Implementation of Audio Navigation for Smart Campus

O. Petrova [1[0000-0002-6499-6017]], G. Tabunshchyk [2[0000-0003-1429-5180]], P. Arras [3[0000-0002-9625-9054]]

[1,2]1Zaporizhzhia National Technical University, Zhukovsky str., 64,Zaporizhzhia, 69063, Ukraine
petrovaoa353@gmail.com, galina.tabunshchik@gmail.com
[3]KU Leuven, Jan De Nayerlaan 5, 2860 Sint Katelijne Waver, Belgium
peter.arras@kuleuven.be

Abstract. The article deals with the task of in-door navigation of visually impaired people. The authors have carried out an analysis of audio navigation software programs such as Google Assistant, Siri and Cortana. Authors suggest a model of the voice navigator, which helps a person to conveniently find the location and build the desired route. Developed software is integrated into the smart-campus solution, which improve the infrastructure of the university.

Keywords: audio navigation, SMART-CAMPUS, BLE, voice navigator, in-door-positioning

1 Introduction

According to statistics nowadays in the European Union, people with disabilities make out about 1/6 of all citizens of working age. In Ukraine, the amount of persons with disabilities is 6.1% of the total population [1]. A person with disabilities faces many problems that are unknown to other people. Mainly this is caused by the restriction of the access of persons with disabilities to social benefits known to the majority of the population, such as shops, pharmacies, underground stations, stations, hairdressers, educational establishments, et cetera [2, 3]. This is due to the fact that in such places there are no special devices for assistance to people with disabilities.
Ukraine is trying to adapt public buildings to this reality by constructing ramps and buttons for the disabled people. At the legislative level, the Government makes changes to the laws that regulate the rights of persons with disabilities, namely, the Laws "On the Basis of Social Protection of Persons with Disabilities in Ukraine" [4], "On Amendments to Some Laws of Ukraine on Increasing Access to the Blind, persons with visual impairments and persons with dyslexia to works published in a special format "[5]," On Amending Certain Legislative Acts of Ukraine on the Protection of the Rights of Persons with Disabilities "[6]," On Amendments to Certain Laws of Ukraine on Education on the organization of inclusive n the voice "[7]. However, there should pass a long period before our country achieves the result that we can see in Europe today. Therefore, the development of audio navigation systems to improve social adaptation of people with visual disabilities is a very important task.

The idea of smart campus based on BLE 4.0 where objects could talk to the students, staff and visitors were described in an number of publications [8,9]. The use of voice for navigation systems could allow visually impaired people: to connect many objects and events; to provide access to the information in the navigation systems; to support new systems of interaction with users, sensors, mobile devices, devices and applications [10].

2 Problem definition

For correctly detection of the location inside the building, it is necessary to determine the current coordinates, compare the position with the cartographic representation, update the location in real time, and check the compliance of the current position with respect to the planned route [11].

Further we will consider a system that uses data from beacons based on BLE 4.0 to identify the current location [112]:

$$S=<X,B,R,Z,K> \tag{1}$$

where X - input data (x_1 - data from sensors x_2 - accelerometer readings, x_3 - gyro readings, x_4 data from beacons, x_5 - voice commands), B - cartographic representation (map represented as matrix [M, N], where M - X, N is the number of points along the Y-axis), R is information about the decisions taken (r_1, r_2 ... r_n), Z-output devices (z_1 - camera, z_2 - audio recording, z_3-phone), K - robot mode (k_1 - autonomous, k_2 controlled).

Positioning methods were developed for this class of systems [13]. However, the task of integrating audio navigation in Smart-Campus systems has not been solved yet. Solving this problem will allow the existing system to be adapted for people with disabilities.

The aim of the work was to develop a voice navigator and integrate it into the indoor positioning and navigation system.

3 An analysis of existing approaches to the implementation of voice navigation

Voice Navigator should help a person to navigate in the building using only voice. However, for the correct information exchange between the user and the application, there should be developed a module which could recognize speech signals.

Problems of voice navigation and speech recognition were investigated by D. Shpakov [14], E. A. Vereshchagina [15], Jen-Tzung Chien [16], Shinji Watanabe [17], Mohamed Afify [18], Chia-Yu [19], Mark D.Skowronski [20].

Automated speech recognition systems can be classified according to many features: by type of language, by a set of dictators, by volume and completeness of the vocabulary that needs to be recognized. By type, the language is divided into discrete and continuous [21]. Discrete language is a language in which pauses between words are much longer than natural pauses inside words. In continuous speech, there are no significant pauses between words. The natural human mode of communication is a continuous language.

Each person has a unique voice, but from a phonetic point of view language consists of many different sounds that have articulation differences. In general, these sounds are called phonemes. But in different words one and the same phonemes may be exaggerated, so there is the notion of alophon - phonemes [22].

For successful speech recognition, areas of the audio signal are considered in a few tens of milliseconds, which are called freemas [22]. The difficulty is that some phonemes are quite similar to one another, but one can solve this problem in terms of "probabilities". Some phonemes are more likely for a given signal, others - less. A acoustic-on-model is being built, which is a function that receives an area of a small audio signal (frame) at the input and outputs the distribution of the probabilities of different phonemes on this frame. On the basis of the acoustic model, one can say with a certain degree of confidence that it was said.

The acoustic model can be built on the basis of such methods and algorithms as neural networks, a model of Gaussian mixtures, dynamic programming [23-26]. In practice, hidden mark models are widely used in practice [27].

.

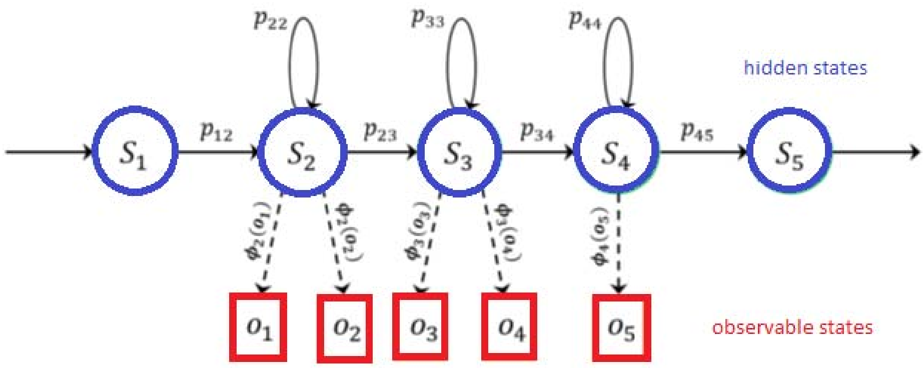

Fig. 1. - Acoustic model [13]

In the system discussed previously, incoming data can come in the form of voice messages. In this case, the task of recognizing audio events will look like this: an audio signal arrives at the audio event detector input, represented by the sequence:

$$\Omega = \{o1, o2, \ldots oM\} \tag{2}$$

where: oi - the value of the sound signal parameter (one of M) taken by the detector at the i^{th} moment of time. The segments of time in which the detector takes off these parameters are states $S = \{s1, s2, ..., sN\}$ of the model λ = (P, Φ, π). Each of these models corresponds to different types of audio events, such as certain words. In order for the system to be able to select the audio event that corresponds most to the initial segment of the audio signal (in other words, to recognize the word), it is necessary to find the fidelity of the appearance of the sequence Ω = $\{o1, o2, ..., oM\}$ for each available models λ = (P, Φ, π). In this way, there is a set of observed states (speech signal) and a probabilistic model that conveys a hidden state (phonemes) and observable quantities.

Thus, the processing of a voice message occurs in a few steps:

Step 1. The input of the system for identifying the current location S is the input data X. One of the input parameters is a voice message x5.

Step 2. A voice message Ω arrives at the audio event detector, which starts with one of the keywords: start navigation, build route, cancel, stop the starting position, destination.

Step 3. The resulting sequence falls into the audio processing block where we get the λ model.

Step 4. This step defines a specific audio event in a probabilistic way. That is, the record is divided into frames and each frame is skipped through the acoustic model. System with machine learning, defines variants of spoken words and context. The accuracy of the results depends on the completeness of the phonetic alphabet of the system. For each sound, a complex statistical model is first constructed that describes the pronunciation of this sound in the language. The system of recognition compares the incoming speech signal with phonemes, and from them they collect words.

Step 5. In this step, the data fall into the next level of the system as a text for decision. The main teams will be: In what building am I? What floor? I need room №? What room do I have a couple of? How do I get to the room №?

Step 6. After receiving the request, the commands will be mapped to the source data, which include: schedule, group lists, placement and maps of the building and each floor, the list of audiences.

Step 7. Next, using the integrated method will determine the current position on the map of the room.

Step 8. In this step, the data is verified using a neuro-fuzzy method of verification [12].

Step 9. After processing the system, we receive z_2 messages and the route is being built [28].

4 Realization of the subsystem of voice indoor navigation

Within the Smart-Campus application, the ability to display the current position of the user inside the building and the search for the shortest path to the specified beacon [9] was implemented. The next step is to modify the Smart-Campus subsystem of voice navigation.

The Smart-Campus, is a system with Bluetooth Low Energy devices and a back-end database with dedicated content management system (CMS). The idea is to find the location from one beacon to the others, for an interactive tour around the campus or to guide visitors to their specific location of interest. To provide navigation, first a map of the building should be provided or developed. Next is showing the appropriate path to another beacon location. This is why the newly developed solution consists of two parts: a map editor and path detection.

The map editor allows creating a map of a floor. You can use a background picture of a known area or develop it from scratch with the easy-to-use editor. The app-user is the client of information related to a certain beacon at a certain location and our solution allow user to get this information in an attractive way on his or her smart phone through a dedicated application. The app itself fetches the information from the server, related to the unique user ID (UUID) the beacon broadcasts on regular basis. On this server the information is added and edited by the beacon owners through the developed CMS. The users can decide on groups of beacons which are allowed to display their information [8].

The voice navigator will help a person find the location of the audience and the body in which it is located. After the audience is found, the navigator will answer the question about the building in which the audience is located, on which floor and construct the map-device from the current position of the user to the required body. Also, the mobile add-on will provide the user with the opportunity to create and manage their class schedules. The timetable will be displayed for the week and the current day. From the schedule, the user will be able to build a route to the required building.

Let us consider software which contains similar functionality: Google Assistant, Siri, Cortana. For analysis following characteristics were selected: dependence on the Internet, speed of operation of the recognizer, understanding the request, number of satisfactory answers to questions, construction of the route, vocabulary, number of supported languages,. The summary is presented in the table 1.

After analyzing the applications, the main characteristics that should have a voice navigator have been highlighted. The voice navigator, for integration into the Smart-Campus must have the following features:

- to record a voice sentence to get an audience that the user is looking for;
- to recognize vocal sentences and convert them to text;

- to formulate a response to the user;
- to issue a voice message about the user's request;
- to determine the location of the user;
- to build a route from the current position of the user to the required body;
- to display the schedule of occupations of the user;
- to add classes to the schedule;
- to enter the name of the class not only through the virtual keyboard but also through speech recognition;
- to edit or delete selected classes from the schedule;
- to get the route to the chosen lesson;
- to display the schedule for the current day;
- to display the list of recent queries.

The interaction diagram for audio navigation is shown at fig.2.

Table 1. Mobile applications comparison

Parameters	Google Assistant	Siri	Cortana	Voice Navigation
Dependence on the Internet	+	+	+	+-
Speed of operation of the recognizer	+	++	+	+
Understanding the request,%	97	92	86	90
Number of satisfactory answers to questions,%	85,5	78,5	52,4	74,3
Construction of the route	+-	++	+	++
Vocabulary	+	++	+	++
Number of supported languages,	9	34	8	8

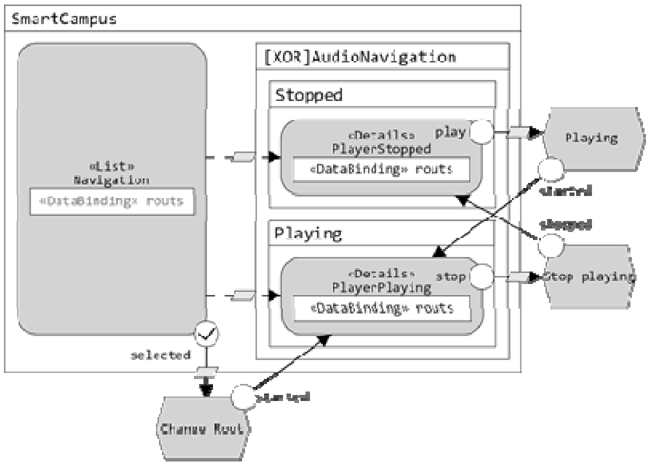

Fig. 2. - Interactions in the voice navigator

For development of the speech recognition, the frame Speech.framework was selected.

First the application is trained with commands which are stored at the local databases.

Fig. 3. – Menu

With each beacon there is connected the voice identification of the location. After the final location is recognised the path is built according to the shortest path algorithm [11]. One of the options is that the user can see the previous voice requests (fig.3).

5 Conclusion

For the in-door navigation system was designed a voice navigator. Integrating the audio navigator into the Smart-Campus system is improving the social adaptation of visually impaired people. Usage of the voice for navigation systems allow user to provide access to information in navigation systems; to connect many objects and events among themselves; to support new user interaction systems, sensors, mobile devices, devices and applications

6 Acknowledgment

The work was partly done within the framework of Erasmus+ [BIOART] Also, the work was carried out within the framework of the agreement on scientific and technical cooperation Agency # 417/156 / 1.4917 dated May 4, 2017 between ZNTU and Limited Liability Company Infocom LTD.

7 References

1. Gnibidenko I.F., Kravchenko M.V., Koval O.M., Novikova O.F.: Social defencing of the population of Ukraine: higher posibilities, per community. K.: View at NAPA; View of "Phoenix". p. 212 (2010)
2. Arras, P., Van Merode, D., Tabunshchyk , G.: Project Oriented Teaching Approaches for E-learning Environment. IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS). pp.317-320 (2017) https://doi: 10.1109/idaacs.2017.8095097
3. Tabunshchyk, G, Parkhomenko, A, Morshchavka, S, Luengo , D.: Engineering Education for HealthCare Purposes: A Ukrainian Perspective. The XIV-th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), Lviv, Polyana, 18-21 April, pp 245 - 249 DOI: 10.1109/MEMSTECH.2018.8365743
4. On the basis of social protection of persons with disabilities in Ukraine: Law of Ukraine 19.12.2017 2249-VIII . (2017)
5. On Amendments to Some Laws of Ukraine on the Expansion of Access to the Blind, Visually Impaired, and Dyslexic Individuals for Works Published in a Special Format: Law of Ukraine 25.12 2015 № 927-VIII . (2015)

6. On amendments to certain legislative acts of Ukraine concerning the protection of the rights of persons with invalidity: Law of Ukraine 18.06 2014 № 1519-VII . (2014)
7. On amendments to some laws of Ukraine on education regarding the organization of inclusive education: Law of Ukraine 5.06.2014 № 1324-VII . (2014)
8. Tabunshchyk, G., Van Merode, D.: Intellectual Flexible Platform for Smart Beacons. In book: Edit by M. Auer, D. Zhutin Online Engineering and Internet of Things, Springer International Publishing, pp. 895-900 (2017) https://doi.org/10.1007/978-3-319-64352-6_83
9. Tabunshchyk, G., Van Merode, D., Goncharov, Y., Patrakhalko, K.: Smart-campus infrastructure development based on BLE4.0. J. Electrotechn. Comput. Syst. 18(94), 17–20 (2015)
10. Speech Recognition. Available at: http://buchuk.domen.uz.ua/index.php?id=realspeaker
11. Petrova, O., Tabunshchyk, G.: Modelling of location detection for indoor navigation systems. IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS),: pp. 961-964. (2017) https://doi: 10.1109/IDAACS.2017.8095229
12. Petrova, O., Tabunshchyk, G., Van Merode, D.: Method for determining the current location in positioning systems and indoor navigation. Electrotechnical and Computer Systems, № 25, pp. 270-278. (2017)
13. Petrova, O., Tabunshchyk, G., Kaplienko, T., Kapliienko, O.: Fuzzy Verification Method for Indoor-Navigation Systems. In: 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET 2018 – Proceedings, Slavske, 20–24 February 2018, pp. 65 – 68 (2018) https://doi: 0.1109/TCSET.2018.8336157
14. Shpakov, D.V.:Voice Recognition in the Sphere of Information Technologies. Young Scientist, №29, pp. 8-11. (2017)
15. Kolesnikova, D. S, Rudnichenko, A. K, Vereshchagina, E.A, Fominova, E.R.: The application of modern speech recognition technologies in the creation of a linguistic simulator to enhance the level of linguistic competence in the field of intercultural communication. Internet journal "Naukovedenie", vol. 9, No. 6. (2017)
16. Chien, J.-T.: Linear Regression Based Bayesian Predictive Classification for Speech Recognition. IEEE Transactions on Speech and Audio Processing, vol. 11, no. 1 January (2003)
17. Watanabe, Sh.: Variational Bayesian Es- tation and Clustering for Speech Recognition. IEEE Transactions on Speech and Audio Processing, vol. 12, no. 4 (2004)
18. Afify, M., Liu, F., Jiang, H.: A New Verification-Based Fast-Match for Large Vocabulary Continuous Speech Recognition. IEEE Transactions on Speech and Audio Processing, vol. 13, no. 4 (2005)
19. Chia-Yu: Histogram-based quantization for Roboust and / or Distributed speech recognition. IEEE Transactions on Audio, Speech And Language Processing, Vol.16, Jan. 1, 2008. (2008)
20. Skowronski, M.D.: Noise Robust Automatic Speech Recognition using a predictive Echo state Network. IEEE transactions on Audio, Speech and Language processing, Vol.15, No.5, June 2007. (2007)
21. Alborova, Zh.V., Rubtsov V.I.: Algorithm and Methods of Speech Recognition. youth scientific and technical weight №FS77 51038, (2016)
22. Rabiner, L. R., Schafer, R. V.: Digital processing of speech signals. Radio and communication, . 496 p. (1981)
23. Subbotin, S.A.: Opt. Mem. Neural Networks 19: 126. (2010) https://doi.org/10.3103/S1060992X10020037

24. Oliinyk, A., Skrupsky, S., Subbotin, S.A.:Parallel Computer System Resource Planning for Synthesis of Neuro-Fuzzy Networks. In: Szewczyk R., Kaliczyńska M. (eds) Recent Advances in Systems, Control and Information Technology. SCIT 2016. Advances in Intelligent Systems and Computing, vol 543. Springer, Cham (2017) https://doi.org/10.1007/978-3-319-48923-0_12
25. Rabcan, J., Rusnak, P., Subbotin, S.: Classification by fuzzy decision trees inducted based on Cumulative Mutual Information. In: 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET 2018 - Proceedings, Slavske, 20-24 February 2018, pp. 208-212 (2018)
26. Leoshchenko, S., Oliinyk, A., Subbotin, S., Zaiko, T.: Using Modern Architectures of Recurrent Neural Networks for Technical Diagnosis of Complex Systems. International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 – Proceedings (2018)
27. Hidden Markov Models, available at: http://www.machinelearning.ru/wiki/images/8/83/GM12_3.pdf