

Audio Database for TIAGo Service Robot

Lacrimioara Grama
Signal Processing Group, Basis of Electronics Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Lacrimioara.Grama@bel.utcluj.ro

Lorena Muscar
Basis of Electronics Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Lorena.Muscar@bel.utcluj.ro

Corneliu Rusu
Signal Processing Group, Basis of Electronics Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Corneliu.Rusu@bel.utcluj.ro

Toma Telembici
Basis of Electronics Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
tomaatelembici@yahoo.com

Abstract—In this paper we present an updated version of the audio database acquired by the TIAGo service robot and by the simulated TIAGo service robot. To the initial database which consists in 1380 audio signals we have added 1920 more acoustic signals. The audio database consists now in 3300 isolated audio events corresponding to 110 classes. All the recorded sound events correspond to the indoor environment, and they are spread into five different scenarios: kitchen, room, appliances, voice and non-verbal. The audio database is intended to be used in order to identify indoor events based on audio signature, especially when elderly or chronically ill people live alone.

Keywords—TIAGo service robot, audio database, indoor environment.

I. INTRODUCTION

Audio analysis, compared to visual analysis, has the advantage of lower amount of data needed to be analyzed, have no limitation in dark or to bright lighting conditions, and is also cheaper. In addition, it provides information on the unfolding of events that are not necessarily in front of the camera lens, it may be behind the camera lens or behind certain obstacles.

The robots are specially programmed and able to automatically perform a complex series of actions. They can be built on the line of human form, the sensors being key components of robots.

TIAGo from Pal Robotics is a robot whose purpose is to work in indoor environments [1]. This service robot combines mobility and manipulation with human-robot interaction capabilities. It can benefit not only using visual information, but also using audio information. Previous results have shown that the TIAGo robot can be used to analyze audio signals for classification purposes and that it can distinguish not only between classes of audio signals, but also between different audio signals in the same class.

The purpose of this paper is to present an updated version of the TIAGo audio database, which can be used for indoor events identification based on audio signature. The rest of the paper is organized as follows: in Section II we describe previous results in processing and analysis of audio signals by TIAGo robot and motivation for updating the audio

database, in Section III we present how the audio signals from the database were recorded, in Section IV we describe our audio database, while Section V is dedicated to conclusions.

II. PREVIOUS RESULTS AND MOTIVATION FOR UPDATING THE AUDIO DATABASE

We have focused our attention on the audio capabilities of the TIAGo and we have obtained good outcomes in [2, 3, 4, 5, 6]. In [2] three types of features have been used (Linear Predictive Coding - LPC, Linear Predictive Cepstral Coefficients - LPCC and Mel-Frequency Cepstral coefficients - MFCC) and eight classifiers were tested (Bayes Network - BN, Quadratic Discriminant Analysis - QDA, Support Vector Machines – SVM, Multilayer Perceptron - MP, k-Nearest Neighbor – kNN, kStar, Fuzzy Lattice Reasoning - FLR, and Random Forests - RF). Overall correct classification rates greater than 99% were attained using MFCC (from 20 to 38) features and SVM as a classifier.

In [3] for 37 MFCC features and 3-kNN, City and Simple Value Difference metric, Inverse Distance voting, Accuracy Based weighting method, we have obtained an overall classification accuracy of 98.25%. By applying resampling before classification the accuracy was increased to 99.21%.

Using the log of the total frame energy and 33 liftering MFCC features together with the 1-kNN, in [4], we have obtained a testing accuracy of 98.55%.

In [5] using 20 MFCC features and 5-kNN we have attained an overall classification rate of 99.27%. In [6] using Moving Picture Experts Group-7 (MPEG7) features together with the Hidden Markov Models (HMM) we have obtained an overall classification accuracy of 96.86%.

Based on the results reported in our previously works, we have concluded that the TIAGo can be used as a service robot for assisting elderly or chronically ill people. That is why the aim of this paper is to improve the number of audio signals in our initial database from [2]. We have focused on improving the number of acoustic signals in the voice part of the database with signals corresponding to medicine names and also with frequently used words in the home environment. The actual database is intended to be used for

the Romanian elderly or chronically ill, this is why the voice scenario comprises only signals in the Romanian language.

Another reason why we have decided to extend audio signal database for the TIAGo service robot is because, until now, most of the researches are done only in the field of the mobility of the robot inside home [7, 8], not on the events recognition based on audio signals.

The initial version of the audio database [2] consists in 1380 audio signals. We have added 1920 more audio signals, in order to improve the database. The audio database consists now in 3300 isolated audio events corresponding to 110 classes. All the recorded sound events correspond to the indoor environment, and they are spread into five different scenarios: kitchen, room, appliances, voice and non-verbal. All the audio signals were recorded with 48 kHz and 16 bit-accuracy.

III. DATABASE RECORDING

The audio signals in the database were recorded with both TIAGo service robot and the simulated service robot as described in [2]. The acquisition block diagram used for the audio recordings using TIAGo is illustrated in Fig. 1.



Fig. 1. Audio signals recording by TIAGo.

As depicted in Fig. 1, just below the head of TIAGo there are stereo microphones that can be used to record audio and process it and a speaker. The stereo microphones are the SuperBeam Stereo Array Microphone from Andrea Electronics [1]; the specifications of the stereo microphone are detailed in Table I, while the specifications for the USB external audio card (PureAudio USB-SA) are presented in Table II.

TABLE I. TIAGO STEREO MICROPHONE ELECTRICAL AND ACOUSTIC SPECIFICATIONS [1]

Mic supply voltage	1.4 – 5.0 VDC
Supply bias resistor	2.2 kΩ – 39.9 kΩ
Operating current (each channel)	0.5 mA
Output impedance at 1 kHz	200 Ω
Max input sound level at 1 kHz, 3% THD	115 dB
Output signal level at THD < 3% @ 1 kHz	24 – 120 mVrms
Sensitivity at 1 kHz (0 dB = 1 V/Pa Vdc = 1.6 V)	-40 – -37 dBV
Frequency response at 3 dB variation noise	20 uVrms
Operating temperature	0 – 70 °C

Recommended operating distance	30.5 – 122 cm
Acoustic signal reduction at 1kHz outside of 30° beamform	15 – 30 dB
Noise reduction	20 – 25 dB

TABLE II. PUREAUDIO USB-SA SPECIFICATIONS [9]

Operation	
Power	USB
Microphone input	2 channel stereo
Audio output	2 channel stereo
Sampling rate	8, 11.025, 16, 22.05, 32, 44.1, 48 kHz
Power supply	
Supply voltage	4.5 – 5.5 VDC
Total power consumption	120 mA
Microphone input	
A/D conversion resolution	16 bit
THD + N	-84 dB
Supply bias resistor	2.2 kΩ @ 3.3 VDC
Frequency response	20 Hz – 20 kHz
Input range	0 – 1.25 Vrms
Dynamic range	95 dB
Record gain range	-6 – 33 dB

In order to record audio signals also with the simulated TIAGo service robot, the arrangement from Fig. 2 was used.



Fig. 2. Audio signals recording by simulated TIAGo.

As depicted in Fig. 2 instead of the service robot the same microphone array from Andrea Electronics were used together with the external digital sound card [9]. The specifications are exactly the ones illustrated in Tables I and II.

No difference could be observed between the signals captured as depicted in Fig. 1 or in Fig. 2. The recorded sounds were chosen to belong to the same scenarios as in [2]. Each data stream was recorded with 48 kHz, 16 bit-accuracy

(such that to contain only one class of acoustic events), in order to be in accordance to the previous audio signals in the database. After the data stream was recorded it was split into audio segments containing only one audio event. No other type of post-processing was realized.

For the voice scenario, in the initial database [2] the words are pronounced by a female, while in the extension of the database presented in this paper the words are pronounced by a male. We wanted to have voice signals from female and male also, because elderly or chronically ill persons from which the words are to be recognized to can have any gender.

IV. AUDIO DATABASE FOR TIAGO SERVICE ROBOT - DESCRIPTION

The initial TIAGo audio database, as presented in [2], is available for free¹. It consists in 1380 audio signals (30 signals/ class, 48 classes) as follows:

- Kitchen scenario – 8 classes ($8 \times 30 = 240$ audio signals)
- Room scenario – 8 classes ($8 \times 30 = 240$ audio signals)
- Appliances scenario – 5 classes ($5 \times 30 = 150$ audio signals)
- Voice scenario – 20 classes ($20 \times 30 = 600$ audio signals)
- Non-verbal scenario – 5 classes ($5 \times 30 = 150$ audio signals)

The total duration of each class from the initial database is illustrated in Table III.

TABLE III. INITIAL DATABASE - AUDIO EVENTS DURATION

Scenario	Class	Total duration [s]
Kitchen	chair	36.959
	tap water	61.818
	drop water	9.504
	shower water	64.520
	porcelain dish	20.846
	cutlery	34.833
	platic bag rush	50.832
	cardboard drop	24.203
Room	page turn	36.171
	velcro	12.461
	zip open	13.653
	zip close	13.088
	door knock	21.841
	door key	13.995
	door open	16.870

¹ TIAGo Audio Database ver.01,
https://sp.utcluj.ro/Sound_Database/TIAGO_Database/TIAGO_ver01.zip

Scenario	Class	Total duration [s]
Appliances	door close	31.075
	washing machine	128.551
	microwave open	19.091
	microwave close	31.824
	microwave alarm	30.757
	toaster alarm	16.659
Voice	unu (one)	21.942
	doi (two)	19.289
	trei (three)	20.302
	patru (four)	24.782
	cinci (five)	24.752
	sase (six)	28.963
	sapte (seven)	28.408
	opt (eight)	17.485
	noua (nine)	26.830
	zece (ten)	27.520
	salut (hello)	24.794
	medicamente (medicines)	40.256
	da (yes)	16.343
	nu (no)	16.639
	dreapta (right)	30.271
	stanga (left)	29.566
	stai (stay)	27.320
	vino (come)	25.752
	du-te (go)	24.251
	TIAGo	26.033
Non-verbal	hand clap	14.409
	finger clap	9.938
	cough	25.587
	laugh	35.233
	whistle	23.105
Total duration [s]		1299.32

The database was updated by another 1920 audio signals (30 signals/class, 64 classes) as follows:

- Kitchen scenario – 1 class ($1 \times 30 = 30$ audio signals)
 - o boiling water
- Room scenario – 3 classes ($3 \times 30 = 90$ audio signals)
 - o door treadmill, power switch, wheelchair
- Voice scenario – all signals are in the Romanian language (for some of them the English correspondence is also given)

- Usual medication – 6 classes: ($6 \times 30 = 180$ audio signals)
 - AlgoCalmin, Calciu, Magneziu, Nurofen, Paracetamol, Vitamina C
- Medication for chronically ill – 32 classes ($32 \times 30 = 960$ audio signals)
 - Aspenter, Atacant, Aspacardin, Amiodaronă, Betaloc, Carvedilol, Propranolol, Betaserc, Indapamid, Nitroglicerină, Captopril, Enalapril, Prestarium, Diaprel, Meguan, Insulină, Euthyrox, Prostamol, Furosemid, Diurex, Spironolactonă, Sintrom, Trombostop, Eliquis, Amlodipina, Cordarone, Keppra, Carbamazepină, Depakine, Pentoxifilin, Ventolin, Miofilin
- Useful things – 21 classes ($21 \times 30 = 630$ audio signals)
 - mască (mask), dezinfecțiant (disinfectant), șerbetel (napkin), batistă (handkerchief), buletin (ID card), lanterna (flashlight), proteză (prosthesis), apă (water), sus (up), jos (down), căciulă (hat), mănuși (gloves), pantaloni (trousers), bluză (blouse), cămașă (shirt), tricou (T-shirt), plăpumă (quilt), pătură (blanket), papuci (shoes), pernă (pillow), șosete (socks)
- Non-verbal scenario – 1 class: snoring ($1 \times 30 = 30$ audio signals)

In Table IV is illustrated the duration of each individual sound which was added to initial database. The total duration of the new added audio signals is 2002.48 seconds.

V. CONCLUSION

In this paper we have presented an updated version for the TIAGo audio database. We have extended the initial database with 1920 audio signals, thus the actual version consists in 3300 acoustic signals, distributed among five scenarios that can be involved in an indoor environment.

The presented database is intended to be used for the help of Romanian elderly or chronically ill, this is why the voice scenario comprises only signals in the Romanian language. However, a part of audio signals are similar for English or other languages, thus our audio database may help in other cases.

Possible developments of the sound database may consider extending the voice scenario with multiple speakers and different speech speeds for the same or different words, in addition to updating some terms in other foreign languages.

ACKNOWLEDGMENT

This work was supported by a grant of the Romanian Ministry of Research and Innovation, CCCDI – UEFISCDI, project number 52/2020, PN-III-P2-2.1-PTE-2019-0867, within PNCDI III.

REFERENCES

- [1] Pal Robotics, TIAGo Handbook, Barcelona: Spain, 2016.
- [2] L. Gramă, C. Rusu, "Adding Audio Capabilities to TIAGo Service Robot," in Proceedings of the 13th International Symposium on Electronics and Telecommunications, Nov. 8-9, 2018, Timisoara, Romania, pp. 263-266, DOI: 10.1109/ISETC.2018.8583897.
- [3] T. Telembici, L. Gramă, "Detecting Indoor Sound Events," Acta Technica Napocensis - Electronics and Telecommunications, vol. 59, no. 2, pp. 13-17, 2018, ISSN: 1221-6542.
- [4] L. Gramă, C. Rusu, "Extending Assisted Audio Capabilities of TIAGo Service Robot," in Proceedings of the International Conference on Speech Technology and Human Computer Dialogue, Oct. 10-12, 2019, Timisoara, Romania, pp. 1-8, DOI: 10.1109/SPED.2019.8906635.
- [5] T. Telembici, L. Gramă, C. Rusu, "Integrating Service Robots into Everyday Life Based on Audio Capabilities," in Proceedings of 14th International Symposium on Electronics and Telecommunications, Nov. 5-6, 2020, Timisoara, Romania, pp. 1-4, DOI: 10.1109/ISETC50328.2020.9301100.
- [6] L.Muscar, L. Gramă, "Audio Events Detection to Help TIAGo to Act as a Medical Robot," in Proceedings of 7th International Conference on Advancements of Medicine and Health Care through Technology Oct.13-15, 2020, Cluj-Napoca, Romania.
- [7] F. Lu, M. Huang, X. Li, G. Tian, H. Wu, W. Si, "Learning and Development of Home Service Robots' Service Cognition Based on a Learning Mechanism," Applied Sciences, vol 10, issue 2: 464, 2020.
- [8] Z. Wang, G. Tian, X. Shao, "Home service robot task planning using semantic knowledge and probabilistic inference," Knowledge-Based Systems, vol. 204, 106174, 2020.
- [9] Andrea Electronics, PureAudio USB Array Microphone. Datasheet available: <https://andreaelectronics.com/array-microphone/>

