High Precision Gesture Sensing via Quantitative Characterization of the Doppler Effect

Haojun Ai[1] aihj@whu.edu.cn
Yifang Men[1] yifangmen@whu.edu.cn
Liangliang Han[2] hllrob@163.com
Zuchao Li[1] whuchaolee@whu.edu.com
Mengyun Liu[3] amylmy@whu.edu.cn

[1]State Key Laboratory of Software Engineering, School of Computer Science, Wuhan University, P. R. China
[2]Aerospace System Engineering Shanghai, Shanghai, P. R. China
[3]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing Wuhan University, P. R. China

***Abstract*—This paper presents a high precision gesture recognition system that leverages the Doppler effect of ultrasound to sense in-air hand gestures. The system can precisely identify a wider variety of gestures than other systems without any modification to consumer laptops. The system recognizes quantitatively detailed and complex movements from the signals reflected by a moving body. A Hidden Markov Model is used to construct a library of independent, discrete gestures. The gestures can be mapped to diverse application actions. Our method can distinguish among similar gestures with slight difference by extracting fewer, more effective features. Our proposed system reduces false positives caused by unintended motions and is versatile and adaptable to multiple device. We implemented a proof-of-concept prototype on a laptop and extensively evaluated the system. Our results show that the system recognizes ten gestures with an average accuracy of 98% and 18 gestures including similar ones with 95% accuracy. The flexibility and robustness on multiple devices highlights its ability to enable future ubiquitous non-contact gesture-based interaction with computing devices.**

I. Introduction

Gesture recognition can help realize natural and intuitive Human-Computer Interaction. Gestures are easy to learn and do not require peripheral equipment such as a mouse or keyboard, and therefore has attracted the attention of many researchers given these advantages. In-air gesture recognition is a promising approach applicable in a variety of scenarios, such as Virtual Reality [1], Motion Sensing Game [2], Smart Home [3,4], and Ubiquitous Computing [5]. In such scenarios, hand gestures can be captured with visual images [6], signals from wearable sensors [7], infrared [8], and ultrasound signals [9, 10, 11]. Visual-based systems have fundamental limitations in that they require sufficient light and create privacy concerns due to the use of cameras. The other two approaches also have limitations as wearable sensor-based systems are inconvenient and infrared techniques need high installation and instrumentation. Ultrasonic-based gesture recognition systems that only use the speaker and microphone, the most ubiquitous components in device such as laptops and mobile phones have been proposed [9, 10, 11] to help overcome the limitations of existing systems as they promise wider range of applications and angles do not need extra equipment, and are not restricted by light conditions. These ultrasonic systems utilize the Doppler shift reflected by a moving human body to recognize motion and gestures.

However, the existing ultrasonic-based gesture recognition system have some limitations, they either use processed Doppler shifts directly as features or leverage empirical matching methods based on the primary motion attributes, and thus are currently unable to estimate gestures precisely, especially for similar motions. For example, when a hand approaches a device at any arbitrary angle at roughly the same velocity. The absolute changes in distance of an object to the device are same, making it challenging to distinguish different gestures using existing methods. These challenges restrict the range of gestures and precision of recognition.

In this paper, we present a high precision gesture recognition system based on quantizing the characteristics of the Doppler Effect. Unlike other systems, our system can distinguish similar gestures precisely, is resilient to changes within the environment, and can operate on consumer laptops without additional sensors. Seven quantitative indicators that are both representative and involve detailed information are extracted from Doppler shifts. Every gesture consists of a successive sequence of features in time. Using Hidden Markov Model (HMM) classification, we achieved an average recognition accuracy of 98% with ten gestures, and an accuracy of 95% with 18 gestures including similar movements in stationary deployments. The main contributions of our work are as follows:

- We present a high precision gesture classification technique leveraging a quantitative characterization of the Doppler effect, and demonstrate the effectiveness of selected properties via experiments.
- We tackle the challenge of distinguishing similar gestures based on the correlation between body parts, which increases the range of gestures identified.

This work was supported in part by the Innovation Fund of Shanghai Aerospace Science and Technology (SAST, 2015014), The Key Technology R&D Program of Hubei Provence (2014BAA153), and SKLSE-2015-A-06. Yifang Men is the corresponding author.

- We describe the adaptation of systems to multiple devices in different environments, and use parameter optimization to reach higher accuracy.

The rest of the paper is organized as follows. Section II reviews the related work on ultrasonic-based sensing technology. Section III describes the operational theory of our proposed system. Section IV presents the method to quantize the Doppler Effect and classify gestures. Section V evaluates its performance in various experiments. Section VI concludes the paper.

II. RELATED WORK

A. Gesture Recognition

As a promising sensing modality, ultrasound has been widely investigated. The tone with 20 kHz, beyond the range of human ear is available in many commercial devices such as laptops, mobile phones, and smart TVs. Gupta et al. [9] proposed Soundwave, a sound-based gesture sensing technique exploiting the Doppler effect to leverage audio hardware found in commercial laptops. They achieved an accuracy of 94% with five gestures. Yang et al. [11] proposed Dolphin, a technique using the speaker and microphone in smartphone to detect gestures. It was able to identify 24 pre-defined gestures at an accuracy of 94% by combining empirical and machine classification. In [12], Raj et al. employed a single transmitter and multiple receivers to recognize different gestures, and explored several novel applications based on ultrasonic Doppler with custom sensors. All their work has been a great inspiration to us.

B. Activity Recognition & Motion Detection

Subsequently, the mounting interest in technologies for healthcare has led researchers to investigate non-contact methods for detecting motion. Watanabe et al. [15] enable activity-aware services with a microphone on the chest and one or more small speakers on the wrists. The sensors detect motion on the basis of the volume of the received sound to determine the distance between the speaker and the microphone using the Doppler effect to estimate the speed of motions. [10] presents various experiments to outline opportunities for activity recognition leveraging ultrasonic techniques and discusses its benefit and limitation. This technology is also used to measure user presence and attention [16]. It is a promising way to realize intelligent environments to monitor the elderly [17].

C. Other Technologies

Recently, more novel research has explored the potential of ultrasonic technology in various scenarios. AirLink [13] allows users to share files between multiple devices with this technique leveraging the Doppler shift caused by a moving hand to identify the direction of hand movement from one device to another. Similarly, DolLink [14] utilizes this technique to enabling impromptu, natural device selection, pairing, and content transfer. Raj et al [12] also developed an application for identifying people based primarily on their gait.

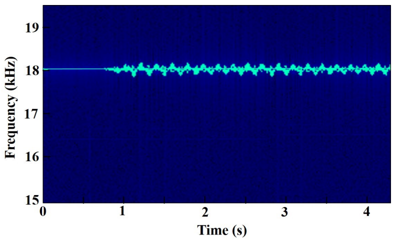

Fig. 1. Frequency-time Doppler profile caused by moving hands. No gesture is performed before second 1. And then the hand moves towards and away from the device alternately.

In contrast to these systems, we introduce an ultrasonic system that expands the ability to recognize more gesture types. We also propose a parameter selection strategy to realize auto-adaptive parameter setting in different environments and for different devices. In contrast to Gupta et al. [9], we leverage quantification of more detailed movement characteristics, by combining machine learning methods and recognize more gestures with higher accuracy. As compared to Dolphin [11], we extracted 6 properties from Doppler shift as eigenvalues for training the classifier, rather than using 60 values directly of ultrasonic vector data. This low dimensional space is more computationally efficient and discriminative.

III. THEORY OF OPERATION

The proposed gesture recognition system works by using a physical principle called Doppler effect, frequency-shift caused by a moving object [18]. The system leverages the speaker (source) of the device to generate a continuous pilot tone, and simultaneously captures it with the microphone (receiver). Since the source and receiver are stationary, the frequency will not change if there is no motion. When the body moves towards the device, it reflects waves, causing a positive shift in frequency. Moving away from the device causes a negative shift in frequency. The shifts sensed by receiver can be measured with the following expression:

$$f = (1 + \frac{\Delta v}{c}) f_0 \tag{1}$$

where f and f_0 are respectively perceived frequency at microphone and original frequency from speaker, c is the speed of sound in air and Δv is the velocity of in-air gesture relative to the device.

Fig. 1 plots the frequency-shifts caused by moving hand. When no gesture is performed, most of the energy is concentrated in the emitted frequency. Shifts occurs when a hand moves towards or away from the device creating the Doppler-effect. The direction of hand movement is one of the most important characteristic attributes in the received signal for gestures recognition.

IV. PROPOSED METHOD

This section presents the method we propose for gesture sensing. As depicted in Fig. 2, three main stages are performed covering Doppler Quantization, Gesture Identification, and Action Mapping.

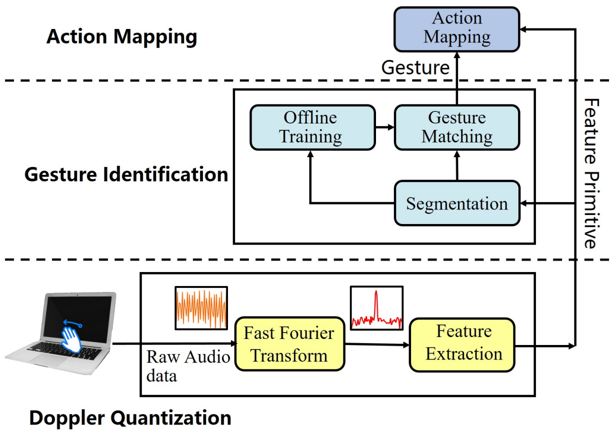

Fig. 2. The system processing components.

A. Doppler effect Quantization

The goal of this stage is to quantize the Doppler effect by extracting features from the raw audio signals. Since the proposed system depends on the speaker to emit a tone and the microphone to capture it, this stage starts with a signal acquisition step and then processes the audio data using Fast Fourier Transform (FFT), followed by a Feature properties quantization step. The details of the different submodules are as follows.

1) Signal Acquisition: The system requires a continuous tone, which can be a pure sine-wave with the highest possible frequency played through the device's speaker. Most laptops and phone speaker system are capable of generate audio up to 22 kHz. We used the pilot tone at 18 kHz because it is inaudible to humans but detectable by almost all standard microphones. We made the microphone sample at 44.1 kHz to meet the conditions of the Nyquist theorem.

Each data frame needed 50 milliseconds for collection purposes. The hand movements in front of a laptop were observed at a velocity up to 3.9 m/sec [9] indicating the duration of a gesture was more than 256 milliseconds. Therefore, when a gesture is performed, at least five frames of raw data were captured.

2) Fast Fourier Transform: The main effect of human hand motions is the frequency-shift occurring in the frequency-domain. Thus it was necessary to transform the original time-domain signal to frequency-domain leveraging the Fast Fourier Transform (FFT). To start with, we leveraged the Hamming window to reduce the amount of spectral leakage, it can be described by the following equation.

$$w(n) = 0.54 - 0.46cos(2\pi \frac{n}{N}), 0 \le n \le N \quad (2)$$

where $N = L - 1$, L is the length of the window. Hence, we computed the FFT with 2048-point Hamming window vectors, yielding 1024-point magnitude vectors with a spectral width of 22.05 kHz. The frequency resolution was 21.5 Hz per bin.

3) Feature Properties: Seven feature properties are extracted from frequency-domain vectors A to quantize the Doppler shift. We define the feature primitive as $P = \{\lambda_1, \lambda_2, ..., \lambda_m\}$, m=7, where m is the number of properties, and λ_i is the quantitative property based on motion characteristics.

a) λ_1 is used to measure the amplitude of the emitted tone, which is usually the highest amplitude and represents the strength of overall signal. λ_1 is related to the size and proximity of the target. If the signal is reflected by human body, there will be a larger amplitude than a signal reflected only from hands. The amplitude increases when user moves closer to the device. It is computed by $\lambda_1 = A(f_0)$ where A is the shift data processed with fft, $A(x_i)$ is the amplitude of frequency x_i, and f_0 is original frequency from speaker.

b) λ_2 represents the energy on the left side of the emitted peak, and λ_3 represents the energy on the opposite side. When the frequency occurs a positive shift, λ_3 will be increased, and a negative shift will increase the value of λ_2 as:

$$\lambda_2 = \sum\nolimits_{i=a}^{b} A(i), \lambda_3 = \sum\nolimits_{i=b}^{c} A(i) \quad (3)$$

where $a = f_0 - \Delta f$, $b = f_0$, and $c = f_0 + \Delta f$. $\Delta f = 300Hz$ can detect movements at a velocity up to 5.7 m/sec, as discussed in Section III.

c) λ_4 is *used* to measure the velocity of movement and is computed by the bandwidth at amplitude θ_v.Change in this property is proportional to the absolute velocity of the target. It is given by $\lambda_4 = \hat{\beta} - \hat{\alpha}$, where $\hat{\alpha} = \arg\min|f_0 - \alpha|$, α is subject to an energy constraint $\forall x_i \in (f_0 - \Delta f, f_0), A(x_i) \le \theta_v$ and $\alpha \in (f_0 - \Delta f, f_0)$ .The right margin $\hat{\beta} = \arg\min|f_0 - \beta|$, β is subject to an energy constraint $\forall x_i \in (f_0, f_0 + \Delta f), A(x_i) \le \theta_v$ and $\beta \in (f_0, f_0 + \Delta f)$.

d) λ_5 represents the direction of movement in this time frame and defined by:

$$\lambda_5 = \sum\nolimits_{i=b}^{c} A(i) - \sum\nolimits_{j=a}^{b} A(j), if\, \lambda_4 \ge \varepsilon, else, \lambda_5 = 0 \quad (4)$$

where ε is the slowest velocity (0.25 m/sec) that can be detected.

e) λ_6 is used to measure the time duration of a gesture indirectly and is the sequence number of the primitive in a gesture sample.

f) λ_7 is defined to quantize slight changes in gestures based on correlations between human body parts. Hand movements also put the arm in motion and the arm often performs different motions in similar gestures, especially the elbow joint. The size of an arm is larger than a hand but the speed of an arm movement is slower. Hence, more signal is reflected by an arm but the frequency shifts slightly. Therefore, the bins near the emitted peak have higher amplitudes. We define the bandwidth at amplitude θ_d (approximately 90% of the maximal amplitude) as the value of this property.

After feature quantization a normalization need to be performed to avoid the uncertainty of ultrasonic intensity.

$$S_t = \sum_{i=1}^{m} P_t(i), F_t(i) = \frac{P_t(i)}{S_t} \quad (5)$$

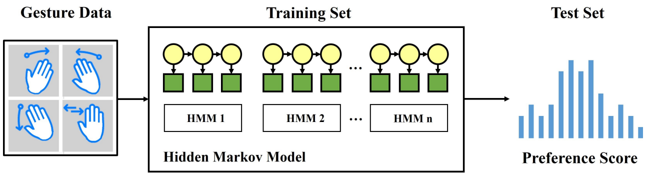

Fig. 3. Framework of the HMM-based gesture classification. One HMM is trained for each gesture and the test motion will be mapped to the gesture which has the highest score.

where F_t is a feature vector at time t, P_t is feature primitive at time t, and m is the number of using properties.

Finally, we present a parameters selection strategy for the threshold θ_v and θ_d. The experiments are performed to explore the relevance between the recognition accuracy and parameters. We find there is an optimization to the highest accuracy for both parameters. Hence, the parameters are initialized based on common practices and we search the optimization using the strategy which is similar to binary search algorithm.

B. Gesture Identification

1) Segmentation: Knowing when a user is generating a gesture is an important question in the normal workings of the system. It helps to eliminate casual actions/noises from the environment and leads to energy-efficiency. We use the properties of feature primitives to judge whether the frequency shifts in a frame time. Primitives are converted to a string sequence: right shifts to positive signs, left shifts to negative signs, and no shifts to zeros. The gesture starts when the primitive is not recognized as zero. If four continuous zeros are detected it means movements have just completed, and a gesture will be quickly extracted. This method can extract gestures precisely and is appropriate for real-time scenarios. Moreover, the system goes into auto-off mode when no gestures are detected for a long time.

2) HMM-based Gesture Classification: Hidden Markov Model (HMM) is a stochastic model which is useful for analyzing a sequence of observations [19]. It is widely used in speech recognition and vision-based gesture sensing system [20].

In this system, feature primitives in order of time form a sequence of vectors $\{F_{t_0}, F_{t_1}, \ldots, F_{t_n}\}$, which is used as observations. Where F_t has been proposed in section A. For instance, a gesture was captured with 20 time frames and each consists of 7 characteristic properties. They form a feature vector of total 140 points:

$$V = [F_{t_0}{}^T, F_{t_1}{}^T, \ldots, F_{t_{19}}{}^T]^T, V \in R^{140\times1}, F_t \in R^{7\times1} \tag{6}$$

The feature vector V as observations is used for training classifier and recognizing gestures. To decide the number of hidden states, the system iterates through various number of states to select the number that provides highest accuracy.

The system constructs an HMM for each gesture using the training samples. Given a total of N gestures, N HMMs $\{HMM_1, HMM_2, \ldots, HMM_N\}$ are learned from gesture samples. Each HMM can be specified using a triplet (π, A, B)

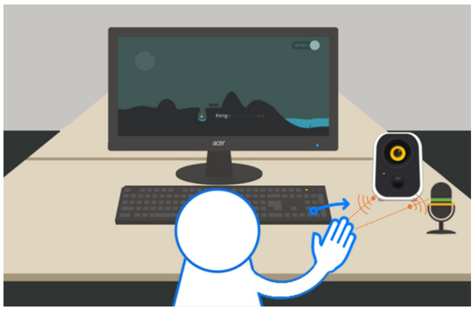

Fig. 4. Component deployment in test environment.

where A is a transition matrix including transition probabilities between the states, B is an emission matrix for the observation symbol probability distributions, and π is the initial state probability. After the initialization of the models for each gesture, the feature vector V extracted from the ultrasonic signal as observations is used for training classifier. The system uses the Baum-Welch algorithm to learn and modify parameters until the models converge to an ideal state.

Given the trained HMMs, we utilize the Forward-Backward algorithm to compute the likelihood of the observation for all the models and the gesture with the highest score will be output. Fig. 3 shows the general procedure in classification.

C. Action Mapping

This is a direct step for mapping application actions to gestures pre-defined in the rich library. As an example, output gestures can act as a game controller. We can make a character walk to the right by a right-swipe gesture while a jump action performed with a tap gesture. Users can also define and train gestures based on own preferences. In addition, the system passes the properties of the gesture to the application for further processing. For instance, the velocity also changes in a time frame and can be used to control the speed of a character's movement in a real-time game.

V. EVALUATION

The first part of the experiments tested how well the method works with similar gestures. Here, we exhaustively interpret these results. In the second part we compared the accuracy of machine and empirical recognition, and discuss their benefits and limitations. In the last part we addressed the flexibility and robustness of the system including the adaption of parameters in different environments and the mistakes when no gesture was performed.

The experiments were conducted in the conference room of laboratory using desktop PCs with an external USB soundcard and microphone. As shown in Fig. 4, user faces the PC with the audio interface deployed right in front of him. The device is above the user in horizontal direction. The speaker generates a pure sine-wave with 18 kHz and the microphone picks up the signals reflected by the moving body. The PC is used to control the operation of system and respond to the gestures. We also tested several laptops, all of which performed similarly in our performance results.

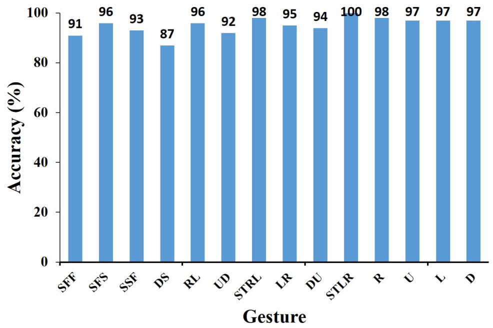

Fig. 5. Similar gestures recognition accuracy. Group 1 contains SFF, SFS, SSF, DS. Group 2 contains RL, UD, STRL. Group 3 contains LR, DU, STLR. Group 4 contains R, U. Group 5 contains L, D.

TABLE I. DEFINITION OF GESTURE

Gesture	Description
$\rightarrow^{\otimes}$ (R)	Swipe your hand to right
$\leftarrow^{\otimes}$ (L)	Swipe your hand to left
$\rightarrow\ \leftarrow^{\otimes}$ (RL)	Swipe your hand to right, then to left
$\leftarrow\ \rightarrow^{\otimes}$ (LR)	Swipe your hand to left, then to right
$\leftrightarrow\mid\leftrightarrow^{\otimes}$ (SFF)	Two handed Seesaw-Fast-Fast: move both hands from center simultaneously in opposite directions at the same time, separating fast and uniting fast
$\leftrightarrow\mid\leftrightarrow^{\otimes}$ $\leftrightarrow\mid\leftrightarrow^{\otimes}$ (DS)	Double-Seesaw: do seesaw action twice with arbitrary velocity
$\leftrightarrow^{\otimes}$ (STLR)	Single-Tap-left-right: move hand to left, then to right quick and slightly which likes a tap action.
$\leftrightarrow\leftrightarrow^{\otimes}$ (DTLR)	Double-Tap: do STLR gesture twice continuously
$\uparrow^{\otimes}$ (U)	Move up your hand
$\downarrow^{\otimes}$ (D)	Move down your hand
$\uparrow\downarrow^{\otimes}$ (UD)	Move your hand up, and then down
$\downarrow\uparrow^{\otimes}$ (DU)	Move your hand down, and then up

We defined a gesture library containing a variety of gestures and two actions. The gestures varied in moving path, velocity, duration, and arm-based property. 12 gestures were chosen from the library (18 gestures in total) and are shown in Table I. The gesture library also contains seesaw gestures such as Seesaw-Fast-Slow (SFS) and Seesaw-Slow-Fast (SSF) that have different velocities than SFF. Tap gestures include STRL and DTRL and these two gestures are different than STLR and DTLR in direction. Moreover; two actions, walking towards (WT) and walking away(WA) from the device are defined in the library.

A. Similar Gestures Recognition Performance

1) Accuracy of Similar gestures: We first evaluated similar gestures detection performance by dividing all of the gestures into five groups and computed the accuracy in each group. One-group gesture can cause similar Doppler shifts over time. For example, UD, RL, and STRL all approach the device first, and then moves away from it. A user performs 200 samples of each gesture for a total of 2800 samples. The results in Fig. 5 show that overall similar gestures detection accuracy averaged over all different groups was 95%.

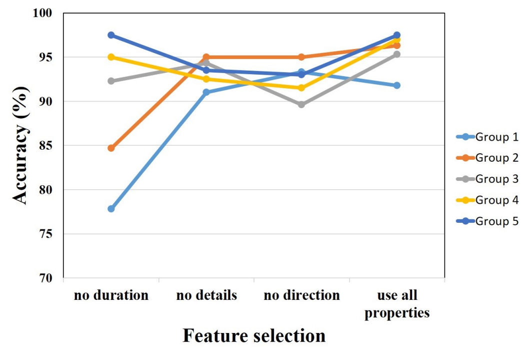

Fig. 6. The accuracy values for five groups containing similar gestures with different feature selection. The first three feature selections all lack a property The last selection uses all properties.

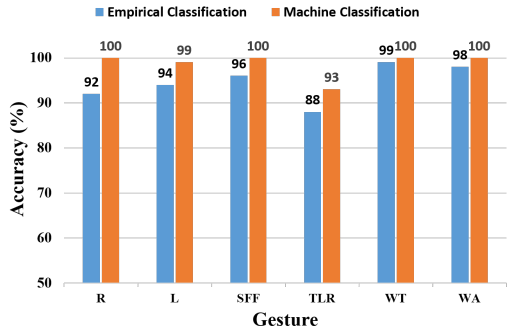

Fig. 7. A comparison of empirical and machine classification accuracy.

Although similar gestures are easily confused, they have slight variations in duration, velocity, and details. Since, it is more convenient for user to swipe from left-to-right (or right-to-left), the Up-Downwards gestures are reluctantly performed. Thus, shifts caused by the former are energy-abundant at higher speed, and therefore, the movement of the arm also plays a role in similar gesture recognition.

2) Selection of Features: Fig. 6 shows the effect of selection of different properties as features in recognition accuracy. We note that accuracy can be further enhanced by combining the properties of duration, direction, and details. Usually, duration is a characteristic of a gesture not a primitive, but we use the index of a primitive sequence λ_6 to act as a property of a primitive. Fig. 6 illustrates that the duration makes all groups more accurate, except group 4 and group 5; because the gestures in these two groups have extremely similar duration, there was no way to distinguish them by this property. In theory, the direction of movement leads to significant changes in the Doppler effect, but this property is linearly dependent to left-energy and right-energy thus increases the redundancy. Leveraging the details improved the recognition accuracy of all groups and shows that our method based on body-correlation is reasonable.

B. Comparison of Classification Methods

We now compare the accuracy of empirical [9] and machine gesture classification. Since the empirical model method cannot recognize similar gestures, we choose six basic gestures to test. From the Fig. 7, we see the machine

TABLE II. ACCURACY IN DIFFERENT ENVIRONMENTS

Environment	Device	θ_v	θ_d	Accuracy
conference room	HP desktop PC	**28**	**37**	**95.6%**
		35	49	89.6%
apartment	HP desktop PC	**29**	**37**	**95.4%**
		28	37	94.7%
conference room	ThinkPad laptop	**35**	**49**	**94.8%**
		28	37	91.2%
apartment	ThinkPad laptop	**35**	**47**	**94.4%**
		35	49	94.2%

classification method with an average accuracy of 98.6% is more effective than the empirical model method, whose average accuracy is 94%.

C. Flexibility and Robustness of the System

1) Adaption of parameters: We evaluated 18 gestures including two actions (walk towards and away from the device) in different environments with various devices. From the Table II, we see that the system performs well without any changes to the algorithms or parameters when it works on the same device in different environments. However, if we convert the system from laptop to a desktop PCs, the parameters are changed to achieve a more accurate result. Hence, the system uses the parameter selection strategy to optimize parameters in the feature extraction step. In Table II, the data in the first column of each item is the result as optimized by system.

2) False-positive rate with no gesture performed: Unintented user motion or the motions of people in the area of the user may be recognized as a gesture by the system. To measure the number of any false gestures over time, we conducted an hour-long test with people present and not present. In the unmanned environment, the false-positive rate was nearly zero. Surrounding noise makes it difficult to reach the standard for any gesture. When users sat in front of the PC, typing or thinking, they were instructed to act normally. An average of 1.5 false gestures occurred per minute, the false gestures detected were mainly tap motions, that moved towards or away from the device in turn, without any velocity constraints. Swipe gestures can be detected only when meeting a set velocity criterion.

VI. CONCLUSION

This paper presents a high precision gesture recognition system that uses the ultrasonic Doppler effect to detect human hand motions. It can distinguish a rich set of gestures without any modification of a laptop. We addressed system challenges including similar gestures recognition, feature primitives extraction, adaptive parameter adjustment for different environments, and reduction of false positive rates.

Extensive evaluation of the proposed system reveals that similar gestures in a dramatically expanded gesture library are detected at an accuracy of 95%. In contrast to other methods, the accuracy of our proposed approach increases to 98% with eight basic gestures using HMM classification. These results show that our system is robust and adaptive to varying environmental conditions, highlighting how it could enable future ubiquitous, non-contact gesture-based human interaction with consumer devices.

REFERENCES

[1] Feng, Zhiquan, et al. "3D-freehand-pose initialization based on operator's cognitive behavioral models." The Visual Computer 26.6-8 (2010): 607-617.

[2] Kulshreshth, Arun, and Joseph J. LaViola Jr. "Exploring the usefulness of finger-based 3D gesture menu selection." Proceedings of the 32nd annual ACM conference on Human factors in computing systems. ACM, 2014.

[3] Adib, Fadel, et al. "Smart homes that monitor breathing and heart rate." Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2015.

[4] Pu, Qifan, et al. "Whole-home gesture recognition using wireless signals." Proceedings of the 19th annual international conference on Mobile computing & networking. ACM, 2013.

[5] Pfeil, Kevin, Seng Lee Koh, and Joseph LaViola. "Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles." Proceedings of the 2013 international conference on Intelligent user interfaces. ACM, 2013.

[6] Aggarwal, Jake K., and Michael S. Ryoo. "Human activity analysis: A review." ACM Computing Surveys (CSUR) 43.3 (2011): 16.

[7] Yatani, Koji, and Khai N. Truong. "BodyScope: a wearable acoustic sensor for activity recognition." Proceedings of the 2012 ACM Conference on Ubiquitous Computing. ACM, 2012.

[8] Liu, Tong, et al. "Compressive infrared sensing for arm gesture acquisition and recognition." Information and Automation, 2015 IEEE International Conference on. IEEE, 2015.

[9] Gupta, Sidhant, et al. "Soundwave: using the doppler effect to sense gestures." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2012.

[10] Fu, Biying, et al. "Opportunities for activity recognition using ultrasound doppler sensing on unmodified mobile phones." Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction. ACM, 2015.

[11] Qifan, Yang, et al. "Dolphin: Ultrasonic-based gesture recognition on smartphone platform." Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on. IEEE, 2014.

[12] Raj, Bhiksha, et al. "Ultrasonic doppler sensing in hci." IEEE Pervasive Computing 2 (2012): 24-29.

[13] Chen, Ke-Yu, et al. "AirLink: sharing files between multiple devices using in-air gestures." Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014.

[14] Aumi, Md Tanvir Islam, et al. "DopLink: using the doppler effect for multi-device interaction." Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. ACM, 2013.

[15] Watanabe, Hiroki, Tsutomu Terada, and Masahiko Tsukamoto. "Ultrasound-based movement sensing, gesture-, and context-recognition." Proceedings of the 2013 International Symposium on Wearable Computers. ACM, 2013.

[16] Tarzia, Stephen P., et al. "Sonar-based measurement of user presence and attention." Proceedings of the 11th international conference on Ubiquitous computing. ACM, 2009.

[17] Cheffena, M. Fall detection using smartphone audio features. Biomedical and Health Informatics, IEEE Journal of PP, 99 (2015), 1–1.

[18] Doppler, Christian. Über das farbige, Licht der Doppelsterne und einigerandererGestirne des Himmels. In Proceedings of the Royal Bohemian Society of Sciences (Part V, Vol 2), Prague, 1842.

[19] Rabiner, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." Proceedings of the IEEE 77.2 (1989): 257-286.

[20] Rabiner, Lawrence R., and Biing-Hwang Juang. "An introduction to hidden Markov models." ASSP Magazine, IEEE 3.1 (1986): 4-16.