DOI: 10.56652/ejmss2021.1.6



*Anna Maria Kamińska*[1] (ORCID 0000-0002-6638-1155), Łukasz Radliński[2]
[1] *Wrocław University of Science and Technology, Faculty of Computer Science and Management, 27 Wybrzeże Wyspiańskiego st., 50-370 Wroclaw, Poland, e-mail: anna.maria.kaminska@pwr.edu.pl, phone: +48 697 53 22 76*
[2] *Wrocław University of Science and Technology, Faculty of Computer Science and Management, 27 Wybrzeże Wyspiańskiego st., 50-370 Wroclaw, Poland, e-mail: 254266@student.pwr.edu.pl, phone: +48 511 604 077*

# The use of speech recognition techniques as a new method of automating communication management in an organization

## Wykorzystanie technik rozpoznawania mowy jako nowej metody automatyzacji zarządzania komunikacją w organizacji

### ABSTRACT

Nowadays, when stationary work has been changed into remote work, the efficiency of it is becoming more and more important. There are many tools to increase it, but all of them offer limited services. What if we could offer entrepreneurs automated real-time creation of documents such as reports, or records of meetings? Such solutions are already present in the market, but are related to a large margin of error. Building a tool that will "teach itself" is now a challenge. The aim of the article is to outline the concept of a method of automating communication management in an organization through the use of speech recognition learning tools. The article will discuss the equality between the traditional form of creating documents and the automated one, as well as the idea and principles of the speech recognition system. This article should be seen as an introduction to the issue of speech recognition. The authors have started work on a tool that will automate the reporting of meetings in the future.

**Keywords:** speech recognition techniques, communication management, communication ecosystem.

### STRESZCZENIE

W dzisiejszych czasach, gdy praca stacjonarna zamieniła się w pracę zdalną, jej efektywność nabiera coraz większego znaczenia. Istnieje wiele narzędzi do jej zwiększenia, ale wszystkie oferują ograniczone usługi. Co by było, gdybyśmy mogli zaoferować przedsiębiorcom zautomatyzowane tworzenie w czasie rzeczywistym dokumentów takich jak raporty czy protokoły? Takie rozwiązania są już obecne na rynku, ale są obarczone dużym marginesem błędu. Stworzenie narzędzia, które „samo się nauczy", jest teraz wyzwaniem. Celem artykułu jest zarysowanie koncepcji metody automatyzacji zarządzania komunikacją w organizacji poprzez wykorzystanie narzędzi do nauki rozpoznawania mowy. W artykule porównano tradycyjną formę tworzenia dokumentów ze zautomatyzowaną, a także omówiono ideę i zasady systemu rozpoznawania mowy. Artykuł należy traktować jako wprowadzenie do zagadnienia rozpoznawania mowy. Autorzy rozpoczęli prace nad narzędziem, które w przyszłości zautomatyzuje raportowanie spotkań.

**Słowa kluczowe:** techniki rozpoznawania mowy, zarządzanie komunikacją, ekosystem komunikacji.

### INTRODUCTION

In the last decade there has been a huge rise in the usage and acceptance of digital technologies. Some may even argue that the COVID 19 pandemic has speeded the adoption of such technologies by several years (Seiler, 2020). This effect is particularly strong in relation to digital communication technologies whose importance had been rising even before the pandemic. According to research Nguyen et al. (2020) 43% of respondents used text messaging more often. This was followed by an increase in voice calls (36%), social media (35%), and video calls (30%).

otrzymano / received: 24.05.2021 | poprawiono / corrected: 09.06.2021 | zaakceptowano / accepted: 18.06.2021

DOI: 10.56652/ejmss2021.1.6

As the usage of digital communication technologies in companies increases, there is an ongoing necessity of creating documentation of digital meetings. Creating such documentation in real-time is barely possible unless there is a specific person assigned to such a task and even then the quality of such documentation is questionable and usually requires verification with recorded meeting. However when we consider usage of voice recognition and text processing technologies, the possibilities of creating documentation are vastly different. Such techniques enable creating real-time documentation of any meeting, conference or job interview. Text processing techniques provide tools for such documentation to be adjusted for the needs of the company i.eg. such documentation can be automatically divided or formatted into templates. That means that the human role in such a process could be only to verify documentation, which vastly reduces the time necessary to prepare such documentation. Speech to text recognition programs are also a very effective way of speeding up the process of creating documentation by humans as most people can speak faster than they can write. Therefore voice recognition and text processing enables not only automating the process but also speeding it up whenever there is the necessity for humans to do it.

**The meaning of communication in an organization**

In the context of organization, we can meet the concept of the so-called *communication ecosystem*. The perfect *communication ecosystem* means that the "right" messages will be communicated: downwards, upwards, and across the organization. The aim is to provide needed information, and that would result in right decisions at the right time. It is important to create and maintain the organization's culture and structures that will support the communication ecosystem by encouraging open and authentic exchange of information within a supportive framework.

There are several roles that are necessary for the effective functioning of the *communication ecosystem* (Figure 1). To have a successful organization it is very important to have responsible "governors", that are developing the vision, mission, and strategy of the organization and communicate it in an appropriate way to the people in the organization. These strategic messages should include: processes and instructions (to assist compliance), basic reporting mechanisms (to provide direction for the organization's management). Different levels of management should efficiently convey specific strategic guidance to lower levels and assure the governing body that the vision, mission and strategy are understood and met. Then, the management supervises and coordinates the activities of people operating in the operational areas (business as usual BAU) and project and program management (PPM). Funded and approved project management activities ensure changes to the BAU. Downstream communication to PPM takes the form of presenting the requirements, goals, and performance standards necessary to implement strategic changes necessary to generate value for the organization. Another element of a perfect communication system is communication to BAU management, which is based on goals, information on results, problems and opportunities. The upstream communication consists of information about the progress against the project objectives and requirements, and all necessary information to ensure that the work is proceeding as expected by the management and executive groups. Communication between the PPM area and the BAU area is focused on effective working relations as well as the transition from PPM to BAU operations. All information about the requirements of stakeholders, activities related to change management, resource conflicts, reporting as well as the realization of benefits are the basic communication in this area. The key to the success of the project is an

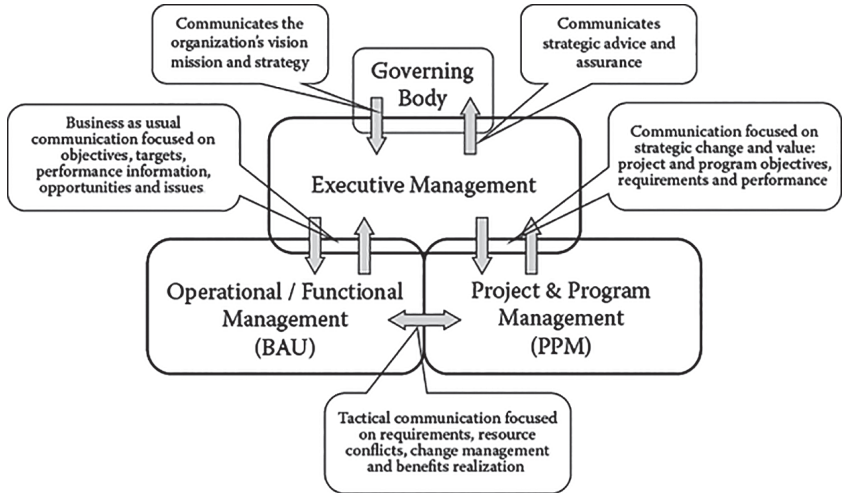


*Figure 1. Roles in the effective functioning communication ecosystem*
*Source: Bourne 2015*

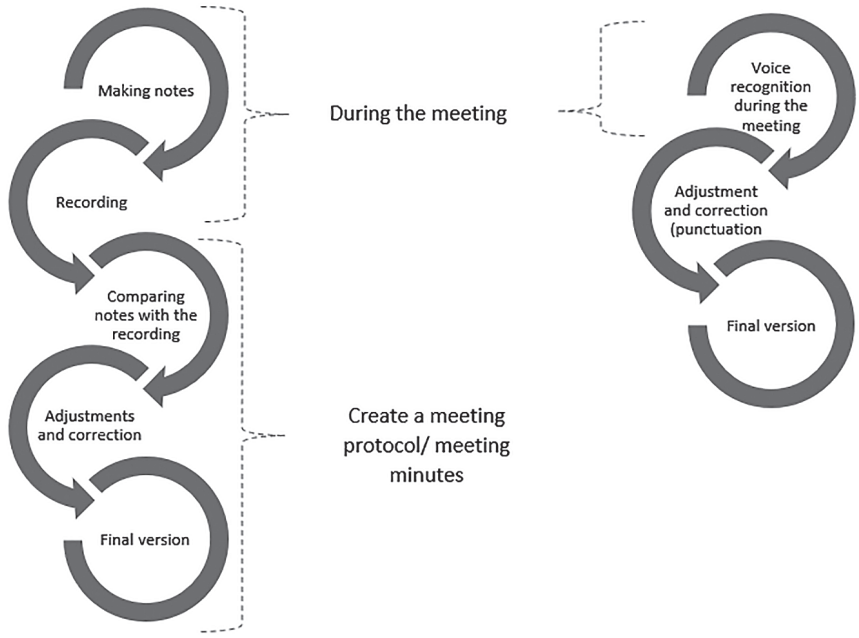


*Figure 2. Process of creating reports/minutes of the meeting: traditional (left), by using STRS (right)*
*Source: Authors' construction*

understanding of how the ecosystem works, which allows for the effective implementation of the organization's vision, mission and strategy, and thus effective operation within it (Bourne, 2015).

The problem of contemporary organizational and communication environments is the gaps in structures and culture, which are crucial for building the described *communication ecosystem*. It is possible to fill these gaps for the improvement of corporate governance structures and processes and strategic project management. Further improvements can be achieved by redefining the boundaries of the project itself and thus redefining the responsibilities of the project manager.

A very important aspect of implementing a voice recognition system into the organization is the significant reduction of the time required to prepare reports and minutes of meetings. This aspect can be crucial for everyday operation in an organization, especially in hi-tech markets, where fast information flow can be the key success factor. Basically the process of creating reports/minutes of the meeting contains two general parts (Figure 2): the first one, during the meeting - creating notes and recording the meeting, and the second one, after the meeting – comparing notes with the recording, making adjustments of the content of the notes and correction, and punctuation and then final validation.

This process lasts a long time, because it requires listening to the recording one more time (or a few times). Additionally, if the person that is making notes is not the stenographer, notes are just an overall view from the meeting. The solution that allows the creation of a report or meeting minutes in real time is using the Speech to Text Recognition System (STRS) (detailed description of the system will be explained in the next chapter of this work).

The desirability of working on speech recognition finds its source in the need to facilitate management and communication in the organization. When such a prosaic matter as preparing minutes of meetings or creating documents by saying them aloud is available to managers, and they do not have to focus on text correction: breaking sentences, changing letters from lower to uppercase or using commas, they focus on work. They get a tool that is their personal trusted assistant to create even and confidential documents. Additionally, the discussed system is a learning system, which means that along with the development of the competences of a given person, the vocabulary and "skills" of the speech recognition system also increase.

## 1. THE TECHNOLOGY AND ITS APPLICATIONS

Over the last few years the field of sound processing and speech processing in particular has greatly expanded as more and more uses for speech recognition can be found. Though commercial usage of speech recognition is commonly known, over the last year speech recognition techniques have got commonly adapted to such fields as e.g. radiology where the field has been greatly developed. "This development coupled with increasing processing power lead to improved accuracy rates and the easier use of natural speech." (Hirschorn & Horii, 2006)

Voice recognition and STRS can be implemented in various different ways with each of them having its pros and cons. The approach analyzed in this article is based on the idea of creating an individual vocabulary base and teaching

DOI: 10.56652/ejmss2021.1.6

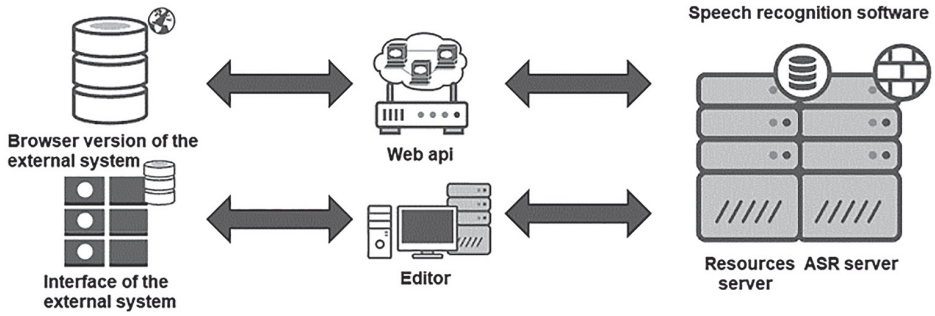


*Figure 3. The structure of speech recognition system*
*Dopisać: Source: Authors' construction*

the program from the ground within the organization where the system is implemented. Though such an approach is more time consuming as opposed to large-scale voice recognition systems with common vocabulary base (usually accessed online), it enables the creation of a more personalized environment which can be vastly more practical and effective.

The process of implementing such a system begins with the designing phase, when the client describes the features and purpose for which the system would be designed. Then the client and program distributor define the database requirements (whether it should contain abbreviations, range of professional vocabulary, colloquial vocabulary, organization-specific vocabulary etc.). Then the client must provide a large amount of teaching data (recordings that ultimately are supposed to be recognized by the system). Usually that process was long and required loads of additional work from the client such as collecting large amounts of data and was time consuming and recording e.g. meetings wasn't common. However nowadays as the COVID pandemic has forced many companies to implement digital communication systems such as Zoom to the working environment it is more common for companies to gather meeting recordings. Moreover workers are more used to working with digital communication systems. Thus nowadays implementation of speech recognition is way easier as companies can quite easily provide necessary data and workers are already accustomed to using technology for communication within the company. Company internal speech recognition system can be adjusted to both an internal network and web network (Figure 3).

For both implementations software consists of two servers:

- Resources server where vocabulary base is stored and updated,
- Automatic Speech Recognition server which is connected to resources server and performs speech recognition.

Software is then connected to Web API or the Editor provided by the developer for easier edition and processing of recognized text. The editor can provide such utilities like time stamping, text editing tools and templates. Furthermore API or the editor can be integrated with external systems such as company internal software or any commercial digital communication software to enable real-time speech recognition. This variety of possible structures enables the creation of complex and efficient communication structures.

## 2. EFFICIENCY AND IMPLEMENTATION PROBLEMS

Modern voice recognition techniques enable precise real-time speech to text recognition without the necessity for a user to speak unnaturally clearly or slowly. "The systems are designed to conform to people's most natural way of communicating and essentially do not require the user to alter this method" (Mehta, 2002). That greatly increases the acceptance of such solutions as users barely notice the usage of such programs. Therefore, one of the biggest concerns about implementation of speech recognition can be neglected as users can no longer be afraid about the necessity for them to adjust their speaking habits.

Implementing speech recognition software for digital communication solves another major issue with speech recognition which is overlapping voices. For most speech recognition software effective processing of overlapping voices is unmanageable. However in digital communication each microphone is operated separately and as users aren't physically in the same room their signals can be processed and recognized separately without overlapping. Furthermore, if users have individual microphones with properly set sensitivity, it will be quite easy to achieve similar efficiency during non-digital meetings.

The main problem about the implementation of such solutions is that most people aren't accustomed to using voice recognition which challenges the company to create new working habits in the company environment. However, since the COVID-19 epidemic has already forced most of the workers to change their working habits, it is a great opportunity for implementing innovative techniques as workers have to adjust to new technologies anyway. The pandemic has also increased usage of digital communication and thus made people more accustomed to speaking to electrical devices. Such circumstances create a good environment for usage of voice recognition as speaking to the computer is quite common during the pandemic and documentation has to be created digitally anyway. In such circumstances it is

way easier to create new social norms in terms of used technology "which are a significant factor when in terms of the adoption of new information system." "Those norms can be further influenced by tweaking the reward system technology training and top management vision and participation" (Simon & Paper, 2007).

Implementation of voice speech to text recognition solution in the company requires creating a vocabulary base specific to a certain language and topic. It's worth mentioning that modern approaches to speech recognition enable not only the recognition of words but also abbreviations, numbers, punctuation and much more. Such an approach enables using such a solution offline since the program doesn't use any external databases such as in e.g. google speech recognition solutions. Of course, during the implementation process there will be the initial period when performance of such a solution won't be as high, since the program will be adapting to the user's speaking habits. However, after the initial period each user will have a tool that is adapted to the user's speaking manner and will be able to recognize the user's speech in real-time with very high accuracy. Speech recognition solutions can not only be implemented as separate programs but also can be integrated with company internal software. An example of such a solution can be the Magic Speech Scribe speech to text recognition software developed by the company Radcomp Ltd. which enables the usage of both an external text editor and internal company software using a desktop widget that integrates voice recognition features with chosen internal software (Radcomp, 2019).

Therefore, the use of the speech recognition system can significantly simplify the management of the organization. Managers, instead of giving casual orders, can quickly make a note and send it to the employee, all meetings can be recorded in an easy and quick way. Thanks to the learning system, document checking can be performed very efficiently and without major errors. The learning system gives many possibilities and ways to personalize the document flow in the organization.

## SUMMARY

While speech recognition technology is being developed with increasing frequency, the field is still in its infancy. Many seemingly simple limitations have not yet been resolved. For example, the success of speech recognition by desktop computers depends on how well-interpreted speech technologies with an underlying processor and operating system are as well as the complexity and availability of the tools required to implement the system(Simon & Paper, 2007).

On the basis of this study, it can therefore be stated:

- the speech recognition system significantly supports the management of the organization and communication
- the COVID-19 pandemic "accidentally" significantly accelerated the work on the speech recognition system.
- creating a well-functioning speech recognition system, contrary to appearances, is not easy, it requires teaching a machine dedicated to a given organization,
- there are companies whose goal is to create dedicated solutions, however, creating a universal tool is still burdened with a mistake.

## REFERENCES

Bourne, L. (2015). *Making projects work. Effective stakeholder and communication management.* CRC Press. Retrieved April 22, 2021, from https://www.routledge.com/Making-Projects-Work-Effective-Stakeholder-and-Communication-Management/Bourne/p/book/9781482206661

Hirschorn, D. S. & Horii, S. C. (2006) PACS Workstation Software. In: K. J. Dreyer, D. S. Hirschorn, J. H. Thrall, & A. Mehta, (Eds.). (2006). *PACS. Guide to the Digital Revolution (pp. 385–431).* New York: Springer.

Mehta, A. (2002), Voice recognition. In: K. J. Dreyer, A. Mehta, & J. H. Thrall (Eds.) PACS. New York: Springer. https://doi.org/10.1007/978-1-4757-3651-9_11

Nguyen, M. H., Gruber, J., Fuchs, J., Marler, W., Hunsaker, A., & Hargittai, E. (2020). Changes in digital communication during the COVID-19 global pandemic: Implications for digital inequality and future research. *Social Media + Society*, 6(3), 1–6.

Radcomp. (2019, March 13) Czym Magic Speech Scribe różni się od innych? [Blog post]. Retrieved April 15, 2021, from https://magicscribesklep.pl/pl/n/3

Seiler, D. (ed.). (2020). How COVID-19 has pushed companies over the technology tipping point–and transformed business forever. McKinsey & Co.Retrieved April 12, 2021, from https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever#

Simon, S. J. & Paper, D. (2007). User acceptance of voice recognition technology: An empirical extension of the technology acceptance model. *Journal of Organizational and End User Computing* 19(1), 24-50. DOI:10.4018/978-1-60566-136-0.ch010