results confirmed that our system recognizes interaction contexts while supporting tap and swipe gestures. Our work will expand interaction contexts by understanding the deeper context where users perform microgestures.

Bibliography

[1] Gu, Y., Yu, C., Li, Z., Li, W., Xu, S., Wei, X. & Shi, Y. Accurate and low-latency sensing of touch contact on any surface with finger-worn IMU sensor. *UIST 2019 - Proceedings Of The 32nd Annual ACM Symposium On User Interface Software And Technology.* pp. 1059-1070, 2019.

[2] Kienzle, W., Whitmire, E., Rittaler, C. & Benko, H. Electroring: Subtle pinch and touch detection with a ring. *Conference On Human Factors In Computing Systems - Proceedings*, 2021.

[3] Sun, W., Li, F., Huang, C., Lei, Z., Steeper, B., Tao, S., Tian, F. & Zhang, C. ThumbTrak: Recognizing Micro-finger Poses Using a Ring with Proximity Sensing. *Proceedings Of MobileHCI 2021 - ACM International Conference On Mobile Human-Computer Interaction: Mobile Apart, MobileTogether*, 2021.

[4] Kannan, S., Jo, W., Parasuraman, R. & Min, B. Material Mapping in Unknown Environments using Tapping Sound. *2020 IEEE/RSJ International Conference On Intelligent Robots And Systems (IROS).* pp. 4855-4861, 2020.

[5] Lopez-Caudana, E., Quiroz, O., Rodríguez, A., Yépez, L. & Ibarra, D. Classification of materials by acoustic signal processing in real time for NAO robots. *International Journal Of Advanced Robotic Systems.* **14**, 2017.

[6] Ono, M., Shizuki, B. & Tanaka, J. Touch & Activate: Adding Interactivity to Existing Objects Using Active Acoustic Sensing. *Proceedings Of The 26th Annual ACM Symposium On User Interface Software And Technology.* pp. 31-40, 2013.

[7] Liang, C., Yu, C., Qin, Y., Wang, Y. & Shi, Y. DualRing: Enabling Subtle and Expressive Hand Interaction with Dual IMU Rings. *Proceedings Of The ACM On Interactive, Mobile, Wearable And Ubiquitous Technologies.* **5**, 2021.

[8] Takahashi, R., Fukumoto, M., Han, C., Sasatani, T., Narusue, Y. & Kawahara, Y. TelemetRing: A batteryless and wireless ring-shaped keyboard using passive inductive telemetry. *UIST 2020 - Proceedings Of The 33rd Annual ACM Symposium On User Interface Software And Technology.* pp. 1161-1168, 2020.

[9] Soliman, M., Mueller, F., Hegemann, L., Roo, J., Theobalt, C. & Steimle, J. FingerInput: Capturing Expressive Single-Hand Thumb-to-Finger Microgestures. *Proceedings Of The 2018 ACM International Conference On Interactive Surfaces And Spaces.* pp. 177-187, 2018.

[10] Tan, Y., Yoon, S. & Ramani, K. BikeGesture: User Elicitation and Performance of Micro Hand Gesture as Input for Cycling. *Proceedings Of The 2017 CHI Conference Extended Abstracts On Human Factors In Computing Systems.* pp. 2147-2154, 2017.

[11] Boldu, R., Dancu, A., Matthies, D., Cascón, P., Ransir, S. & Nanayakkara, S. Thumb-In-Motion: Evaluating Thumb-to-Ring Microgestures for Athletic Activity. *Proceedings Of The Symposium On Spatial User Interaction.* pp. 150-157, 2018.

[12] Wen, H., Rojas, J. & Dey, A. Serendipity: Finger gesture recognition using an off-the-shelf smartwatch. *Conference On Human Factors In Computing Systems - Proceedings*. pp. 3847-3851, 2016.

[13] Yeo, H., Wu, E., Lee, J., Quigley, A. & Koike, H. Opisthenar: Hand poses and finger tapping recognition by observing back of hand using embedded wrist camera. *UIST 2019 - Proceedings Of The 32nd Annual ACM Symposium On User Interface Software And Technology*. pp. 963-971, 2019.

[14] Laput, G. & Harrison, C. Sensing fine-grained hand activity with smartwatches. *Conference On Human Factors In Computing Systems - Proceedings*, 2019.

[15] Meier, M., Streli, P., Fender, A. & Holz, C. TapID: Rapid touch interaction in virtual reality using wearable sensing. *2021 IEEE Virtual Reality And 3D User Interfaces (VR)*. pp. 519-528, 2021.

[16] Yoon, S., Zhang, Y., Huo, K. & Ramani, K. TRing: Instant and customizable interactions with objects using an embedded magnet and a finger-worn device. *Proceedings Of The 29th Annual Symposium On User Interface Software And Technology*. pp. 169-181, 2016.

[17] Parizi, F., Whitmire, E. & Patel, S. AuraRing: Precise electromagnetic finger tracking. *Proceedings Of The ACM On Interactive, Mobile, Wearable And Ubiquitous Technologies*. **3**, 1-28, 2019.

[18] Xu, X., Gong, J., Brum, C., Liang, L., Suh, B., Gupta, S., Agarwal, Y., Lindsey, L., Kang, R., Shahsavari, B., Nguyen, T., Nieto, H., Hudson, S., Maalouf, C., Mousavi, J. & Laput, G. Enabling Hand Gesture Customization on Wrist-Worn Devices. *Conference On Human Factors In Computing Systems - Proceedings*, 2022.

[19] Melcer, E., Astolfi, M., Remaley, M., Berenzweig, A. & Giurgica-Tiron, T. CTRL-labs: Hand activity estimation and real-time control from neuromuscular signals. *Extended Abstracts Of The 2018 CHI Conference On Human Factors In Computing Systems*. pp. 1-4, 2018.

[20] McIntosh, J., McNeill, C., Fraser, M., Kerber, F., Lochtefeld, M. & Krüger, A. EMPress: Practical hand gesture classification with wrist-mounted EMG and pressure sensing. *Conference On Human Factors In Computing Systems - Proceedings*. pp. 2332-2342, 2016.

[21] Iravantchi, Y., Goel, M. & Harrison, C. BeamBand: Hand gesture sensing with ultrasonic beamforming. *Conference On Human Factors In Computing Systems - Proceedings*, 2019.

[22] Gong, J., Yang, X. & Irani, P. WristWhirl: One-Handed Continuous Smartwatch Input Using Wrist Gestures. *Proceedings Of The 29th Annual Symposium On User Interface Software And Technology*. pp. 861-872, 2016.

[23] Zhang, C., Xue, Q., Waghmare, A., Meng, R., Jain, S., Han, Y., Li, X., Cunefare, K., Ploetz, T., Starner, T., Inan, O. & Abowd, G. FingerPing: Recognizing fine-grained hand poses using active acoustic on-body sensing. *Conference On Human Factors In Computing Systems - Proceedings*, 2018.

[24] Amesaka, T., Watanabe, H., Sugimoto, M. & Shizuki, B. Gesture Recognition Method Using Acoustic Sensing on Usual Garment. *Proceedings Of The ACM On Interactive, Mobile, Wearable And Ubiquitous Technologies*. **6**, 1-27, 2022.

[25] Siddiqui, N. & Chan, R. A wearable hand gesture recognition device based on acoustic measurements at wrist. *Proceedings Of The Annual International Conference Of The IEEE Engineering In Medicine And Biology Society, EMBS*. pp. 4443-4446, 2017.

[26] Zhou, B., Aiskovich, M. & Guven, S. Acoustic Sensing-based Hand Gesture Detection for Wearable Device Interaction, 2021.

[27] Oh, S., Park, C., Jeon, Y. & Choi, S. Identifying Contact Fingers on Touch Sensitive Surfaces by Ring-Based Vibratory Communication. *UIST 2021 - Proceedings Of The 34th Annual ACM Symposium On User Interface Software And Technology*. pp. 208-222, 2021.

[28] Sim, J., Noh, H., Goo, W., Kim, N., Chae, S. & Ahn, C. Identity Recognition Based on Bioacoustics of Human Body. *IEEE Transactions On Cybernetics*. **51**, 2761-2772, 2021.

[29] Siddiqui, N. & Chan, R. Multimodal hand gesture recognition using single IMU and acoustic measurements at wrist. *PLoS ONE*. **15**, 2020.

[30] Laput, G., Xiao, R. & Harrison, C. ViBand: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. *UIST 2016 - Proceedings Of The 29th Annual Symposium On User Interface Software And Technology*. pp. 321-333, 2016.

[31] Kubo, Y., Koguchi, Y., Shizuki, B., Takahashi, S. & Hilliges, O. AudioTouch: Minimally invasive sensing of micro-gestures via active bio-acoustic sensing. *Proceedings Of The 21st International Conference On Human-Computer Interaction With Mobile Devices And Services, MobileHCI 2019*, 2019.

[32] Deyle, T., Palinko, S., Poole, E. & Starner, T. Hambone: A bio-acoustic gesture interface. *2007 11th IEEE International Symposium On Wearable Computers*. pp. 3-10, 2007.

[33] Wolf, K., Naumann, A., Rohs, M. & Müller, J. A taxonomy of microinteractions: Defining microgestures based on ergonomic and scenario-Dependent requirements. *Human-Computer Interaction–INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part I 13*. pp. 559-575, 2011.

[34] Taylor, B. & Bove Jr, V. Graspables: grasp-recognition as a user interface. *Proceedings Of The SIGCHI Conference On Human Factors In Computing Systems*. pp. 917-926, 2009.

[35] Sharma, A., Salchow-Hömmen, C., Mollyn, V., Nittala, A., Hedderich, M., Koelle, M., Seel, T. & Steimle, J. SparseIMU: Computational Design of Sparse IMU Layouts for Sensing Fine-Grained Finger Microgestures. *ACM Trans. Comput.-Hum. Interact.*, 2022.

[36] Chan, E., Seyed, T., Stuerzlinger, W., Yang, X. & Maurer, F. User elicitation on single-hand microgestures. *Conference On Human Factors In Computing Systems - Proceedings*. pp. 3403-3414, 2016.

[37] Feix, T., Romero, J., Schmiedmayer, H., Dollar, A. & Kragic, D. The GRASP Taxonomy of Human Grasp Types. *IEEE Transactions On Human-Machine Systems*. **46**, 66-77, 2016.

[38] Sharma, A., Roo, J. & Steimle, J. Grasping microgestures: Eliciting Single-hand Microgestures for Handheld Objects. *Conference On Human Factors In Computing Systems - Proceedings*, 2019.

[39] Sharma, A., Hedderich, M., Bhardwaj, D., Fruchard, B., McIntosh, J., Nittala, A., Klakow, D., Ashbrook, D. & Steimle, J. Solofinger: Robust microgestures while grasping everyday objects. *Conference On Human Factors In Computing Systems - Proceedings*, 2021.

[40] Zhang, Y. & Harrison, C. Tomo: Wearable, Low-Cost Electrical Impedance Tomography for Hand Gesture Recognition. *UIST 2014 - Proceedings Of The 27th Annual ACM Symposium On User Interface Software And Technology.* pp. 167–173, 2015.

[41] Dementyev, A. & Paradiso, J. WristFlex: Low-power gesture input with wrist-worn pressure sensors. *UIST 2014 - Proceedings Of The 27th Annual ACM Symposium On User Interface Software And Technology.* pp. 161-166, 2014.

[42] Li, G., Rempel, D., Liu, Y., Song, W. & Adamson, C. Design of 3D microgestures for commands in virtual reality or augmented reality. *Applied Sciences (Switzerland).* **11**, 2021.

[43] Rudolph, J., Holman, D., De Araujo, B., Jota, R., Wigdor, D. & Savage, V. Sensing Hand Interactions with Everyday Objects by Profiling Wrist Topography. *Sixteenth International Conference On Tangible, Embedded, And Embodied Interaction.* pp. 1-14, 2022.

[44] Asakura, T. & Iida, S. Hand gesture recognition by using bioacoustic responses. *Acoustical Science And Technology.* **41**, 521-524, 2020.

[45] Chan, L., Chen, Y., Hsieh, C., Liang, R. & Chen, B. CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions Through a Fisheye Ring. *Proceedings Of The 28th Annual ACM Symposium On User Interface Software & Technology.* pp. 549-556, 2015.

[46] Jiang, S., Kang, P., Song, X., Lo, B. & Shull, P. Emerging wearable interfaces and algorithms for hand gesture recognition: A survey. *IEEE Reviews In Biomedical Engineering.* **15** pp. 85-102, 2021.

[47] Sagayam, K. & Hemanth, D. Hand posture and gesture recognition techniques for virtual reality applications: a survey. *Virtual Reality.* **21**, 91-107, 2017.

[48] Gong, J., Gupta, A. & Benko, H. Acustico: Surface Tap Detection and Localization Using Wrist-Based Acoustic TDOA Sensing. *Proceedings Of The 33rd Annual ACM Symposium On User Interface Software And Technology.* pp. 406-419, 2020.

[49] Mujibiya, A., Cao, X., Tan, D., Morris, D., Patel, S. & Rekimoto, J. The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. *Proceedings Of The 2013 ACM International Conference On Interactive Tabletops And Surfaces.* pp. 189-198, 2013.

[50] Iravantchi, Y., Zhang, Y., Bernitsas, E., Goel, M. & Harrison, C. Interferi: Gesture Sensing Using On-Body Acoustic Interferometry. *Proceedings Of The 2019 CHI Conference On Human Factors In Computing Systems.* pp. 1-13, 2019.

[51] Pan, S., Ramirez, C., Mirshekari, M., Fagert, J., Chung, A., Hu, C., Shen, J., Noh, H. & Zhang, P. Surfacevibe: vibration-based tap & swipe tracking on ubiquitous surfaces. *2017 16th ACM/IEEE International Conference On Information Processing In Sensor Networks (IPSN).* pp. 197-208, 2017.

[52] Chen, W., Guan, M., Huang, Y., Wang, L., Ruby, R., Hu, W. & Wu, K. Vitype: A cost efficient on-body typing system through vibration. *2018 15th Annual IEEE International Conference On Sensing, Communication, And Networking (SECON).* pp. 1-9, 2018.

[53] Zhang, Y., Kienzle, W., Ma, Y., Ng, S., Benko, H. & Harrison, C. ActiTouch: Robust touch detection for on-skin AR/VR interfaces. *Proceedings Of The 32nd Annual ACM Symposium On User Interface Software And Technology.* pp. 1151-1159, 2019.

[54] Huang, D., Zhang, X., Saponas, T., Fogarty, J. & Gollakota, S. Leveraging Dual-Observable Input for Fine-Grained Thumb Interaction Using Forearm EMG. *Proceedings Of The 28th Annual ACM Symposium On User Interface Software & Technology*. pp. 523-528, 2015.

[55] Nandakumar, R., Iyer, V., Tan, D. & Gollakota, S. Fingerio: Using active sonar for fine-grained finger tracking. *Proceedings Of The 2016 CHI Conference On Human Factors In Computing Systems*. pp. 1515-1525, 2016.

[56] Harrison, C., Tan, D. & Morris, D. Skinput: Appropriating the Body as an Input Surface. *Proceedings Of The SIGCHI Conference On Human Factors In Computing Systems*. pp. 453-462, 2010.

[57] Zhang, C., Hersek, S., Pu, Y., Sun, D., Xue, Q., Starner, T., Abowd, G. & Inan, O. Bioacoustics-based human-body-mediated communication. *Computer*. **50**, 36-46, 2017.

[58] Huang, D., Chan, L., Yang, S., Wang, F., Liang, R., Yang, D., Hung, Y. & Chen, B. DigitSpace: Designing Thumb-to-Fingers Touch Interfaces for One-Handed and Eyes-Free Interactions. *Proceedings Of The 2016 CHI Conference On Human Factors In Computing Systems*. pp. 1526-1537, 2016.

[59] Heo, S., Annett, M., Lafreniere, B., Grossman, T. & Fitzmaurice, G. No Need to Stop What You're Doing: Exploring No-Handed Smartwatch Interaction.. *Graphics Interface*. pp. 107-114, 2017.

[60] Saggio, G., Santoro, A., Errico, V., Caon, M., Leoni, A., Ferri, G. & Stornelli, V. A Novel Actuating–Sensing Bone Conduction-Based System for Active Hand Pose Sensing and Material Densities Evaluation Through Hand Touch. *IEEE Transactions On Instrumentation And Measurement*. **70** pp. 1-7, 2021.

[61] Profita, H., Clawson, J., Gilliland, S., Zeagler, C., Starner, T., Budd, J. & Do, E. Don't mind me touching my wrist: a case study of interacting with on-body technology in public. *Proceedings Of The 2013 International Symposium On Wearable Computers*. pp. 89-96, 2013.

[62] Yeo, H., Flamich, G., Schrempf, P., Harris-Birtill, D. & Quigley, A. Radarcat: Radar categorization for input & interaction. *Proceedings Of The 29th Annual Symposium On User Interface Software And Technology*. pp. 833-841, 2016.

[63] Oh, S., Park, C., Jeon, Y. & Choi, S. Identifying Contact Fingers on Touch Sensitive Surfaces by Ring-Based Vibratory Communication. *The 34th Annual ACM Symposium On User Interface Software And Technology*. pp. 208-222, 2021.

[64] Amento, B., Hill, W. & Terveen, L. The Sound of One Hand: A Wrist-Mounted Bio-Acoustic Fingertip Gesture Interface. *CHI '02 Extended Abstracts On Human Factors In Computing Systems*. pp. 724-725 , 2002.

[65] Viegas, S., Patterson, R., Hokanson, J. & Davis, J. Wrist anatomy: incidence, distribution, and correlation of anatomic variations, tears, and arthrosis. *The Journal Of Hand Surgery*. **18 3** pp. 463-75, 1993.

[66] Analog Devices. Programmable Frequency Sweep and Output Burst Waveform Generator, 2017.

[67] Ōura Health Oy. Oura Ring, 2021.