# Gesture Spotting Using Wrist Worn Microphone and 3-Axis Accelerometer

*Jamie A Ward [1], Paul Lukowicz [2] and Gerhard Tröster [1]*

*(1) Swiss Federal Institute of Technology (ETH)*
*Wearable Computing Lab, Zurich*
*(2) University of Health Sciences, Medical Inf. and Tech. (UMIT)*
*Tirol, Austria*

Wearable Computing Lab. ETH Zürich

UMIT ::: CSN

UMIT ::: CSN

- Motivation
- Overview of experiment
- Sound & acceleration
- Classifier combination
- Evaluation
- Conclusion

Wearable Computing Lab. ETH Zürich

UMIT ::: CSN

- **Motivation**
- Overview of experiment
- Sound & acceleration
- Classifier combination
- Evaluation
- Conclusion

Wearable Computing Lab. ETH Zürich

UMIT CSN

# Wearable Activity Recognition

- Assembly and maintance a key wearable application
- Activity recognition and tracking
  - Automatic manuals
  - Hazard prevention, etc.

"Recognizing Workshop Activity
Using Body Worn Microphones and Accelerometers"
*Pervasive ='04*

wearit@work

Wearable Computing Lab. ETH Zürich

UMIT CSN

- Motivation
- **Overview of experiment**
- Sound & acceleration
- Classifier combination
- Evaluation
- Conclusion



# Woodshop Scenario

- 9 woodwork tool activities
- Fixed sequence:
  ..
  'get saw from drawer'
  'cut wood with saw'
  'put saw back in drawer'
  ..

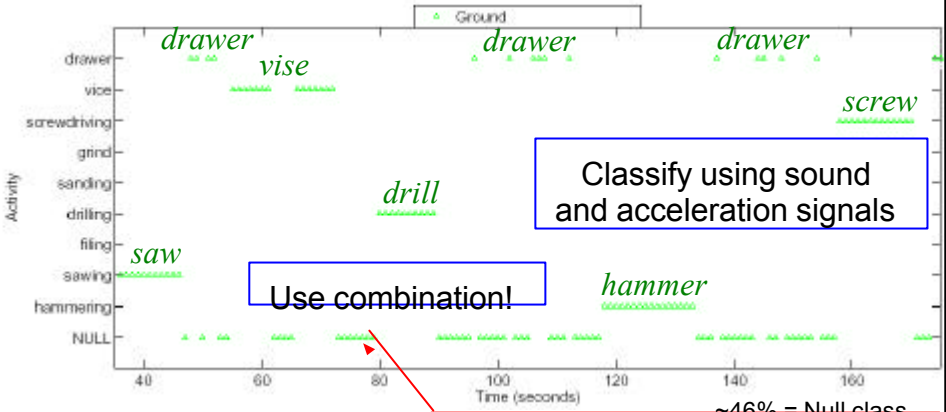
# Woodshop Scenario

- 9 woodwork tool activities
- Fixed sequence:
  ..
  'get saw from drawer'
  'cut wood with saw'
  'put saw back in drawer'
  ..
- Wrist mounted microphone & accelerometer (3-axis)
- 5 subjects
- 20 experiment sets

Wearable Computing Lab. ETH Zürich

UMIT CSN

# Typical sequence 'ground truth'




Segmentation problem:
Difficult to model 'Null' activity

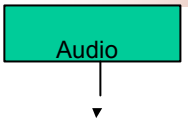
Wearable Computing Lab. ETH Zürich

UMIT CSN

- Motivation
- Overview of experiment
- **Sound & acceleration**
- Classifier combination
- Evaluation
- Conclusion



# Classification of Sound

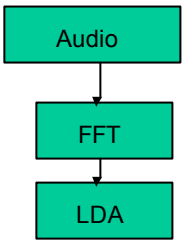



- Frame-by-frame
  - Moving window over continuous data (100ms)

Wearable Computing Lab. ETH Zürich
UMIT CSN

# Classification of Sound

- Frame-by-frame
  - Moving window over continuous data (100ms)
  - Spectrum pattern matching
  - Linear Discriminant Analysis



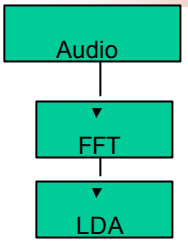

( 256 dim. FFT vector )

( 9 classes
=> 8 dim. LDA vector )



# Classification of Sound

- Frame-by-frame
  - Moving window over continuous data (100ms)
  - Spectrum pattern matching
  - Linear Discriminant Analysis
  - Distance from class means



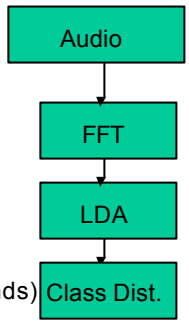

( 256 dim. FFT vector )

( 9 classes
=> 8 dim. LDA vector )

Wearable Computing Lab.
ETH Zürich
UMIT CSN

# Classification of Sound

- Frame-by-frame
  - Moving window over continuous data (100ms)
  - Spectrum pattern matching
  - Linear Discriminant Analysis
  - Distance from class means
- Larger moving window (2 seconds)
  - Mean of LDA distances
  - Classify using Min. Distance



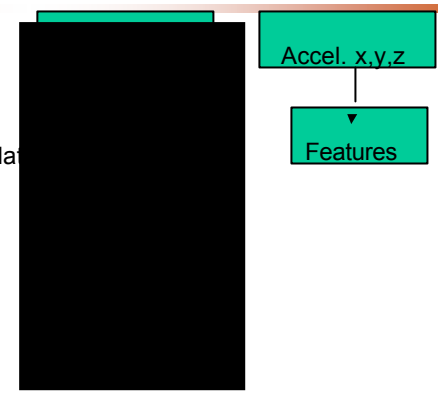

( 256 dim. FFT vector )

( 9 classes
=> 8 dim. LDA vector )



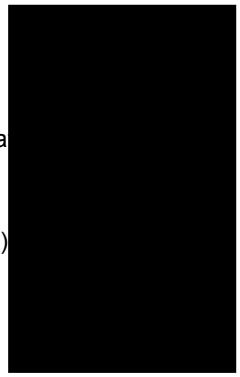
# Classification of Acceleration

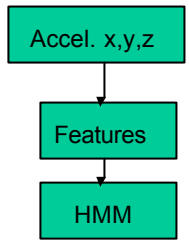
- Features
  - Peak count and mean
  - Mean and variance of raw da



# Classification of Acceleration

- Features
  - Peak count and mean
  - Mean and variance of raw da
- Moving window (2 seconds)
  - Hidden Markov Model (HMM)
  - Class likelihoods

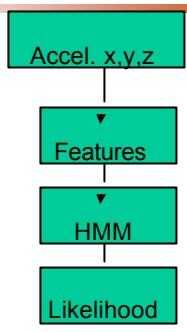




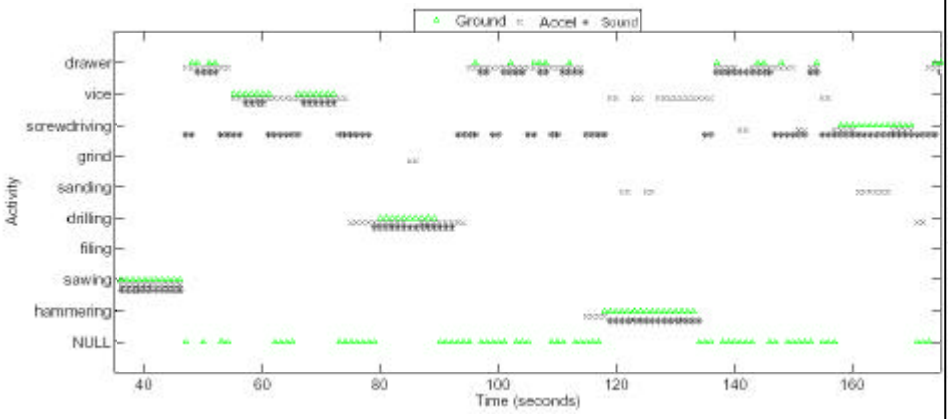



# Classification of Acceleration

- Features
  - Peak count and mean
  - Mean and variance of raw da
- Moving window (2 seconds)
  - Hidden Markov Model (HMM)
  - Class likelihoods
  - Classify using Max. Likelihoo



# Typical prediction output: sound and acceleration




Wearable Computing Lab. ETH Zürich

UMIT CSN

# Confusion matrix from all sound data

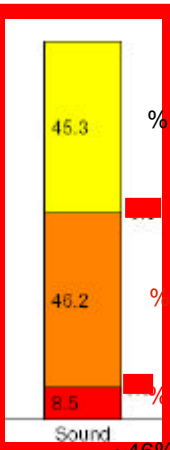
| Class | Total(s) | hammer | saw | file | drill | sand. | grind | screwdr. | vise | drawer | *NULL* | %Correct |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| hammer | 195.7 | **168.5** | | | | | | 19.6 | 6.7 | 0.8 | | 86.12 |
| saw | 306.4 | | **267.2** | 14.0 | | 2.0 | | 10.5 | 12.7 | | | 87.21 |
| file | 304.6 | | | **238.8** | | 34.4 | 1 | 16.1 | 2.5 | 2.8 | | 78.40 |
| drill | 241.5 | | | | **226.5** | 12.0 | 2.0 | 1.0 | | | | 93.79 |
| sand. | 313.0 | | 6.0 | 13.9 | | **258.0** | 2.0 | 21.7 | 0.9 | 10.5 | | 82.42 |
| grind. | 277.7 | | | | | 2.9 | **274.5** | 0.3 | | | | 98.86 |
| screwdr. | 260.4 | | | | | | | **249.0** | 9.8 | 1.6 | | 95.64 |
| vise | 678.1 | | 0.3 | | | | | 101.8 | **571.2** | 4.8 | | 84.24 |
| drawer | 658.8 | | | | | 0.7 | | 163.6 | 22.7 | **471.8** | | 71.61 |
| *NULL* | 2777.5 | 5.5 | 8.8 | 7.3 | 111.5 | 18.0 | 67.5 | 1360.8 | 506.4 | 691.7 | | 0 |

Zero true negatives

Wearable Computing Lab. ETH Zürich

UMIT CSN

# Errorbar summary from all sound data




% True positive time

% False positive time

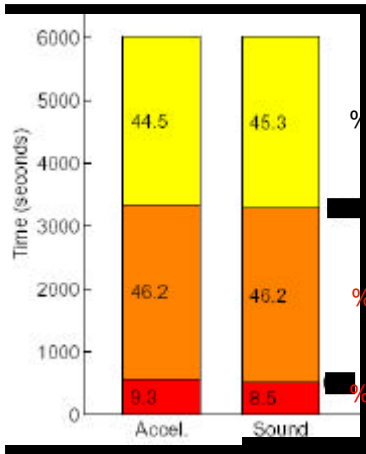
%

~46% ground truth total == Null class

Wearable Computing Lab. ETH Zürich

UMIT CSN

# Errorbar summary from Acceleration & Sound




%

% False positive time

% Substitution time

Wearable Computing Lab. ETH Zürich

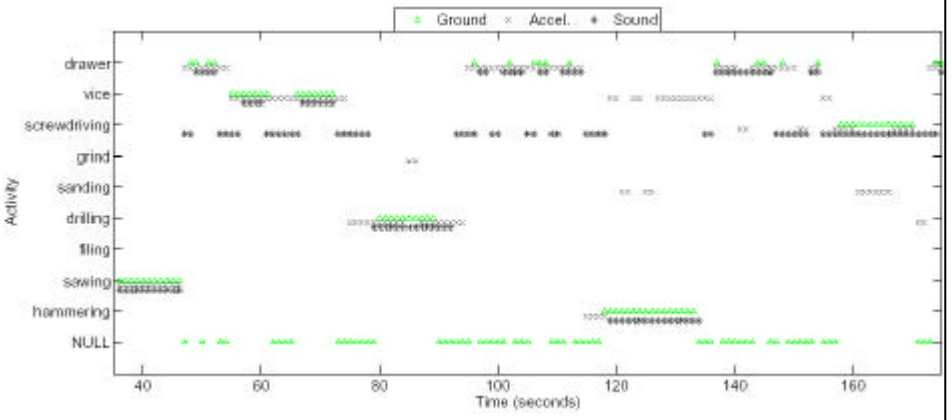
UMIT CSN

- Motivation
- Overview of experiment
- Sound & acceleration
- **Classifier combination**
- Evaluation
- Conclusion

Wearable Computing Lab. ETH Zürich

UMIT CSN

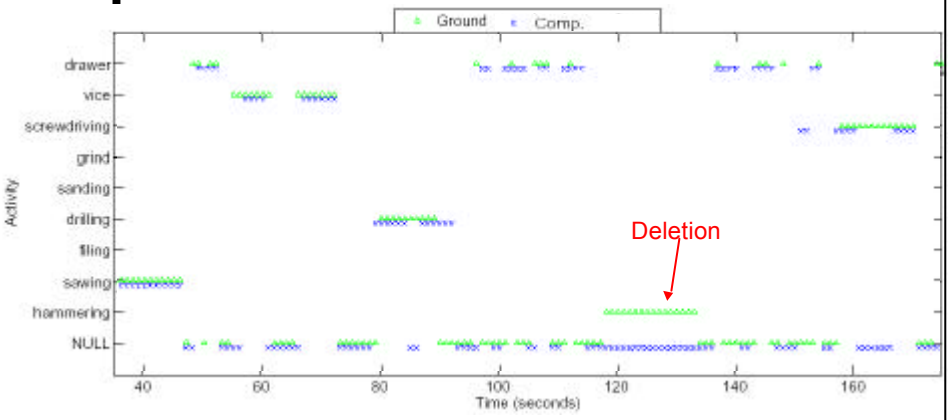
# Typical prediction output: sound and acceleration




Wearable Computing Lab. ETH Zürich

UMIT CSN

# Combination #1: Comparison of top choices



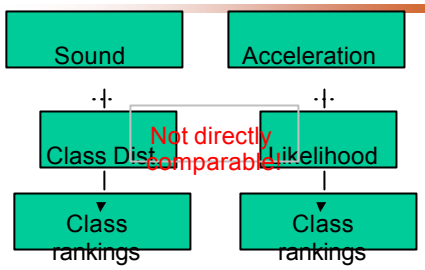

Wearable Computing Lab. ETH Zürich

UMIT CSN

# A 'softer' combination using class rankings

e.g. over a 2 second window..

| class | *hammer* | vise | file | drill | .. |
|---|---|---|---|---|---|
| **sound rank** | *1* | 3 | 2 | 4 | .. |
| **accel. rank** | 2 | *1* | 4 | *3* | .. |



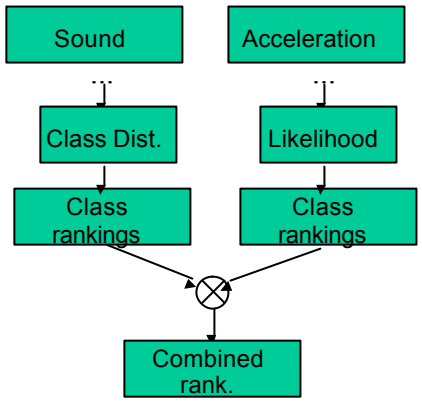

Wearable Computing Lab. ETH Zürich

UMIT CSN

# A 'softer' combination using class rankings

e.g. over a 2 second window..

| class | *hammer* | vise | file | drill | .. |
|---|---|---|---|---|---|
| **sound rank** | *1* | 3 | 2 | 4 | .. |
| **accel. rank** | 2 | *1* | 4 | *3* | .. |
| comb. rank | 1 | 2 | 3 | 4 | .. |



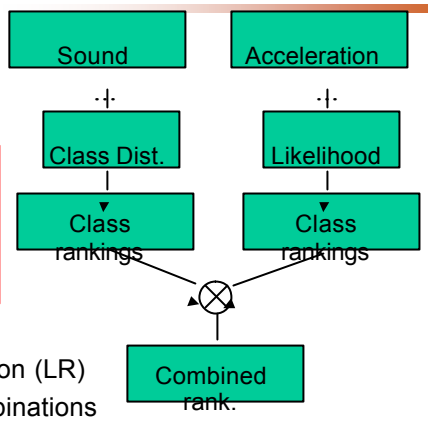

Wearable Computing Lab. ETH Zürich

UMIT CSN

# A 'softer' combination using class rankings

e.g. over a 2 second window..

| class | *hammer* | vise | file | drill | .. |
|---|---|---|---|---|---|
| **sound rank** | *1* | 3 | 2 | 4 | .. |
| **accel. rank** | 2 | *1* | 4 | *3* | .. |
| comb. rank | 1 | 2 | 3 | 4 | .. |



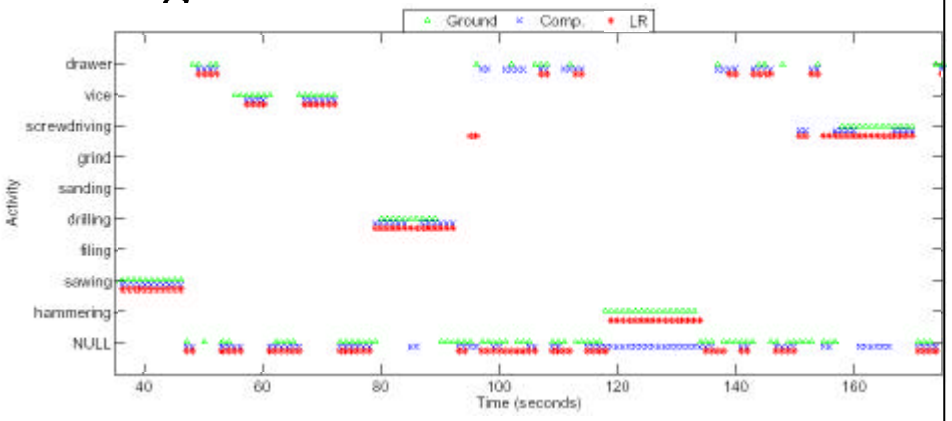

- Method based on Logistic regression (LR)
  - account for common rank combinations

"Decision combination in multiple classifier systems"
*- Ho et. al, TPAMI '94*

Wearable Computing Lab. ETH Zürich

UMIT CSN

# Combination #1: ranking fusion using LR

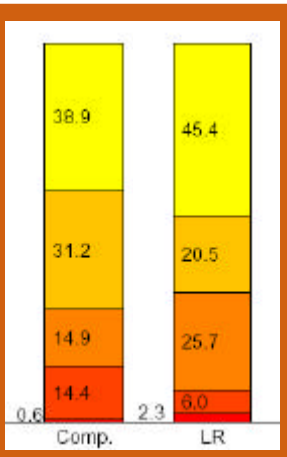





- Motivation
- Overview of experiment
- Sound & acceleration
- Classifier combination
- **Evaluation**
- Conclusion

Wearable Computing Lab.
ETH Zürich
UMIT CSN

# Results for Comparison and LR




% True positive time

% True 'Null' time

% False positive time

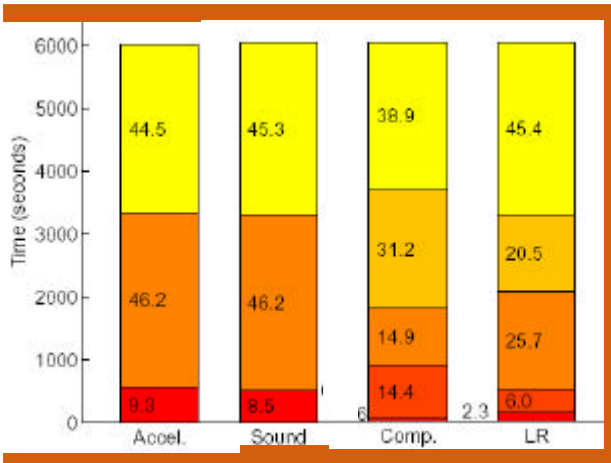
% False Null' time
% Substitution time

~46% ground truth total == Null class

Wearable Computing Lab.
ETH Zürich

UMIT ::: CSN

# Results for Comparison and LR




% True positive time

% True 'Null' time

% False positive time

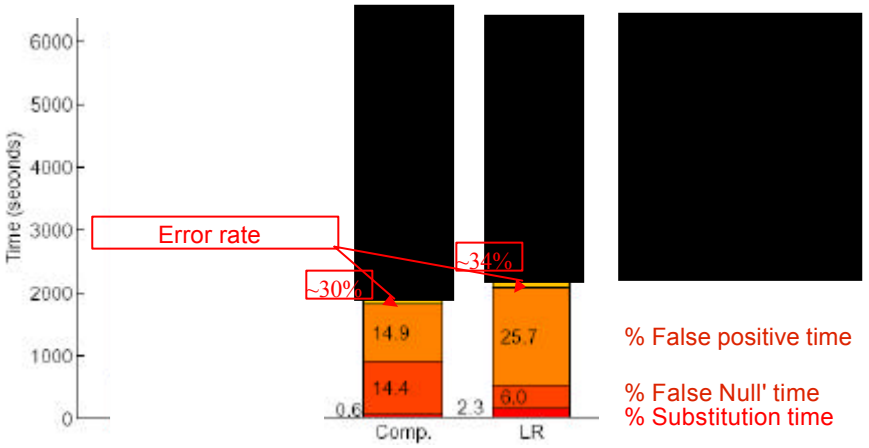
% False Null' time
% Substitution time

~46% ground truth total == Null class

Wearable Computing Lab.
ETH Zürich

UMIT ::: CSN

# Results for Comparison and LR

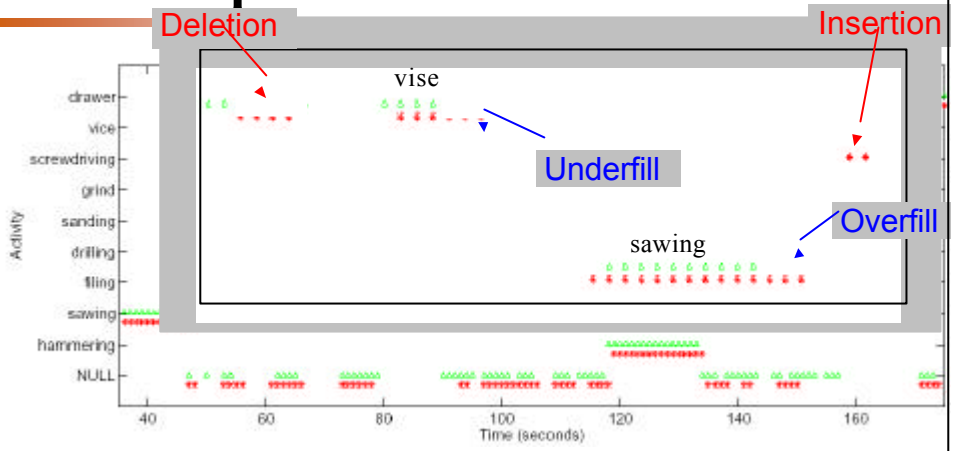



~46% ground truth total == Null class

Wearable Computing Lab. ETH Zürich

UMIT CSN

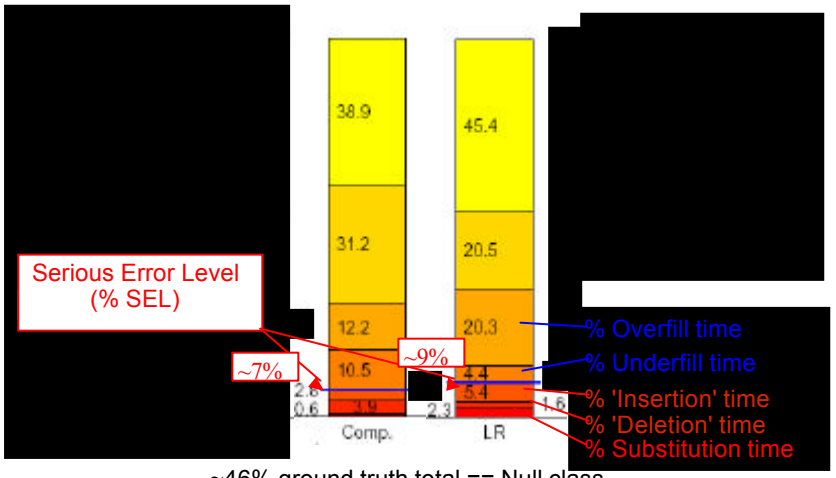
# Closeup of errors




Wearable Computing Lab. ETH Zürich

UMIT CSN

# Results for Comparison and LR




~46% ground truth total == Null class



- Motivation
- Overview of experiment
- Sound & acceleration
- Classifier combination
- Evaluation
- **Conclusion**

Wearable Computing Lab. ETH Zürich
UMIT CSN

# Conclusion

- Single wrist-worn microphone and accelerometer sufficient to recognise hand and tool activities
  - provided activity has corresponding motion and sound
  - combine sound and acceleration classifiers to detect *Null*
- *Time-based evaluation of continuous recognition*
  - confusion matrix inacurate for certain evaluation criteria
  - suggest use of 'minor error' categories: Overfill and Underfill

Wearable Computing Lab. ETH Zürich UMIT CSN

# Conclusion

- Questions?

## Gesture Spotting Using Wrist Worn Microphone and 3-Axis Accelerometer

*Jamie A Ward (ETH Zürich)*
*Paul Lukowicz (UMIT Austria)*
*Gerhard Tröster (ETH Zürich)*

Wearable Computing Lab. ETH Zürich UMIT CSN

Wearable Computing Lab.
ETH Zürich

UMIT ::: CSN