Expanding functionality of the autonomous voice control system

Vladimir Zhigalov[1*]

[1]Moscow Aviation Institute MAI (National Research University), Volokolamsk highway, 4, 125993, Moscow, Russia

Abstract. The article describes the main advantages of voice control systems, notes the shortcomings of the first implemented voice control systems, their functional limitations, and limited use in comparison with modern voice assistants, and also formulates the shortcomings of modern voice control systems. It is noted that with the expansion of functionality, the scope of voice control systems is expanding significantly. The recommendations of leading companies that implement voice technologies, indicating the need to continue research and development in this area. Due to complex calculations in speech recognition, there is a need to use high-speed computers, which makes it difficult to create compact and autonomous voice control systems that work without an Internet connection. At the same time, the growing need for autonomous voice control systems requires a reduction in computational costs, an increase in the speed of calculations by simplifying the methods and algorithms of these systems, implemented in software and hardware in the form of autonomous electronic devices. Methods are distinguished that allow performing recognition with sufficient reliability, such as hidden Markov models and the method for determining mel-frequency cepstral coefficients. In addition, methods are proposed to reduce computational complexity by reducing the number of similar-sounding words included in the command and the commands themselves, and to increase the speed of calculations, the use of electronic components is proposed in the implementation of computational algorithms and their parts. The operability and the possibility of practical application of ways to expand functionality and reduce computational complexity for autonomous voice control systems have been tested and confirmed experimentally, and some of them correspond to the recommendations of leading companies implementing voice technologies.

1 Introduction

The voice control system is an important part of the voice interface, often considered to be its backbone. A voice interface that allows objects to be controlled not only by hand through manual elements or controls, but also by voice, provides a number of important

* Corresponding author: vladivan25@yandex.ru

advantages, as it allows all necessary manual operations to be performed simultaneously, which significantly increases the overall functionality of systems capable of receiving and executing voice commands. In addition, voice control systems, being a convenient, simple, natural and more comfortable for a person form of interaction with the controlled object, increase the speed of the operations performed, reduce the labour required, for example, when entering commands from the keyboard. The need to create speech recognition systems and systems with voice interface has existed for a long time, but their practical implementation for a long time was impossible due to the lack of the necessary technical, or rather computing power, and software.

Since the creation of a speech recognition system is a rather complex task, the first works in this direction were devoted to the development of speech command recognition systems and voice control systems. With a limited number of commands and orientation to a particular speaker, such systems used rather simple speech processing methods and allowed to obtain satisfactory results of command recognition and control. To improve the quality of recognition, the first systems used pre-training and tuning to the voice of a particular speaker. The recognition task was limited to a set of telephone subscribers' digits or the selection of a short command containing only one word from a specific, limited set. In addition, strict requirements were imposed on the timbre of the speaker's voice, intonation and speed of command pronunciation, which had to correspond to a previously recorded reference. There were known cases when even insignificant changes in the voice timbre, for example, due to a throat disease of the announcer, resulted in unsatisfactory recognition or errors in recognising commands. At the same time, strict requirements were imposed on the quality and conditions of recording of the standards, as well as on the environmental conditions during the pronunciation of the command, i.e., on the ambient noise level which should be minimal or better absent as well as during the recording of the command standard. It is clear that such strict limitations significantly reduced the confidence in these systems from the users, and failure to comply with the requirements led to malfunctioning and failures in recognition. It became possible to remedy the situation only after the transition to more advanced recognition methods. So, at present, voice or virtual assistants and assistants successfully enough recognise commands consisting of a single word, answers to questions asked by them in the form of "yes" or "no", as well as short phrases formed according to the rules defined by templates and these assistants. However, the speech and command recognition capabilities of even modern systems are very limited, which reduces the reliability of recognition and creates certain inconveniences and sometimes problems in speech control.

2 Methods and Materials

Nowadays, the use of voice to control objects has become widespread, and it is possible to distinguish areas in which voice technologies have been used successfully and for quite a long time. These include:

- telephony, in which the use of answering machines has made it possible to process automatically hundreds of thousands of calls per day, saving both the caller's time and the telephone company's resources,
- household appliances, enabling voice control of various household devices, such as lighting systems, hoovers, switches, receivers, TV sets, etc., as well as a variety of gadgets and computers, and even cars, where modern voice control systems are triggered either by precisely defined commands or by the presence of keywords in the commands,
- medical equipment that uses voice control for people with disabilities, such as wheelchair control, etc.

It should be noted that with the improvement of the quality and reliability of voice interfaces their application areas are constantly increasing and, in addition to the known, in which they have successfully proved themselves, there is a need to use voice interfaces to control not only household but also complex technical devices. It is known that often the number of control elements of complex systems is quite large and when it is necessary to simultaneously set control actions with the help of manual commands, the operator finds himself in a rather difficult situation, especially in the absence of experience and skills in controlling such systems. Therefore, the operator is often forced to undergo special training to reduce the number of possible errors in control, although training cannot completely eliminate all of his potential errors. The consequences of errors are also important. In the best case, if there is an error, it can be quickly corrected, for example by issuing a second correct command, cancelling the erroneous one. However, unfortunately, this is not always possible when controlling complex systems performing critical operations in a number of areas, e.g., nuclear power, aviation, space technology, medicine, etc. In addition, the presence of a significant number of control elements of complex systems often leads to the inaccessibility of individual elements and creates inconvenience in the management and operation of these systems. Therefore, an important issue for modern complex systems is the creation of a convenient and simple operator-system interaction interface that reduces potential operator errors. Therefore, an interface augmented with voice or speech control is generally considered superior to the previously used manual interface.

However, despite the fact that modern voice interfaces and voice control systems are much more reliable and functional than their predecessors, some performance deficiencies inherent in the first voice control systems remain, the main ones being.

1. The requirement to use the internet,
2. The requirement to be noise free,
3. Limited number of working commands in a compact design of the control device,
4. The need for relatively high speed of command processing and recognition devices when using a voice interface with a significant number of realisable functions.

It is expedient to describe them in detail.

1. Internet availability allows solving complex tasks in real time, including recognition of continuous speech using Google and Yandex recognition systems, however, in this case, in the absence or disconnection of the Internet without an autonomous voice control system it is not possible to recognise and execute commands after their recognition. Although for a number of tasks, failures and malfunctions of the control system, including voice control, are unacceptable. Internet loss is possible in a lift, basement, router failure, etc., which will lead to the impossibility of voice control of, for example, a wheelchair. Therefore, such devices should have a free-running voice control system. When using standalone devices with limited computational capabilities, to ensure acceptable recognition quality and speed, the number of commands that the voice interface of these devices supports should not be large. Hence, to ensure smooth and reliable operation of the voice control system, it is necessary to develop voice control interfaces for devices that can recognise and execute voice spoken commands when the internet is disconnected, offline, so this task is important and relevant.

2. The presence of noise during the operation of voice control and speech recognition systems is a significant problem not only for the first, primitive, but also for modern voice interfaces. In order to reduce the number of possible errors and violations of the reliability of voice command recognition, developers offer various private ways to reduce ambient noise to create more ideal working conditions for their systems, which cannot fundamentally solve the problem of noise removal. First of all, it is recommended to limit or reduce noise exposure if it cannot be completely eliminated. For example, when voice controlling devices in the car and external noise is present, developers recommend closing

doors and windows, as well as the sunroof and repeating the command. Similar recommendations are offered by systems with significantly greater computational capabilities. For example, if there is uncertainty in the identification of a word or command, the system may suggest repeating the word or command or selecting the desired command from a list compiled by the system of the most appropriate utterances after it has been recognised. It is clear that in order to remove noise more effectively, it is necessary to filter it using methods recommended for cleaning signals before recognising speech messages [1], which can also be used to remove noise in the pronunciation of voice commands.

3. The need to limit the number of commands, to a certain extent, contradicts the idea of creating a "friendly" user interface, since a "friendly" interface assumes the assignment and pronunciation of commands in a fairly arbitrary form, limited only by the rules of grammar of the Russian language. It is clear that in the absence of the Internet to implement an autonomous voice interface, including a system of command recognition and voice control, identifying commands in arbitrary form, is currently not possible. The only possible option may be a compromise between a freer than, for example, template form of command assignment and the minimum number of commands required, while the number of words in the commands should also be minimised. Limiting the number of words and the number of commands is necessary to reduce the stored reference base and to obtain acceptable performance of the voice control system.

4. The recently observed increase in the implemented voice interface functions leads to an increase in the number of commands and, as a rule, to an increase in the number of words in commands, i.e., to an increase in the complexity of commands. The increase in the volume of processed commands and their complexity inevitably leads to the expansion of the base of standards against which it is necessary to compare the commands when recognising them, the number of comparisons also increases. Therefore, the computational power used earlier may be insufficient for efficient operation of more modern voice interfaces. Consequently, an increase in performance and computing power is a natural, inevitable process, like upgrading a computer after a certain time of its operation and switching to a newer operating system.

The trend of creating a "user-friendly" interface has now been significantly developed and new features are being added to the voice interface, allowing the user to dictate commands in a more convenient form. It is possible for example, to allow the user to pronounce extraneous words together with commands, which is realised in modern software voice control systems of Windows OS by Microsoft. For example, it is possible to use the GARBAGE rule, which allows the user to match sequences of words in speech input that are not defined in a given grammar. Such a feature would allow users to speak additional words that are not identified as commands. For example, "gimme", "and", "ugh", "maybe", etc. The ability to recognise partial phrases contained in commands is also useful. In this case, the recognition of commands should be performed by keywords.

It should be noted that the task of voice control is not trivial and, despite the significant positive results obtained in recent years in the development of voice interfaces, for their stable and more reliable operation in accordance with the recommendations of Microsoft [2] must fulfil a number of conditions. Examples include the following:

- do not use words with similar sounds,
- use several limited grammars with a small number of words instead of one large grammar volume,
- if the grammar contains specialised terms with unfamiliar pronunciation, the variants should be identified to improve recognition accuracy,
- when necessary, training should be used, etc.

This indicates the need for continued work to improve voice control systems, especially autonomous systems that lack the ability to use the Internet in recognising commands,

where the probability of misidentification of commands may be greater. A known way to reduce errors, proven to be practically effective, is to increase the naturalness and ease of interaction with the voice control system.

In order to provide more convenient and natural interaction, it is advisable that the system does not require the user to memorise sequences of words in commands and accurately reproduce them when voicing. Thus, the use of rigid templates by recognition systems when constructing commands consisting of several words does not allow in a number of cases to reliably identify commands in which the correct word sequences are violated in accordance with the rules of the Russian language, which leads to the need for the user to memorise and accurately reproduce word sequences in complex commands. An attempt to violate the required sequence of words in commands, i.e. their rearrangement, often leads to a violation of the reliability of recognising complex commands. Such disadvantage, of course, are deprived of recognition systems that use complex models of the Russian language, and a fairly free (not template) construction of phrases, which are focused on the recognition of fused speech, but they solve more complex problems than the identification of commands, use bases of significant volumes and powerful processing facilities (both software and hardware, including special servers) with high speed. For solving a number of practical tasks it is inexpedient and impossible to use such equipment for many reasons, including power consumption, size, weight and cost. Thus, when solving practical tasks, it is useful to use a system for recognising complex (but limited number of) commands that allows word permutations in them and does not require increased computational capabilities and special server equipment for implementation.

The application of modern voice recognition technologies is based on complex computations and places high demands on the resources of computing devices. Therefore, developers of mobile devices began to move the main computations, including speech recognition, to Internet servers. In this case, applications on users' local devices sent fragments of speech via the Internet and after their processing on servers received the results of their recognition. This approach has yielded positive results in mobile phones and is currently actively used by Apple for Siri and Google for Voice Search [3, 4]. The complexity of automatic speech recognition under severe limitations of computational resources is also noted in [5].

A variety of voice-activated devices have been successfully used in medicine for a relatively long time to facilitate the lives of disabled people, people with limited physical capabilities, patients after major operations and traumas, and those in medical institutions for rehabilitation, as well as elderly people. Considerable attention, both in our country and abroad, has recently been paid to the safety and development of automated life support systems due to the increase in elderly people and physical disabilities [6]. Since this problem is more urgent abroad, they are actively developing systems that provide not only safety, but also independence of sedentary people and create more comfortable living conditions for them [7]. When developing such systems that automate household operations, they try to create more user-friendly interfaces and install them, including standard household equipment, more often using voice control.

When developing modern voice interfaces, they are endowed with the ability to distinguish extraneous sounds, knocks and noises from speech commands, and some security-oriented sound processing systems control the occurrence of emergency situations, such as a fall of a person, knocks and noises in case of burglary, fire, leaks, etc. Thus, in [8] a system for detecting the sound of a person falling, based on the method of three-dimensional localisation of the sound source is described. Thus, modern voice control systems under development should be able to distinguish voice commands from extraneous noises and sounds, for recognising which, if necessary, systems with other purposes are used. In addition, they must be able to distinguish between words that sound like

commands spoken during a conversation, for example, and the commands themselves, which is an even more difficult task. Although a simple solution may be to utter special control words that initialise the voice control systems before issuing commands.

The presence of a significant amount of literature devoted to the creation of local voice recognition devices, currently developed in the form of stand-alone, software implemented devices operating without access to the Internet, and in the form of stand-alone electronic recognition boards, indicates a significant relevance of the development of voice recognition and voice control systems. An example of a popular, open source and free software product that is used to implement voice recognition and voice control systems is the rather large CMU Sphinx project [9], which is capable of creating its own dictionary and grammar in the form of a sequence of words that form commands in voice control. However, the recognition quality it provides depends on the correct spelling of the transcription of the words included in the commands. Its disadvantages also include the difficulty of Russifying the acoustic model, as well as the complexity of its installation. In addition to software implementations, there are currently available on the market ready-made solutions in the form of autonomous and functionally complete electronic devices, such as Voice Recognition Module (VRM) [10]. However, the device requires pre-training in the form of recording up to 80 voice commands, but at the moment of using the device it is able to recognise and process only seven commands out of 80, which indicates the limited size of its vocabulary.

It is known that hardware implementation, i.e. implementation in the form of independent, autonomous electronic devices allows to increase the speed of computation. Therefore, in the hardware implementation of voice control and speech recognition devices formed at least two directions: the implementation in the form of electronic devices of individual functions of the recognition system, such as algorithms for the selection of speech features in the form of small-kepstral coefficients on programmable logic chips (PLC) [11, 12] and full hardware implementation in the form of electronic devices and blocks of recognition algorithms. Another reason for the use of hardware implementation is the complexity of realisable algorithms such as hidden Markov model algorithms, forward algorithms and Viterbi algorithms. At the same time, if it is necessary to use them in the case of recognition of a large number of commands and continuous speech, the general idea of simplification of computations during recognition can be implemented in relation to these algorithms [13], as well as their parts.

It should be noted that the need to create compact devices that perform speech command recognition without access to the Internet has led to the creation of such systems not only for civilian but also for military purposes. Thus, the Israeli concern Israel Aerospace Industries has developed the REX robot, which is controlled only by voice commands [14], following the military man who controls it, and can transport ammunition, military equipment, food and other inventory.

3 Results

Voice control systems improvement is connected with the expansion of the realisable number of functions, i.e., the use of additional functions, such as rearrangement of words in commands, use of keywords, by which recognition is performed, allowing the presence of extraneous words and sounds in commands, performing recognition of a command by individual words included in it or a set of words, allowing partial matching. This will make it possible to abandon the use of previously popular rigid templates, allowing the user to pronounce commands in a more natural, free and convenient for him form without prior memorisation, i.e., will provide more "friendly" voice interface. It is important, however, that the voice control device be compact and self-contained, as noted earlier. The need for a

compact, self-contained, autonomous voice-controlled device with sufficiently high speed to reliably recognise speech commands has given rise to a search for methods to increase the speed of command identification without significantly reducing the quality of recognition. One of the known ways to increase the speed was to reduce the complexity and amount of computation, i.e., to simplify computational procedures. However, the need to increase the "friendliness" and expand the functions of the voice interface on the one hand, and the need to ensure the required reliability of recognition on the other, do not allow simplification to abandon the methods that significantly determine this required reliability. Thus, all simplifications and adjustments of methods made for this purpose must not degrade the performance of voice control systems and reduce the quality of recognition. A compromise solution in this situation may be to reduce the number of commands performing similar functions, which will reduce the number of possible, including unlikely sequences of words in commands and will allow to abandon complex methods of analysing such unlikely sequences. Thus, for example, there is no need to check all possible combinations of words in commands when recognising commands but to recognise individual words, and identify them as key words, which will allow to correctly recognise a large number of similar commands, even in the presence of extraneous words and sounds in them. Identification of words in commands should be performed sequentially, checking each word separately as a key word. In this case, a command should be considered recognised when all the words included in it are consistently identified. In a more complicated case, if there is a possibility of invalid sequences in a command in the form of a pair of words, it is possible to use a matrix that considers only valid combinations of words and generates an error message otherwise. Such a simplified approach will eliminate the use of often used in speech and command recognition, but more complex hidden Markov models (HMM) [15, 16], which require higher speed and longer computation time. When using HMM, the speech signal is represented as a probabilistic process taking into account time-varying [17]. The language phonemes that make up words are modelled by HMM state chains. Then each command word can be represented by a separate HMM, or a larger HMM can be used to represent a command consisting of several words. To obtain higher quality and accuracy of recognition with a significant number of commands and words included in them, more complex structures of speech recognition systems can be used, in which neural networks are used together with HMM [18, 19]. However, their implementation requires powerful computers with large amounts of memory, which are currently very difficult to implement in the form of compact, stand-alone devices without Internet connection. It should be noted that in order to ensure the reliability of recognition, the correctness and appropriateness of simplifications of methods, algorithms and their specific implementations should be tested experimentally on the full set of commands of the voice control system.

4 Discussion

Rejecting complex methods and using simplifying techniques will allow reducing the total number of commands recognised and used as reference standards and performing voice command recognition on local computers in stand-alone mode without accessing powerful servers via the Internet. It should be noted that at the present time in the recognition of commands as benchmarks are used not the whole commands and not the words included in the commands, as before in the first systems but sounds and their characteristic features and coefficients, which are often called acoustic models. It is known that the speech signal is most accurately represented in the frequency domain [20] by features in the form of linear prediction coefficients and mel-frequency cepstral coefficients (MFCC). Linear prediction methods are based on calculating the features or coefficients of the current sample of the

speech signal based on a linear combination of the features of previous samples. Using the MFCC method, the signal characteristics of the spoken sound are extracted [21]. For this purpose, the logarithm of the spectrum of the input signal is calculated in the frequency domain, to which either the inverse discrete Fourier transform or the discrete cosine transform is further applied. Before calculating the logarithm for modelling the human perception of a speech signal, the spectrum of the input signal is processed by a set of mel-filters with bandwidths selected on a mel-scale, the resolution of the filters deteriorating when moving to higher frequencies. Since with similar, and in some cases better recognition quality, the MFCC method is implemented simpler than the linear prediction method, and the speed of the MFCC method is ensured by the use of fast Fourier transform in the calculation of both forward and inverse discrete Fourier transforms, so, as a rule, the MFCC method is used more often.

It should be noted that it is reasonable to use simplifying corrections for a significant number of solved problems, because the procedure of speech recognition with high reliability without the use of techniques to reduce computational complexity, simplifying and often eliminating individual processing steps, is a complex process. This process in its entirety usually involves the use of [22]: speech signal processing and feature extraction tools, an acoustic model, a language model and a decoder in the sequence shown in Figure 1.

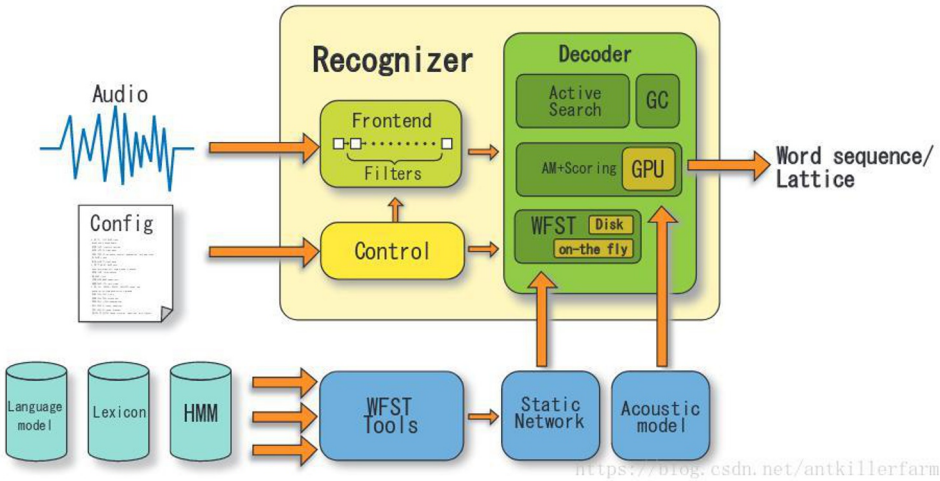

Fig. 1. General structure of speech recognition, *Source: https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/39654/1/MP-SS1-4.pdf*

5 Conclusions

Consequently, when expanding the functionality of the developed modern, compact, and autonomous voice control system, it is necessary to be guided by the following recommendations:

- to use for realisation methods and algorithms that provide recognition of voice commands without connection to the Internet, i.e., work in autonomous mode;
- to give preference to compact variants of device realisation when designing and creating;
- to apply methods of extending functional capabilities, for this purpose:

a) to allow rearrangement of words in commands,

b) to include in the commands key words, by which to perform recognition, allowing the presence of extraneous words and sounds in the commands,

c) to perform recognition by individual words included in the command, to consider the command identified after recognising all words of the command, regardless of the sequence of recognition, and in the presence of inadmissible sequences of words in the commands to use the recognition matrix;

- to use simpler methods of calculations, if possible, as well as to apply methods of reducing the computational complexity of the procedures of processing, recognition and identification of voice commands, using simplifying changes in the implemented methods and algorithms, for this purpose:

a) to reduce the number of similar-sounding words and the commands in which they are included and, if possible, to reduce the total number of commands,

b) to use special control words for initialisation of the voice control system, if necessary;

- to use methods and means of reducing and suppressing noise during the pronunciation of commands;

- to apply methods, algorithms and hardware implementations on electronic components, providing high speed of computational operations.

The feasibility of using the methods and recommendations formulated and listed in this article in the implementation of autonomous voice control systems has been tested and confirmed experimentally, and some of them, as noted above, correspond to the recommendations of leading firms in the field of development and creation of voice interfaces and voice control systems such as Microsoft, Google, Yandex, Apple.

In conclusion, it should be noted that the set of methods and tools proposed in this paper is determined by the problem to be solved and is not the only one, however, the proposed approach, including the recommendations described herein for the correction of methods and algorithms, tested practically, has proved their validity and can be used in the development and implementation of autonomous voice control systems and voice interfaces, so they are of practical interest.

References

1. V. Zhigalov, E3S Web of Conf. **381**, 10 (2023) https://doi.org/10.1051/e3sconf/202338101024
2. Microsoft documentation. https://learn.microsoft.com/ru-ru/windows/mixed-reality/design/voice-input
3. M. Pinola, Speech recognition through the decades: how we ended up with Siri // PCWorld (2011).
4. B. Johnson, How Siri works, How Stuff Works. URL: http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/siri2.htm.
5. A. Schmitt, D. Zaykovskiy, W. Minker, Int. J. of Speech Tech. **11**, 63-72 (2008)
6. F. Portet, et al. ,Personal and Ubiquitous Computing **32(1),** 1–18 (2011)
7. M. Chan, E. Campo, D. Esteve, J. Fourniols, Maturitas 64, 2, 90–97 (2009).
8. M. Popescu, Y., Li M. Skubic, M. Rantz, *An acoustic fall detector system that uses sound height information to reduce the false alarm rate*, Proc. of 30th Annual Intern. IEEE EMBS Conf., p. 4628–4631 (2008)
9. Overview of the CMUSphinx toolkit. cmusphinx.github.io: Open-source speech recognition toolkit resource. https://cmusphinx.github.io/wiki/tutorialoverview/ (access date: 29.06.2020).

10. Introduction To Voice Recognition With Elechouse V3 and Arduino. instructables.com: Place that lets you explore, document and share your creations: http:// www.instructables.com/id/Introduction-to-Voice-Recognition-With-Elechouse-V/ (access date: 28.06.2020)
11. V. V. Ngoc, J. Whittington, J. Devlin, *Real-time Hardware Feature Extraction with Embedded Signal Enhancement for Automatic Speech Recognition* Speech Technologies, Intech, pp. 29-54 (2011)
12. M. Bahoura, H. Ezzaidi, *Hardware implementation of MFCC feature extraction for respiratory sounds analysis* 8th Workshop on Systems, Signal Processing and their Applications, pp. 226-229 (2013)
13. Mosleh M., Setayeshi S., Mehdi Lotfinejad M., Mirshekari A., *FPGA implementation of a linear systolic array for speech recognition based on HMM* The 2nd International Conference on Computer and Automation Engineering (ICCAE) **3**, 75-78 (2010)
14. A. Egozi, IAI's REX robot to face first operational trial (Israel Defence, 2012). URL: http://www.israeldefense.com/?CategoryID=411&ArticleID=1515.
15. Y. A. Alotaibi, International Journal of Speech Technology **15**, 1, 25-32 (2011)
16. I. Patel, Y.S. Rao, *Speech recognition using hidden markov model with MFCC-subband technique* International Conference on Recent Trends in Information, Telecommunication and Computing, pp. 168-172 (2010)
17. M.A. Anusuya, S.K. Katti, Int. J. of Computer Science and Inf. Security **6,** 3, 181-205 (2009)
18. Dahl G.E., Yu D., Deng L., Acero A., IEEE Transactions on Audio, Speech, and Language Processing **20**, 1, 30-42 (2012)
19. Mohamed A., Dahl G.E., Hinton G., IEEE Transactions on Audio, Speech, and Language Processing **20**, 1, 14-22 (2012)
20. M.A. Anusuya, S.K. Katti, International Journal of Speech Technology, **14**, 99-145 (2011)
21. Jacob Benesty, M. Mohan Sondhi, Yiteng Huang, Springer handbook of speech processing (Springer, 2008)
22. Dixon, Paul R.; Oonishi, Tasuku; Iwano, Koji; Furui, Sadaoki, Recent Development of WFST-Based Speech Recognition Decoder // : APSIPA ASC 2009 : Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, 138-147.