IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Speech recognition in flight simulator

To cite this article: S Rustamov *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **459** 012005

IOP Conf. Series: Materials Science and Engineering **459** (2019) 012005 doi:10.1088/1757-899X/459/1/012005

Speech recognition in flight simulator

S Rustamov[1,3], E Gasimov[1], R Hasanov[1], S Jahangirli[2], E Mustafayev[3], D Usikov[2]

[1]ADA University, School of Information Technologies and Engineering, 11 Ahmadbey Aghaoglu St, Baku, Azerbaijan

[2]Azerbaijan National Aviation Academy, Azerbaijan

[3]Institute of Control Systems of ANAS, Azerbaijan

srustamov@ada.edu.az, ADA University

Abstract. The article is devoted to the investigation of speech recognition in Flight Simulator cockpit. We have done research and developed software for speech recognition in Flight Simulator with limited vocabulary in C# from scratch. The speech recognition system works in real time and offline mode in Windows platform. We used Mel Frequency Cepstral Coefficients and Linear Predictive Coefficients feature extraction algorithms and trained the system by multilayer artificial neural networks. The User Interface of platform is highly functional and allows users to update system parameters through the interface. Our Speech Recognition system results are compared to the results of Microsoft Speech SDK and received satisfactory achievement.

1. Introduction

Virtual Reality (VR) creates an opportunity for users to interact with three-dimensional digital representations of products through natural human means. Through combining speech interface with a feasibility to interact directly with objects in the virtual environment leads to a multimodal interface wherein users are able to interact with the Virtual Reality surroundings by the means of physical or speech commands [1].

Two approaches are available for applying speech recognition to a virtual environment: speech of fully interactive and speech of "command and control". Notwithstanding the fact that fully interactive speech makes possible to recognize numerous words and phrases, it requires a speaker-dependent system to be applied. The speech of "command and control" needs just a minor set of predefined commands and allows using speaker-independent methods [9]. As the illustrated approach demands only a limited vocabulary, we have chosen the command and control approach combined with speaker-independence. The main purpose of the article is to realize the system with a menu of speaker-independent, command and control speech recognition for Flight Simulator which one of the VR applications. As experiment we recognize 41 Flight checklist commands. The major function of the flight deck checklist is to ensure that the crew will properly configure the airplane for any given segment of flight.

2. Related works

There are many researches have been investigated for speech recognition.

D Doye et. al. suggested non-linear time alignment model for the Marathi language. As feature extraction algorithms they used Mel Frequency Cepstral Coefficients (MFCC), Linear Frequency Cepstral Coefficients (LFCC) and Linear Prediction Coefficient (LPC) [12].

In [13] authors used DTW approach based on MFCC features for the Speech Recognition of Tamil database.

Elyes et. al. suggested hybrid approach of SVM/HMM for Arabic ASR based on triphones modelling. They showed SVM/HMM hybrid model gives better results than HMMs and MLP/HMM models [14].

J. Padmanabhan et.al used Gaussian mixture model and HMM for automatic speech recognition [15].

In [16] authors proposed sparse auto-encoder using Convolution Neural Network.

In [17] authors suggested hybrid HMM/DTW approach by using kernel adaptive filters for speech analysis and recognition.

Aida-zade et al. applied multi-structured neural networks to improve SR for Azerbaijani language [20][22].

The authors from [19] developed End-to-End Speech Recognition with Recurrent Neural Networks. They trained SR from raw data without intermediate phonetic representation. The suggested system is based on a combination of the deep bidirectional LSTM and the Connectionist Temporal Classification objective function. They achieved 27.3% WER on the Wall Street Journal corpus with no prior linguistic information, 21.9% with only a lexicon of allowed words, and 8.2% with a trigram language model. Combining the network with a baseline system further reduces the error rate to 6.7%.

Xiang-Lilan et. al. introduced merged-weight dynamic time wrapping algorithm and showed this approach gives better result than DTW [20].

3. Feature extraction

The following tasks have been carried out for calculation of characteristics of speech.

Cleaning noise from speech. As speech signal varies in low frequency, for cleaning high frequency noise of it, speech is filtered in first order high frequency filter.

Identification of the end points of speech. Identifying the start and end points of pronounced speech is one of the main problems in Speech Recognition. VAD (Voice Activation Detection) is one of the methods used to identify the end points of speech signal. Our system provides user with three parameters to identify the end points of speech accurately.

1. The number of blocks that speech is divided to identify the end points of speech. In our system, for silence 100 partitions have been spared. Having high value helps to get the pauses between speech's phonemes.

2. The number of blocks in which environmental sounds occur. In our system, 5 blocks have been spared for silence (30msecs).

3. Speech separation coefficient. Changing this coefficient according to speech recording sensitivity, we can test the correctness of identification of end points of speech signal.

Dividing words which establish speech into classes, and finding their average length: As the recognition system that we create is realized with neural networks model, the input of neural networks should be collection of data taken in equal length. That's why speech parts with different length are resulted in matching length. Lagrange's interpolation method is used for this problem.

Speech framing. According to the practices, speech signal keeps its stationarity in some interval, for approximately 20 ms. Using speech signal's quasi stationary characteristics, to analyse speech, it is divided into frames. In the next steps, before applying discrete Fourier transformation, speech signal is multiplied by Hamming window. After Hamming window, in some part of speech signal information decreases considerably. That's why, to avoid the risk of losing useful information of speech signal, frames are approximately overlapped.

Feature extraction from speech frames. In our system, two types of features extraction algorithms are used in parallel: Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC) features.

Cleaning speech signal from channel effect. While recording speech signal some noise occur, which is called channel effect. During system training and afterward usage of it different channel effects decreases the system recognition accuracy. To avoid this, cepstral mean subtraction method is used.

4. Recognition module

Depending on the pronunciation of speech, system divides speech samples into classes according to their length: short and long. Artificial neural networks is used to train the system. Separate network is created for each class. The outputs of neural networks are number of words in class, and input of neural networks are number of features of speech signal. The neural network that we use is fully connected graph and trained by conjugate gradient method.

As mentioned above, in the speech pre-processing block, its MFCC and LPCC features are calculated. Training and recognition are done based on both two features. Results are compared based on recognition according to both two features.

Recognition process is done in two steps:

1. Parallel recognition processes are done in MFCC-based and LPC-based subsystems.
2. Calculated recognition results are compared in MFCC-based and LPC-based subsystems, and speech recognition system accept the result of two subsystem.

5. Experiment results and Conclusion

We have done research and developed software for speech recognition in Flight Simulator with limited vocabulary in C# from scratch. As experiment we recognize 41 Flight checklist commands. The major function of the flight deck checklist is to ensure that the crew will properly configure the airplane for any given segment of flight. The list of checklist commands given in the table 1. We processed approximately 33000 speech samples from 80 people and applied artificial neural networks to train collected data. We divide commands into two classes (long and short) and trained neural networks for each class. Some words included in both classes. At first, the system checks uttered command length, then call the appropriate neural network for recognition.

Table 1. The list of checklist commands

Preflight inspection	Cockpit
Towbar	Cockpit checklist completed
Weight and balance	One
AC documents	Two
Circuit Breakers	three
Seats & Belts	Four
Cabin doors	Five
Fuel selector	Six
All switches	seven
Completed	eight
On	nine
Off	zero
Closed	decimal
Checked	removed
Fuel Shutoff Valve	aboard
Shut-off cabin heat	adjusted & locked

Alternate air door	locked
Battery+Main bus	open
Fuel Quantity	In
Fuel Temperature	Sufficient
Flight controls	

The speech recognition system works in real time and offline mode in Windows platform. We used Mel Frequency Cepstral Coefficients and Linear Predictive Coefficients feature extraction algorithms and trained the system by multilayer artificial neural networks. The User Interface of platform is highly functional and allows users to update system parameters through the interface. Our Speech Recognition system results are given in table 2 and 3. We tested Microsoft Speech SDK for our data set and received 89.6% correct recognition, 2% error recognition and 8.4% rejection rate. It is not fair to compare directly our results with Microsoft Speech SDK due to data collection conditions. Our collected data and testing data were recorded in same environments and by the same microphone.

Table 2. Recognition result of the training samples

	Accuracy (%)	Rejection (%)	Error (%)
Short words with MFCC	99.86	0.14	0.00
Short words with LPC	98.85	1.12	0.03
Long words with LPC	99.97	0.03	0.00
Long words with MFCC	99.97	0.03	0.00

Table 3. Recognition result of the test samples

	Accuracy (%)	Rejection (%)	Error (%)
Short words with MFCC	94.85	4.36	0.79
Short words with LPC	93.86	4.56	1.58
Long words with LPC	97.56	1.72	0.72
Long words with MFCC	97.56	1.72	0.72

Acknowledgement
This work has been carried out in Center for Data Analytics Research at ADA University.

References

[1] Scott Mc and Tomas A 2007 A Speech Interface to Virtual Environments. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.39.9070&rep=rep1&type=pdf
[2] Rami M, Svitlana M, Vyacheslav V L, Nataliya V B 2017 Speech Recognition Systems: A Comparative Review *IOSR Journal of Computer Engineering (IOSR-JCE)* **19** 5 pp 71-79
[3] Aman A, Sonu K M, Rinaz Sh, Chandraketu K G, Prakhar M, Soudamini P 2016 A Survey Paper on Acoustic Speech Recognition Techniques *Int. Journal of Recent Advances in Engineering & Technology* **4** 7
[4] Dinesh K D, Yogesh R 2015 A Survey Paper on Automatic Speech Recognition by Machine *Int. Journal of Computer Science and Information Technologies* **6** 3 pp 2918-2922
[5] Suma S and Ramakrishnan K 2013 An Efficient Speech Recognition System *Computer Science & Engineering Int. Journal* **3** no 4
[6] José M C, Manuel D, José M P 2012 Automated Speech Recognition in ATC Environment. *ATACCS'2012*

[7] Neha Ch, Gangwar R, Rajeev B 2015 Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey. *Int. Journal of Computer Applications* **131** no 11

[8] Shane S, Jonathan R, Takaaki H, Shinji W, John R H 2018 End-to-End Multi-Speaker Speech Recognition *IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*

[9] Denis V D, Judy M V 2002 Implementing Speech Recognition in Virtual Reality *Mechanical Engineering Conference Presentations, Papers, and Proceedings*

[10] Ayushi Y V, Krina A S, Yesha A K, Nirali P 2017 Survey paper on Different Speech Recognition Algorithm: Challenges and Techniques *Int. Journal of Computer Applications* **175** no 1

[11] Kelvin V, Kevin Ch, and Endra J 2017 Virtual Reality Flight Simulator *Internetworkıng Indonesıa Journal* **9** no 4

[12] D D Doye, T R Sontakke, Smita N 2015 The Nonlinear Time Alignment Model for Speech Recognition System *IETE Journal of Research, Taylor & Francis* pp 271-275

[13] Kavitha R, Nachammai N, Ranjani R, Shifali J 2014 Speech Based Voice Recognition System for Natural Language Processing *Int. Journal of Computer Science and Information Technologies* **5** 4 pp 5301-5305

[14] Elyes Z, Yassine A, Faiez G 2014 Hybrid continuous speech recognition systems by HMM, MLP and SVM: a comparative study *Int. Journal of Speech Technology* **17** 3 pp 223-233

[15] Jayashree P and Melvin J P K 2015 Machine Learning in Automatic Speech Recognition: A Survey *IETE Technical Review, Taylor & Francis* pp-1-13.

[16] Qirong M, Ming D, Zhengwei H, and Yongzhao Z 2014 Salient Features for Speech Emotion Recognition Using Convolutional Neural Networks *IEEE transactions on multimedia* **16** no 8.

[17] Siva Prasad Nandyala and T. Kishore Kumar 2014 Hybrid HMM/DTW based Speech Recognition with Kernel Adaptive Filtering Method *Int. Journal on Computational Sciences & Applications* **4** no 1

[18] Aida-Zade K, Ardil C, Rustamov S 2007 Investigation of combined use of MFCC and LPC Features in Speech Recognition Systems *World Academy of Science, Engineering and Technology Int. Journal of Computer and Information Engineering* **1** pp 2647-2653

[19] Alex G, Navdeep J 2014 Towards End-to-End Speech Recognition with Recurrent Neural Networks *Proc. of the 31 st Int. Conference on Machine Learning, Beijing, China. JMLR: W&CP* **32**

[20] Xiang-Lilan, Zhang, Zhi-Gang, Luo, Ming Li 2014 Merge-Weighted Dynamic Time Warping for Speech Recognition *Int. Journal of Computer Science and* Technology **29** 6 pp 1072-1082

[21] Aida-Zade K, Rustamov S 2008 The Principles of Construction of the Azerbaijan Speech Recognition System *The 2nd Int. Conf. "Problems of Cybernetics and Informatics"* **1** pp 183-186

[22] Aida-Zade K, Rustamov S, Mustafayev E 2009 Principles of Construction of Speech Recognition System by the Example of Azerbaijan Language *Int. Symposium on Innovations in Intelligent Systems and Applications* pp 378-382

[23] Ayda-zade K, Rustamov S 2005 Research of Cepstral Coefficients for Azerbaijan speech recognition system *Transactions of Azerbaijan National Academy of sciences."Informatics and control problems* **25** no 3 pp 89-94.