Accessible Driver Announcement in Public Transportation: A Solution based on Speech-to-Text and Display

Abhijeet Sapkota
Department of Computer Science
OsloMet – Oslo Metropolitan University
Oslo, Norway
sapkotaabhijeet10@gmail.com

Raju Shrestha
Department of Computer Science
OsloMet – Oslo Metropolitan University
Oslo, Norway
ORCID: 0000-0001-6409-8893

***Abstract*—Today, the information in public transport, especially when an unforeseen event such as signal failure, platform change, etc. occurs, is largely provided by making an audio announcement. This, in many situations, is hardly understandable due to various factors, and it rarely helps people with hearing impairment. This work attempts towards an accessible driver announcement in public transportation based on a speech-to-text conversion and on-screen displays. Incorrect speech-to-text conversions are checked and corrected using a GloVe (Global Vector) based algorithm so that passengers get correct information. User testing of the system with different people including hard of hearing shows the effectiveness of such a solution. This work has been done, by taking the current system in public transportation in Oslo, Norway as a reference. However, the proposed solution and workflow can be applied to any mode of public transportation and anywhere in the world.**

Keywords—universal design, accessibility, speech-to-text, public transportation, hearing impaired, information display

I. INTRODUCTION

The socio-economic development, prosperity, and extendibility of an urban metropolitan area are generally associated with the efficiency and accessibility of the public infrastructure therein. The primary focus in the development of the transport sector is mainly on infrastructure, environmental friendliness (less pollution), fuel efficiency, and reduced traffic congestion [1].

Public transport services rely heavily on their information system not only for its smooth operation but also for the management of unexpected events, delays, and other situations. Information about such an event is provided to people through websites, mobile applications, news portals, and various other means. These mediums are only effective in providing information once the condition has been known to the operators and get it updated on time [2]. The study conducted in this work found that conveying information about unforeseen events to passengers on-board in an understandable manner has been a special challenge in public transport services around the world. To inform the passengers about an unexpected situation that arises due to a sudden and unpredicted event such as platform change, system failure, and various emergencies like a fire, emergency evacuation, disaster, accidents, etc., drivers make announcements through speakers and provide appropriate instructions. Since information is generally provided in audio form, it hardly helps people with hearing impairment. Most of the time, even normal people have had difficulties understanding such announcements because of various reasons such as poor audio quality and noise. Moreover, foreigners living in the area and tourists face difficulty understanding because of the local language primarily used in the announcements.

This work is based on a case study conducted mainly on the three types of public transportation: bus, tram, and metro in Oslo, the capital city of Norway. The study finds that metros in Oslo are equipped with light-emitting diode (LED) display screens, which show the next station information only (see Fig. 1). Buses and trams are equipped with display screens, which display information about upcoming stops along with estimated time to arrive (see Fig. 2).

Fig. 1. A photo of a LED display panel in a metro in Oslo.

Fig. 2. Display screen showing upcoming stops and estimated arrival time in public buses.

978-0-7381-1120-9/20/$31.00 ©2020 IEEE
DOI 10.1109/iSCI50694.2020.00010

The main objective of this research is to propose a universally designed information system to make driver announcements in public transportation accessible for all passengers including the hard of hearing and hearing-impaired. The proposed solution is based on speech-to-text conversion and text information display. Incorrect speech-to-text conversions are corrected using a Global Vector (GloVe) based algorithm [3]. The system also incorporates a language translation module so that information is displayed in both the local language (Norwegian) and the English language, thus making the information accessible to both local and foreign passengers.

The next section introduces related works. Section III presents the results of our study on the current state of accessibility of information in public transportation. The proposed speech-to-text and display-based system and its development are described in Section IV and V respectively. Section VI presents and discusses system testing, user testing, and performance evaluation and results. Finally, we conclude the article in section VII.

II. RELATED WORKS

Development and improvement in the public transport system are not sufficient if we do not consider the need for all kinds of passengers. Cho et al. [2] developed a digital terrestrial television broadcast system, which provides support for audio signals during emergencies and information is provided through broadcasters via on-screen text display. Their advanced emergency alerting system has the ability to wakeup portable receivers to notify the users in case of acute emergencies. The system also provides emergency preparation guidelines and post-event information such as animated weather maps or escape routes and delivers extended and redundant information by multiple means such as broadcast and broadband, which increases the likelihood of distribution of emergency information for everyone. The system can override the broadcasting of basic information as necessary in case of an emergency alert. Internet, mobile application, SMS, websites are widely used to provide information regarding pre-trip and at-terminal information, which are also displayed at stations or platforms. On the other hand, drivers and/or staff on-board announce information about the next station. Nevertheless, the survey with passengers found that an appropriate amount of information was lacking in the overall journey.

Gnap et al. [4] described some basic features of an information display system that are situated at the front and back of a vehicle. Signage and information used in these systems should be displayed properly with correct alignment and management within the vehicle itself. For individuals with disabilities, accessing information from these systems is crucial to reduce extra difficulties during their travel. To address these difficulties, real-time information might be the only solution that enables them to travel freely. The authors opinioned that some transportation experts follow dedicated benchmarks for displaying information. For instance, London Transport's standard recommends placing display screens 20m apart, and letters ought to be around 75mm, which is calculated based on 10mm letter height for each meter of distance with no letter under 22mm. Øksenholt and Aarhaug [5] evaluated various aspects of the Norwegian public transport system in the Oslo region. They found that individuals with visual impairment encounter challenges locating stations during arrival and departure. Drivers have difficulty realizing that some of them who do not use a stick or a guide dog at the stop have a problem getting into the vehicle and then cannot make a correct action. Besides, people with visual impairment may need help from the driver or staff to get off at the correct stop. Overall, the authors emphasize a need for standards and guidelines for universal design throughout the information system in the transportation industry in Norway.

Papangelis et al. [6] introduced a smartphone application with SMS service that provides real-time passenger information (RTPI) during a disruption. The application allows users to validate information through crowdsourcing information about the transportation service they are using. They opined that smartphones, SMS services, e-mail services, websites, and community displays are the most suitable channels for conveying information during transportation disruption. Further investigation showed that passenger behavioral responses are influenced and shaped by several aspects of information they get during a disruption such as quality of information and passenger's past disruption experiences. The article also describes that incomplete and potentially highly inaccurate information can lead to a high level of uncertainty and inconvenience to the passengers during the travel.

Recently, human speech recognition has gained popularity and use in a wide range of applications. Speech recognition enables any machine or program to identify any words and phrases of spoken language and convert them into machine-readable data. Technological development has made speech recognition capability within any electronic devices feasible. Husnjak et al. [7] discussed the possibilities of speech recognition in smart terminal devices. Their analysis was mainly based on the transportation environment that produced various potential advantages and disadvantages of such a system. There are basically two different architectures that one can follow while implementing a speech recognition system, online and offline. Gazetić [8] compared online cloud-based and offline speech recognition architectures. An online cloud-based speech recognition system communicates with a remote server to recognize the received audio data. This requires an uninterrupted wireless network connection (cellular or wifi) to connect to the server. Also, network performance affects cloud-based speech recognition [9]. An offline speech recognition system operates within the device itself. Disadvantages with this architecture are that it not only has relatively small vocabularies but also has limitations with memory, computation power, and power consumption.

Offline speech recognition is suitable when predefined terms are used to command the execution of certain tasks.

Even though there is a significant improvement in speech recognition technology in recent years, the technology is not yet perfect as errors can occur in speech-to-text conversion. Algorithms such as GloVe can be used to detect and correct such errors [3]. This algorithm scans through the words in a given sentence against the predefined word vectors to detect and correct missing and incorrect words in a sentence. The word vectors also help to relate the structure of phrases, sentences, and similarity between sentences. The accuracy of the algorithm is determined by the size of the word vectors used to train the system. Rodríguez [10] used the GloVe algorithm to compare Spanish phrases to analyze the accuracy of the system.

III. STATE OF INFORMATION ACCESSIBILITY IN PUBLIC TRANSPORTATION

We have conducted a study on the state of accessibility of information, particularly provided through driver announcement in the current public transportation in Oslo. Nineteen participants including hard of hearing and international students were interviewed. More than 65 percent of the participants consider public transportation as their primary means of transportation. Almost 60 percent of the participant depends on the on-screen displays to get information during the travel. Fig. 3 shows the result of the study in terms of the difficulty in understanding the announcements made by drivers in public transportation. Five out of the nineteen responded that they often have difficulties understanding announcements made by the driver. This indicates the necessity of an accessible system for passengers that helps access correct information given by the drivers through announcements.

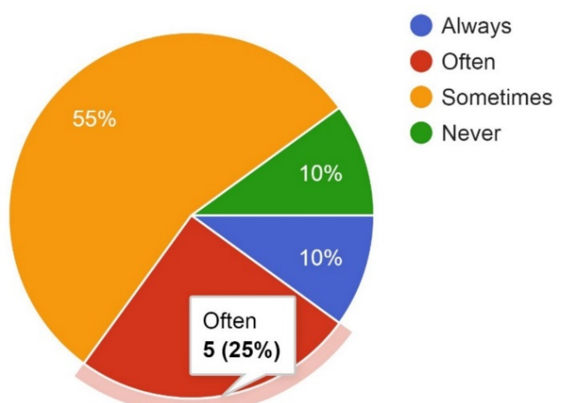


Fig. 3. Results on difficulties understanding driver announcements in public transportation.

The participants mentioned various problems, which hinder them from understanding driver announcements; among them, noise and poor language speaking skills are the most common problems. Most of the participants also criticized the quality of audio from the speakers that are installed inside buses and trams. One international participant described it as ***"****The audio from the speakers is too low. It is hard to hear what the driver is saying. Often the drivers only say it once, and also not in English. And, there is no written message*". Few other international participants explained the language barriers that prevent them from understanding the announcement. This shows a need for a system that provides information in the English language as well. Most of the participants mentioned that it would be very helpful if display panels highlight announcements in real-time both in Norwegian (local) and English (international) languages. They also recommended that the drivers should check if the display panels were turned on and make sure they are showing the correct information about the next station and other vital information.

IV. PROPOSED SYSTEM

Fig. 4 depicts the system architecture of the proposed system, which consists of three main modules: speech-to-text conversion module, GloVe algorithm-based error detection and correction module, and display system. The modules are as described below.

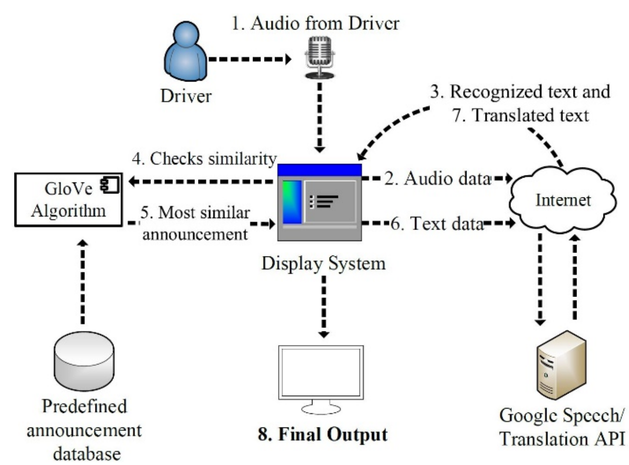


Fig. 4. The system architecture of the proposed system.

A. Speech-to-Text conversion

The speech-to-text conversion module takes the driver announcement captured with a microphone as an audio input and converts it into text. In our implementation, we have used a cloud-based Google's speech-to-text conversion API as it is one of the best performing options available at the moment and easy to use. We assume that an uninterrupted wireless internet connection is available. The audio data is sent to a Google server as a chunk of digital audio data. Google also provides a language translation functionality via its translation API, and it can be used to translate a text in one language to another. A microphone with a noise cancellation feature can be used to minimize the effect of noise in the audio data.

B. GloVe algorithm-based error detection and correction

This module checks the similarity between the text converted from the audio and the list of predefined announcements. The algorithm selects the most similar announcement from the list. This not only enables the system to detect incorrect words but also to detect missing words, and accordingly return a correct and complete sentence with meaningful information before displaying it on the screen. We have discussed the performance of this module in terms of conversion speed or accuracy in section VI.C below.

C. Display system

Finally, the display system displays the correct text output from the error detection and correction module on the display screen(s), both in Norwegian and English languages. This enables passengers including those with hard of hearing as well as others who have difficulty understanding the audio announcement to get the information in text form, thus making driver announcements accessible.

V. DEVELOPMENT OF THE SYSTEM

A working prototype system is developed in Python programming language using the rapid application development (RAD) approach [11]. Google speech recognition API is used for speech-to-text conversion, and Google translation API is used for translation from Norwegian to English and vice versa. The GloVe-based error detection and correction module is developed, which detects and corrects any errors in the recognized text returned by Google speech recognition API.

The system was trained with word vectors consisting of four hundred thousand words. The system uses predefined announcement text in various situations to generate an accurate output text. Table I shows some examples of the predefined announcements. In case of an incorrect or incomplete text, the module generates a complete text with the help of these predefined announcements.

Fig. 5 illustrates the display interface of our system. For the passengers, the display interface is similar, but without elements 3, 4, and 5. Different elements in the interface are described below.

TABLE I. EXAMPLES OF PREDEFINED ANNOUNCEMENTS (PAs)

PA1	Attention passengers, we cannot move the train unless you keep out of the way of the closing doors. Stand clear.
PA2	This train has been delayed due to essential engineering work. We will keep you informed. Please stay calm.
PA3	Because of a security alert, the next station has been closed and evacuated. Therefore, this metro will not stop at the next station.
PA4	This metro and the station are now being evacuated. Please leave the metro now and get the staff's help to exit the station.

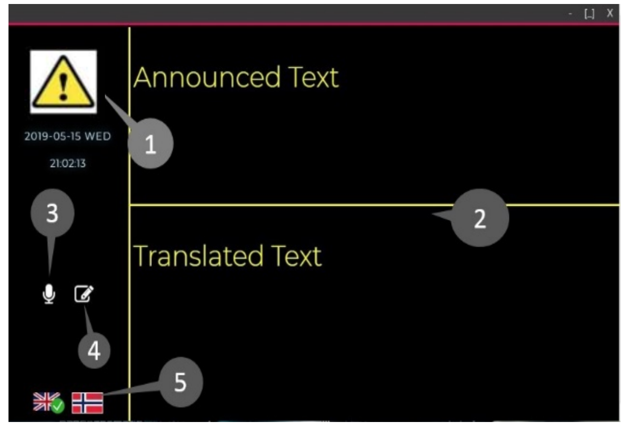


Fig. 5. Display interface for announcement text.

(1) ***Icon, date, and time***: This element displays date and time. It also has an icon that will change once the driver starts to make an announcement, which can be very useful to grab the attention of passengers during the travel.
(2) ***Text display area***: The text display area is divided into two different sections. The upper section displays the text in the local language, and the lower section displays the text in English.
(3) ***Mic button***: Once this button is clicked, the program takes the audio input and sends it to the server.
(4) ***Text correction button***: Clicking this button starts the text checking and error corrections.
(5) ***Language option icons***: Two icons with British and Norwegian flags provide options to make an announcement either in English or Norwegian.

The interface is designed, taking into account the recommendation made by the Norwegian Association for Hard of Hearing (Hørselshemmedes Landsforbund, HLF), to have better visibility and understanding while displaying the text. Black background and yellow text are used considering the people with color vision deficiency (or color blind).

Universal Design Consideration

Considering the diversity of passengers accessing information during their travel, it is always difficult to perfectly figure out the requirement of an individual. Each passenger may have a different way of pursuing the information through the proposed system. Therefore, the system is developed following the standard universal design principles so that all the passengers traveling at the same time receive the same necessary information with ease [12]. The following universal design principles are taken into account while designing and developing the system. **a)** ***Equitable use*** - The interface design of the system has high contrast that helps the users with a weak vision to distinguish and read the text displayed on the screen. **b)** ***Flexibility in use*** - Any screen size will not affect the appearance of the text due to the automatic sizing feature of the text within the layout. The size of the text will increase or decrease based on the size of the screen available within public transport. **c)** ***Perceptible information*** - Besides having adequate text contrast and sizing, the information provided by this system explains the same and necessary information to the passengers. **d)** ***Low physical effort*** - The information display should be placed within a better viewing angle from the passenger's seat. This will allow passengers to read information without much physical effort.

VI. TESTING, EVALUATION, AND RESULTS

Testing and evaluation of the system were done in 3 parts: system testing, user testing, and performance evaluation.

A. System Testing

The system was tested for its usability and functionality of different features. Testing was divided into three parts. First, the test was conducted in two different scenarios without using text correction, one in a studio environment with a broadband internet connection, and then in a simulated environment with an internet connection using a cellular network. Then, the text correction module was tested.

Studio environment: In this environment, the system was connected to broadband Internet, and the testing was done in a reasonably quiet room. Fig. 6 shows an example display output from the system when an announcement was made in English. The recognized English text and translated Norwegian text was reasonably good.

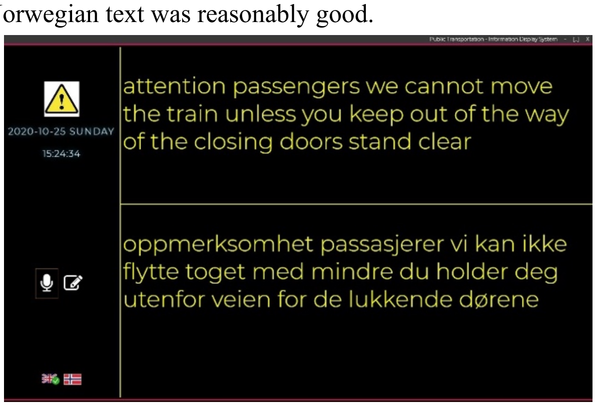


Fig. 6. Sample output from the system when tested in the studio environment.

Simulated environment: The system was tested in a simulated environment with noises, which try to mimic a real public transportation environment. In this scenario, the system was connected to the Internet through a mobile network. Fig. 7 shows a sample output in this scenario. We can see some errors from the speech-to-text conversion module, and the accuracy of the recognized text was lower compared to the results from the studio environment.

Testing error detection and correction module: The module was tested to validate if this can correct inaccurate or incomplete sentences output from the speech-to-text module. Table II shows four example test announcements (TA). Each of them corresponds to the predefined announcements as shown in Table I. Fig. 8 shows a sample output from the module while feeding an incomplete test announcement.

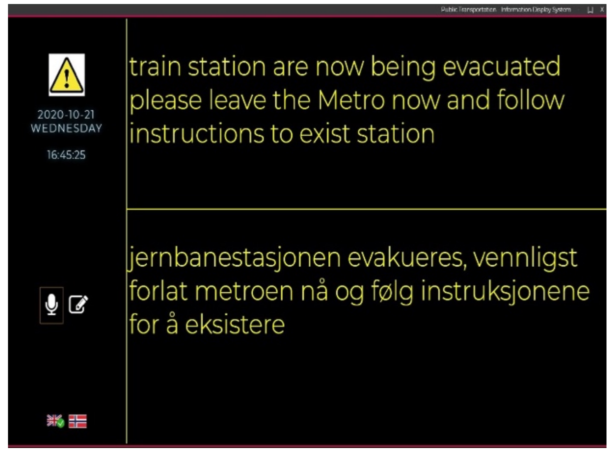
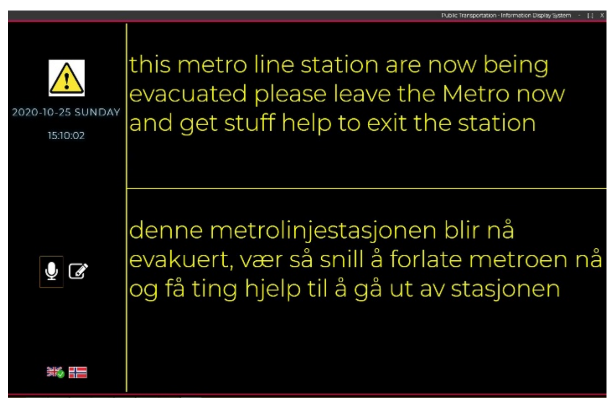


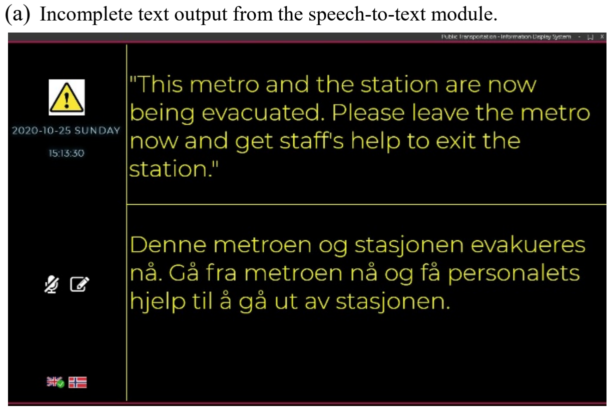
Fig. 7. Sample announcement text displayed in the simulated environment.

TABLE II EXAMPLE OF TEST ANNOUNCEMENTS (TAs)

TA1	We cannot move the train doors are closing stand clear.
TA2	The train has been delayed due to essential work. We will keep you informed.
TA3	Social Security alerts Nexus station has been closed and evacuated so this metro will not stop in nexus station.
TA4	This metro line station is now being evacuated. Please leave the Metro now and get stuff help to exit the station.



(a) Incomplete text output from the speech-to-text module.



(b) Corrected complete text from the GloVe module.

Fig. 8. A sample result showing an incorrect/incomplete text in (a) and resulting corrected text in (b).

We see that the module generates a complete and meaningful text by picking the most matched sentence from the list of predefined announcements.

B. User Testing

The system was demonstrated to the same 19 participants who were interviewed in our study conducted on the state of accessibility of information in public transportation above in section III. The participants were asked to test and evaluate the system in terms of visibility, readability, and meaningfulness of the displayed text generated by the system. Fig. 9 shows the results of the user testing in terms of the percentage of participant's responses on visibility, readability, and meaningfulness.

The results show that more than 70 % of the participants found the system providing meaningful information, and more than 60% of participants responded as good visibility. However, less than 50% of them found the text to be readable. As per some participants, the reason behind this is because of the long text in most of the test cases. This could be improved by shortening texts along with the text correction, possibly using a more advanced semantic analysis method. We have considered this as a possible future work.

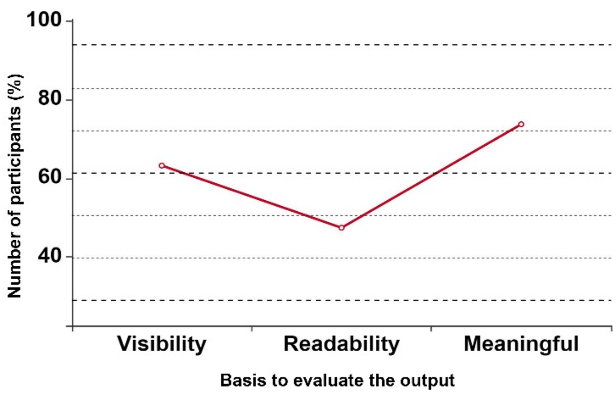


Fig. 9 User testing results in terms of visibility, readability, and meaningfulness of the output from the system.

The feedback provided by the participants was very useful to further improve the system. Some of them have recommended using a darker background and white-colored font as it has a higher color contrast level. They also suggested using larger and contrasting symbols and pictograms that can change its dimension based on screen size. On the contrary, some have also mentioned using a completely white background. One of the participants suggested that the terms such as "It is an emergency" could be made larger, and the rest of the message could be displayed on the next line. Participants also suggested having a function that detects the language automatically. Messages for frequently occurring situations such as traffic delay, the red signal could be standardized. Implementing a QR code that opens a website or the text message followed by the contact information to the concerned authority was a unique and new feature suggested by one of the participants.

Most of the participants found the real-time announcement display function on the system very remarkable. Participants also provided positive feedback on error correction and language translation features.

C. Performance Evaluation

This sub-section describes the overall performance of the system based on speech-to-text conversion accuracy and speed. Real-time driver announcements are compared with the predefined announcements list based on the structure and similarity of the sentences. Fig.10 shows the similarity measures of the four TAs and the corresponding PAs listed in Table I and Table II respectively. The brighter background represents a higher percentage match, and the darker represents a lower percentage match. The Glove-based error detection and correction algorithm selects the most similar sentence which has the highest similarity percentage. We see that this functionality has shown improvement in the accuracy of the information that is displayed to the passenger.

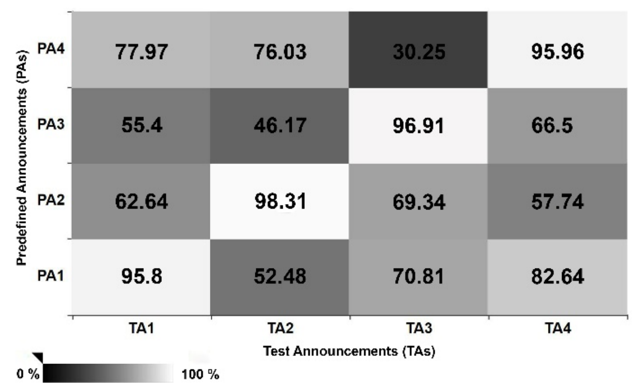


Fig. 10 Similarity measures between the test announcements and the predefined announcements in percentage.

Fig. 11 shows speed of the system in terms of execution time for speech-to-text conversion and text correction. On average, the total execution time is less than 2 milliseconds.

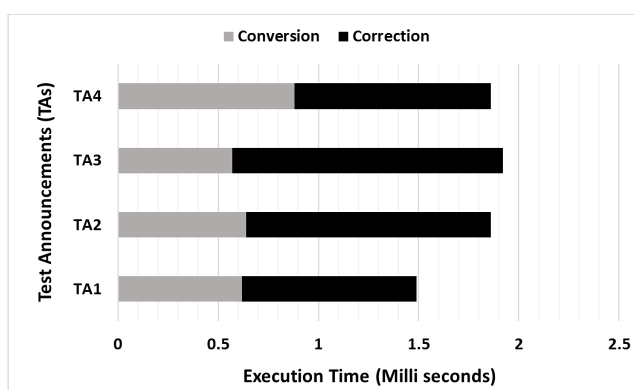


Fig.11 Execution time for a complete speech-to-text, correction, and display of the test announcements.

VII. CONCLUSION

The case study we have done on public transportation in Oslo found various barriers in the current information system, especially for the people with hard of hearing and foreigners when the information is given through drivers announcing certain unforeseen events such as signal failures, delays, and emergencies. To address this, a solution based on speech-to-text conversion and translation of the local language to English and displaying on the screens is proposed to make information easily accessible to the passengers. The proposed system was also integrated with a GloVe-based algorithm to correct the errors made while converting driver announcements to text. The algorithm helped to fix incorrect, incomplete, and confusing information. The proposed solution is found to be effective in improving the accessibility of the information. We believe that such a system will be helpful for people with hard of hearing for traveling independently in public transportation.

ACKNOWLEDGEMENTS

The authors would like to thank the Norwegian Association for Hard of Hearing (Hørselshemmedes Landsforbund, HLF) for their support by providing participants and also by giving valuable feedback throughout the research work.

REFERENCES

[1] M.D. Marchi, J. Eriksson, and A.G. Forbes, *TransitTrace: Route planning using ambient displays*, in Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2015, ACM: Seattle, Washington. p. 1-4.

[2] Y. Cho, B. Bae, and Y. Kim. *Design and Implementation of Emergency Alert Gateway utilizing ATSC 3.0 terrestrial UHD broadcasting*. 2020 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2020, pp. 1-2, doi: 10.1109/ICCE46568.2020.9043029.

[3] J. Pennington, R. Socher, and C. Manning. *Glove: Global vectors for word representation*. in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014. p 1532-1543.

[4] J. Gnap, V. Konečný, B. Poliaková, and M. Gogola. *Improving of information for passengers of urban public transport in Košice*, in Attractive Urban Public Transport for Accessible Cities. 2014, University of Zilina. p. 86.

[5] K.V. Øksenholt and J. Aarhaug. *Public Transport and People with Disabilities - the Experiences of Non-users*. in European Transport Conference, Bercelona, Spain. Association for European Transport (AET). 2016.

[6] K. Papangelis, S. Sripada, D. Corsar, N. Velaga, P. Edwards, and J.D. Nelson. *Developing a Real Time Passenger Information System for Rural Areas*. Information and Interaction for Health, Safety, Mobility and Complex Environments. Lecture Notes in Computer Science, vol 8017. Springer, Berlin, Heidelberg, 2013. https://doi.org/10.1007/978-3-642-39215-3_19

[7] S. Husnjak, D. Perakovic, and I. Jovovic, *Possibilities of using speech recognition systems of smart terminal devices in traffic environment.* Procedia Engineering, 2014. **69**: p. 778-787.

[8] E. Gazetić, *Comparison Between Cloud-based and Offline Speech Recognition Systems.* Master Thesis, 2017, Technical University of Munich, Germany.

[9] M. Assefi, M. Wittie, and A. Knight. *Impact of network performance on cloud speech recognition*. in Computer Communication and Networks (ICCCN), 2015 24th International Conference on. 2015. IEEE.

[10] I.S. Rodríguez, *Text similarity by using GloVe word vector representations*. Master thesis, 2017, Polytechnic University of Valencia, Spain.

[11] M. Hofmann, J. Burke, and J. Pearlman, *Clinical and Maker Perspectives on the Design of Assistive Technology with Rapid Prototyping Technologies*, in Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility. 2016, Association for Computing Machinery: Reno, Nevada, USA. p. 251–256.

[12] M. Allen, C. Berg, J.Q. Dawson, and N. Leveridge *Insights from the Application of Universal Design Principles to Support English Language Learners*, in Proceedings of the 23rd Western Canadian Conference on Computing Education. 2018, Association for Computing Machinery, New York, NY, USA, Article 10, 1–6. doi:https://doi.org/10.1145/3209635.3209646