Adding Audio Capabilities to TIAGo Service Robot

Lacrimioara GRAMA , Corneliu RUSU

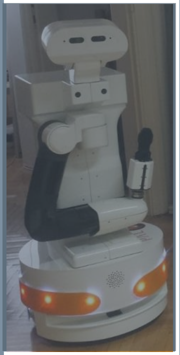
Technical University of Cluj-Napoca
Faculty of Electronics, Telecommunications and Information Technology
Basis of Electronics Department
Signal Processing Group

ISETC'18

This work was supported by two grants of the Romanian National Authority for Scientific Research and Innovation, CNCS/CCCDI-UEFISCDI: partially by project number PNIIIP2-2.1-BG-2016-0378, 54BG/2016, and partially by project number PNIII-P2-2.1-PED-2016-1608, 222PED/2017, within PNCDI III.

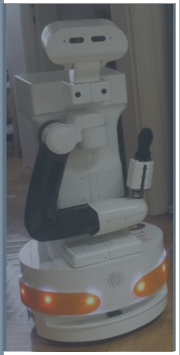
Outline

- Research Aim
- Who is TIAGo?
- Acquisition and Processing
- Database
- Features Extraction
- Classification
- Results
- Conclusion



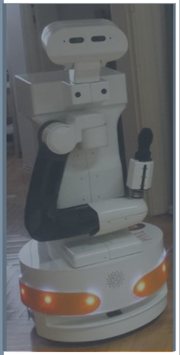
Research Aim

- To solve the problem of context awareness based on acoustic analysis for the service robot TIAGo, from PAL Robotics
- Service or social robots that share a space with people require the capacity to detect and track humans and recognize their activities
- Nowadays camera hardware and computer algorithms allow robots to process visual data from the world
  - Unfortunately, computer vision
    - Is not the cheapest solution
    - Needs a lot of information to be processed
    - Depends on the ambient lightning
    - Cannot detect information behind the scene $\Rightarrow$ doesn't detect what is happening after some obstacles
      - This might be crucial for some human activities



Research Aim

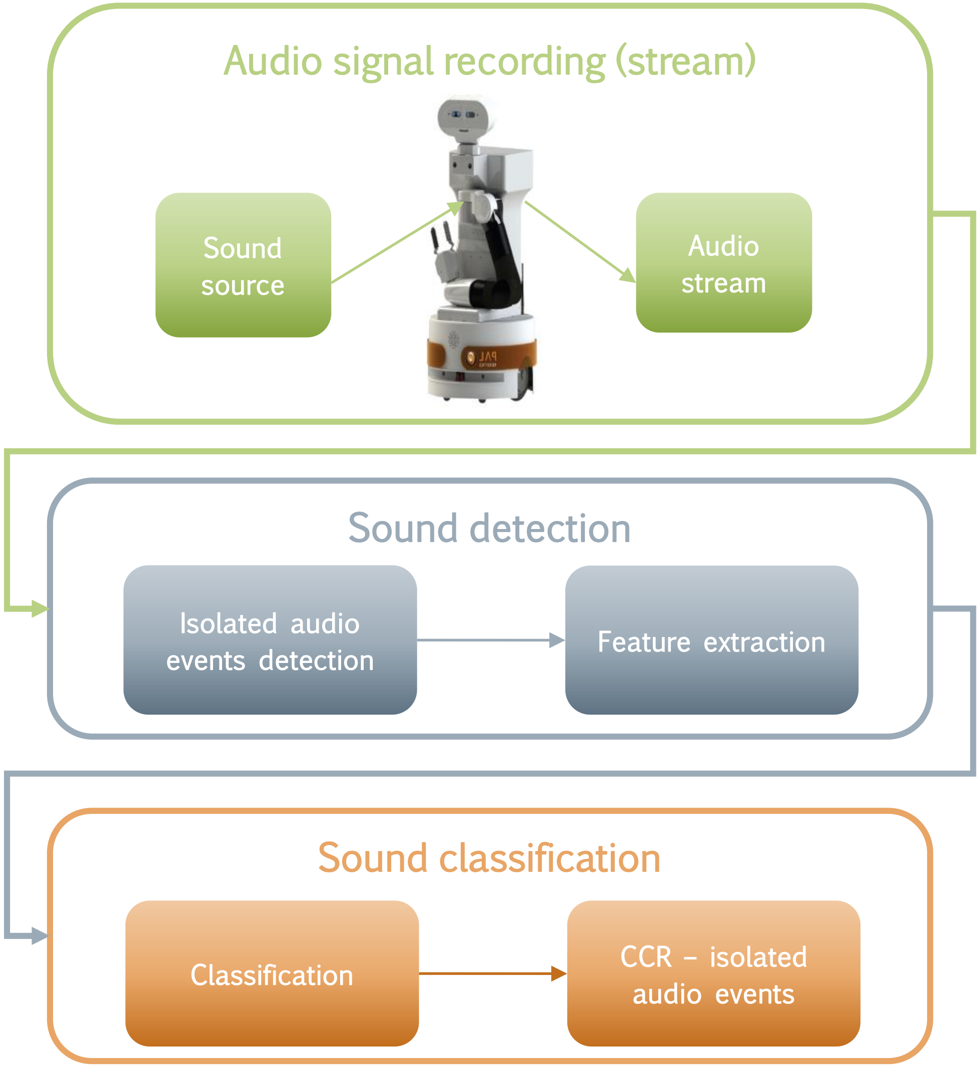
- Audio appears to be a well suited sensory modality for this task
  - Since many human activities produce very characteristic sounds from which a robot can infer the corresponding human actions
- Sound environment analysis is an important, complementary simpler mean, compared to vision
  - Can help the robot to achieve a good understanding of the context
- For example, during the interaction with the human user the robot will be aware of the activities the human is involved into, i.e. walking, sleeping and talking
- The robot should at all times be aware of the environment in the house



Who is TIAGo?

- TIAGo is a service robot designed to work in indoor environments
- People interested especially on ambient assisted living or in some easy tasks from industry can find TIAGo a suitable platform to develop their applications
- He combines mobility, perception, manipulation and human-robot interaction abilities
- Yet, researches on TIAGo have not been focused extensively on audio applications
- In this work we have presented how TIAGo's capabilities can be extended in such a way that he might handle tasks received through audio signals

Acquisition and Processing



- Main purpose: to recognize isolated audio events that can occur in the environment in a home
  - Short duration (0.55 ÷ 3.7 s), as a result of events such as
    - opening/closing of the microwave oven door
    - fall of an object on the floor
    - hand clap
    - opening/closing of a door with the key, etc.

Sounds Acquisition

- The audio signals were recorded with both the TIAGo robot and the simulated TIAGo robot
- The robot was simulated with a 2-microphone array connected to a laptop, identical to the robot one

  - There was no difference between the two types of signals
  - The USB-SA Array Microphone offers a USB digital sound card with a stereo array microphone for Andrea AudioCommander's enhancement software
  - The USB-SA external sound card eliminates computer noise floor problems

Sounds Acquisition

- The audio data stream was recorded to contain only a certain class of sound events
  - It was subsequently split into signals containing only one acoustic event
  - Except to time-domain isolation of the sound event, no post-processing such as noise elimination or normalization has been performed
- The recorded sound events were chosen to belong to the home environment

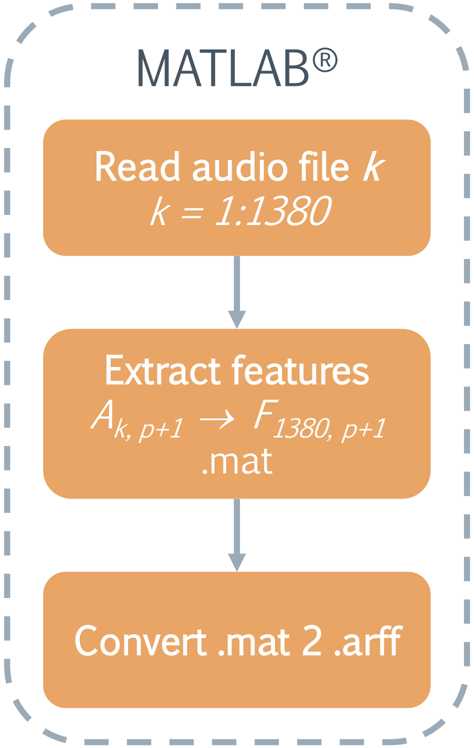
TIAGo Database

Kitchen (8)	Room (8)	Appliances (5)	Non-verbal (5)	Voice (20)
•chair	•page turn	•washing machine	•hand clap	•1
•tap water	•Velcro	•microwave open	•finger clap	•2
•drop water	•zip open	•microwave close	•cough	•3
•shower water	•zip close	•microwave alarm	•laugh	•4
•porcelain dish	•door knock	•toaster alarm	•whistle	•5
•cutlery	•door key			•6
•plastic bag rush	•door open			•7
•cardboard drop	•door close			•8
				•9
				•10
				•salut (hello)
				•medicamente (medicines)
				•da (yes)
				•nu (no)
				•dreapta (right)
				•stanga (left)
				•stai (stay)
				•vino (come)
				•du-te (go)
				•TIAGo

- 1380 audio events
  - 46 sound classes
  - 5 scenarios
  - 30 sounds/class
  - 48 kHz, 16-bit accuracy

Features Extraction

- Frames: 25 ms + 60% overlap
- Features vector
  - LPC: $\mathbf{A}_k = [\sigma_k^2 \quad a_{k,1} \quad a_{k,2} \quad \dots \quad a_{k,p}]$
    - σ_k^2 – prediction error variance
    - $a_{k,i}$ – last p LPC coefficients
  - LPCC: $\mathbf{A}_k = [\ln\sigma_k^2 \quad c_{k,1} \quad c_{k,2} \quad \dots \quad c_{k,p}]$
    - $c_{k,i}$ – LPCC coefficients computed based on σ_k^2
  - MFCC: $\mathbf{A}_k = [E_k \quad C_{k,1} \quad C_{k,2} \quad \dots \quad C_{k,p}]$
    - E_k – signal energy
    - $C_{k,i}$ – MFCC coefficients



- Features matrices $\mathbf{F}_{1380\times(p+1)} = \begin{bmatrix} \mathbf{A}_{1,p} \\ \mathbf{A}_{2,p} \\ \vdots \\ \mathbf{A}_{1380,p} \end{bmatrix}$
    - p = 10:2:38 – orders for the prediction filter (LPC, LPCC)/ number of cepstral coefficients (MFCC)

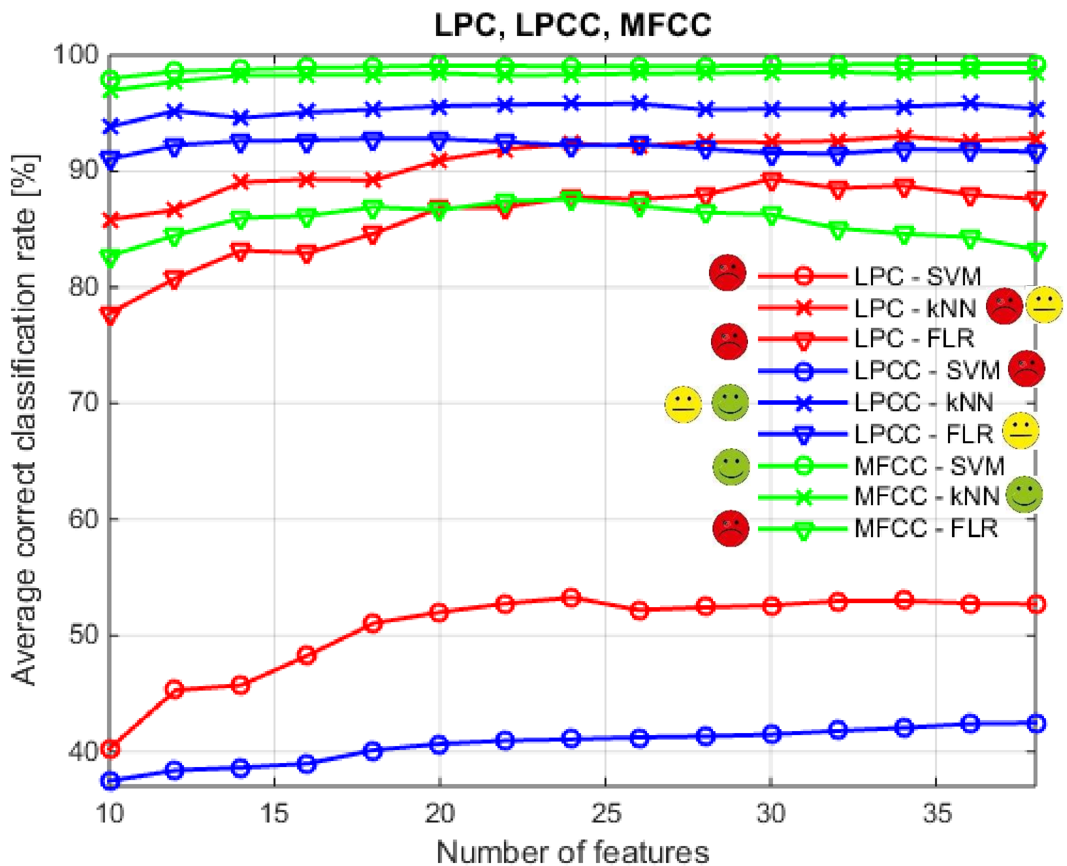
Classification

WEKA
The University of Waikato

- WEKA – 8 classifiers
  - Open source software issued under the GNU General Public License
  - A collection of machine learning algorithms for data mining tasks
  - Tools for data pre-processing, classification, regression, clustering, association rules, and even visualization
    1) Bayesian Network (BN)
    2) Quadratic Discriminant Analysis (QDA)
    3) Support Vector Machines (SVM)
    4) Multilayer Perceptron (MLP)
    5) k-Nearest Neighbor (kNN)
    6) KStar
    7) Fuzzy Lattice Resoning (FLR)
    8) Random Forests (RF)

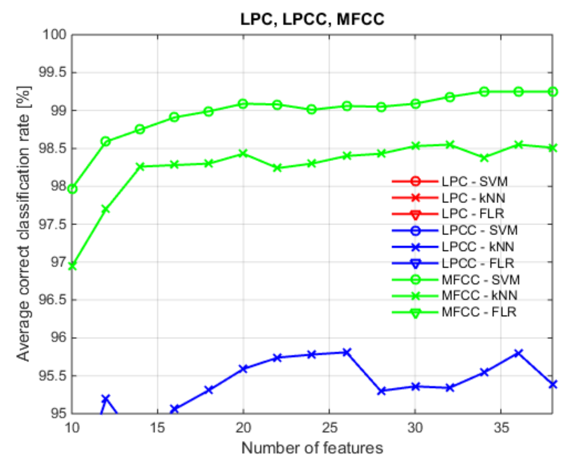
Results

- The results obtained represent the average of the 10 runs (100 experiments/classifier) for each classifier

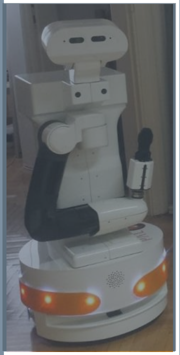


Results

- For audio signal detection problems we need a high CCR and a low Std.Dev.
- Only MFCC fulfil the requirements
  - The highest CCR and lowest Std.Dev.
    - SVM and kNN classifiers
    - CCR is greater than 96.5%

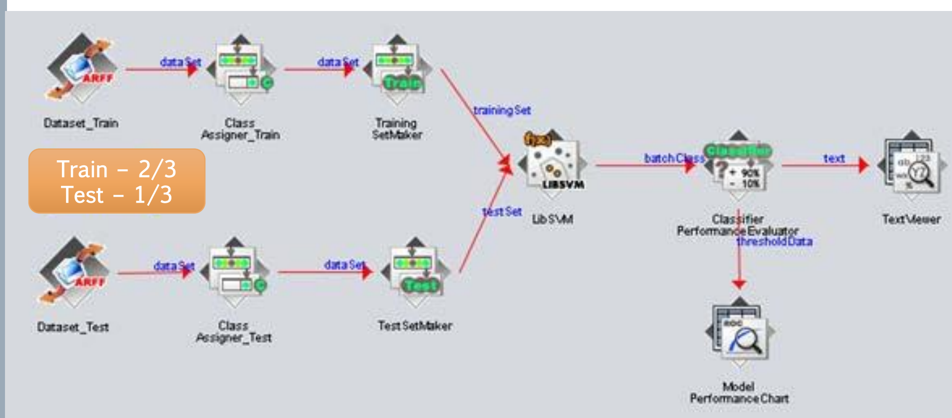


- In order to obtain a CCR greater than 99% we need at least 21 characteristics for SVM
- To improve the accuracy of the classification, a grid search algorithm was applied to SVM, and in the case of kNN the optimal value for *k* was searched



Real Life Scenario

› We split the data from each class into
  - Training data (2/3)
  - Testing data (1/3)
  - SVM – CCR[%] (Std.Dev.)



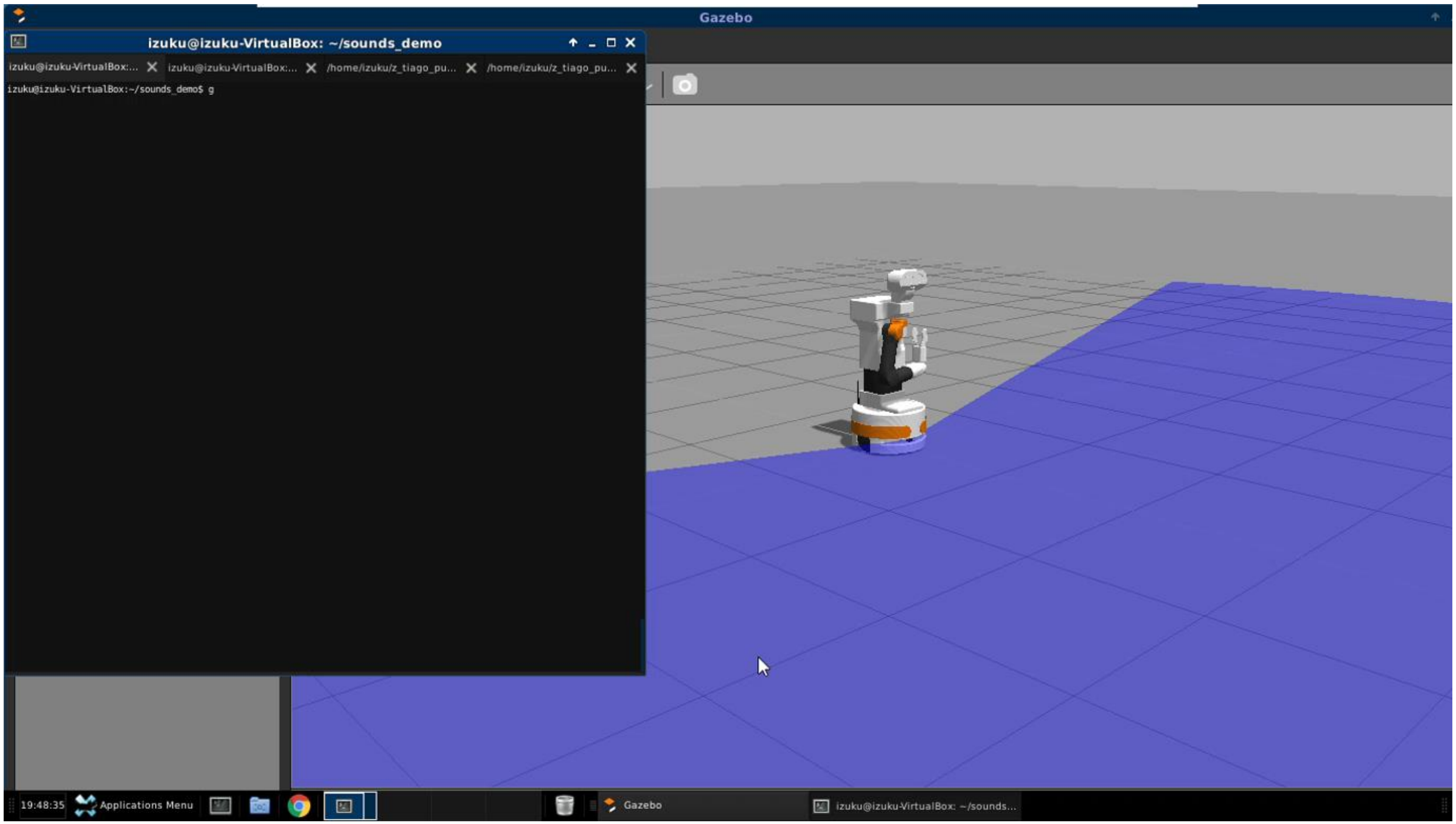
Features	Training	Testing
MFCC-20	100 (0.08)	99.35 (0.06)
MFCC-22	100 (0.09)	99.13 (0.05)
MFCC-24	100 (0.09)	99.35 (0.06)
MFCC-26	100 (0.08)	99.35 (0.06)
MFCC-28	100 (0.09)	99.35 (0.06)
MFCC-30	100 (0.11)	99.35 (0.06)
MFCC-32	100 (0.11)	99.35 (0.05)
MFCC-34	100 (0.11)	99.35 (0.06)
MFCC-36	100 (0.11)	99.35 (0.06)
MFCC-38	100 (0.11)	99.35 (0.05)

Conclusion

- We have shown how the capabilities of the TIAGo robot can be extended in such a way that it might handle tasks received through audio signals
- The audio data stream was recorded to contain only a certain class of sound events and a total of 30 audio signals were obtained for each 46 classes
- For each audio signal, three types of features have been extracted: LPC, LPCC, and MFCC
- From the 8 tested classifiers, the best results were obtained for SVM, kNN and FLR
- On the features extraction side, it is best to use the MFCC coefficients
- As a classifier, either SVM or kNN can be used in terms of credibility, specialized literature and computation time

Demo

› Gazebo simulator



Adding Audio Capabilities to TIAGo Service Robot

Lacrimioara GRAMA , Corneliu RUSU

Technical University of Cluj-Napoca
Faculty of Electronics, Telecommunications and Information Technology
Basis of Electronics Department
Signal Processing Group

ISETC'18

This work was supported by two grants of the Romanian National Authority for Scientific Research and Innovation, CNCS/CCCDI-UEFISCDI: partially by project number PNIIIP2-2.1-BG-2016-0378, 54BG/2016, and partially by project number PNIII-P2-2.1-PED-2016-1608, 222PED/2017, within PNCDI III.