

Speech Technologies Market in Azerbaijan

Ali Abbasov
National Academy of Sciences

Rauf Fatullayev
Voicedocs Software GmbH

Abulfat Fatullayev
Institute of Control Systems

Speech technology as a direction of artificial intelligence (AI) helps to increase labor productivity, free a person from tedious, monotonous and sometimes hazardous to health work, and in some cases ensures transparency in the activities of state, municipal and business structures. The purpose of this article is to study the speech technology market in Azerbaijan. This market still remains unexplored and does not attract well-known companies of relevant software products. Our goal is to attract the attention of such manufacturers, on the one hand, and, on the other hand, to bring to the attention of state, municipal and business structures the benefits of using AI in their activities.

Keywords: speech technologies, automatic speech recognition, speech analysis, speech-to-text transcription, speech technologies market, call-centre, the Azerbaijani language

INTRODUCTION

Speech is the main form of communication between people, and the use of human speech for communication between a person and a machine is one of the human natural dreams.

The emergence and development of speech technologies (ST) is connected precisely with the realization of this dream: the development of computer programs that would allow people to communicate with different devices (computers, cars, home appliances, etc.) in a natural language.

Beginning in the first half of the 20th century, scientists and engineers from Bell Laboratories (USA) proposed a model for speech analysis and synthesis (Dudley, 1939; Dudley, et al, 1939), and in the early 1950s Bell Laboratories developed the "Audrey" system that recognized numbers. After 10 years, IBM demonstrated the "Shoobox" system, which understood only 16 words¹.

In the late 60s of the last century, an approach to speech recognition based on dynamic programming was developed (Vinitzyuk, 1968). In the 70s of the 20th century, speech recognition systems made a big leap with the sponsorship from the US Department of Defense. The DARPA project was one of the largest in the history of speech recognition. The "Harpy" system created within the framework of the project understood about a thousand words.

Since the 80s of the last (20th) century, a new statistical method based on Hidden Markov Models (HMM) had been proposed (Rabiner, 1989). The use of HMM had raised the accuracy of speech recognition to a new level.

Since the mid-2000s, with the advent of computers based on powerful processors and GPUs (Graphic Processor Units), Deep Learning has gained popularity.

The use of artificial neural networks in speech recognition (one of the most important directions of ST) has helped to obtain even greater accuracy in recognition, and this approach, at the moment, is considered the most promising. Artificial neural networks, in addition to speech recognition, are successfully used in solving such complex problems that did not succumb to an effective solution earlier, such as speech understanding, speaker identification, computer vision, machine translation, face recognition, etc. (Ciresan, et al, 2012).

Currently ST are widely used in automotive, healthcare, banking, financial services, insurance, education, government & public services, legal, consumer electronics, military & defense, etc.

GLOBAL ST MARKET

The most common speech technologies of our time are speech analytics, speech recognition, speaker identification, intelligent virtual assistant, media monitoring, medical transcription, voice biometrics and many others. Below we have provided a forecast of the development of some popular speech technologies in the near future.

In subsequent years, the global *Speech Analytics* market size to grow from USD 1.5 billion in 2020 to USD 3.8 billion by 2025, at a CAGR of 20.2%³.

According to the market research report "*Speech and Voice Recognition Market by Technology Global Forecast to 2025*", the market is expected to grow from USD 7.5 billion in 2018 to USD 24.9 billion by 2025, at a CAGR of 18.72%⁴.

Another fast-growing speech technology market is the *Intelligent Virtual Assistant* (such as Apple's Siri). The global intelligent virtual assistant market size is expected to reach USD 45.1 billion by 2027, expanding at a CAGR of 34.0%, according to a new report by Grand View Research, Inc.⁵

According to a new market report pertaining to the Global *Media Monitoring* tools market published by transparency market research, the global media monitoring tools market is projected to reach **USD 12,0 billion** by 2030. The media monitoring tools market is projected to expand at a CAGR of **14.1%** from 2020 to 2030⁶.

The global *medical transcription*⁷ software market size was valued at USD 1.32 billion at 2019 and is projected to reach USD 4.89 billion by 2027, exhibiting a CAGR of 17.8% during the forecast period⁸.

It can be stated that only on these directions the speech technology market will reach about USD 90 billion.

ACTORS OF THE AZERBAIJANI SPEECH TECHNOLOGIES MARKET

The Azerbaijani information technology market has its own characteristics. The Azerbaijani language is the state language of the Republic of Azerbaijan, and all internal document circulation is conducted in this language. Therefore, software products (including software based on speech technologies) supplied for the Azerbaijani market must support this language. In recent years, software manufacturers have localized their products into Azerbaijani (for example, Microsoft supplies all the latest versions of Windows with the Azerbaijani localization).

Although Azerbaijani language is also available in dictation and transcription applications of some market giants (for example, Google), they are not popular in Azerbaijan due to security reasons. In private conversations with many representatives of government and business structures (for whom it is very profitable to use speech applications), almost everyone confirms that speech technologies, where there is a need for their use, increase labor productivity several times.

However, they do not want their information to leave the country, which is inevitable when using "global" applications (information is processed outside the country). Everyone wants to have their own servers for storing and processing their own information. Another reason for the unpopularity of speech technologies was called recognition inaccuracy.

For these and other reasons ST is still a novelty for Azerbaijani state and private organizations (Fatullayev, et al. 2020).

However, the computerization of society sooner or later leads to the use of more and more new technologies (including speech technologies). The following table (Table 1) shows the "speed" of growth of computerization in Azerbaijan in 2015-2019.

Although the number of computers per thousand inhabitants of the country is much lower than the average (159), 741 out of a thousand inhabitants are computer users. Considering that, on average, every resident has a smartphone, we can say that the computers have firmly entered life in Azerbaijan.

TABLE 1
THE NUMBER OF COMPUTERS IN ENTERPRISES AND HOUSEHOLDS⁹

№	Name	2019	2018	2017	2016	2015
1	Number of computers in enterprises (thousand pieces)	269.7	265.6	263.2	261.4	260.3
2	Number of computers in households (thousand pieces)	1578.5	1557.9	1 537.4	1 516.2	1 493.8
3	Number of computers per 100 people	15.9	15.8	15.7	15.5	15.4
4	Number of computer users per 100 people	74.1	72.5	71,7	70,8	69,9
5	Number of mobile phone subscribers (thousand)	10750.3	10339.7	10127.0	10189.0	10697.3

How Does This Fact Affect the Spread of Speech Technologies?

In significant cases, people use computers to enter and correct text documents, and ASR speeds up and greatly simplifies the process of entering and correcting documents. The increase in the number of computers and mobile phones increases the number of potential users of ASR apps (the law of the dialectic "Transition of quantity into quality").

From this point of view, office workers of all government and business structures in the country are potential users of speech technologies.

Popular applications among the users who work a lot with documents are the dictation and transcription applications. The dictation app allows users to enter text to the computer without using a keyboard, and the transcription app allows users to convert the speech recorded beforehand (audio files) to text.

Traditionally, the main users of the dictation and transcription programs are journalists, physicians, court workers, advocates, investigators, students, and other categories of people who spend a lot of time at the computer, preparing various documents. The dictation and transcription apps increase the productivity of such employers three times.

TABLE 2
NUMBER OF POTENTIAL CUSTOMERS FOR ASR IN AZERBAIJAN

Target group	Number	Link
Journalists (Media)	25000 (5343)	https://azvision.az/news/223983/--azerbaycanda-kiv-lerin-sayi-aciqlandi--.html
Physicians and paramedical workers	89500	https://www.stat.gov.az/source/healthcare/
Courts (Number of court cases in 2020)	566 (12422)	https://courts.gov.az/en/main/judges
Advocates	1844	https://report.az/daxili-siyaset/azerbaycanda-vekillerin-sayi-aciqlanib11-012021/
Teachers at universities and colleges	20000	https://www.stat.gov.az/source/education/
Students	220000	https://www.stat.gov.az/source/education/
Total	356910	

According to the Statistical Office of the Republic of Azerbaijan, it is possible to approximately estimate the number of potential ASR users (data for 2018-2020) in Azerbaijan (Table 2).

Only six groups of the potential users are included in the Table 2 (Total >350 thousand users), but even this incomplete list of potential users indicates a sufficient vastness of the customer market for the speech technologies.

Adding the office employers of all state and private structures not listed in Table 2 to these numbers, it turns out that hundreds of thousands pages of documents are prepared per month in Azerbaijan.

Another important application that is in great need in Azerbaijan is speech analysis. Many state and private organizations have call centers. As elsewhere in the world, commodity or service producers understand how important it is to have feedback with customers. Based on public opinion, adjustments are made to the quality and list of the manufactured products (services), thereby increasing the competitiveness of the products (services).

In those organizations where there are large archives of telephone conversations for statistical processing of this archive, ST, in particular, speech analysis can be used.

According to the Decree No. 50 of the Cabinet of Ministers (February 25, 2015), the rules for the organization of Call centers were approved in order to (a) *timely meet the requests of citizens*, and (b) *ensure transparency and accessibility* of government agencies. This document states that in Call centers, *periodic monitoring should be carried out in order to measure and improve the quality of service, to identify and prevent recurrence of shortcomings in the future*¹⁰.

One of the popular Call centers in Azerbaijan is the Call Center (195-1 *Çağrı mərkəzi*) of the Ministry of Economy. According to the reports¹¹ for 2016-2020, the number of citizens who contacted the Call Center of the Ministry of Economy is shown in Table 1.

From Table 1 it turns out that for the previous 4 years (2016-2019) and for the ten months of the 2020 year (January-October 2020), a call database has already been formed in this call center with a duration of more than 100 thousand hours (Note the average duration of one call is about 4 minutes. Hereinafter we will use this average value).

TABLE 3
THE NUMBER OF CALLS TO THE CALL CENTER OF THE MINISTRY OF ECONOMY

Period	Calls (*1000)	Answered calls (*1000)	Unanswered calls (*1000)	Percentage of unanswered calls in all calls	Duration of talk time (min)	Total duration of talk time (*1000 hours)
1	2	3	4	5	6	7
2020 January- October	961.6	417.6	544.0	56.6	4.00	27.8
2019	674.2	377.0	297.2	44.1	4.5	28.3
2018	354.3	256.3	98.0	27.7	3.58	15.3
2017	414.6	312.5	102.1	24.6	3.60	18.8
2016	316.3	220.4	95.9	30.3	3.88	14.3
Total						104.5

What conclusions can be drawn from this table? First, the number of calls, with a small exception, increases from year to year (column 2).

Secondly, as the number of calls increases, so increases the number of unanswered calls too. In 2020, more than half of calls (56.6%) were not answered (column 4). This means that point "a" is not fulfilled for the majority of citizens who applied to the call center.

And finally, due to the increase in the number of answered calls, the possibility of "manual" analysis of telephone calls is lost and therefore points "b" and "c" are also violated.

That is, call centers do not fulfill their purpose. To solve their problems, call centers must keep numerous staff, expensive equipment and create appropriate working conditions for the staff.

Example: «Since August 2018, the Call Center has been operating on a new office with a total area of about 2000 square meters. The center is equipped with tables, computers, the necessary software..., headphones with a silencer function, a microphone and other modern technical and other equipment, The Call Center has launched the Cisco UCCE (Cisco Unified Contact Center Enterprise) IP telephony system, which allows applying the latest innovations in the organization of services in this area. In total, 90 employees and 10 team leaders will work in the office ... »¹²

Hundreds of thousands of USD are needed per year to maintain such a call center. Using Intellectual Virtual Assistant together with Speech Analysis will reduce these costs at least several times.

The Call Center (146) of the Ministry of Education, according to the most conservative estimates for 2014-2018, has already a database of about 600 thousand phone calls¹³ (about 40 thousand hours).

And according to the "ASAN SERVICE" message¹⁴ in April 2019, the Call Center (108) received the 3 millionth call. It can be assumed that the "ASAN SERVICE" archive of telephone conversations reached and exceeded 200 thousand hours.

The Call Center (155) of the Ministry of TCHT handles about 100 thousand calls every month¹⁵. Then, from the opening date (June 2018) until the end of 2019 (18 months), the volume of telephone conversations will reach 120 thousand hours.

In call centers of other organizations, for example, the General Prosecutor's Office (961), the Ministry of the Interior Affairs (102), etc. of course, there are also such large archives of telephone call records, and the size of these archives is increasing every day.

Mobile operators (Azercell, Bakcell and Nar), banks and other financial institutions that are very interested in the use of speech technology can be added to this list.

"Manual" processing of such databases is an impossible task. The only solution is to use ST: to

transcribe calls into text and provide the necessary statistical studies.

But all statistical analysis systems work only with data in text format; these systems do not have the ability to analyze audio/video files "directly". First, audio files of the telephone conversations must be transcribed to text.

Thus, we can say that the main potential actors in the Azerbaijani speech technology market are mainly large government and business structures.

CONCLUSIONS

The research gives only an approximate idea of the volume of the Azerbaijani speech technologies market. Lack or scarcity of official sources do not make it possible to comprehensively research the market and make accurate forecasts.

But as we have seen in several examples, the use of speech technology increases labor productivity, quality of service, and it significantly reduces the cost of maintaining the organization itself.

We have identified several reasons for the lack of development of the Azerbaijani speech technology market:

- a) lack of support for the Azerbaijani language (Fatullayev, et.al, 2008-2010),
- b) lack of guarantees for information security, and
- c) low accuracy of speech recognition.

Solving the first two problems is not difficult, because the process of developing speech technologies for the Azerbaijani language is not much different from the development process for other languages, and the information security problem can be avoided by creating local versions of the applications.

As for the quality of speech technologies, despite the development of new and more accurate methods of speech recognition, the quality of recognition still lags behind the recognition accuracy of the "recognizer number 1" - the human brain. This is a global problem.

But, with the joint efforts of scientists and developers from all over the world, the problem of improving the quality of speech recognition is being solved step by step.

The boom in speech recognition applications indicates that the time for speech recognition has come and we can expect a huge number of useful speech-based applications in the near future.

The main conclusion of this paper: The Azerbaijani market needs speech technologies and is ready to consider relevant proposals.

ENDNOTES

- 1 https://www.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html
- 2 In this paper transcription means converting audio/video files into text
- 3 Speech and Voice Recognition Market - Global Opportunity Analysis and Industry Forecasts (2019-2025).
- 4 <https://www.meticulousresearch.com/product/speech-and-voice-recognition-market-5038>
- 5 <http://www.marketsandmarkets.com/PressReleases/speech-voice-recognition.asp>
- 6 <http://www.grandviewresearch.com/press-release/global-intelligent-virtual-assistant-industry>
- 7 <https://www.transparencymarketresearch.com/pressrelease/media-monitoring-tools-market.htm>
- 8 <https://www.fortunebusinessinsights.com/industry-reports/medical-transcription-software-market-101572>
- 9 <https://www.azstat.org/portal/tblInfo/TblInfoList.do#>
- 10 <http://www.e-qanun.az/framework/29632>
- 11 <https://www.taxes.gov.az/az/page/hesabat>
- 12 <http://www.vxsida.gov.az/media/shares/12compressed.pdf>
- 13 <https://versus.az/article/az/4694>
- 14 <https://asan.gov.az/az/media/press-releases/asan-xidmet-in-108-cagri-merkezine-3-milyonuncu-zeng>
- 15 <http://www.mincom.gov.az/az/view/news/326/neqliyyat-rabite-ve-yuksektexnologiyalar-nazirliyinin-vahid-chagri-merkezi-yaradildi-155>

REFERENCES

- Ciresan, D., Meier, U., & Schmidhuber, J. (2012, June). Multi-column deep neural networks for image classification. *IEEE Conference on Computer Vision and Pattern Recognition Journal*, pp. 3642–3649.
- Dudley, H. (1939). The Vocoder. *Bell Labs Record*, 17, 122–126.
- Dudley, H., Riesz, R.R., & Watkins, S.A (1939). A Synthetic Speaker. *J. Franklin Institute*, 227, 739–764.
- Fatullayev A., & Ibrahimov, B. (2009). Analysis and choice of sampling frequency for preliminary processing of speech signals. Digital signal processing. *Proceedings of the 11th International Conference “DSP and its applications”*, pp. 237–240. Moscow.
- Fatullayev, R., Abbasov, A., & Fatullayev, A. (2008, November 20–21). “Dilmanc” is the 1st MT system for Azerbaijani. *Proceedings of SLTC 2008: Swedish Language Technology Conference*, pp. 63–64. Stockholm, Sweden.
- Fatullayev, R., Abbasov, A., & Fatullayev, A. (2010, September 6-8). HMM-based large vocabulary continuous speech recognition system for Azerbaijani. *The 3rd International Conference “Problems of Cybernetics and Informatics”*, pp. 23–26. Baku, Azerbaijan.
- Rabiner, L. (1989, February). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognitions. *Proceedings of the IEEE*, 77(2), 257–289.
- Vintsyuk, T.K. (1968, January–February). Speech Discrimination by Dynamic Programming. *Kibernetika*, 4(2), 81–88.