HEAR-YOUR-ACTION: HUMAN ACTION RECOGNITION BY ULTRASOUND ACTIVE SENSING

Risako Tanigawa and Yasunori Ishii
Panasonic Holdings Corporation

1. Introduction

Action recognition

- has wide applications: healthcare, behavior monitoring etc ….
- has mostly been realized by visual clues such as RGB videos. → Privacy is not protected

Conventional privacy concern recognition methods:

- Sensing with radio waves such as Radio Frequency and Wi-Fi signals
  → Radio waves can be interfered with by electronic devices.
- Acoustic sensing (passive and active)
  Passive sensing → Voice can be a potential identifier for individuals.
  Active sensing → Privacy preserved method but NOT well investigated

→ We propose a new task for human action recognition by ultrasound active sensing.

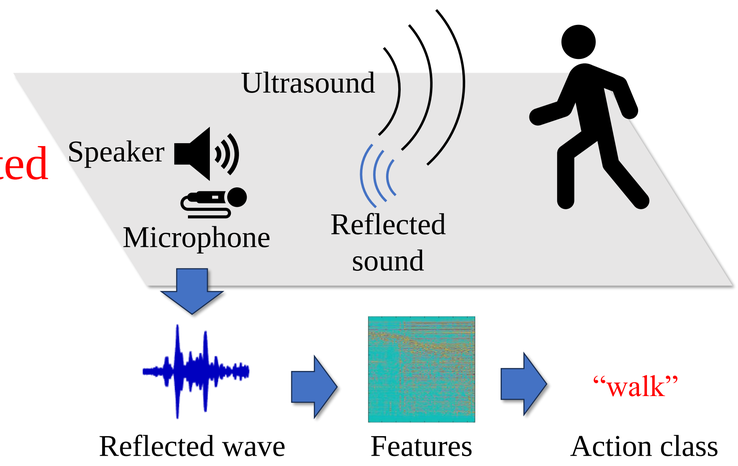


Fig. 1 Concept diagram of our work. Action classes are estimated based on reflected ultrasound. Ultrasound was chosen due to inaudible to humans.

2. Proposed method

Sensing system (Fig. 2)

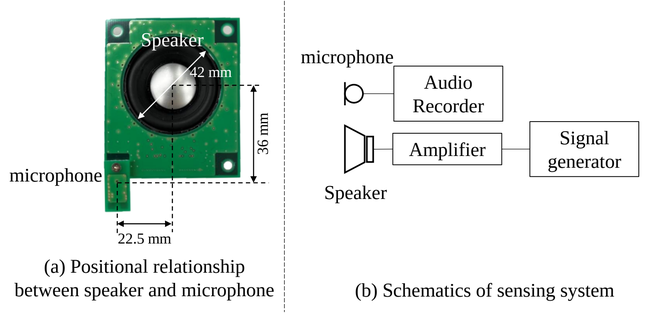
- Tweeter emitted chirp signal from 20 kHz to 40 kHz at interval of 11.8 ms.
- MEMS microphone received ultrasound and sampling frequency was set to 96 kHz.
- Measurable range was from 0.30 m to 2.0 m from the sensor.

Feature Extraction

- We focus on extracting features by observing the changes in the propagation characteristics of ultrasound associated with human movements.
- Two types of features:

(1) Time-series reflected waves
- Extracting reflected wave for each period of the chirp signal ($\boldsymbol{y}_i$: i-th reflected wave)
- Concatenating the waves as $F_{\mathrm{ref}} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N]$

(2) Time-series envelopes of reflected waves
- Calculating envelope of the reflected wave to eliminate the influence of phase ($\hat{\boldsymbol{y}}_i$: i-th envelope of reflected wave)
- Concatenating the envelopes of reflected waves as $F_{\mathrm{env}} = [\hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_2, \ldots, \hat{\boldsymbol{y}}_N]$



(a) Positional relationship between speaker and microphone

(b) Schematics of sensing system

Fig. 2 Schematic diagram of our sensing system.

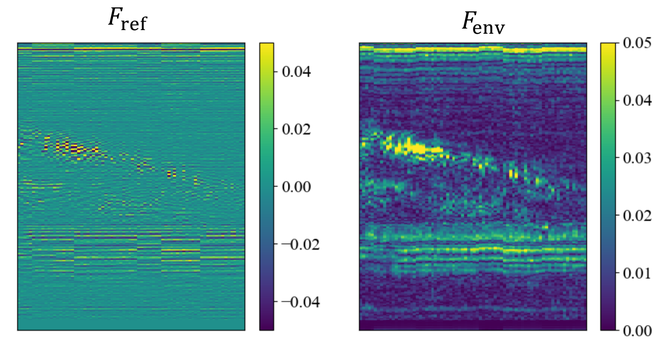


Fig. 3 Example features of `walking` class.

3. Dataset

- We created a dataset for ultrasound action recognition because there was no existing dataset
- Data was recorded in three different rooms (Fig. 4)
  Ra: anechoic chamber, Rb: Room without furniture, Rc: Room with furniture
- Each data includes a single subject and a total number of subjects are four, aged from 23 to 28.
- They continuously performed one action for about one minute.
- 8 action classes:
  `hand-waving, throwing, kicking, picking-up, walking, lying-down, sitting, standing`
- The total duration is 2,004 seconds.

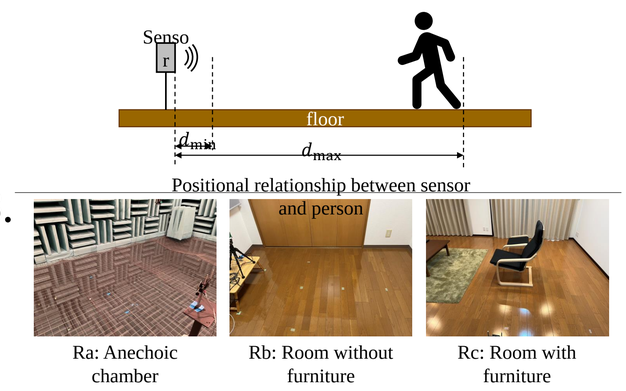


Fig. 4 Schematic diagram of data acquisition condition and pictures of room.

4. Experimental result

- Action classification were performed with SVM [1] and VGG [2].
- We evaluated 7 different conditions by changing the pairs of data used for training and evaluation.
- Experimental conditions and accuracy results are in Table 1.
  - Model comparison: VGG was better than SVM except for No. 4 with F_{env}
  - Feature comparison: F_{env} had 11.6 points higher than F_{ref} in SVM.
- Performance tended to depend on room conditions and subjects.
  - No. 1, 2 (same-room-same-subject): reached 99.8% in the best
  - No. 3-6 (same-room-different-subject): depended on subjects
  - No. 7 (different-room): low accuracy

No.	Ra #1	Rb #1	Rc				Accuracy [%]			
							SVM		VGG	
			#1	#2	#3	#4	Fref	Fenv	Fref	Fenv
1	T/E	-	-	-	-	-	78.4	**85.0**	**97.9**	94.6
2	-	T/E	-	-	-	-	97.3	**98.7**	98.3	**99.8**
3	-	-	E	T	T	T	**81.1**	80.1	89.3	**89.5**
4	-	-	T	E	T	T	52.9	**73.5**	**72.1**	60.0
5	-	-	T	T	E	T	18.1	**43.8**	46.7	**51.2**
6	-	-	T	T	T	E	0.8	**35.2**	40.2	**49.6**
7	T	T	E	E	E	E	**20.7**	14.1	**22.7**	22.2
Average							49.9	**61.5**	**66.7**	**66.7**

Table 1. Experimental conditions and accuracy results. The character "T" and "E" represents the data used for training and evaluation, respectively. #1~#4 represents subject ID.

5. Conclusions

- We proposed a new task of human action recognition using ultrasound.
- We confirmed that action can be estimated with high performance under simple conditions.
- We will consider feature extraction methods that are robust to environmental conditions.

References

[1] C. Cortes et al., *Mach. learn.*, 273–297, 1995.
[2] K. Simonyan et al., arXiv, 1409.1556, 2014.