BARTON: Low Power Tongue Movement Sensing with In-ear Barometers

Balz Maag*, Zimu Zhou*, Olga Saukh+ and Lothar Thiele*

*ETH Zürich

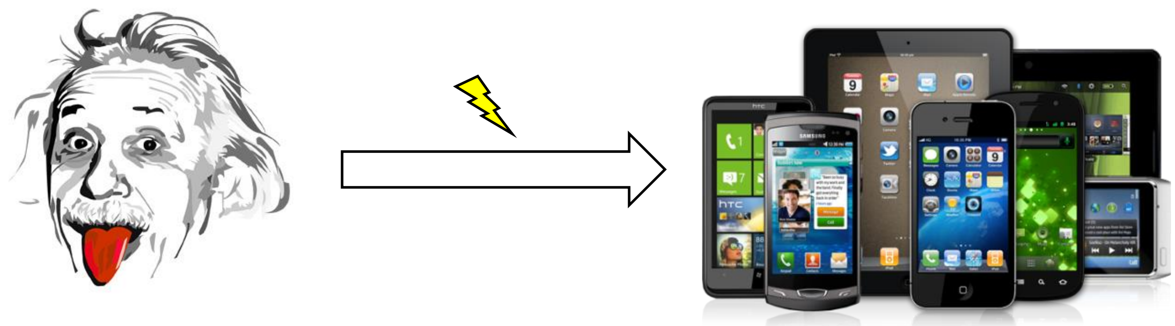
+TU Graz, CSH Wien

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

COMPLEXITY SCIENCE HUB VIENNA

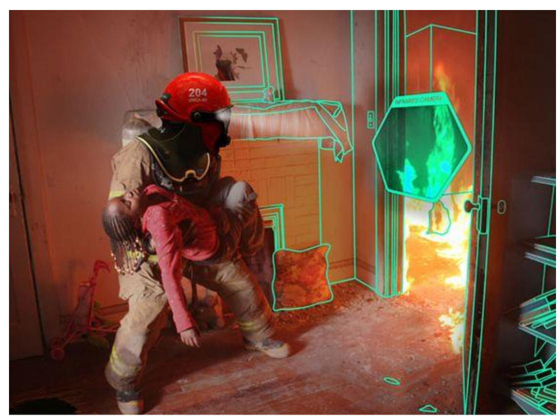
Tongue Movement Sensing

- Hands-free computer-human interaction



- Popular applications:

Handicapped support



Firefighters

Sports

Related Work

- Wireless sensing
  - Li et al.: Radar mounted on shoulders
- Tongue-mounted sensors
  - Sahni et al.: Magnetic sensors
- Skin-mounted sensors
  - Zhang et al.: EMG electrodes attached to neck
  - Rahman et al.: Microphones attached to neck

Intrusive
Limited practicability
Complex hardware

- In-ear sensing
  - Vaidyanathan et al.: Microphone in ear-canal
  - Mace et al.: Microphone in ear-canal

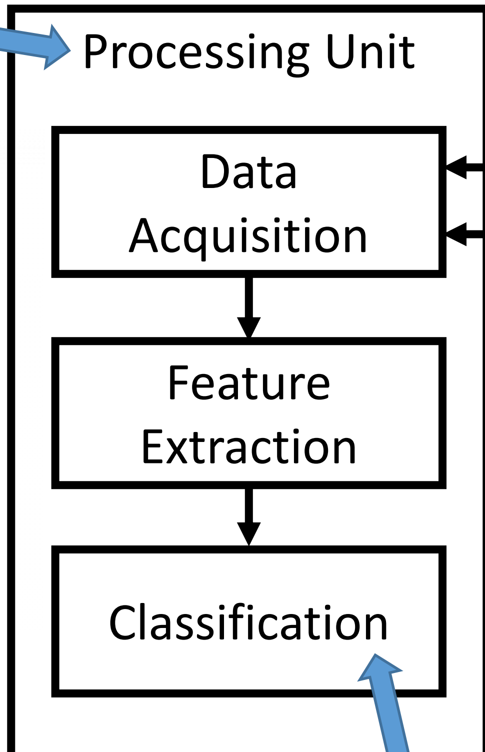
Vulnerable to interference (e.g. music)
Complex signal processing
High power consumption

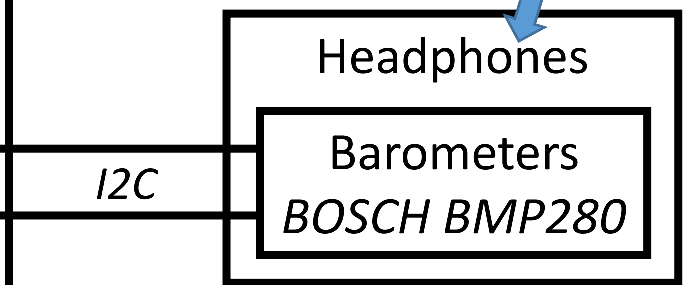
Our goal: A **practical** and **robust** tongue movement sensing system with **low power consumption** using **in-ear barometers**

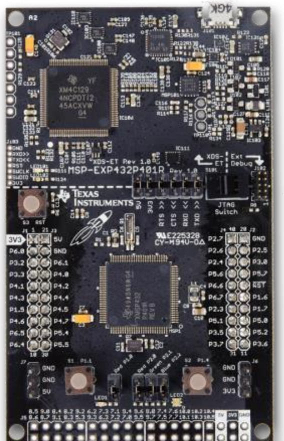
BARTON Design

***Practical**: COTS hardware and ordinary headphones*

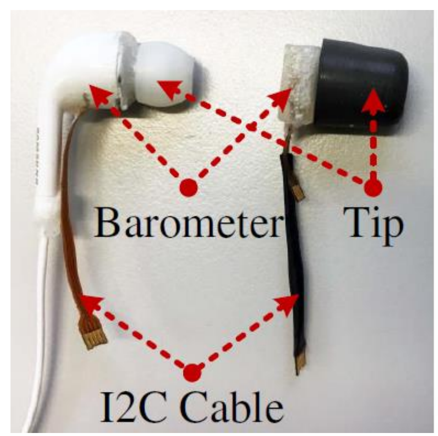
***Low-power**: Simple signal processing*

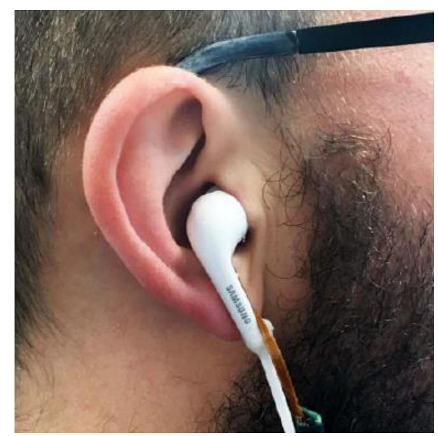






TI MSP432 Launchpad

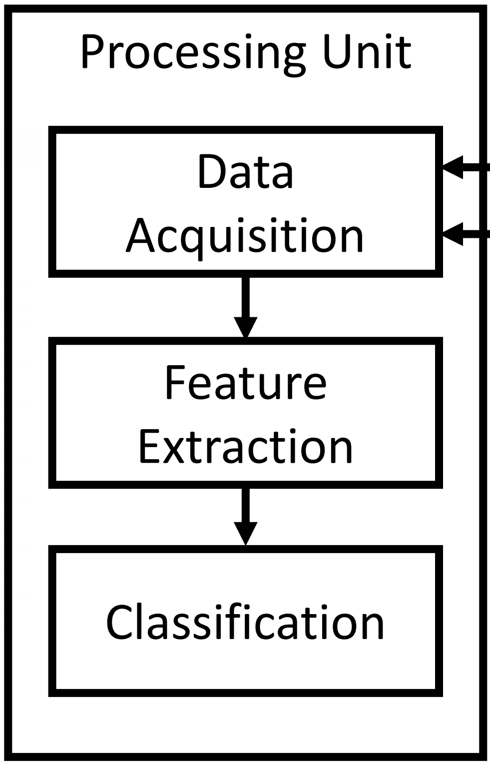


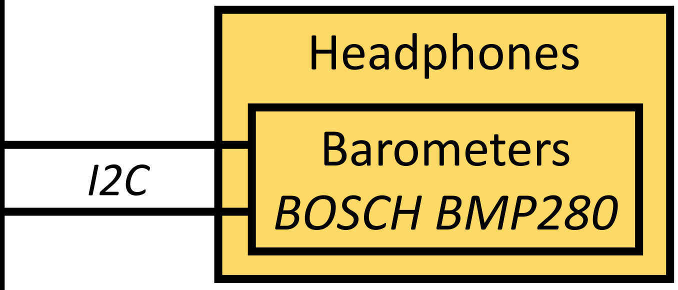


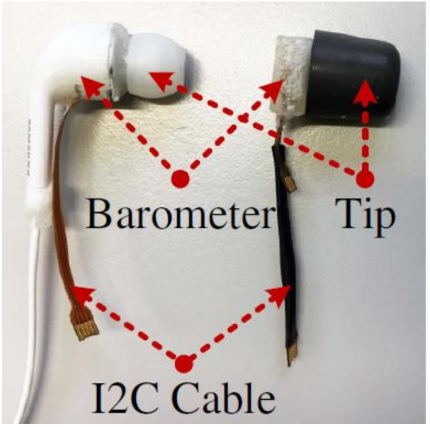
***Robust**: Resilient to head movement and music interference*

BARTON Design

TI MSP432 Launchpad

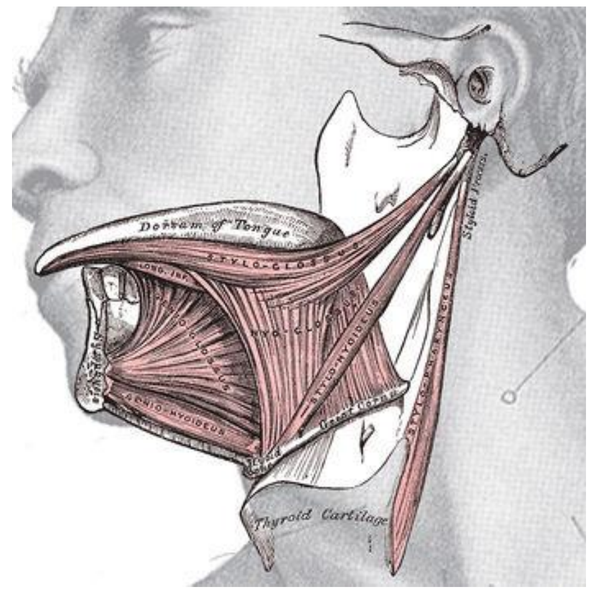






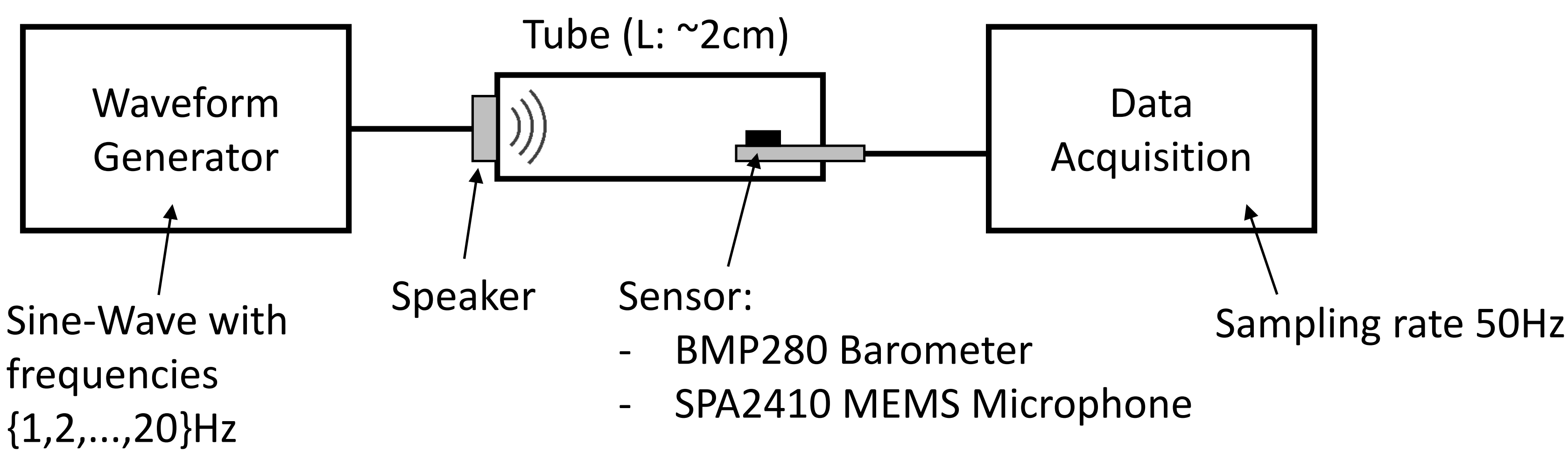
In-Ear Sensing

- Tongue movements (e.g. left, right, forward) lead to deformation of the ear-canal
- In-ear pressure changes with frequencies between [3, 6] Hz
- Challenges:
  - Capture the low-frequency pressure changes
  - Distinguish tongue movements from interferring activities such as head movements, music etc.

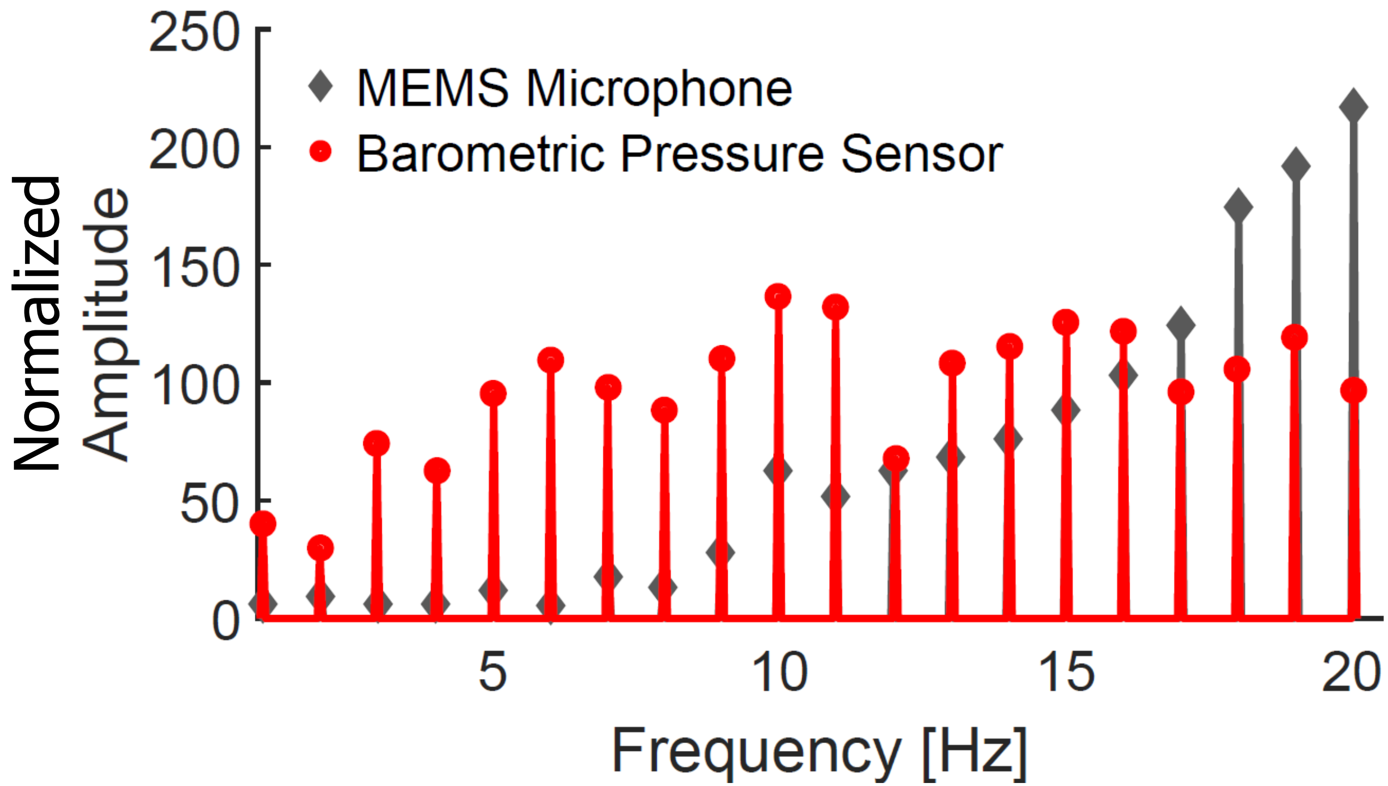


Barometers vs. Microphones

- Drawbacks of MEMS microphone based solutions:
  - High sampling rates (>8kHz) and frequency-space signal processing leads to high power consumption
  - MEMS microphones are not designed to capture pressure changes with frequencies <100Hz
- Barometers can capture low-frequent pressure changes at low sampling rates (≤50Hz)
- Experiment:

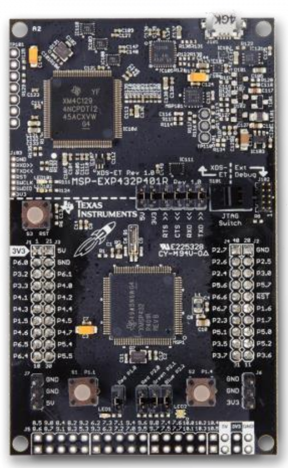


Frequency Response

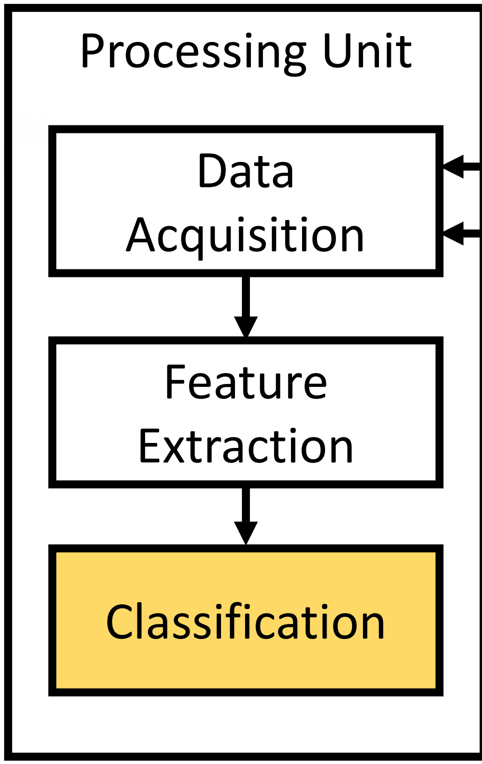


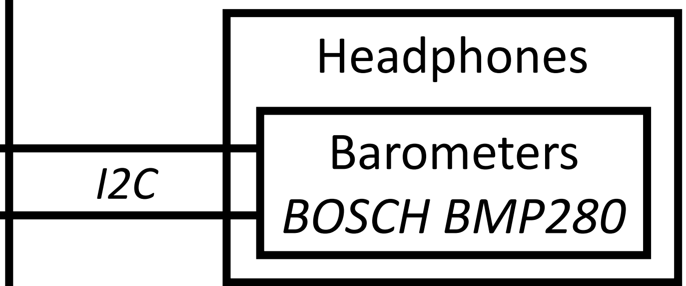
- Microphone frequency response drops linearly and cuts off at around 10Hz
- Barometer is not suffering from a linear frequency response drop and, most importantly, is more sensitive to frequencies below 10Hz

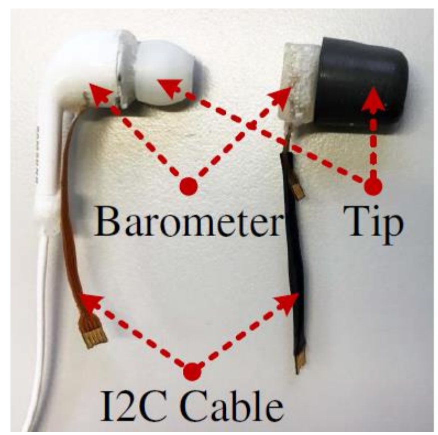
BARTON Design

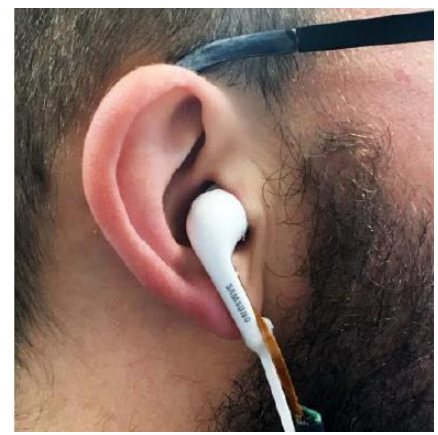


TI MSP432
Launchpad



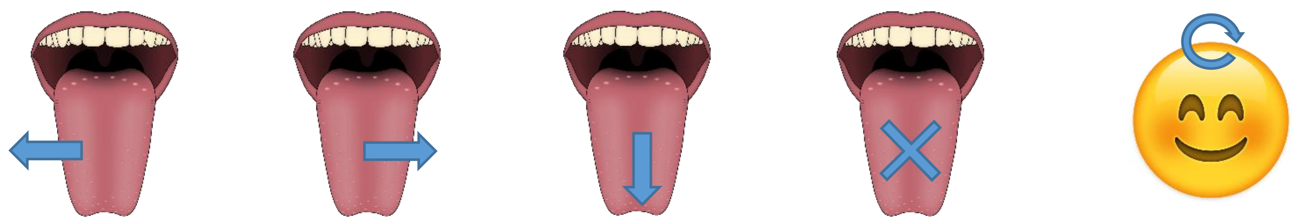






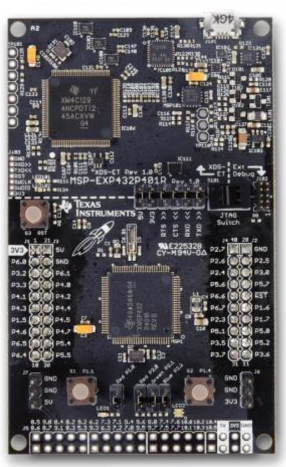
Classification

- 5 classification classes:

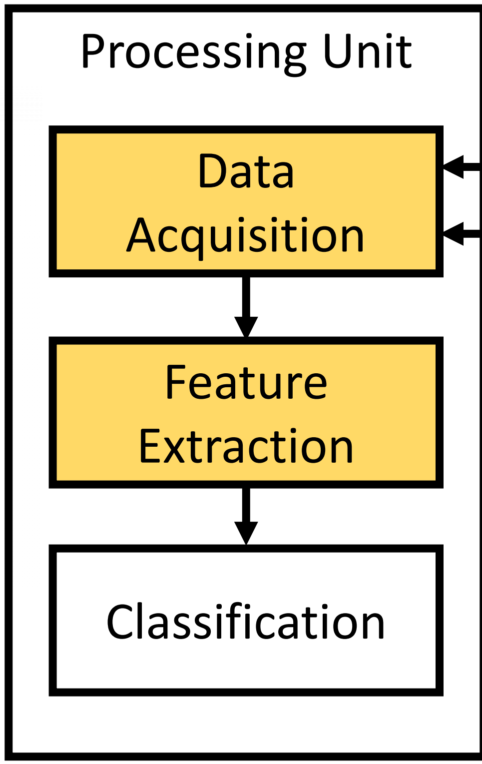


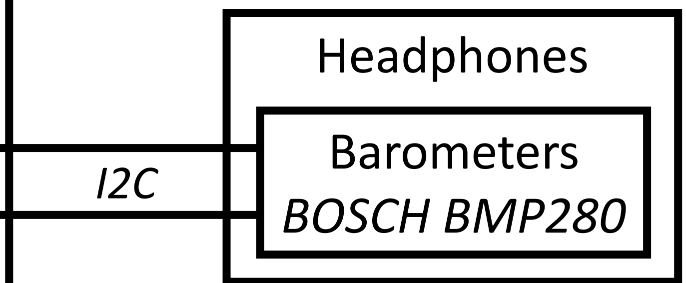
- 3 Simple Classifiers
  - Support vector machines (SVM)
  - k-Nearest Neighbors (KNN)
  - Decision Tree (DT)

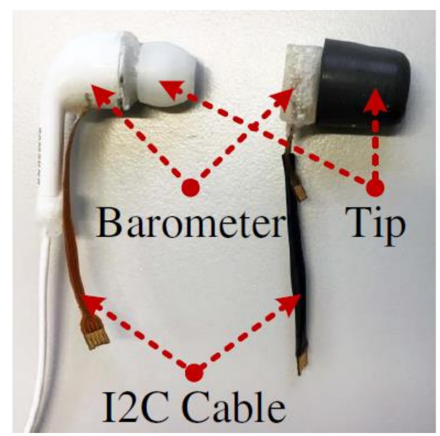
BARTON Design

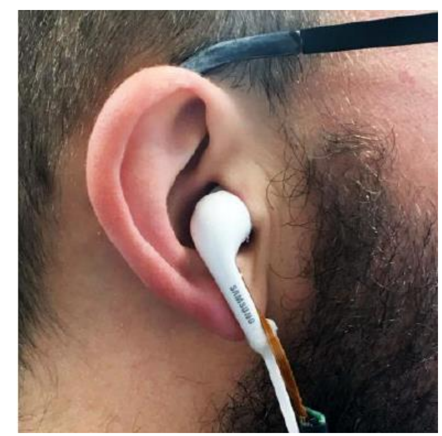


TI MSP432
Launchpad

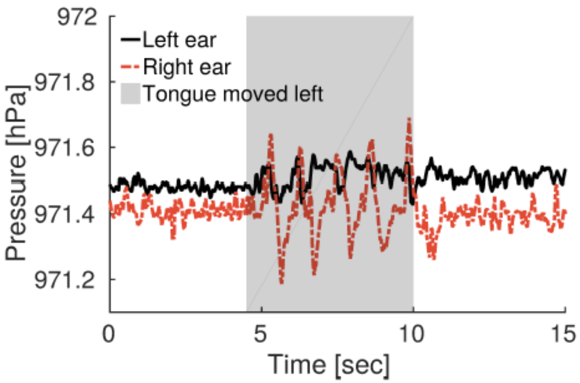




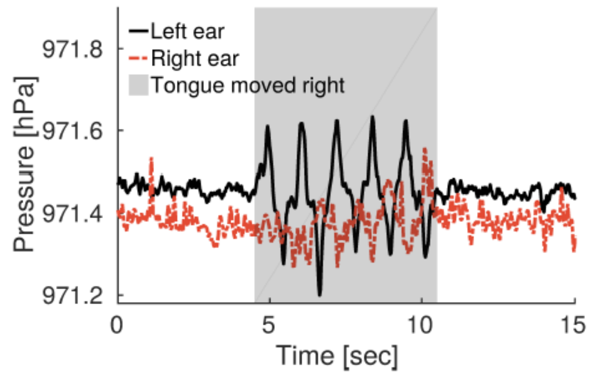




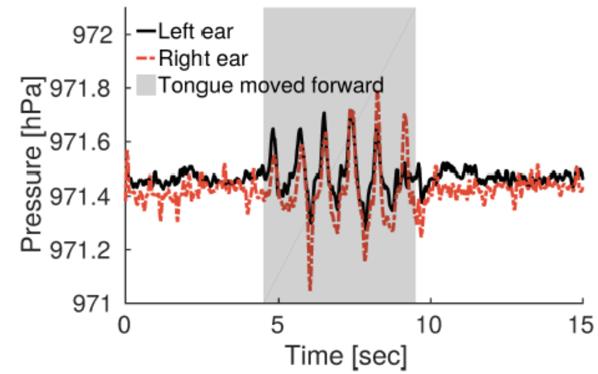
Feature Extraction



Tongue left (L)



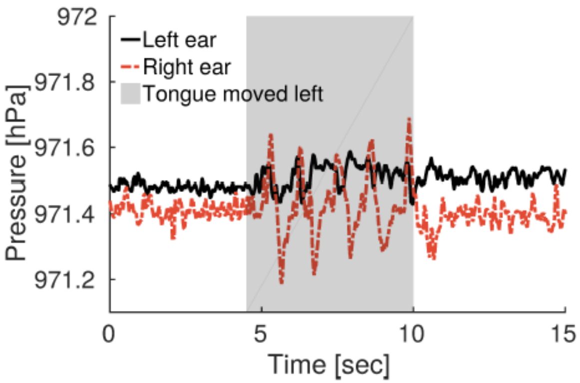
Tongue right (R)



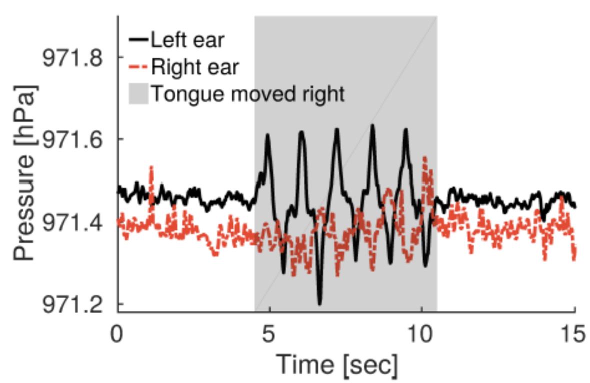
Tongue forward (F)

- Pre-processing of the two pressure data streams:
  - Sliding window *W* with 50% overlap
  - Detrending (subtracting mean) of samples within window
- Only temporal features to reduce computational complexity

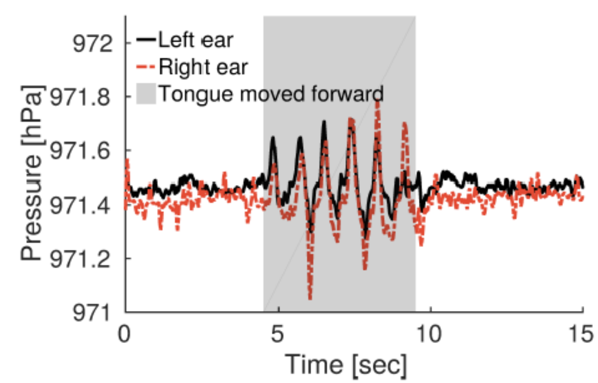
Feature Extraction



Tongue left (L)



Tongue right (R)



Tongue forward (F)

- Important features:
  - *min(W), max(W), min(diff(W)), max(diff(W)):*
    Moving the tongue to the left causes peaks in the right ear canal, and vice-versa
  - *covariance(W):*
    Moving the tongue forwards causes strong correlations between the pressure signals
- Feature set refinement: Sequential feature selection algorithm
  - Additonal features: *root-mean-square(W), covariance(diff(W))*

Evaluation: Classification Accuracy

	L	R	F	H	I
L	99.1	0.0	0.2	0.7	0.0
R	0.0	98.0	0.7	1.4	0.0
F	0.5	0.7	88.3	6.5	4.1
H	0.7	1.1	7.0	89.0	2.3
I	0.0	0.9	2.7	1.1	95.3

SVM
mean acc.: 94%

	L	R	F	H	I
L	98.9	0.0	0.5	0.7	0.0
R	0.0	95.9	3.4	0.7	0.0
F	0.2	6.3	80.6	10.6	2.3
H	1.4	0.9	15.8	78.6	3.4
I	0.0	0.2	2.3	3.2	94.4

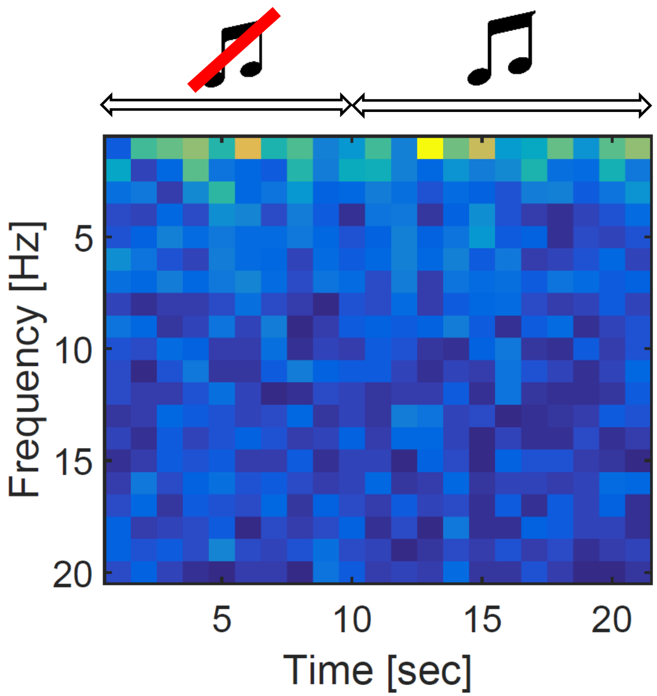
KNN
mean acc.: 91%

	L	R	F	H	I
L	98.6	0.0	0.0	1.4	0.0
R	0.5	95.5	1.4	2.0	0.7
F	0.2	1.4	83.6	11.7	3.2
H	0.7	1.6	11.3	83.3	3.2
I	0.0	0.2	2.9	2.7	94.1

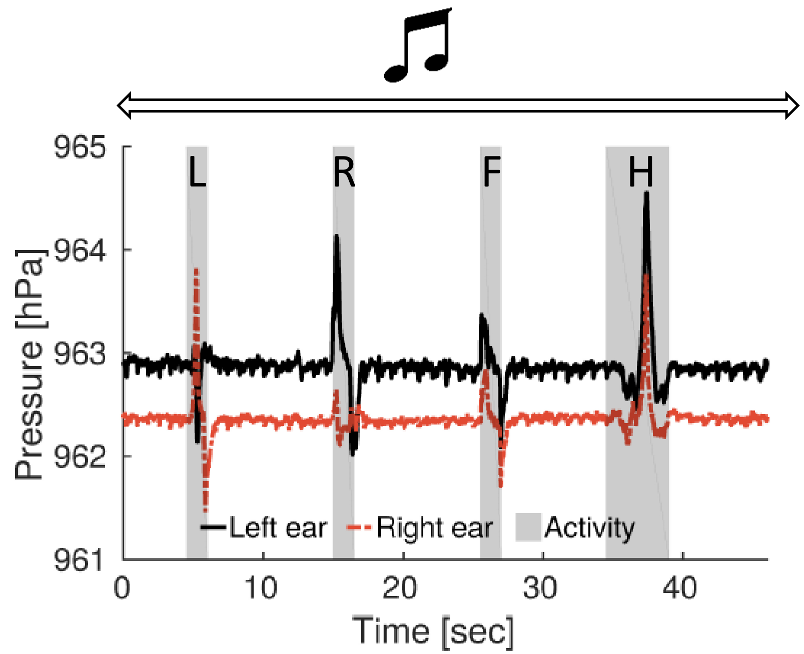
DT
mean acc.: 89%

- 40Hz sampling rate, 1sec sliding window
- 444 elements for each class
- 10-fold cross-validation

Robustness to music



Spectogram during Idle



Pressure signals while playing music

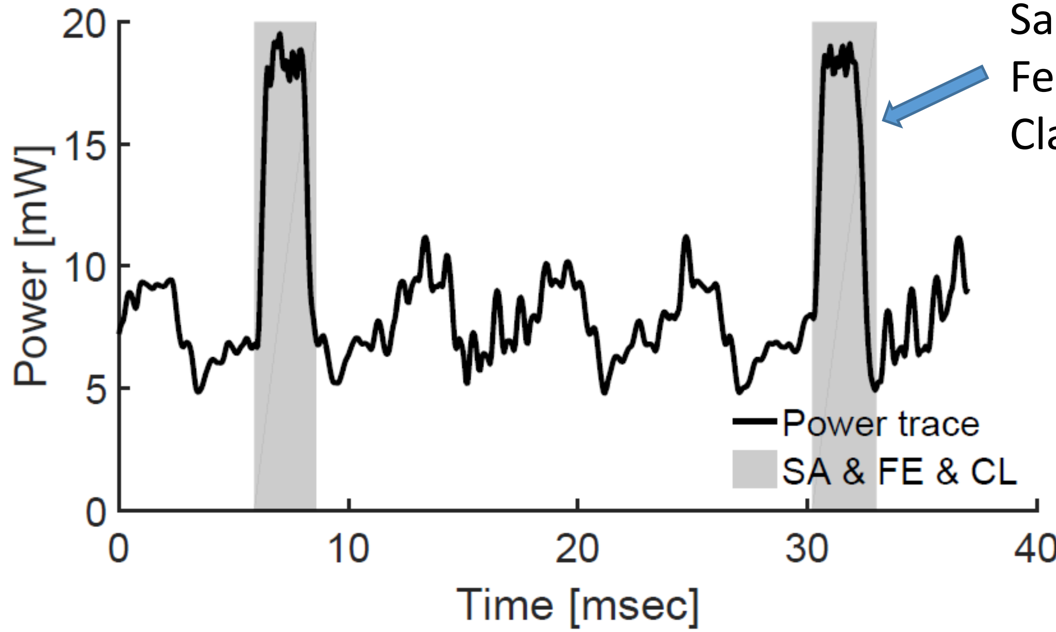
- Music played by Huawei P8 smartphone at 75% volume
- BARTON is not affected by music played through the headphones

Robustness to different users

Activity	User 1	User 2	User 3	User 4	User 5
Left (L)	97%	54%	91%	61%	45%
Right (R)	90%	67%	72%	50%	82%
Forward (F)	100%	87%	63%	20%	100%
Head mov. (H)	93%	64%	93%	100%	56%
Idle (I)	100%	97%	96%	100%	100%
Overall	**96%**	**72%**	**87%**	**66%**	**70%**

- 5 users (1, 2: female, 3, 4, 5: male)
- 2-fold cross-validation with 12 to 38 elements per class for each user
- Reasons for different accuracies:
    1. Features of the tongue movements are different for each user
    2. Headphone tips need to fully enclose the ear canal (particularly a problem for user 4 and 5)

Power



Sampling ~1.5msec
Feature extraction ~0.14msec
Classification ~0.1msec

- Decision tree implementation using TI-RTOS
- Sampling rate 40Hz, 1sec sliding window
- Average power consumption: 8.4mW
  - 44x more efficent than micro-phone based implementation by Rahman et al.

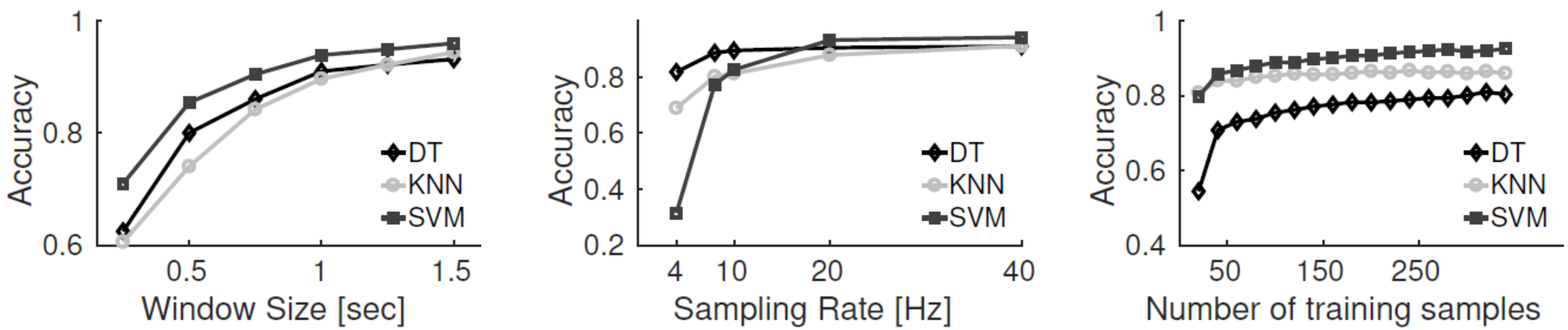
Conclusion

- We present BARTON, a tongue movement sensing system that...
  - is **practical** due to the integration of barometers into ordinary headphones
  - is **robust** against interference like head movement and music
  - with **low-power** consumption due to simplistic signal processing and classification
- Evaluations show that BARTON can achieve up to 94% classification accuracy of three major tongue movements while consuming 44x less power compared to related work

Thank you!

Questions?

Backup: Classification Robustness



- **Window size** of at least 1 sec achieves good accuracies for all the three classifiers.
  - Tongue activities have an average duration of 1 sec while the shortest and longest samples last for 0.75 sec and 1.25 sec respectively.
- A low **sampling rate** improves power efficiency but the coarse sampled pressure signals may decrease the classification accuracy.
- BARTON requires a minimal of 40 **samples** for each tongue movement to yield a reasonable classification accuracy of 85% for both the SVM and KNN classifiers. The decision tree needs over 300 samples to reach an accuracy over 80%.