Appliance Control by TDoA-based Localization and Gesture Recognition

Ting-Hui Chiang*, Yan-Ann Chen†, Chun-Ting Chang*, Ling-Jyh Chen‡ and Yu-Chee Tseng*
*Dept. of Computer Science, College of Computer Science, National Chiao Tung University, Taiwan
†Dept. of Computer Science and Engineering, Yuan Ze University, Taiwan
‡Institute of Information Science, Academia Sinica, Taiwan

***Abstract*— Smart home is one of the major *Internet of Things (IoT)* applications. When realizing a smart home, designing an intuitive interaction for appliance control has become a crucial issue. A control intention to an appliance comprises two main meanings, identification and control command. But, previous works seldom deliver these two meanings at the same time. To fully convey the control intention, this paper proposes a scheme of intuitive appliance control that tracks and recognizes hand moving trajectory and hand gestures respectively by exploiting acoustic information. We utilize a $3D$ positioning method based on *time difference of arrival (TDoA)* of acoustic signals to detect the trajectory of a hand movement for appliance identification and design a gesture recognition method by Doppler effect for control command. We validate the effectiveness of the proposed scheme via real-life experiments. The results show that average locating errors to the targets are $5.24°$ and $10.35°$ in the horizontal and vertical planes respectively and the accuracy of gesture recognition is 96.66%. Since human's eyesight is horizontal to the ground, locating an object in the vertical view is error-prone. We also provide a scroll gesture to assist the appliance selection.**

***Index Terms*—Acoustic signal processing, Doppler effect, gesture recognition, indoor localization, smart home.**

I. Introduction

Internet of Things (IoT) utilizes Internet- and sensing-enabled devices to make the world smart. Building a smart home is one of the major IoT applications. For controlling home appliances, e.g., television and air conditioner, in a traditional home, we have to approach the control panel or use the corresponding remote controller. Designing an intuitive and effective way for appliance control is one of the important problems to create a smart home. The works which are focused on this topic can be categorized into 4 types of controlling methods, automatic, user-interface (UI) based, pointing-based, and gesture-based control system. First, automatic systems [1] control home appliances to satisfy users' needs. But, in some cases, user intervention is still needed to fulfill some fine-grain controls. Second, in user-interface (UI) based control systems [2], users can control appliances through the designed UI on the specific device, e.g., smart phone, tablet, and touch screen device. Third, the pointing-based control systems [3] identify the user's intention for the control by analyzing the direction of a pointing action. Finally, in gesture-based control systems [4], user performs some specific gestures for the appliance control. These systems can recognize user's gestures by vision-based or radio-based methods. For doing a control in an environment with dozens of appliances, we observe that there are two essential information in the intention of controlling. The first information is the appliance which we are going to interact with and the second one is the control command we are going to issue. However, no systems which are pointing-based or gesture-based can deliver the two information for controlling an appliance at the same time. We believe that this is the first work to fully consider the intention of controlling.

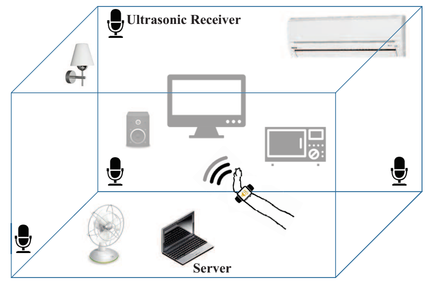

Fig. 1. The scenario of the proposed appliance control scheme in a smart home.

In this work, we exploit acoustic information to realize an intuitive appliance control scheme by tracking and recognizing hand gestures in a smart home environment. Fig. 1 shows our system scenario where a user wears the smart device with an ultrasonic emitter on his/her hand and we deploy ultrasonic receivers (at least 4) near the corner of each wall. A user can simply perform a push gesture toward an appliance and then utilize the predefined gestures to control the appliance. A home server is used to collect received signals, compute the controlling intention, and issue the control command. For selecting the appliance, the user performs a push gesture for pointing to an appliance where he is going to control. While the hand moves toward to the appliance, the ultrasonic receivers in the environment will receive a series of acoustic signals. We exploit *time difference of arrival (TDoA)* technique between different receivers to calculate the moving trajectory of the user's hand. Then the system picks out

978-1-5386-6009-6/18/$31.00 ©2018 IEEE

the appliance for the further control. For commanding the appliance, the user utilizes one of the predefined gestures to control the selected appliance. We analyze Doppler effect of received acoustic signals to recognize the control command. So we define 6 types of gesture as the control commands by combining hand motions of approaching and leaving the selected appliance. We conduct real-life experiments for validating the performances. The experimental results show that the average positioning errors to the locations of selecting targets in the horizontal and vertical planes are $5.24°$ and $10.35°$ respectively and the accuracy of gesture recognition is 96.66%. Since human's eyesight is horizontal to the ground, locating an object in the vertical view is error-prone. Therefore, we provide a scroll gesture for users to pick the appliance.

The remainder of this paper is organized as follows. We survey some related work in Section II. Section III presents our proposed appliance control scheme. We validate our proposed system by real-life experiments in Section IV. Section V concludes this paper.

II. Related Work

A. TDoA-based Localization

Trilateration scheme is one of famous technique in indoor localization. Trilateration schemes can exploit time of arrival (ToA) [5] or time difference of arrival (TDoA) [6] to derive locations. For TDoA-based localization, there are many applications with different source signals, such as visible light [7], radio frequency (RF) signals [8], acoustic signals [6]. Reference [6] utilizes the linear chirp signals to estimate TDoA.

B. Gesture Recognition

Gesture recognition approaches have been studied intensively. In the general these approaches can be categorized as vision-based, inertial measurement units (IMU)-based, RF-based and Doppler-based system. Vision-based system [9] can identify the micro movement, but the system need line-of-sight (LoS) condition and have privacy concern. In the RF-based system, reference [10] does not need users wear devices, but the systems need specialized hardware, Intel 5300 NIC. Doppler effect has been applied on activity/gesture recognition in the several recent works [4]. WiSee [4] proposes a system, which extracts Doppler shifts from wireless signals and maps Doppler shifts to gestures to recognize gesture.

III. Appliance Control Scheme

In this paper, we propose a mechanism to intuitively deliver the intention including the selection and command for controlling a home appliance with hand gestures. Fig. 2 shows the workflow of our system. We divide the flow into two main phases which express selection and command intentions respectively. In the

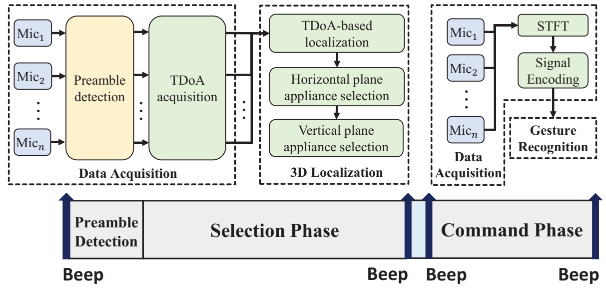

Fig. 2. System workflow.

selection phase, ultrasonic receivers acquire acoustic signals from the user's wearable device. A centralized server calculates TDoA between each pair of receivers. Then we track the $3D$ trajectory of the user's push gesture by the TDoA information. Finally, we can determine the appliance where the user wants to control by comparing distances between the position of each appliance and the trajectory vector. In the command phase, we further utilize Doppler effect caused by user's gestures to understand the control intention. The basic concept of Doppler effect is that the received frequency of acoustic signal will vary due to the relative moving between the acoustic emitter and the receiver. Command gestures are composed of hand motions of approaching or leaving the selected appliance.

In the following, we present acoustic signals adopted in this system and our proposed mechanisms for selecting and commanding the appliance respectively.

A. Acoustic Signals

In our system, we make user's wearable device transmit acoustic signals via an ultrasonic emitter. In the selection phase, we define two types of chirp, preamble-chirp and TDoA-chirp, for time synchronization and TDoA computation respectively. The preamble-chirp is from $23.5kHz$ to $22kHz$ and the TDoA-chirp is from $17kHz$ to $19kHz$. The chirp duration of these two types are both $50ms$. User's wearable device transmits the preamble-chirp once at the beginning of this phase and then repeats the TDoA-chirp until the end of this selection phase. For separating the two phases, we add a short time gap between them. On the other hand, user's wearable device transmits $22kHz$ beep continuously during the period of the command phase. We present the transmitted signals with respect to the two phases. Below, we discuss how these phases work.

B. Selection Phase

When a user is trying to select an appliance, the wearable device transmits the predefined chirps to the receivers in the environment.

1) Data Acquisition:

a) Preamble Detection: We use preamble-chirp to trigger this appliance control system. While ultrasonic receivers receive signals, we apply a matched filter to compare received signals with the preamble-chirp. If we obtain a sharp peak after calculating the cross-correlation, the appliance control system is triggered and execute TDoA acquisition.

b) TDoA Acquisition: After the preamble, all of the ultrasonic receivers will get a series of TDoA-chirps. We also apply a matched filter to compare received signals with the TDoA-chirp. Let $M_i = \{M_i^1, M_i^2\}$ be the i^{th} pair of 2-combinations of the receivers. Then the difference of arrival time from an ultrasonic emitter to receivers M_i^1 and M_i^2 are $\hat{t}_i$. Since $\hat{t}_i$ may be affected by noises, we utilize Z-score method for outlier detection. The $\hat{t}_i$ is an outlier if $Z = (\hat{t}_i - \mu)/\sigma > T_{outlier}$, where μ and σ are respectively the mean and the standard deviation of a series of TDoA measured by the pair M_i and $T_{outlier}$ is a pre-defined threshold.

2) 3D Localization: TDoA is one of the most popular localization techniques. In $2D$ geometry, TDoA exploits the definition of hyperbola for localization where the difference of the distances from two focuses of the hyperbola to any point on the hyperbola is a constant. If we know the difference of the distances from two known positions to a target, the target may lie on the hyperbola formed by the two positions as the focuses and the distance difference as the constant distance. After collecting distance differences from pairs of known positions, we can locate the target by intersecting the formed hyperbolas. In our system, we need to locate the target in $3D$ geometry. Therefore, we calculate the location of the user's wearable device, the ultrasonic emitter, by intersecting hyperboloids formed by different pairs of receivers. Assume that each ultrasonic receiver's location is known. To form a hyperboloid of the receiver pair M_i, we first calculate the difference distance by the measured TDoA information. Let $d(.,.)$ be the distance computation of two locations in Euclidean space. We formulate the computation of the distance difference by $|d(L_{M_i^2}, L_s) - d(L_{M_i^1}, L_s)|$ where $L_{M_i^2}$ and $L_{M_i^1}$ are the location of M_i^2 and M_i^1 respectively and L_s is the location of the user's wearable device. Finally, we can use the difference of the distance and the location of M_i^1 and M_i^2 to form the hyperboloid.

In reality, intersecting hyperboloids formed by some receiver pairs may not exist an intersection due to TDoA estimation errors. We formulate the TDoA-based trilateration problem as a non-linear least square (NLLS) problem as follows.

$$L_s^* = arg \min_{L_s} \sum_{i=1}^{n_p} (\hat{t}_i \cdot v_s - |d(L_{M_i^2}, L_s) - d(L_{M_i^1}, L_s)|),$$

where n_p is the number of receiver pairs and v_s is the speed of sound. Then we apply iterative Gauss-Newton algorithm to solve the NLLS problem and obtain the location of the user's wearable device L_s^*.

3) Appliance Selection: For appliance selection, we obtain a series of locations, namely the trajectory, by the $3D$ localization while doing a push gesture toward the appliance. In $3D$ Euclidean space, we project the direction of the obtained trajectory onto the horizontal and vertical planes.

- Step 1: On the horizontal plane, we apply linear regression to the projected trajectory. We use the starting point of the trajectory as the center to form a circular sector with a predefined angle where the regression line bisects the sector. Then we make the sector toward the direction of the trajectory. Next, we get a set of appliances S_h within the circular sector.
- Step 2: On the vertical plane, we apply the method as step 1 to the vertical plane for getting a set of appliances S_v.
- Step 3: The selected appliance set $S = S_h \cap S_v$.

Since determining the vertical location of an object by human eyes is error-prone, we keep all the appliances in S if $|S| > 1$. Instead, we provide a scroll gesture for users to pick the appliance in S in the command phase.

C. Command Phase

We separate the selection and command phases by a short gap without transmitting any acoustic signal for reducing the interference. The wearable device plays a beep sound to inform user that the command phase is started and then emits $22kHz$ signals continuously until the second beep sound. User can perform predefined gestures to control the selected appliance during this period of time.

1) Gesture Definition and Data Acquisition: In Fig. 3, we define 6 types of command gesture according to combinations of hand movements of approaching and leaving the selected appliance. The concept of our gesture recognition is to observe frequency shifts, which is caused by Doppler effect, when performing these types of gesture. Here, Doppler effect is that the frequency of received signals by a receiver will increase when the acoustic emitter, the wearable device, is approaching the receiver. On other hand, the frequency of received signals will decrease when the wearable device is leaving the receiver. Therefore, to obtain the frequency shifts, we use *short-time Fourier transform (STFT)* to convert received signals, which is from the ultrasonic receivers deployed in the smart home, to the frequency domain.

Next, we define 3 states of frequency shifts which are positive effect, negative effect, and zero effect. These 3 frequency shifts represent that the frequency of a received signal is increased, decreased, and stable respectively. To determine those states, we extract mean and standard deviation of a received signal in the frequency domain. We consider the mean as the baseline and we then set thresholds for positive effect and negative effect to the mean plus one standard deviation and the mean minus one standard deviation respectively. When

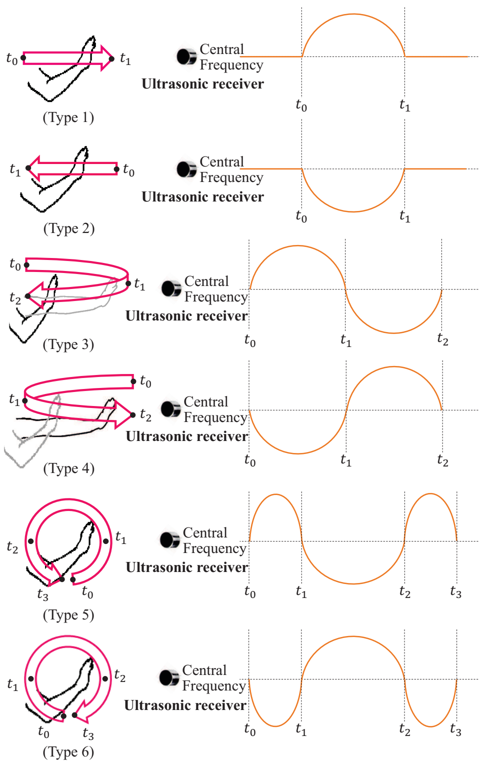

Fig. 3. Predefined 6 types of gesture.

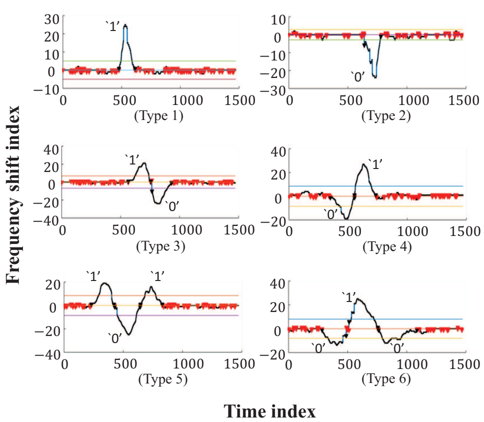

Fig. 4. Encoded sequences of the predefined 6 types of gesture.

the shift exceeds one of the thresholds, we regard that the state of frequency shift is changed to the positive effect or negative. Otherwise, we regard the state as zero effect. Finally, we encode positive and negative effect state to '1' and '0' respectively as shown in Fig. 4.

2) Gesture Recognition: After the data acquisition, we get a sequence of '1' and '0'. In Fig. 4, we observe that the encoded sequences are unique and different across these 6 types of gesture. Thus we define the encoded sequences of these 6 types of gesture to '1', '0', '10', '01', '101', and '010' respectively. Then we map encoded sequences to the sequences of predefined gesture. That is, if we obtain encoded sequence '101', this sequence is classified to the type 5 gesture. Now we can perform gesture recognition by matching the encoded

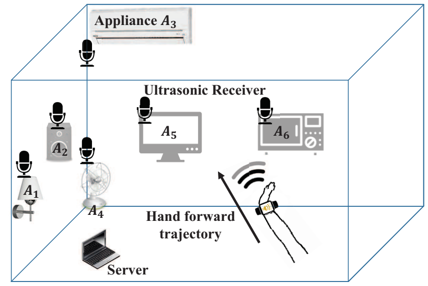

Fig. 5. The deployment of ultrasonic receivers and appliances.

sequences of received signals with the predefined set of encoded gesture sequences. In this method, we can overcome the problem of gestures with variable speeds. Since we segment one frequency shift into a state, the duration of the gesture will not affect the change of state. To sum up, we can apply these types of gesture for users to scroll the appliances, turn on/off, and tune volume, etc.

IV. Field Trial

We conduct field experiments in a meeting room to evaluate the performance of our system. We have implemented the system proposed in Section III. The ultrasonic receiver comprises a Raspberry Pi 2 with a USB stereo sound card (ICUSBAUDIOMH) and a USB wireless adapter where an omnidirectional microphone (MAX9814) is connected to the sound card. The wearable device is composed of a Raspberry Pi 2 with a 7.1 channel sound card where a directional acoustic emitter (Matsushita $0D24K2$) is connected to the sound card. Then we deploy 6 ultrasonic receivers on two perpendicular walls and each receiver has one appliance next to it as shown in Fig. 5. Let A_i denote the appliance i, $i = 1..6$. In the following, we study the effectiveness of the appliance selection and the gesture recognition respectively.

A. Appliance Selection Experiment

In this experiment, we evaluate the pointing accuracy of the selecting action by our system. First, the user pushes the hand with the wearable device toward an appliance after hearing a beep sound. The wearable device emits a preamble-chirp immediately after the beep and then the device emits TDoA-chirps while the push action. We set frequency changes of preamble-chirp and TDoA-chirp from $23.5kHz$ to $22kHz$ and from $17kHz$ to $19kHz$ respectively. The duration of these chirps are all 50 ms. All the ultrasonic receivers will receive TDoA-chirps and join the $3D$ localization for computing the trajectory of the user's push gesture. To show our system performance, we separately measure direction errors of computed trajectories on the horizontal and vertical planes as shown in Fig. 6.

We ask the participant to select 6 designated appliances for 20 times each by the above process. Table I and Table II shows the errors, which is θ_{XY} and θ_{XZ}

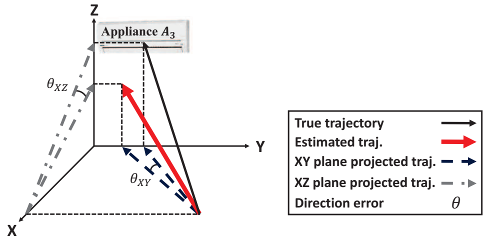

Fig. 6. The example of the projected trajectory on horizontal and vertical planes.

in Fig. 6, between the projected trajectories of the estimated and true trajectories on the horizontal and vertical planes respectively. From Table I, we can see that the mean errors of A_1 and A_6 are higher than that of others. The reason is that the emitting range of the directional transmitter is restricted to about $30°$ to $40°$ from the centroid. The received acoustic signals near the border of the transmission range are weak. Therefore, when we select the appliance near the edge of the testing field, like A_1 and A_6, more receivers reside in the border of the transmission range. From Table II, we can see that the mean errors are all larger than $10°$. The reason is that pointing an object in the vertical view is error-prone since human's eyesight is horizontal to the ground. When an appliance is higher or lower than the height of the participant's head, parallax may incur some pointing errors to the push gesture. For dealing with this problem, we provide a scroll gesture for users to pick up the appliance among a possible set in the command phase.

TABLE I
ANALYSIS OF DIRECTION ERROR ON THE HORIZONTAL PLANE

Appliance	Mean (degrees)	STD	Maximum	Minimum
A_1	12.01	4.99	22.39	2.64
A_2	7.83	4.51	16.42	0.73
A_3	6.1	4.23	14.5	1.44
A_4	5.25	4.15	13.06	0.53
A_5	5.24	3.26	11.8	0.48
A_6	11.82	7.46	33.17	0.47

TABLE II
ANALYSIS OF DIRECTION ERROR ON THE VERTICAL PLANE

Appliance	Mean (degrees)	STD	Maximum	Minimum
A_1	10.35	10.19	33.1	0.09
A_2	14.71	17.68	50.58	1.08
A_3	16.73	20.91	61.68	0.13
A_4	15.53	12.86	53.33	0.81
A_5	25.05	19.36	59.62	1.68
A_6	19.38	16.16	67.43	1.01

B. Gesture Recognition Experiment

In this experiment, we evaluate the accuracy of gesture recognition by our system. We define 6 types of gesture to control appliance as shown in Fig. 3. We ask the participant to perform each gesture for 30 times. Table III shows that the recognition accuracy of each gesture is above 93%. It shows that we can differentiate gestures by our designed encoding sequence extracted from frequency shift. The errors may be caused by the finger vibration and the short gap between the approaching and leaving movements.

TABLE III
ACCURACY OF GESTURE RECOGNITION

Gesture type	Gesture code	Accuracy rate (%)
Type 1	'1'	95.12
Type 2	'0'	95.23
Type 3	'10'	96.55
Type 4	'01'	96.66
Type 5	'101'	93.1
Type 6	'010'	94.73

V. CONCLUSION

This paper addresses that an intuitive appliance control should convey the intentions of appliance selection and control in a smart home. We have proposed the intuitive control scheme to track and recognize hand gestures by analyzing acoustic information. Our scheme tracks the moving trajectory of a push gesture in terms of TDoA-based positioning for appliance selection and recognize pre-defined gestures in terms of Doppler effect for appliance control. From these experiments, we have shown that our system can accurately locate the moving trajectory of a push gesture for appliance selection. To enhance this system, we could consider using high frequency radio frequency to perform our system for avoiding noises and adding more control gestures.

ACKNOWLEDGEMENT

Y.-C. Tseng's research is co-sponsored by ITRI, MOST 105-2218-E-009-003, MoE Project "Center for Open Intelligent Connectivity", and Academia Sinica AS-105-TP-A07.

REFERENCES

[1] R. A. Ramlee, M. Leong, R. S. Singh, M. M. Ismail, M. A. Othman, H. A. Sulaiman, M. H. Misran, and M. M. Said, "Bluetooth remote home automation system using android application," *IJES*, vol. 2, no. 1, pp. 149–153, 2013.

[2] J.-h. Roh and S. Jin, "Device control protocol using mobile phone," in *IEEE ICACT*, 2014, pp. 355–359.

[3] M.-S. Pan and C.-J. Chen, "Intuitive control on electric devices by smartphones for smart home environments," *IEEE Sensors J.*, pp. 4281–4294, 2016.

[4] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *ACM MobiCom*, 2013, pp. 27–38.

[5] J. He, Q. Wang, Q. Zhang, B. Liu, and Y. Yu, "A practical indoor ToA ranging error model for localization algorithm," in *IEEE PIMRC*, 2011, pp. 1182–1186.

[6] Y.-T. Wang, R. Zheng, and D. Zhao, "Towards zero-configuration indoor localization using asynchronous acoustic beacons," *IEEE/IFIP EUC*, 2016.

[7] S.-Y. Jung, S. Hann, and C.-S. Park, "TDoA-based optical wireless indoor localization using LED ceiling lamps," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, 2011.

[8] H. Xiong, Z. Chen, W. An, and B. Yang, "Robust TDoA localization algorithm for asynchronous wireless sensor networks," *Int. J. of Distributed Sensor Networks*, vol. 11, no. 5, pp. 598 747:1–598 747:10, 2015.

[9] J. Kim, K. Chung, and M. Kang, "Continuous gesture recognition using HLAC and low-dimensional space," *Wireless Personal Commun.*, vol. 86, no. 1, pp. 255–270, 2016.

[10] W. He, K. Wu, Y. Zou, and Z. Ming, "Wig: Wifi-based gesture recognition system," in *IEEE ICCCN*, 2015, pp. 1–7.