深圳大学
SHENZHEN UNIVERSITY

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

AcouDigits: Enabling Users to Input Digits in the Air

Yongpan Zou[†], Qiang Yang[†], Yetong Han[†], Dan Wang[†], Jiannong Cao[‡], Kaishun Wu[†]

[†]*College of Computer Science and Software engineering, Shenzhen University*
[‡]*Department of Computing, Hong Kong Polytechnic University*

@Kyoto

PerCom 2019

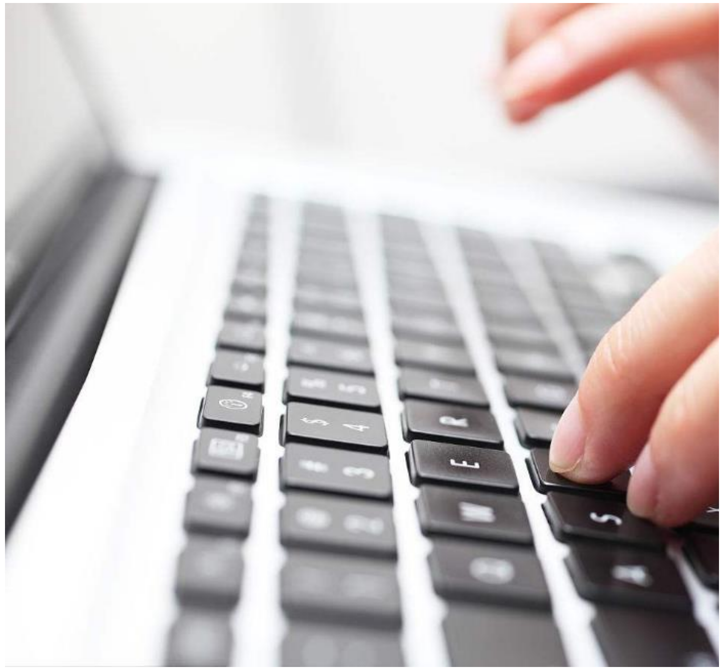
Outline

AcouDigits – motivation

Traditional interaction interface - Keyboard

Smartphone

Table computer



PC

AcouDigits – motivation

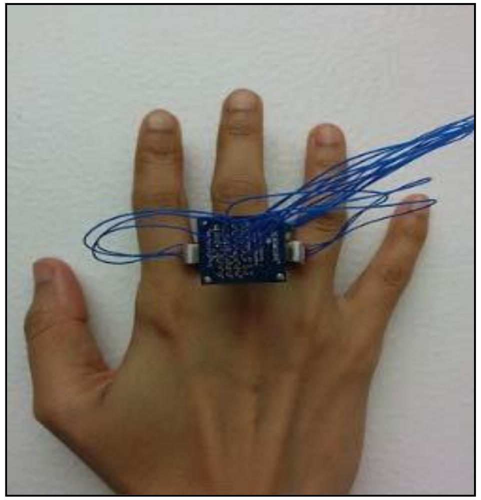
For new smart devices? **Small screen size / no screen!**

Smart watch

Smart glass

Smart home

AcouDigits - related work

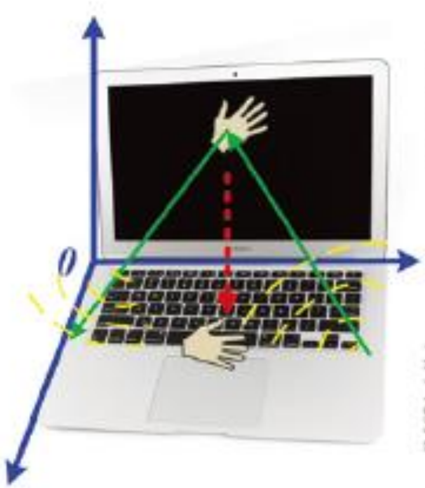


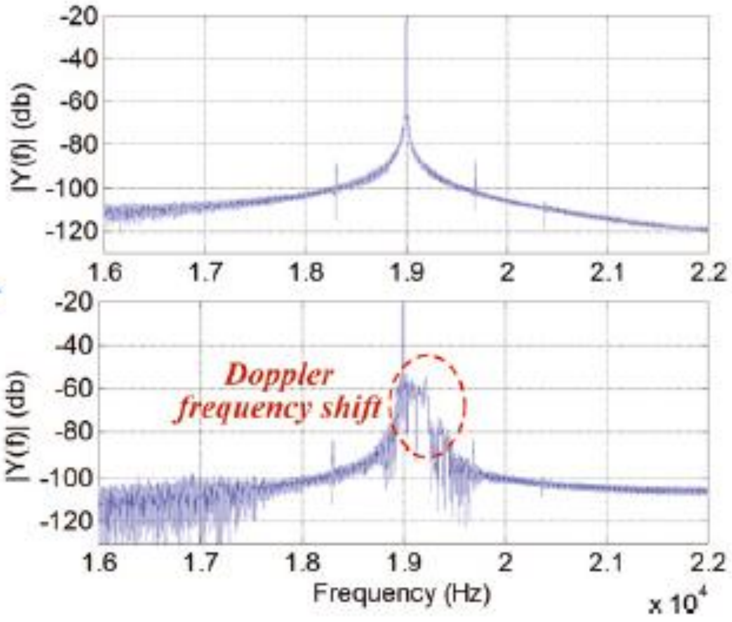
Keyboard | RF | speech recognition | IMU

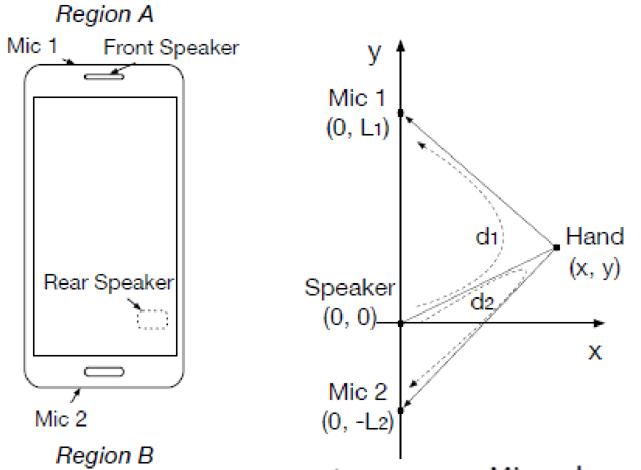
Small | *Unstable/Device* | *Privacy concern* | *Wearing device*

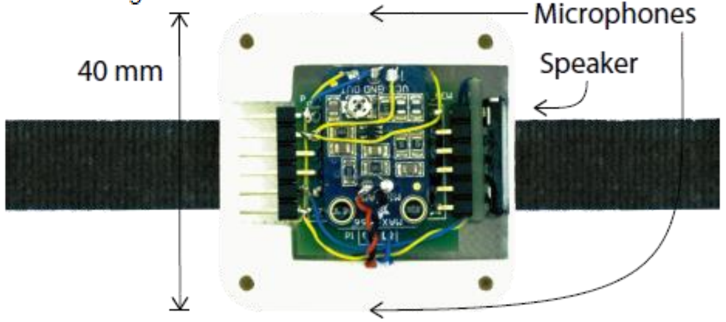
1. L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in Proceedings of ACM MobiSys, 2015.
2. J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: virtual touch screen in the air using rf signals," in Proceedings of ACM SIGCOMM, 2014.
3. S. Nirjon, J. Gummeson, D. Gelb, and K.-H. Kim, "Typingring: A wearable ring platform for text input," in Proceedings of ACM MobiSys, 2015.
4. C. Amma, M. Georgi, and T. Schultz, "Airwriting: Hands-free mobile text input by spotting and continuous recognition of 3d-space handwriting with inertial sensors," in Proceedings of IEEE ISWC, 2012.

AcouDigits - related work









Hand gesture recognition

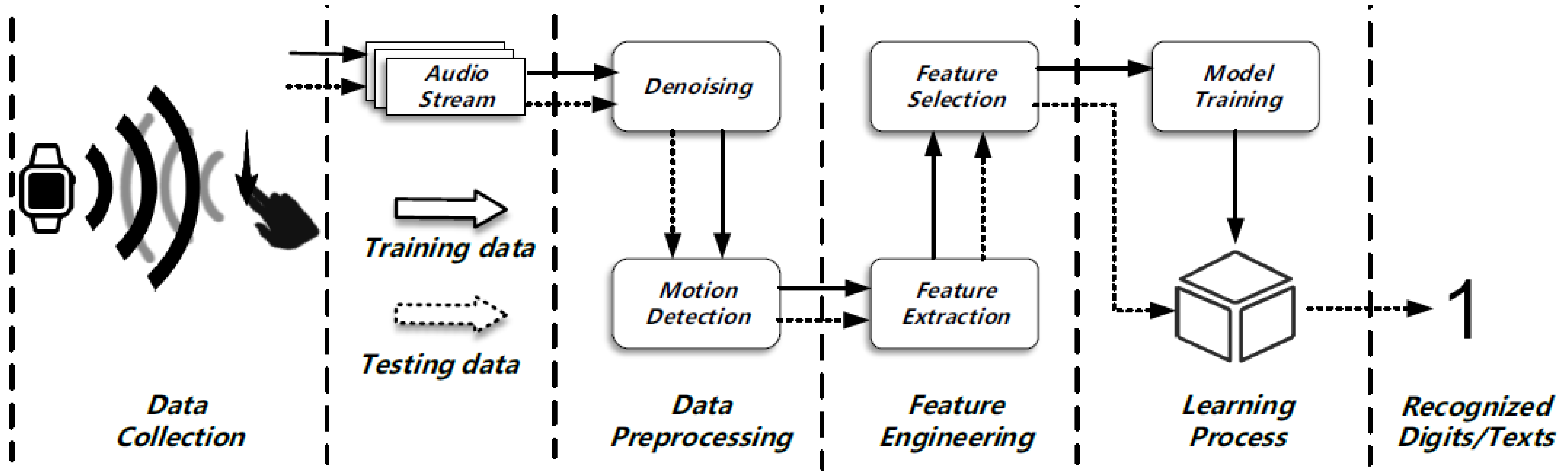
Coarse-grained HAND gesture

Acoustic finger tracking

Two microphones are required

1. *S. Gupta, D. Morris, S. Patel, and D. Tan, "Soundwave: using the Doppler effect to sense gestures," in Proceedings of ACM CHI, 2012.*
2. *W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in Proceedings of ACM Mobicom, 2016.*
3. *W. Mao, J. He, and L. Qiu, "CAT: high-precision acoustic motion tracking," in Proceedings of ACM Mobicom, 2016.*

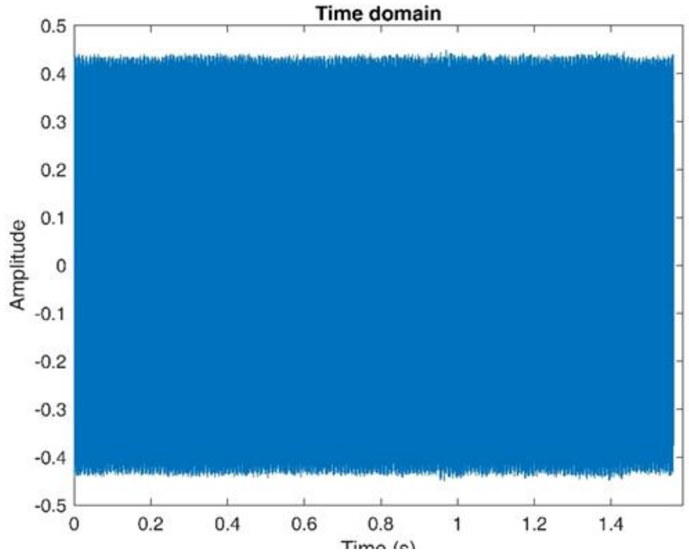
AcouDigits - workflow

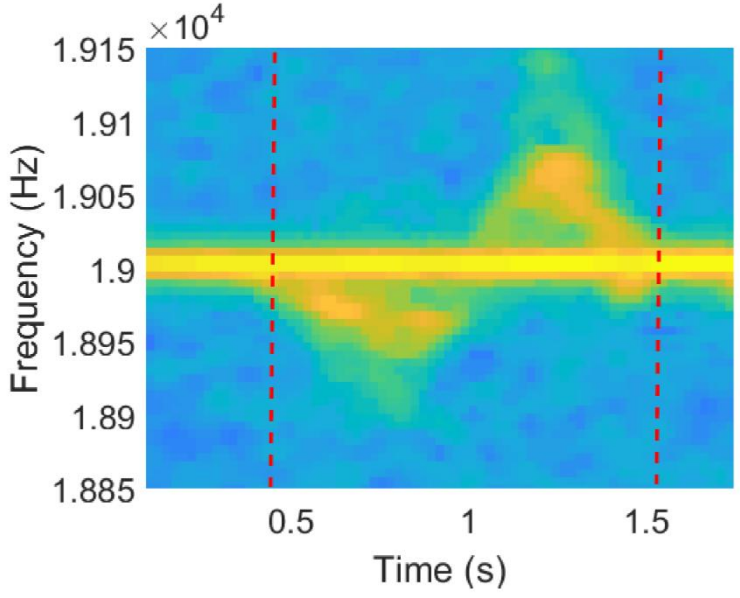


19 KHz

AcouDigits - Data preprocessing

- **Denoising**
  - Bandpass filter: [18850; 19150]
  - Direct path: Bandstop filter
- **Event Detection**
  - Continuous 4 frequency bins exceed a threshold: *Active*
  - Segment: Continuous 4 frequency bins less than a threshold: *End*





Doppler Effect

$$\boxed{\Delta f = f_0 \cdot |1 - \frac{v_s \pm v_f}{v_s \mp v_f}|}$$

f_0, the frequency of emitted signals
v_s, the speed of sound
v_f, the velocity of finger motion

AcouDigits - Data preprocessing

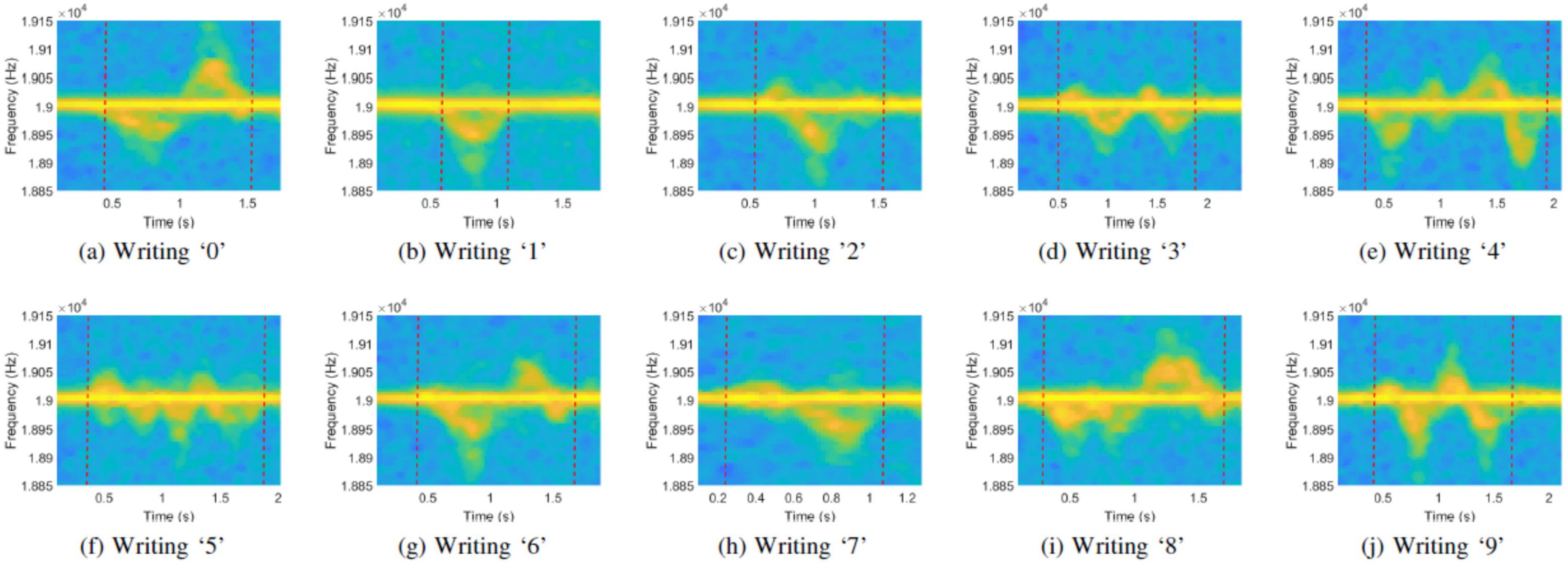
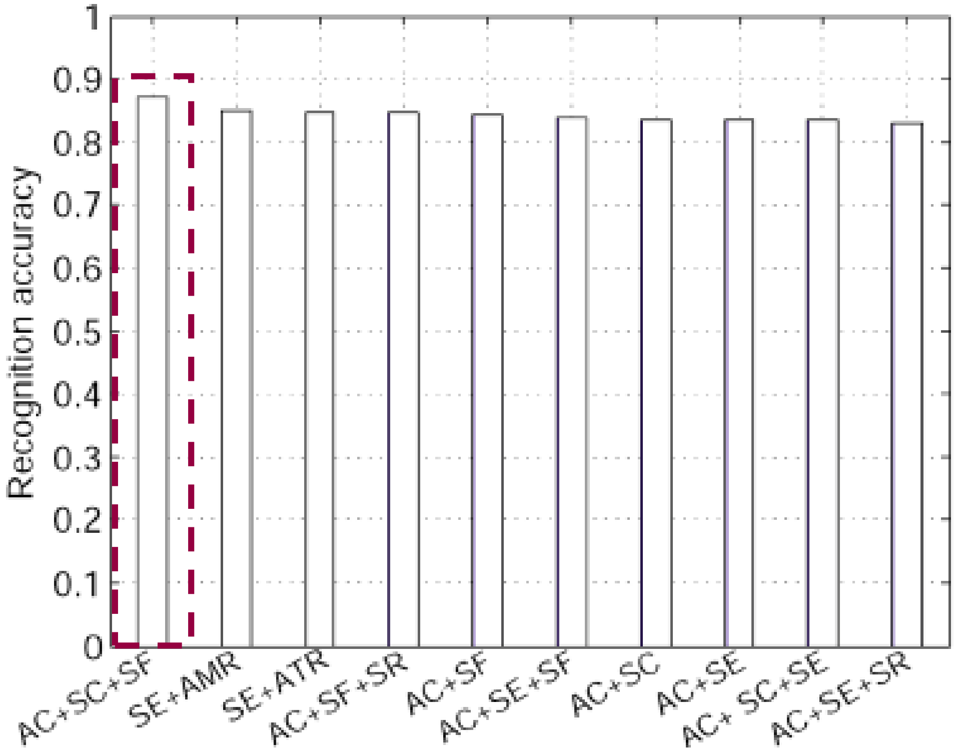


Fig. 3. Illustrations of spectrograms and writing activity detection.

AcouDigits – feature engineering

Feature vector: Mean value and variance of AC, SC, SF

Feature domain	Feature	Description
Time domain	Root mean square (RMS)	The energy in an acoustic frame
	Zero crossing rate (ZCR)	The point where acoustic samples change signs
	ATR	The average value of top k RMSs
	Above α-mean ratio (AMR)	The ratio of high-energy frames in a window
	AC	Auto-correlation coefficients
Frequency domain	Spectral entropy (SE)	The flatness indicator of acoustic spectrum shape
	Spectral flux (SF)	The stability reflector of acoustic events
	Spectral rolloff (SR)	Indicator of a frame's spectral energy distribution
	Spectral centroid (SC)	The balance point of the spectral energy distribution

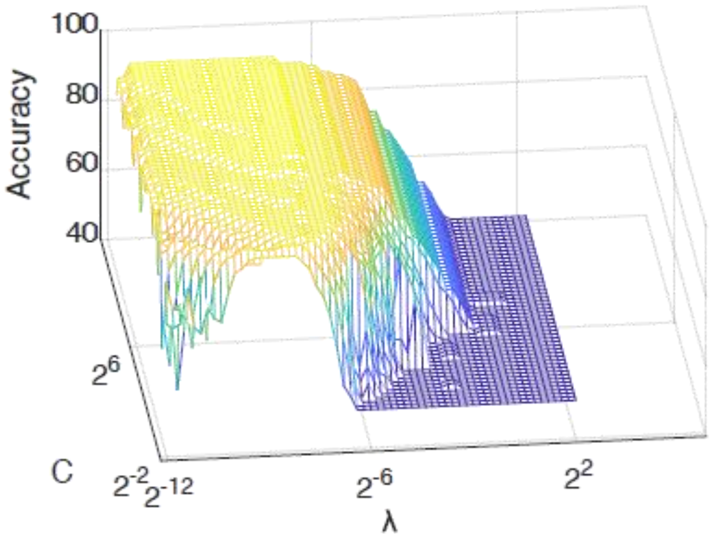


Feature selection **(Wrapper method)**

10-fold cross validation

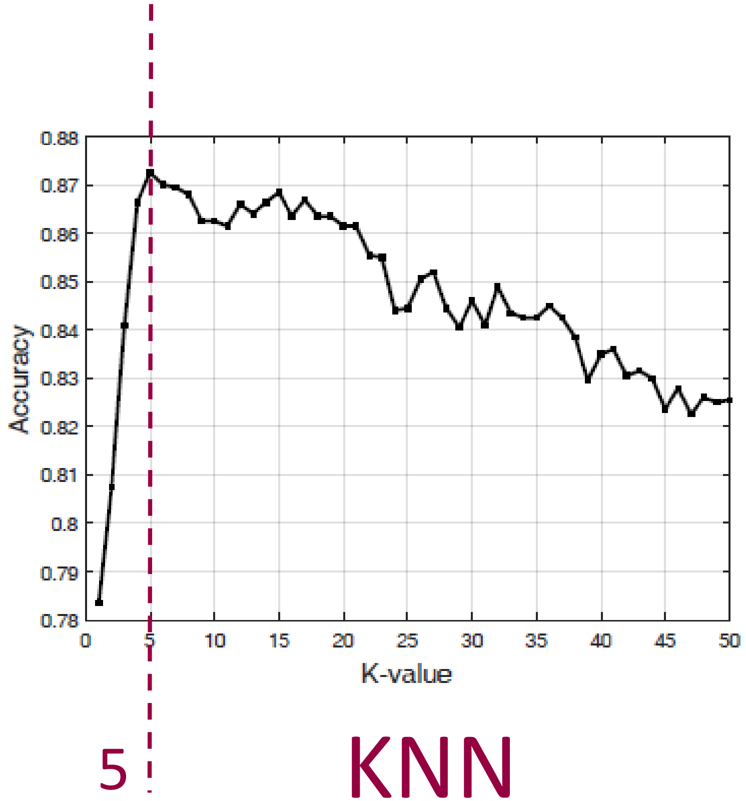
AcouDigits – Model training

- **SVM**
  - RBF kernel
  - C (penalty coefficient): 2^{10}
  - Γ (kernel function coefficient): 2^{-10}



SVM

- **KNN**
  - K=5



KNN

AcouDigits – Model training

ANN

PARAMETER SETTINGS OF ANN MODEL

Parameters	Value
Number of layers (L)	2
Number of nodes (N)	10
Training function (f)	Levenberg-Marquardt algorithm
Activation function (ϕ)	$\phi_1 = \frac{2}{1+e^{-2n}} - 1$ $\phi_2 = \frac{e^n}{\sum e^n}$

PERFORMANCE OF DIFFERENT TRAINING FUNCTIONS

Training functions	trainlm	trainbr	trainbfg	trainrp	trainscg	traincgb	traincgf	traincgp	trainoss	traingdx	traingdm	traingd
Training accuracy	94.80%	98.60%	77.10%	86.90%	83.60%	81.50%	82.20%	84.00%	81.40%	78.60%	16.90%	6.90%
Testing accuracy	92.80%	90.00%	77.00%	87.70%	81.00%	80.70%	82.30%	83.30%	81.00%	75.70%	20.00%	5.30%
Time(s)	19	266	3	1	1	1	1	1	1	1	2	2

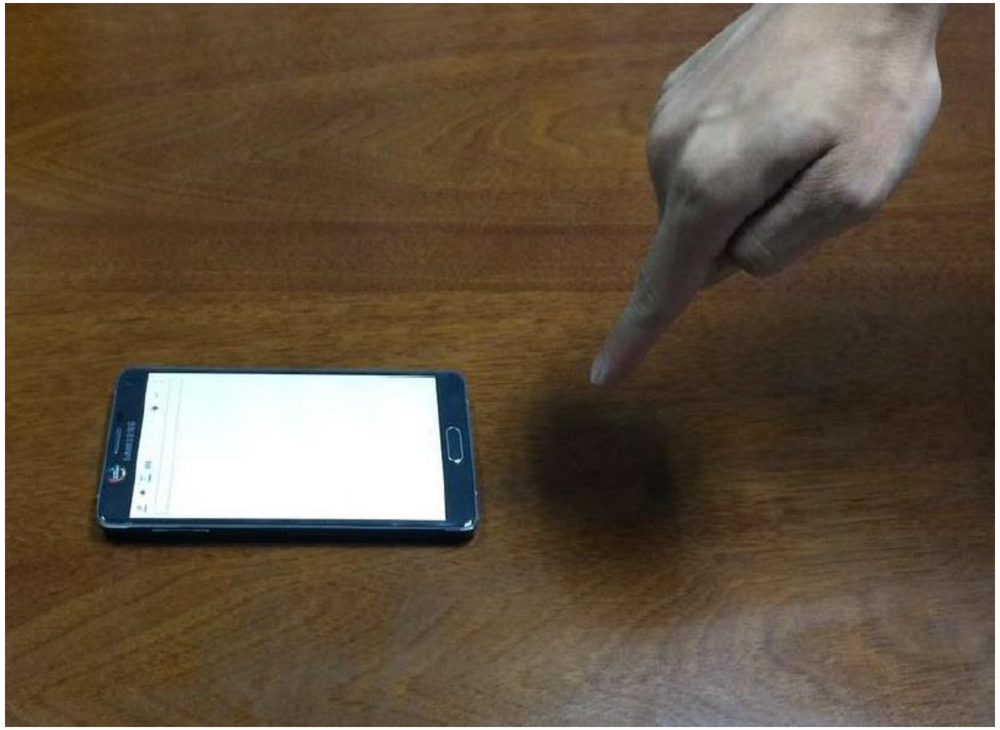
PERFORMANCE OF DIFFERENT ACTIVE FUNCTIONS

Active fuctions	compet	elliotsig	hardlim	hardlims	logsig	netinv	poslin	purelin	radbas	radbasn	satlin	satlins	softmax	tansig	tribas
Testing accuracy	9.80%	90.40%	9.00%	10.30%	88.30%	20.00%	84.30%	90.60%	86.30%	89.00%	73.30%	88.70%	90.30%	92.70%	65.00%

AcouDigits – experiment

Setup

- Samsung Galaxy S5
- Emitting: 19 KHz
- Sampling: 44.1KHz
- Distance:2-16cm



10 digits X 6 participants X 200 repetitions = 12,000
10 digits X 6 participants X 8 dis intervals X 50 repetitions = 24,000

8 distance intervals: 2-4-6-8-10-12-14-16cm

AcouDigits – evaluation

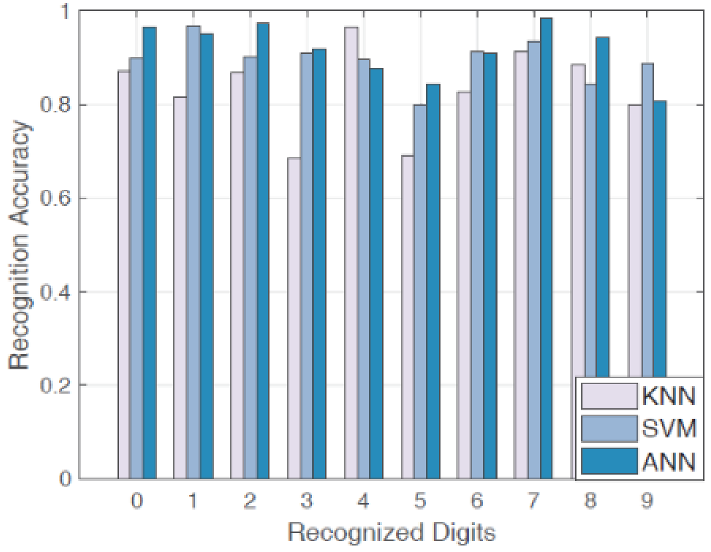


Fig. 8. The overall performance of AcouDigits for KNN, SVM and KNN models.

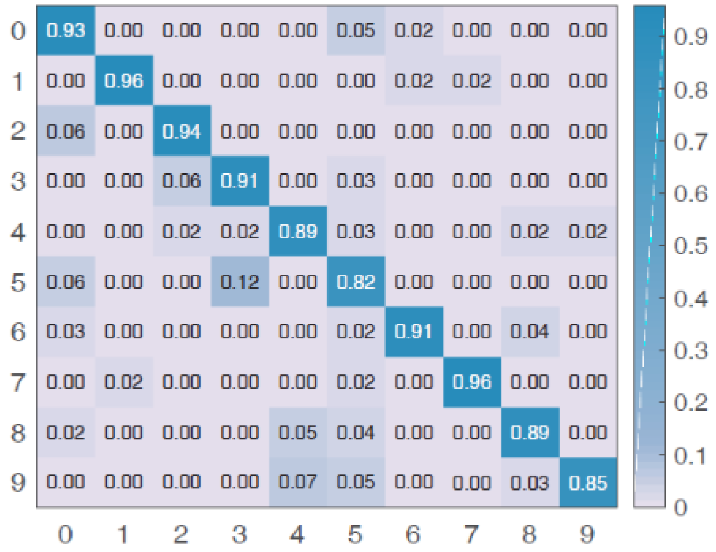


Fig. 9. The confusion matrix of AcouDigits while averaging the performance of SVM and ANN models.

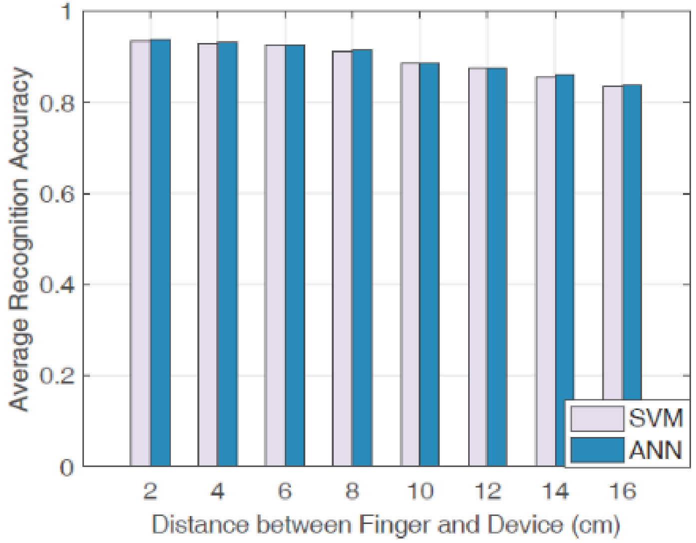


Fig. 10. The performance of AcouDigits for different distances between the finger and device.

Recognition Performance

- The overall recognition accuracy of SVM and ANN models are **89.5%** and **91.7%**, and are higher than that of KNN by 6.3% and 8.5%, respectively.

Safe Distance

- Within 8 cm, the performance remains acceptable with an accuracy no less than 91.5%.

AcouDigits – evaluation

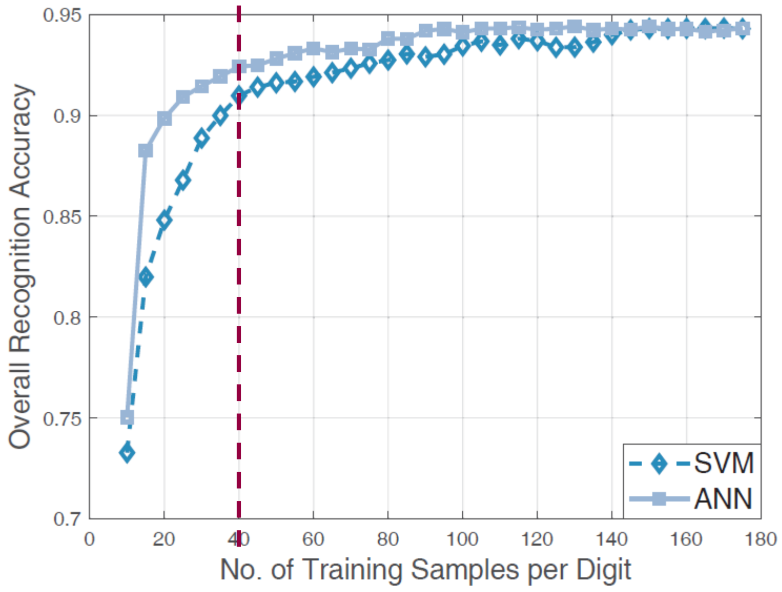


Fig. 11. The performance of AcouDigits for different numbers of training samples.

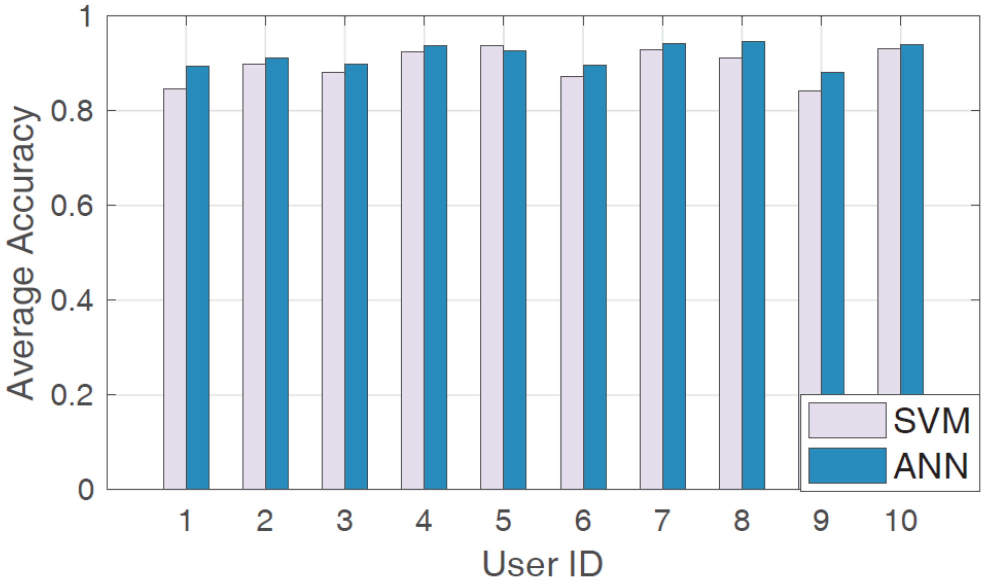


Fig. 14. The performance of AcouDigits for different participants in the experiments.

Training Overhead

- When the number of training samples exceeds **40**, the recognition accuracy increases much more slowly and remains nearly constant.

User Diversities

- The recognition accuracy varies from (**84.2%**, **88.0%**) to (**94.8%**, **95.2%**) with (0.14%, 0.06%) variance among different participants due to different writing habits.

AcouDigits – evaluation

Cross-person performance

- Training AcouDigits with one participant's data and testing it with another one's data.
- Randomly selected 5 pairs
- The average accuracies for SVM and ANN are 75.4% and 78.0%, respectively.

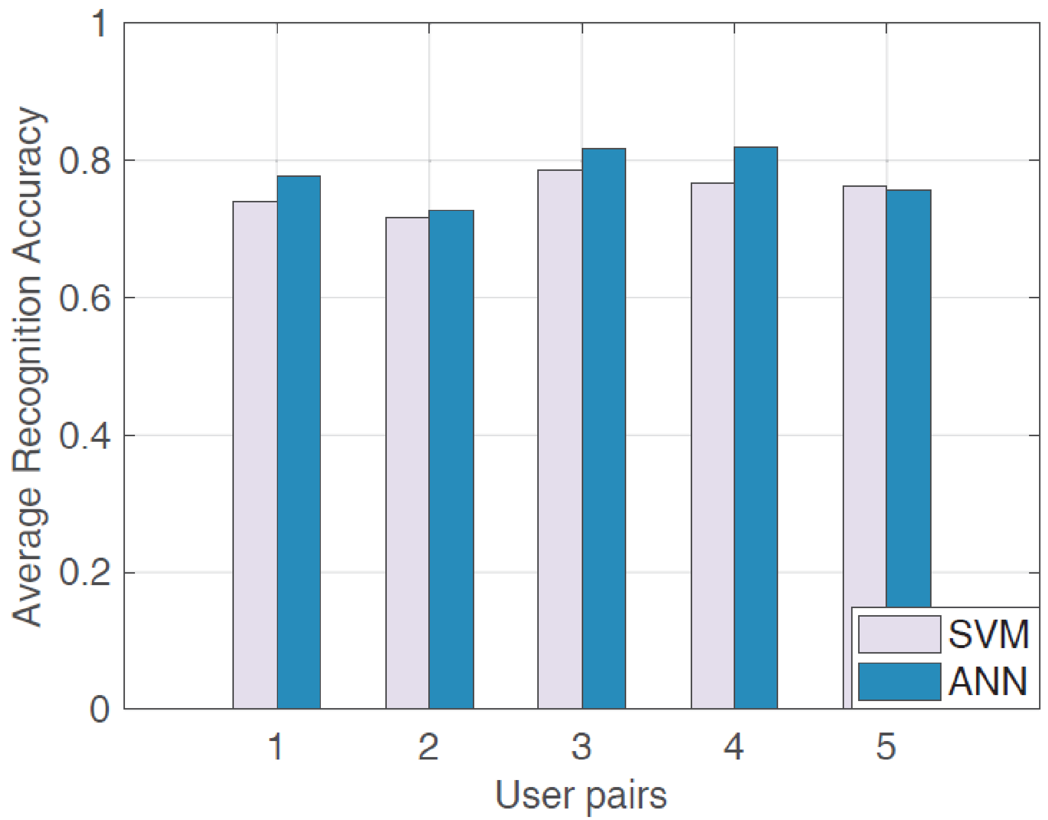


Fig. 12. The average accuracies of the selected five training-testing pairs.

AcouDigits – evaluation

A Direct Extension to English Letters

- **6** (participants)×**26** (letters)×**100** (repetitions) =**15600**
- use **ANN** as the learning model
- The average accuracy in recognizing **26** letters is **87.4%**
- Several letters have very similar writing patterns

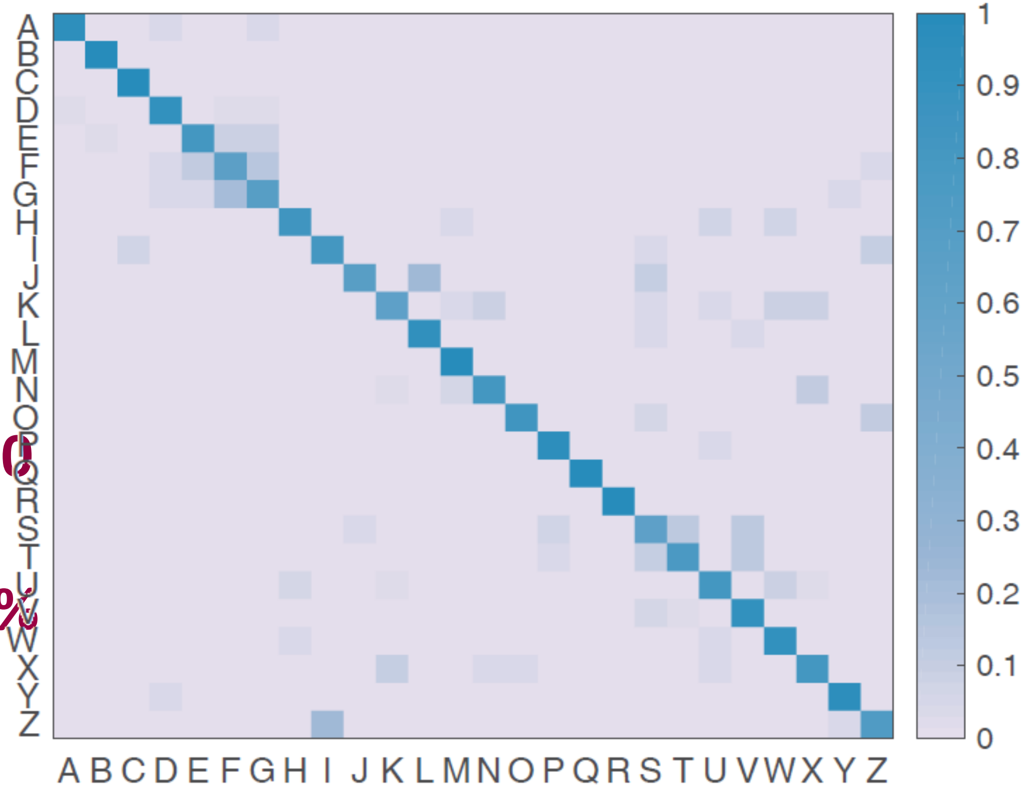


Fig. 13. The performance of AcouDigits in recognizing uppercase English letters.

AcouDigits – Conclusion

- We propose a novel interface that enables users writing digits and alphabets in the air without wearing any additional devices.
- By careful model selection and parameters tuning, AcouDigits can achieve up to **91.7%** recognition accuracy for digits.
- We extend AcouDigits to recognize 26 English letters, and can achieve an accuracy up to 87.4%.

AcouDigits – Further work

Deep learning-based [***ongoing extension***]

We transform acoustic signals to spectrograms, and using CNN to recognize digits and letters, which can achieve 94.9% accuracy.

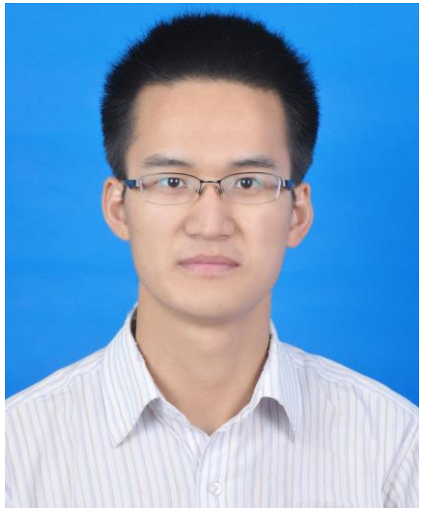
Writing anywhere [***ongoing extension***]

With the data produced by Data Augmentation at different location of devices, more robust AcouDigits can be trained, and user can writing digits at any location around the device.

Training-free text input [***new work under review***]

By decomposing English letters to basic strokes and modeling their intrinsic characteristics, we can input text without any user-training overload.

THANKS



https://yongpanzou.github.io/
yongpan@szu.edu.cn
College of Computer Science and Software Engineering
Shenzhen University