IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 14, 2021, accepted July 22, 2021, date of publication August 6, 2021, date of current version August 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3102502

Call Transcription Methodology for Contact Center Systems

MIROSŁAW PŁAZA[1], (Member, IEEE), ŁUKASZ PAWLIK[2], (Member, IEEE), AND STANISŁAW DENIZIAK[1], (Member, IEEE)

[1]Faculty of Electrical Engineering, Automatics and Computer Science, Kielce University of Technology, 25-314 Kielce, Poland
[2]Altar Sp. z o.o., 25-528 Kielce, Poland

Corresponding author: Mirosław Płaza (m.plaza@tu.kielce.pl)

This work was supported by European Union's Smart Growth Operational Program 2014–2020, under Agreement POIR.04.01.04-00-0079/19.

ABSTRACT Nowadays, one of the key areas of research on contact centre systems is their automation. The main element that influences the possibility of automation of contact centre processes is the call transcription methods implemented by automatic speech recognition (ASR) systems. Such systems enable developing intention recognition methods and, consequently voice bots. The current solutions used in ASR systems for many less popular languages do not guarantee a fully satisfactory transcription quality for hotline voice calls. This is due to the unique characteristics of the sound signal generated there, whose quality parameters differ significantly from those of studio recordings. The paper presents a comparative study of selected speech recognition systems that were additionally supplemented with elements of preprocessing of sound recordings and postprocessing of originally produced transcriptions. As for preprocessing, the following methods were tested: separation of the client and agent channels into two independent signals, training of ASR systems, and audio signal correction. With regards to postprocessing, on the other hand, tests were performed for inarticulate sounds, normalization of standard phrases (e.g. numbers, dates, times, etc.), and identification of close-sounding phrases and foreign language phrases, and lemmatization. Based on the research conducted and the analyses performed, a new method of call transcription intended specifically for contact center systems was proposed. The research conducted for this paper was based on the Polish language model, for which major problems are observed with the quality of automatic contact center call transcriptions.

INDEX TERMS Automatic speech recognition, call centre, contact centre, transcription, word error rate.

I. INTRODUCTION

The development of call/contact centre (CC) systems observed in recent years is largely focused on the automation of many of processes performed there. Automation of routine tasks, on the one hand, makes it possible to reduce costs generated by CC systems and, on the other hand, is important for social reasons, as it relieves agents from the need to perform tedious, repetitive tasks. This greatly improves the comfort of agents' work, thus preventing turnover in this difficult job [1]. Despite the existence of many different communication channels at CCs (e.g. e-mail, chat, and social media), traditional phone calls continue to be one of the dominant channels. The technological development of this area in the context of automation is particularly important because it has a significant impact on the quality of customer service and the costs incurred [2]. In recent years, intelligent virtual assistants taking the form of chat bots and voice bots have become very popular in CCs [3]. The bots currently implemented in CC systems can perform many repetitive and routine actions with good results. Voice bots can also make phone calls directly to customers by themselves, thus completely resolving some issues [4]. The main component that is used in the design of voice bots is intention recognition methods. These methods are in turn built using ASR (Automatic Speech Recognition) systems [5]. There are many different ASR systems on the market today, both free and offered on a commercial basis. These systems are characterized by many different

The associate editor coordinating the review of this manuscript and approving it for publication was Tariq Umer.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

parameters that determine the possibility of their use in the development of intention recognition methods. Table 1 lists popular ASR systems along with parameters that are important from the standpoint of their potential applications in CC systems.

As described in the literature, the effectiveness of speech recognition by known ASR systems is up to 95% [6]. However, this value is usually achieved for studio-quality recordings (e.g. movies and radio broadcasts). The audio signals typically found in CC systems are extremely varied and often distorted. Such signals usually have a sampling rate of only 8 kHz, a 16-bit resolution, and the PCM (Pulse-Code Modulation) format. In addition, calls are often made using the VoIP (Voice over Internet Protocol) technology, usually recorded in mono format without separate channels of the client and the agent, so that the utterances of the interlocutors overlap. These channels are usually characterized by different parameters, e.g. different amplitude or different noise level coming from the interlocutors' surroundings [7]. For the aforementioned audio signal parameters in the English language model, the automatic speech recognition effectiveness achieved was determined in paper [8] to be 86%. This is a limit value for the ability to effectively use the received automatic transcriptions in English-speaking CC systems. Automatic transcription is more difficult and even less effective for language models that are not very popular, for example Polish.

Issues related to the possibility to improve the quality of transcriptions performed for dedicated CC systems with a complex language corpus have not yet been described in the world literature. These issues motivated the authors to conduct research aimed at finding solutions that will ensure the possibility to improve the quality of automatic transcriptions for calls made on CC hotlines. As a result of that research, a new transcription method intended specifically for CC systems was developed, in which the integrated ASR systems were supplemented with selected recording preprocessing algorithms and/or postprocessing algorithms for the original automatic transcriptions prepared. Given the unique characteristics of the acoustic signals described herein, it was assumed that both preprocessing of an acoustic signal before it is transcribed and a properly defined process of correction of the original transcriptions made are likely to improve their quality. The preprocessing and postprocessing algorithms used are described in detail in section III of the paper. The solutions presented herein focus on improving the effectiveness of ASR systems for calls in the Polish language. Nevertheless, the recommended method can be generalized for other language models.

Based on an evaluation of the parameters summarized in Table 1, it was determined which ASR systems could potentially be used to build a transcription method intended for CC systems that support the Polish language. Ten popular ASR systems developed by different companies were selected for evaluation. The key criteria that were analysed were native support of the Polish language model, licensing method, ability to work in real time, available APIs, software delivery model (cloud or on premises), and access to advanced, transparent tools for monitoring, reporting, and accounting for the services offered (which is very important from the standpoint of business applications). Since the solution presented in this paper is intended for transcription of calls in CC systems conducted in the Polish language, a natural selection criterion was the need for native support of the Polish language model. From the point of view of the target applications in CCs, the following solutions have no or insufficient support for that language: Amazon Transcribe, Facebook wav2letter, IBM Watson Speech to Text, and Mozilla Deep Speech. Therefore, the aforementioned systems were not considered during the research presented later in this paper. Another criterion that is important from the point of view of the target application of the solution was the possibility to run it in a secure and popular cloud environment such as Microsoft Azur, Google Cloud, Amazon AWS, IBM Cloud, or Oracle Cloud Infrastructure [19]. It was also important to provide easy access to advanced tools for monitoring, reporting, and cost management of automated transcriptions. An additional aspect was that the ASR system needed to provide the possibility

TABLE 1. Chosen Automatic Speech Recognition Systems.

No.	Name of ASR system	Polish language model	License	Real time streaming	API	Secure and popular cloud platform	Advanced Reports, Billing	Reference
1.	Microsoft Speech to Text	Yes	Commercial	Yes	HTTP	Native cloud platform	Yes	[9]
2.	Nuance ASR	Yes	Commercial	Yes	HTTP	Requires installation	No	[10]
3.	Google Speech-to-Text	Yes	Commercial	Yes	gRPC	Native cloud platform	Yes	[11]
4.	VoiceLab ASR	Yes	Commercial	Yes	HTTP	Requires installation	No	[12]
5.	Techmo ASR	Yes	Commercial	Yes	gRPC	Requires installation	No	[13]
6.	PrimeSpeech	Yes	Commercial	Yes	MRCP	Requires installation	No	[14]
7.	Mozilla Deep Speech	Not enough	Open Source	Yes	Native clients	Requires installation	No	[15]
8.	Amazon Transcribe	No	Commercial	Yes	HTTP	Native cloud platform	Yes	[16]
9.	Facebook wav2letter	No	Open Source	No	Native	Requires installation	No	[17]
10.	IBM Watson Speech to Text	No	Commercial	Yes	HTTP	Native cloud platform	Yes	[18]

to train the language model by using the available built-in internal learning mechanisms. As a result of the analyses performed and based on the above criteria, it was determined that two solutions, Microsoft Speech to Text and Google Speech-to-Text, denoted later in this paper by their acronyms MST and GST, respectively, would be considered for further research.

The main contribution of this paper is the following:

- a preprocessing methodology, dedicated to CC input signals, including separation of voice channels, CC-oriented training of ASR tools, and audio signal correction and
- a postprocessing methodology, improving the quality of transcription, using a text correction method, identification of close-sounding and foreign words, and a lemmatization algorithm.

The above methodologies were integrated with existing ASR tools, creating a novel transcription system for CCs. The experimental results showed that our method significantly improves the quality of transcription for CC-quality signals in comparison with existing solutions.

Above methodologies were integrated with existing ASR tools, creating a novel transcription system for CC. Experimental results showed that our method significantly improves the quality of transcription for CC-quality signals in comparison to existing solutions.

Section II describes the methodology of the research conducted and the testing environment for the ASR performance improvement methods being developed. Section III presents the developed preprocessing and postprocessing methods aimed at improving the speech recognition performance. Section IV presents a complete ASR system intended specifically for CC systems. Section V presents the experimental results obtained. The effect of individual preprocessing and postprocessing algorithms on the efficiency of automatic transcription is also described. These elements were examined in the context of improvement of the effectiveness of automatic transcriptions originally performed by the ASR systems selected for the research. The paper ends with conclusions and a description of the directions for future work.

II. RESEARCH METHODOLOGY

The main research work was divided into two main parts. The first part was research on the preprocessing elements, while the second part was research on the postprocessing elements. With regards to preprocessing, the effects of channel separation for individual interlocutors (client/agent), the ability to train the studied ASR systems, and the ability to correct the parameters of the audio signal (noise neutralization, normalization of agent and client channel volume levels) were investigated in succession. With regards to postprocessing, the following were examined: text correction possibilities, along with normalization of selected elements. The problem of close-sounding words, words from outside the Polish language corpus, and the mechanisms of lemmatization were also examined.

The aim of the research carried out for the above-mentioned preprocessing and postprocessing algorithms was to determine the potential possibility to use them as components of a new transcription method intended specifically for CC systems supporting the Polish language. The basic objective of the work was to increase the quality of automatic transcriptions for calls made to CC hotlines. The work ultimately made it possible to identify the mechanisms that work with ASR systems to best enhance the quality of transcriptions. A flowchart of the research methodology is shown in Figure 1.

The quality of automatic transcriptions, as recommended in the literature [20], is determined using the WER (World Error Rate) and LER (Letter Error Rate) metrics. From a CC perspective, the WER metric is most useful for measuring the effectiveness of ASR systems [21]. It is based on the Levenshtein's distance edit metric for words [22]. WER parameters can be determined by comparing transcriptions of recordings from a given ASR system with reference transcriptions made manually by a human. This is done using the following formula [23]:

$$WER = (sw + dw + iw)/nw \tag{1}$$

where: nw - number of words in the reference transcription, sw - number of words substituted, dw - number of words deleted, and iw - number of words inserted required to transform the output transcription into the target (among all possible transformations, that one that minimizes the sum $iw + sw + dw$ is selected). Note that the number of correctly recognized words, cw, is determined by the following formula:

$$cw = nw - dw - sw \tag{2}$$

The LER metric, which is calculated for specific letters as the ratio between their incorrect transcriptions and the total number of their occurrences, also provides important information. This metric is defined by the following formula [24]:

$$LER = (s + d + i)/n \tag{3}$$

where: n - total number of letters, s - substitutions, d - deletions, and i - number of character insertions.

The paper presents comparative studies and analyses for both metrics described above. The results obtained are described in detail in section V. In the initial phase of the work, a first database of sound recordings was created, which included 754 real conversations between a customer and an agent on a CC hotline. These samples were used during work on the development and optimization of a transcription method for CC systems supporting the Polish language. The archived conversations were conducted during actual campaigns performed by a large commercial CC system. Additionally, a second database of 300 recordings was created for the final verification of the developed method.

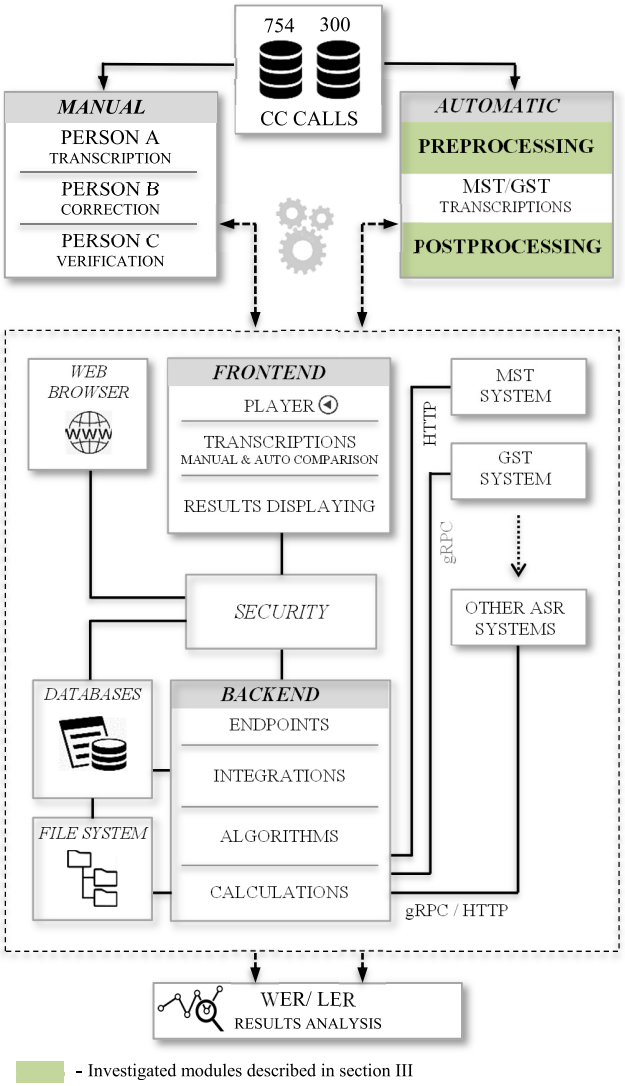


FIGURE 1. Research methodology flow chart.

For testing purposes, three main topics of the calls were selected: a) invoices and payments; b) technical information; and 3) contracts and amending annexes. For the purpose of the research, it was assumed that the duration of each full conversation should be no less than 3 minutes and no more than 20 minutes, which made it possible to create a set of 100 hours 53 minutes and 41 seconds of recordings in total (first database: 77 hours 54 minutes and 30 seconds; second database: 22 hours 59 minutes and 11 seconds). All samples for research and verification were selected randomly and each was anonymized before being transferred for research purposes (all data identified as sensitive according to the General Data Protection Regulation (GDPR) was removed).

The recordings databases required a testing environment that would automate their processing and analysis, and guarantee repeatability of results. In this environment, algorithms supporting the processes used in preparation of manual reference transcriptions (human-made transcriptions) were introduced; the ASR systems selected in the first section were integrated; algorithms enabling the comparison of the reference transcriptions with automatic transcriptions were developed; algorithms calculating the quality of transcriptions were introduced; and a number of other functionalities were developed as the transcriptions evolved. The prepared software tools enabled efficient work with the recordings database, ensured easy performance of multiple comparative tests on large data sets, and facilitated changes to test parameters.

The testing environment that was developed was used in both manual and automatic transcription processes. For this purpose, a recordings player tool was used extensively, the functionalities of which enabled presenting the acoustic spectrum divided into the client and agent channels, adjusting the playback speed, easily navigation between recordings, and automatically marking the duration of individual utterances of the interlocutors recorded in the transcription files. The muting functionality for a selected channel and the ability to zoom in on the spectrum were also often used. The manual call transcripts provided reference data for comparison with data automatically generated by the studied ASR systems. It should be emphasized that correct performance of manual transcriptions is an extremely difficult, costly, and time-consuming process; however, it is crucial for reliability of the results of the research. Therefore, the manual transcription process involved as many as three steps. In step one, the original transcription was prepared by the first person (person A). Step two involved proofreading of the completed original manual transcription. This proofreading was done by a second person, marked in the figure as person B. In the third step, another person (person C) verified the transcription proofread by person B. A linguist specializing in the Polish language participated in the work on the manual transcription, its proofreading, and its verification. The transcriptions made were saved in a relational database for further processing. The manual transcription process carried out in this manner guaranteed their very high quality, which was necessary for further research.

With the reference transcriptions available, the next step involved an examination of the impact of the preprocessing and postprocessing elements on the transcription quality achieved by the automated MST and GST systems. The direct integration of the ASR systems selected for the testing in the test environment greatly simplified the tasks related to processing, comparing, and archiving of the data obtained as a result of automatic transcription. The environment itself enables eventual integration of other ASR systems as well. From the end user's point of view, the functions implemented in the *FRONTEND* block play a major role in the environment. That block is responsible for the usability and ergonomics of work. An important element of the block is the integrated recordings player, which provides the numerous aforementioned important functionalities that support transcription processes. Thanks to the built-in comparison tools for manual and automatic transcriptions and the possibility

to present the results in real time, the block also supports postprocessing processes. Implementation of all functionalities of the environment is possible by using the *BACKEND* block elements. In that block, individual endpoints are made available using the REST API. The block is responsible for data persistence in the *DATABASES* and the *FILESYSTEM*, as well as for data management and consistency. *BACKEND* also provides integration with ASR systems and postprocessing algorithms. Integration with the GST system was ensured by using an API (Application Programming Interface) based on the gRPC (Remote Procedure Calls) protocol. The MST system, on the other hand, was integrated using an API based on the HTTP protocol. The environment has embedded databases in which appropriate structures were created to enable effective cooperation of the *DATABASES* block with other elements of the system. In the *DATABASES* block, a manual reference transcription and a number of automatic transcriptions prepared (depending on different parameters) are stored for each sound file. The histories of the WER and LER metrics tests and the elements used by postprocessing mechanisms are also stored. Each recording is classified according to the topic of the conversation, its duration, the sampling rate, and the bit resolution. Such sets are stored in the *FILESYSTEM* block. Direct access to each set is only possible for authenticated and authorized users from within the testing environment. This block is also used for presentation to end users of temporary automatic transcriptions which, depending on the system administrator's decision, are eventually saved in the *DATABASES* block. Even though they are anonymized and do not contain personal data, address, or contact details, the actual phone call recordings collected in the system should still be protected with due diligence. Therefore, it was necessary to use security mechanisms to identify and manage access to the resources, which was implemented in the *SECURITY* block. The testing environment can be operated from any web browser.

III. METHODS FOR IMPROVING TRANSCRIPTION

This section describes the methods for preprocessing input data and postprocessing output data, which affect the quality of speech transcription performed for samples obtained from CC systems that support the Polish language. The research focused on improving the transcription quality of the ASR systems selected for testing in the first section.

A. PREPROCESSING METHODS

The purpose of input data preprocessing is to improve transcription quality by adjusting the speech signal processing method and the transcriptor according to the specific characteristics of the CC system. The approach presented in this paper involves the use of three methods: (1) separation of the client and agent channels, (2) training of the ASR systems, and (3) audio signal correction.

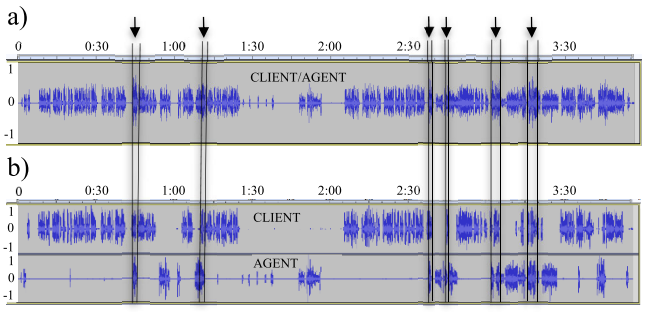


FIGURE 2. A part of a spectrum of a conversation on a CC hotline: a) combined interlocutors' channels, b) separated interlocutors' channels.

1) SEPARATION OF CHANNELS

In real CC systems, the archived calls are usually stored as optimally compressed signals. Depending on the service provider, sometimes it is a mono format, sometimes it is a stereo format. From the point of view of ASR systems, the quality of analysed signals plays a very important role. The utterances of the interlocutors often overlap, making the correct transcription impossible. Therefore, in step one, a decision was made to check the impact of separation of the agent and client channels into two independent data streams by performing independent transcriptions for each of them. This required recording the conversations in a stereo format. In addition, the data was time-stamped to indicate the start time of subsequent utterances so that the correct order could be maintained in the target transcription. Figure 2 shows a part of the spectrum of one of the actual conversations examined in the course of the study. Figure 2(a) shows the signal in which the client and agent channels are superimposed, while Figure 2(b) shows their separated spectra. Note that in the rather short excerpt of a conversation that is shown, there are many places where the client and the agent speak simultaneously. These places are marked with arrows.

Considering the unique characteristics of conversations conducted in CCs, the case described herein in which the client and the agent speak simultaneously is repeated many times in almost every call. From the standpoint of the ASR systems studied, the GST solution has built-in mechanisms that separate the two channels and it handles this problem relatively well. The MST solution, on the other hand, does not have integrated options for optimized separation of the interlocutors' utterances and treats the stereo file provided for transcription as a regular mono file. Consequently, the client and agent data streams need to be stored in separate files and then undergo separate parallel transcriptions to get the right results. Such an operation significantly improves the quality of transcription. The detailed results of this preprocessing element are shown in Figure 8 in the column labelled SEP.

2) TRAINING OF ASR SYSTEMS

The next step in the examination of the preprocessing mechanisms was the implementation of the teaching processes

TABLE 2. Training parameters for the MST system.

Name	Number of samples	Calls duration	WER improvement
	[pcs]	[h:m:s]	[%]
MODEL 1	20	05:29:52	0.03
MODEL 2	50	13:24:40	0.18
MODEL 3	100	27:09:23	0.98
MODEL 4	158	41:47:16	1.00

pcs = number of recordings, h = hours, m = minutes, s = seconds

provided in different forms by both ASR systems studied. The MST solution makes it possible to create one's own custom model of a given language, based on a selected sample of recordings and their reference transcriptions. On the other hand, the learning mechanism in the GST solution is based on the use of a common set of prepared model phrases that are most frequently repeated in the utterances of agents and clients. Both solutions have been tested. It is worth noting that for ASR systems that do not have an integrated learning module, such a component can be implemented as an additional custom component.

The studied MST solution has the Custom Speech toolkit [25], which enables teaching for many different languages, including Polish. Both integrated basic language models and user-created custom models can be used for this purpose. Additionally, each of the supported languages has at least one base model. Custom models are created by teaching the selected basic model on the basis of sound recording samples provided and their reference transcriptions containing specific vocabulary relevant from the standpoint of further applications. Teaching of the base model of the MST system was done by creating custom models that used different amounts of data in the learning process. Table 2 contains a list of the models that have been prepared.

Training was carried out in four models, with the amount of teaching data increased for each from 20 to 158 recordings, together with their reference transcriptions. Recordings were selected from the samples collected in the first (primary) database. Verification of the teaching processes, on the other hand, was performed on a set of 30 test stereo recordings with the client and agent channels separated. Automatic transcriptions of the test recordings were made and the WER metric was calculated for each model. As shown in Table 2, increasing the amount of teaching data to 158 resulted in a 1% improvement in the transcription quality. Further training did not improve the results.

During the research, the various components of the WER index listed in Section II were analysed, namely *sw* (substitutions), *dw* (deletions), *cw* (recognitions), and *iw* (insertions). It can be seen that after the training process, the greatest improvement is in the *sw* component and the *dw* index, which is consistent with the expectations and confirms the training effect. For example, for model 4 the *sw* component improved from 1,886 (10.10%) to 1,774 (9.50%), the *dw* component - from 635 (3.40%) to 541 (2.90%), and the *cw* component - from 16,152 (86.5%) to 16,358 (87.6%). On the other hand, the *iw* component slightly deteriorated from 803 (4.30%) to 821 (4.40%). This is reasonable, because training does not eliminate the recognition of inarticulate sounds as words. When evaluating the quality of machine learning in terms of word classification, the effectiveness of the training can be expressed with a confusion matrix, where the following assumptions are made, respectively: *True Positive* (TP) = *cw*; *False Negative* (FN) = *dw*; *False Positive* (FP) = *sw* + *iw*, and *True Negative* (TN) = 0. The tools used do not provide data on correctly recognized inarticulate sounds, as this is irrelevant to transcription. Therefore, statistics on the TN set are not available, as well as not relevant. The confusion matrix is presented in Table 3.

TABLE 3. Confusion Matrix.

		PREDICTED VALUE	
		POSITIVE	*NEGATIVE*
ACTUAL VALUE	*POSITIVE*	16152 / 16358	635 / 541
	NEGATIVE	2689 / 2595	0

In contrast, the GST system was tested using the speech context function [26]. For this purpose, two sets of 100 most frequent phrases were prepared, which most often caused problems in transcription. One set comprised phrases that occurred in the client channel, while the other comprised phrases that occurred in the agent channel. The phrases varied in length from 11 to 96 characters. The sets were used in the teaching process of the GST system. Once the systems were prepared in this manner, automatic transcriptions were made for a set of selected 30 test recordings. The detailed test results are presented in section V in columns labelled AI/ML.

3) AUDIO SIGNAL CORRECTION

A very important factor affecting the automatic transcriptions made by ASR systems is the quality of the audio signal. This quality can be improved by reducing noise, clicking sounds, other unwanted background noise. It may also be advisable to normalize the volume levels for both separated channels. Other widely available filters can also be used. The audio processing mechanisms are today very well recognized and implemented internally in many ASR systems, including the MST and GST systems studied. However, when integrating ASR systems that do not optimize the aforementioned parameters, it may be necessary to correct them before transcription is performed. Therefore, an audio signal correction module that optimizes the aforementioned factors is recommended as another preprocessing element. Figure 3 shows the illustrative structure of the audio signal correction block.

The tests of the recommended module with reference to the MST and GST systems did not demonstrate the need for its application, which is confirmed by the test results shown in Figure 9 in the columns labelled DEN and NORM. However, the recommended module may be needed for

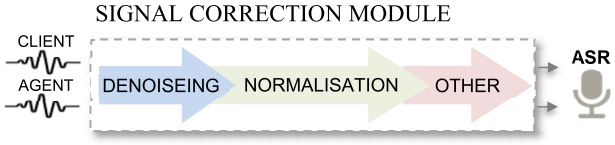


FIGURE 3. Signal correction module.

potential implementations other than the MST and GST systems described in detail in this paper, where the audio signals may not be sufficiently optimized by the internal components of these systems. This aspect, however, requires further research, which was not advisable for this paper.

B. POSTPROCESSING METHODS

Postprocessing methods involve optimizing the final result by improving the output text prepared by the ASR system. The recommended solution used the following methods: 1) text correction module, 2) identification of close-sounding phrases and foreign language phrases, and 3) lemmatization.

1) TEXT CORRECTION MODULE

A detailed comparative analysis of the transcriptions made with the preprocessing mechanisms and the manual transcriptions showed that the quality of automatic transcription is strongly affected by inarticulate sounds that commonly appear in CC conversations. From the semantic point of view, these sounds are irrelevant to the understanding of the conversations carried out in CCs. However, they negatively affect the quality of automatic transcription, as their transcription is usually incorrect. Therefore, a decision was made to eliminate them at the postprocessing stage. Table 4 contains a list of sounds of this type that occurred most frequently in the processed database of client-agent conversations.

TABLE 4. The most common human inarticulate sounds in database.

No.	Inarticulate sounds in Polish pronunciation	Number of appearances
1.	"hm," "hmm," "hmmm," "mhm"	762
2.	"aha," "haha"	458
3.	"ee," "eee," "eeee"	252
4.	"yy," "yyy," "yyyy"	110
5.	"em," "emm," "emmm"	49
6.	"mm," "mmm"	19
7.	others inarticulate sounds	93

As part of the module described in this section, the way such elements as numbers, amounts, dates, and times are transcribed was also normalized. Moreover, the impact of lowercase and uppercase letters, and punctuation marks on the quality of transcription was eliminated. The algorithm of the text correction module is shown in

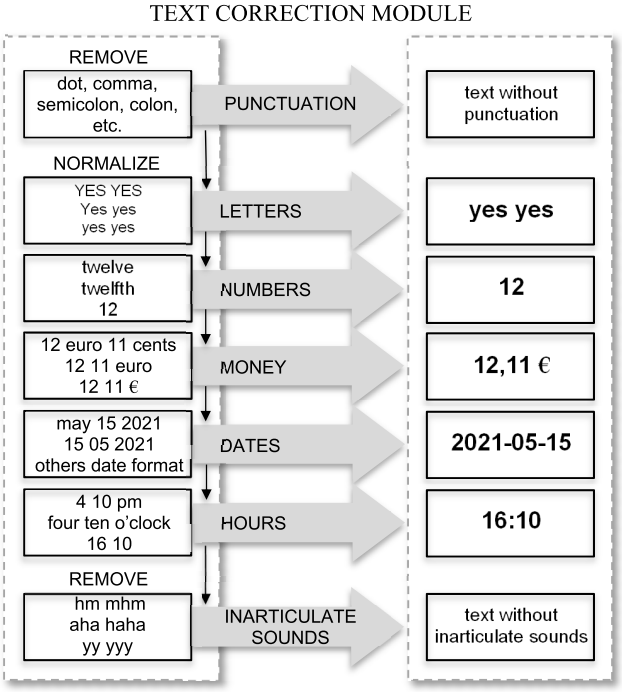


FIGURE 4. Text correction module.

Figure 4. The results of the examination of this element of postprocessing shown in section V are denoted in the graphs as COR.

2) CLOSE-SOUNDING AND FOREIGN WORDS MODULE

From the point of view of the quality of the transcriptions studied, close-sounding words are also a serious problem. An example for the Polish language is the "number *IMEI*" phrase, which was usually changed to "number *e-mail*" in the course of a transcription. It is obvious that in this case the correct phrase is "number *IMEI,*" because the proper connotation for e-mail is an address, not a number. In view of the above, for the conversation topics selected for the research presented in this paper, it can be assumed that the misspelled phrase "number *e-mail*" should be replaced with the phrase "number *IMEI*." Consequently, for the conversation topics identified in section II, a database was built that comprised similar-sounding words/phrases that were often confused by the ASR systems studied. The use of such a dataset had a positive impact on the quality of the transcriptions made. The algorithm of the close-sounding and foreign words module is shown in Figure 5.

Another problematic element for ASR systems is words from outside the Polish language corpus (e.g., names of companies, products, brands, or Polish words borrowed from other languages, mainly English). Many such words are transcribed by ASR systems phonetically, e.g.: instead of "*iphone,*" the system makes a phonetic transcription in Polish into "*ajfon*" or "*ifon.*" These elements can be corrected on an ongoing basis using the studied post-

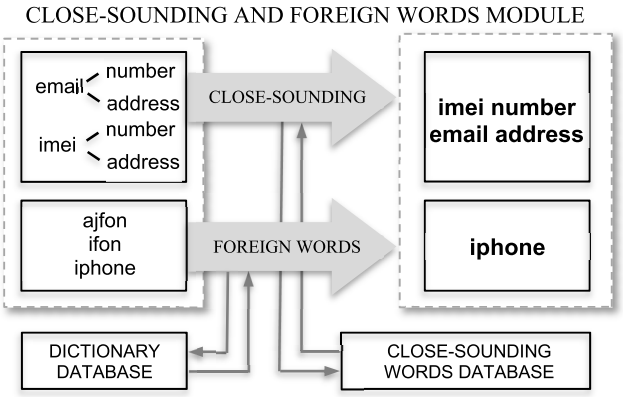


FIGURE 5. Close-sounding and foreign words module.

processing algorithms. A separate dictionary was created in the form of a database that contained words that are problematic for the conversation topics considered in the research. This database was used for real-time automatic correction of incorrectly recognized words. It can be continually expanded depending on the conversation topics. The tests for this module shown in section V are denoted as NCS.

3) LEMMATIZATION MODULE

Due to the complicated nature of the Polish language associated with numerous complex inflections of individual words, a decision was made to develop, implement, and examine the influence of a lemmatization algorithm on the quality of transcriptions. As it is known, this algorithm reduces a given word to its base form [27], [28], which, for example, can be used in algorithms for multi-criteria categorization of text data or algorithms for grouping identical statements used in the design of virtual assistants. A publicly available library [29] was used for the lemmatization. The results of the examination of the lemmatization module are denoted as LEM.

IV. TRANSCRIPTION METHOD FOR CC SYSTEMS

Figure 6 shows the recommended automatic transcription method designed for use CC systems that support the Polish language. In the initial phase, the signal recorded during the conversation between a client and an agent undergoes preprocessing tasks (1). These tasks are performed before the data is sent for further analysis by the embedded intelligent ASR systems (2). The selected ASR systems create transcripts of the conversation, which then go to the block that performs postprocessing algorithms (3). That block is responsible for optimizing the automatic transcription performed by the ASR systems. The result of its operation is the target transcript (4).

The first action is to make the recordings directly in a stereo format. Such recordings contain two separate channels for each of the interlocutors (client and agent). The audio stream from each channel is saved in the WAV (Waveform Audio File Format) format. According to the recommended method, this task should always be performed first. Then, depending on the needs, the signal can be separated into two independent channels for each interlocutor and be forwarded to other components performing preprocessing tasks. The preprocessing modules described herein can be used as necessary as potential additional elements to optimize the transcription quality. These elements can be switched on or off depending on the needs and the ASR system used. If the preprocessing elements are switched off, the stereo signal can be forwarded directly to the selected ASR system.

In the second phase, after the selected preprocessing tasks are completed, the received audio signal is forwarded to one or more ASR systems. As explained earlier, two solutions were selected for the purpose of achieving the objectives set in the study: MST and GST. However, this does not restrict the possibility to use the recommended method for applications other than transcription for CC systems and does not restrict the use of other ASR systems. The result of the operation of each of the ASR systems is a

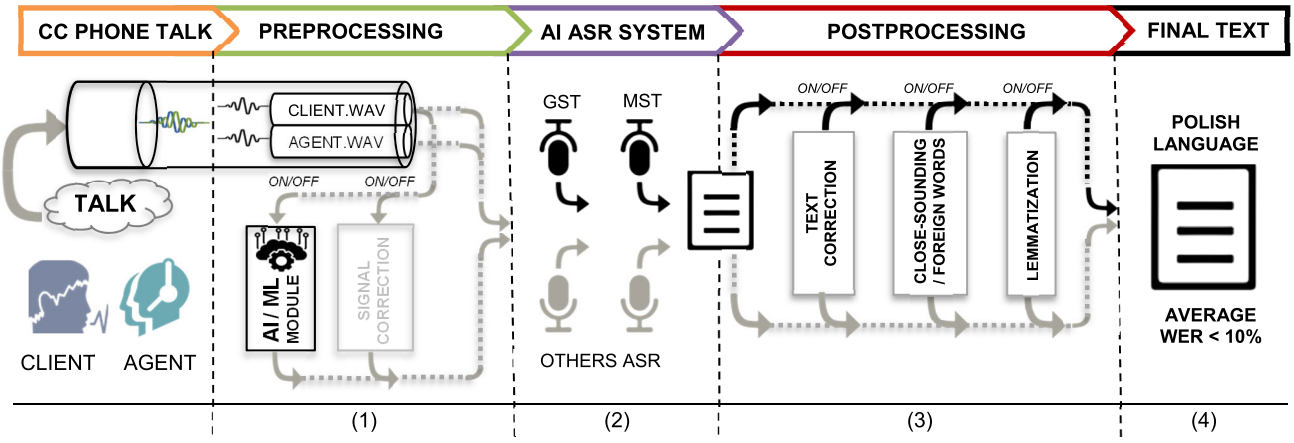


FIGURE 6. Automatic transcription method designed for CC systems that support the Polish language.

transcript of the sound recording made in a text form, which is then processed by the postprocessing elements in the third phase.

In the transcription method developed, sets of algorithms were prepared and implemented to optimize the transcripts, taking into account the conversation topics selected for the study. The necessary functionalities of these algorithms were selected on the basis of comparative analyses performed for the reference transcriptions and the automatic transcriptions made. A comparative study of both transcriptions made it possible to identify the weakest features of the ASR systems studied in terms of correctness of transcription of conversations conducted in CCs in the Polish language. This study enabled identification of the postprocessing elements that needed to be implemented to dramatically improve transcription quality. The first of these elements is the text correction module and the second is the module responsible for optimizing transcription from the point of view of close-sounding words and words coming from outside the Polish language corpus. The last block implemented is the lemmatization module. These blocks are described in section III. The study also made it possible to determine the best order in which the postprocessing algorithms should be initiated in order to optimize the obtained test results. Figure 7 shows a diagram that describes the optimal sequence of operation of individual modules and algorithms. First, the algorithms implemented in the text correction module are initiated to eliminate the impact of upper- and lower-case letters and punctuation marks on the quality of the transcriptions made by the integrated ASR systems. In the next step of the operation of the algorithms in this module, the transcription of numbers, amounts, dates, and times is normalized. After these operations are completed, the algorithms are initiated that remove characters created as a result of transcription of inarticulate sounds that are irrelevant to the meaning of the sentence and that often occur in the narratives of the interlocutors. The second module executed contains mechanisms that optimize the correctness of transcription for words that are similar in terms of pronunciation, but have completely different semantic meanings. This is the first action performed in this module, after which the system optimizes the transcription of foreign words. The last of the post-processing modules is lemmatization. This algorithm enriches the postprocessing block and is useful due to the complicated structure of the Polish language. However, the use of lemmatization is not always justified and depends on the purpose for which transcriptions are made. The algorithm execution sequence described above provided the best transcription quality for the recommended method. However, if the method needs to be adapted to other languages, it is possible to easily change this sequence.

The method developed provides a satisfactory level of automatic transcription for conversations conducted in CC voice channels between clients and agents. Of note is the fact that this level is suitable for its further use in intelligent client intent recognition solutions.

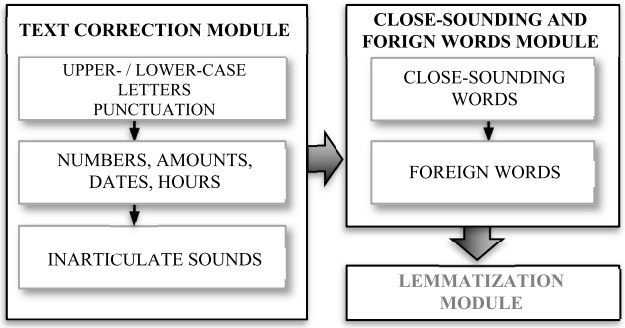


FIGURE 7. Postprocessing algorithms execution sequence.

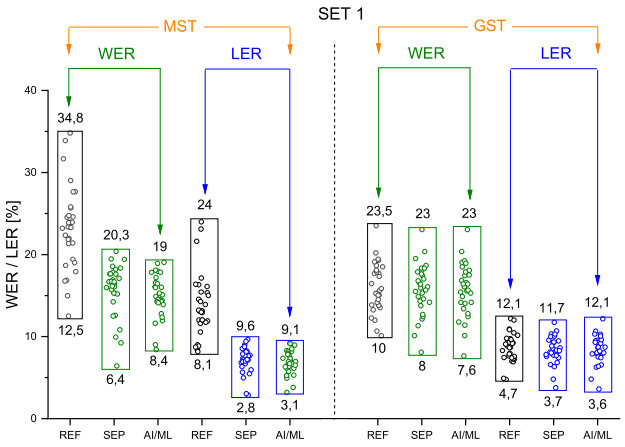
V. RESULTS AND ANALYSIS

This section describes the results obtained for the preprocessing and postprocessing components described earlier. This study formed the basis for the development of a new transcription method, presented in section IV, intended for CC systems that support the Polish language. They can also be very helpful if the recommended method is to be adapted for use with other languages.

A. RESULTS OF PREPROCESSING ELEMENTS

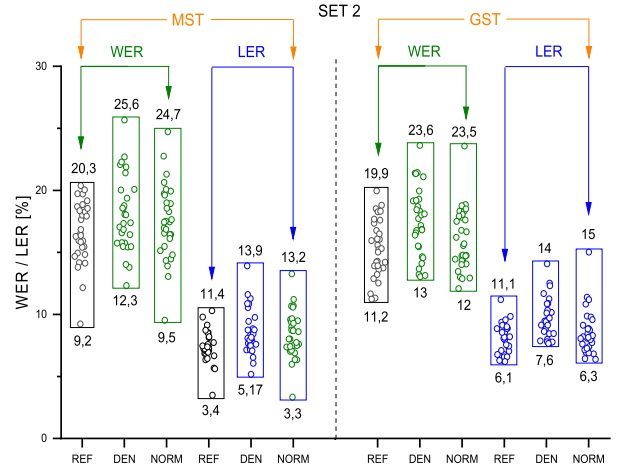
Figures 8 and 9 show the impact of the preprocessing elements used on transcription quality expressed by the WER and LER metrics. In order to evaluate the effectiveness of the tested solutions, two data sets containing 30 recordings each of conversations conducted during real CC campaigns were created from the pool of recordings related to the ''invoices and payments'' topic. The impact of separation of the interlocutors' channels and the possibility to train the ASR systems on transcription quality was examined using the first set of recordings. That set was selected in a representative manner to include recordings where the calculated original values of the WER metric ranged from a minimum value to a maximum value. The second set, on the other hand, was used in the tests performed on recordings with corrected acoustic signal parameters. That set contained recordings in which the quality of the sound, as heard by humans, was the worst. This was due to the distinct noises, crackling sounds, and sounds made by cars or animals which were present in the recordings, the different volume levels of the client's and agent's channels, as well as many other unwanted sounds which interfered with the conversations. During the testing of this preprocessing element, all the aforementioned interferences were eliminated.

The prepared set of recordings was automatically transcribed in the MST and GST systems. The tests performed on the MST system indicate that the values of the WER metric for primary signal transcription range from 12.5% to 34.8%. When separated agent and client data are fed to the transcriptor, this metric improves significantly and ranges from 6.4% to 20.3%, respectively. A very large improvement is shown in this regard by the LER metric. For primary signal tests, the value of this metric ranged from 8.1% to 24%.



LEGEND: REF - All preprocessing elementss are OFF, SEP - Channel separation component is ON, AI/ML - Learning component is ON

FIGURE 8. The impact of separation of the client and agent channels the teaching processes on the transcription quality of the MST and GST systems.



LEGEND: REF - All preprocessing modules are OFF, DEN - Denoising component is ON, NORM - Normalization component is ON

FIGURE 9. The impact of noise reduction and normalization of the client and agent channel volume levels on the transcription quality of the MST and GST systems.

Preprocessing consisting in channel separation improved the above results to the level of 2.8% to 9.3%. Analogous tests were performed for the second system selected for tests as part of this research, the GST system. The system has an internal mechanism for automatic extraction of separate tracks for the client's and agent's signals from stereo files. The tests of this preprocessing element show that the values of the WER metric for automated transcriptions made from the primary signal range from 10% to 23.5%. When the agent and client channels were separately fed to the transcriber, this metric improved slightly and ranged from 8% to 23%. This improvement is negligible due to the functioning, well-optimized interlocutor channel separation mechanism that is integrated into the GST system. Therefore, only one stereo data stream can be fed when performing transcriptions using this system. For the LER metric, the GST system originally reached values between 4.7% and 12.1%. The tested preprocessing element did not affect the LER metric, as the average value of LER was 8.3% in both cases. The corresponding arithmetic mean values for the tests described herein are summarized in Table 2. An analysis of the results shows an improvement in the quality of automatic transcription when two separate data streams were fed to the MST system. Therefore, when this system is used in CCs, it is necessary to adopt the preprocessing element described as the base solution. The GST solution, on the other hand, handles transcriptions of stereo files quite well without having to separately feed the signals they contain.

TABLE 5. Average WER and LER metrics for preprocessing solutions.

METRIC	ASR	SET 1			SET 2		
		REF [%]	SEP [%]	AI/ML [%]	REF [%]	DEN [%]	NORM [%]
WER	*MST*	23,2	15,7	14,7	15,8	15,4	17,4
	GST	15,9	15,7	15,7	15,9	16,8	15,7
LER	*MST*	13,8	7,1	6,8	7,4	7,0	7,4
	GST	8,3	8,3	8,3	9,0	9,1	8,0

The next step in the examination of the preprocessing mechanisms was the implementation of the teaching processes described in section III, which are provided in different forms by both ASR systems studied. An analysis of the effect of the teaching mechanisms on the quality of the transcription made using the MST system indicated further possibility of its optimization. The results obtained for the WER metric range from 8.4% to 19%, which is a significant improvement over the reference values. On the other hand, compared to the values obtained in the first preprocessing step, the transcription quality improved by another 1% on average. Nevertheless, training of the GST system did not bring the expected results and the average values of both metrics were very similar to the reference results. Taking the above into account, it can be concluded that for both systems, the Polish language models provided by the ASR system vendors are at a high level. Nevertheless, the learning process adopted in the MST system had quite a significant impact on the tested quality of the automatic transcriptions performed. Described above research results are presented in Figure 8.

The last preprocessing module tested was the audio signal correction module. The test results showing the impact of this module on the level of the WER and LER metrics for the GST and MST systems are shown in Figure 9.

Although in the analysed set of recordings the improvement in sound quality was significantly noticeable to humans, the transcription quality results are not improved. It can be assumed that the implemented ASR systems already have built-in and well-developed correction mechanisms for audio signals that are optimized for the transcriptions performed. Therefore, from the point of view of the systems analysed, it is most important to make sure that the stereo source recording is of the best possible quality.

Table 5 summarizes the averaged WER and LER metrics values for the preprocessing elements discussed herein. An analysis of the data shows that in terms of the possibility to use the examined solutions for CCs serving clients in

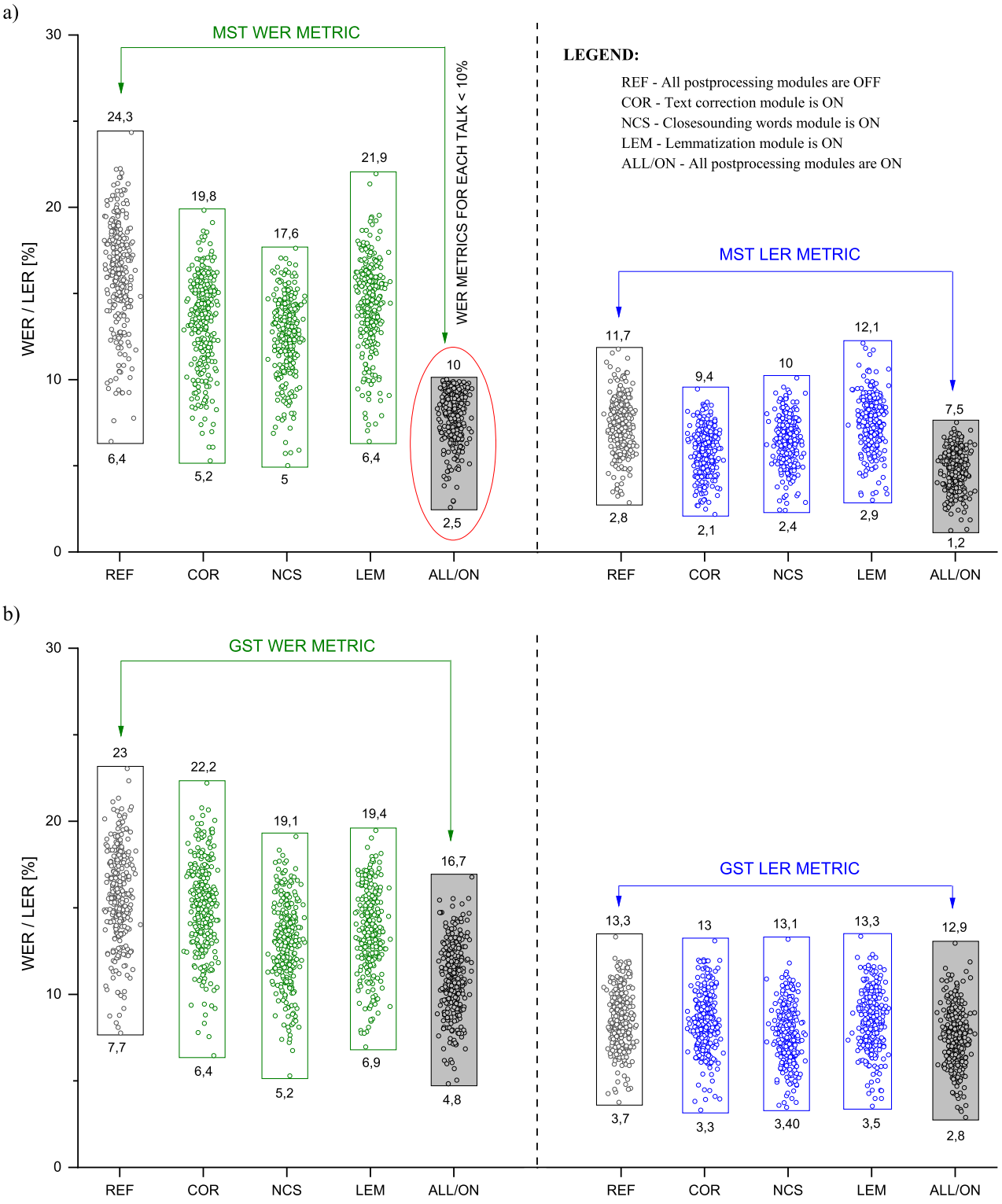


FIGURE 10. Impact of post-processing elements on transcription quality of integrated systems: a) MST and b) GST.

Polish, slightly better results were achieved by using the MST system.

B. RESULTS OF POSTPROCESSING ELEMENTS

Examinations of individual postprocessing elements were performed for a set of 300 randomly selected recording samples. The samples covered the three conversation topics declared in the section II: invoices and payments; technical information; and contracts and amending annexes. One hundred recordings were selected for each topic. Figure 10 shows the results of examination of the effect of the applied postprocessing elements on the transcription

TABLE 6. Average WER and LER metrics for postprocessing solutions.

METRIC	ASR	REF [%]	COR [%]	NCS [%]	LEM [%]	ALL/ON [%]
WER	*MST*	16,2	13,4	12,5	14,3	8,0
	GST	15,6	15,1	12,9	13,6	10,8
LER	*MST*	7,3	5,9	6,4	7,5	4,8
	GST	8,5	8,3	7,7	8,4	7,5

quality for the tested ASR systems, expressed by the WER and LER metrics. Figure 10a shows the results obtained from tests of the MST system, while Figure 10b summarizes the data from tests of the GST system. The postprocessing elements studied later were the individual modules discussed in section III.

The text correction module was examined first. As shown in Figure 10a, the module optimizes the WER metric to between 5.2% and 19.8% and the LER to between 2.1% and 9.4%. By using the block to optimize transcription performed by the GST system, it is possible to improve the WER metric to the level between 6.4% and 22.2% and the LER metric to the level between 3.3% and 13%, as shown in Figure 10b. The *close-sounding and foreign words* block improved the transcription quality compared to the reference values for the WER metric for the MST system from 5% to 17.6% and for the GST system from 5.2% to 19.1%. The impact of this module on the LER metric was not significant. The lemmatization module for both transcriptions (reference and automatic) resulted in the WER metric ranging from 6.4% to 21.9% in the case of the MST system and from 6.9% to 19.4% in the case of the GST system. The LER metric in both cases was at similar levels. The research shows that the lemmatization algorithm can be used for the Polish language, while for other languages its implementation requires additional research on the corpus of those languages. Particular attention should be paid to the average values shown in Table 6, which indicate that for the examined sets of recordings, the average transcription quality was at the level of 10.8% for the GST system and 8% for the MST system. Such very good results enable optimal use of transcriptions in the CC industry.

As can be seen from the research results shown in Figure 10, the post-processing algorithms used play a very important role in improving transcription quality. For both the MST and GST solutions, a very clear improvement in the WER metric over the reference values given in Table 6 is usually observed for each of the studied modules. It should also be noted that preprocessing for the MST solution requires an agent and client channel separation module.

VI. CONCLUSION

In this paper, a novel methodology of transcription dedicated to CCs was proposed. By adding the preprocessing and the postprocessing steps to the standard ASR system, the quality of transcription was significantly improved. The main novel steps that were applied in the methodology are separation of the interlocutors' channels, training of the ASR tool using CC datasets, audio signal correction, text correction, identification of close-sounding and foreign words, and a lemmatization algorithm.

The purpose of the research was to recommend an effective method of transcription of telephone conversations between client and agents conducted on a CC helpline in a Polish language model. In the research, special attention was paid to the low quality of audio signals, which is the main cause of problems with satisfactory level of automatic transcriptions performed by known ASR systems. In addition, it was pointed out that the English language models in ASR systems are much better refined and optimized than those of other less popular communication languages. As a result, the currently popular ASR systems used in CCs serving clients in different languages do not meet the expectations. On the other hand, a high enough quality of transcription is very important for further use of the conversation transcripts.

As shown in this paper, the quality of automatic transcriptions can be significantly improved by appropriate application of the preprocessing and postprocessing mechanisms described herein. For the random representative data sample comprising 300 voice calls, the average score for the WER metric in the case of the MST system reached 8%, which is a very good value in terms of further applications of the transcription method. The main advantage of the developed solution is the aforementioned high quality of the automatic transcriptions obtained. This allows the proposed method to be used to work directly in real time. Better transcription quality, in turn, translates into more possibilities to use the developed method, which consequently increases the effectiveness of the algorithms that use it and optimizes the profits achieved and costs incurred. Because the fees paid for the use of some ASR systems may be a certain limitation, the method was developed in such a way that it can be integrated with any ASR solutions operating in both non-gratuitous and free models. A rather difficult task, which constitutes another limitation of the proposed method, is proper configuration of the *close sounding and foreign words* module. The database for this module needs to be built for each subject-specific campaign separately and has to be adapted directly to the ASR system used, which is usually very labor-intensive.

In future work, it is planned to use the solution proposed in this paper to build a new method for recognizing the intentions of clients calling CC hotlines. The known intention recognition mechanisms rely solely on text data. Therefore, in order to recognize callers' intentions, it is necessary to make automatic transcriptions of voice calls. It is expected that improving the quality of transcription for real-time conversations will also increase the effectiveness of intention recognition. This will allow for further implementation of the solution in question in the target intelligent assistant (voicebot) mechanisms. Furthermore, research is planned on a new method for recognition of clients' emotions intended for direct use in CC systems. Using AI/ML algorithms, emotions

can be detected from both audio signals (phone calls) and text (chat conversations or transcripts of phone calls). Therefore, it is planned to use the proposed transcription method as one of the components of the emotion recognition method, which will support the process of recognition of emotional behaviour of clients and agents in audio channels. The high transcription quality level that will be achieved will therefore make it possible to build effective methods for recognizing intentions and emotions, which is the direction of our further research. We expect the results of our further work to improve the effectiveness and quality of conversations conducted by bots.

Moreover, the proposed method can be used in future CC systems. It is predicted [30] that the key definition of the term ''Customer'' will change by 2025. This will be driven by the rapid development of the Internet of Things (IoT) in many areas [31]–[34] as well as video technology [35]. Thanks to which items equipped with smart sensors capable of mutual communication will also be able to contact CC hotlines on their own. Hotlines of this type will be served mainly by smart bots which will ensure an adequate level of support.

REFERENCES

[1] A. Khalid, A. Sarfaraz, S. Ahmed, and F. Malik, ''Prevalence of stress among call center employees,'' *Pakistan J. Social Clin. Psychol.*, vol. 11, no. 2, pp. 58–62, 2013.

[2] M. Płaza and Ł. Pawlik, ''Influence of the contact center systems development on key performance indicators,'' *IEEE Access*, vol. 9, pp. 44580–44591, 2021, doi: 10.1109/ACCESS.2021.3066801.

[3] Z. Liu, C. Long, X. Lu, Z. Hu, J. Zhang, and Y. Wang, ''Which channel to ask my question? Personalized customer service request stream routing using deep reinforcement learning,'' *IEEE Access*, vol. 7, pp. 107744–107756, 2019, doi: 10.1109/access.2019.2932047.

[4] C. B. Nordheim, A. Følstad, and C. A. Bjørkli, ''An initial model of trust in chatbots for customer service—Findings from a questionnaire study,'' *Interacting Comput.*, vol. 31, no. 3, pp. 317–335, May 2019, doi: 10.1093/iwc/iwz022.

[5] G. Suciu, A. Pasat, T. Uşurelu, and E.-C. Popovici, ''Social media cloud contact center using chatbots,'' in *Proc. Int. Conf. Future Access Enablers Ubiquitous Intell. Infrastruct.*, in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 2019, pp. 437–442, doi: 10.1007/978-3-030-23976-3_39.

[6] J. Markof, ''From your mouth to your screen, transcribing takes the next step,'' The New York Times, Jun. 2020.

[7] B. Smagowska, ''Noise at workplaces in the call center,'' *Arch. Acoust.*, vol. 35, no. 2, pp. 253–264, May 2010. Accessed: May 1, 2021. [Online]. Available: https://acoustics.ippt.pan.pl/index.php/aa/article/view/250/240

[8] A. Narayanan, A. Misra, K. C. Sim, G. Pundak, A. Tripathi, M. Elfeky, P. Haghani, T. Strohman, and M. Bacchiani, ''Toward domain-invariant speech recognition via large scale training,'' Aug. 2018, *arXiv:1808.05312*. Accessed: May 1, 2021. [Online]. Available: http://arxiv.org/abs/1808.05312

[9] Microsoft. (2021). *Speech to Text | Microsoft Azure*. [Online]. Available: https://azure.microsoft.com/en-us/services/cognitive-services/ speech-to-text/

[10] Nuance Communications. (2020). *Automatic Speech Recognition Software For Customer Self Service*. [Online]. Available: https://www.nuance.com/omni-channel-customer-engagement/voice-and-ivr/automatic-speech-recognition.html

[11] Google Cloud. (2021). *Speech-to-Text: Automatic Speech Recognition*. [Online]. Available: https://cloud.google.com/speech-to-text

[12] VoiceLab. (2021). *Intent Recognition Use Cases*. [Online]. Available: https://voicelab.ai/#products

[13] GitHub. (2021). *Github Repository—Dictation Client*. [Online]. Available: https://github.com/techmo-pl/dictation-client

[14] Primespeech. (2017). *Primespeech ASR Server™ Telecom—Speech Technologies Ready to Prime*. [Online]. Available: https://primespeech.pl/oferta/primespeech-asr-server-telecom/

[15] GitHub. (2021). *Github Repository—DeepSpeech*. [Online]. Available: https://github.com/mozilla/DeepSpeech

[16] Amazon. (2021). *Amazon Transcribe—Speech to Text—AWS*. [Online]. Available: https://aws.amazon.com/transcribe/

[17] Facebook. (2018). *Wav2letter*. [Online]. Available: https://ai.facebook.com/tools/wav2letter/

[18] IBM. (2019). *Watson Speech to Text*. [Online]. Available: https://www.ibm.com/pl-pl/cloud/watson-speech-to-text

[19] N. Drake. (Nov. 27, 2020). Best cloud computing services of 2021: For digital transformation. TechRadar. [Online]. Available: https://www.techradar.com/best/best-cloud-computing-services

[20] N. Zeghidour, N. Usunier, G. Synnaeve, R. Collobert, and E. Dupoux, ''End-to-end speech recognition from the raw waveform,'' 2018, *arXiv:1806.07098*. [Online]. Available: http://arxiv.org/abs/1806.07098

[21] Y.-Y. Wang, A. Acero, and C. Chelba, ''Is word error rate a good indicator for spoken language understanding accuracy,'' in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Nov./Dec. 2003, pp. 577–582, doi: 10.1109/ASRU.2003.1318504.

[22] A. Zgank and Z. Kacic, ''Predicting the acoustic confusability between words for a speech recognition system using Levenshtein distance,'' *Electron. Electr. Eng.*, vol. 18, no. 8, pp. 81–84, Oct. 2012, doi: 10.5755/j01.eee.18.8.2628.

[23] D. Klakow and J. Peters, ''Testing the correlation of word error rate and perplexity,'' *Speech Commun.*, vol. 38, nos. 1–2, pp. 19–28, 2002, doi: 10.1016/s0167-6393(01)00041-3.

[24] Google. (2021). *2,3 Computing Error Rates—Text Digitisation*. [Online]. Available: https://sites.google.com/site/textdigitization/qualitymeasures/computingerrorrates

[25] Microsoft. (2021). *Custom Speech Overview—Speech Service—Azure Cognitive Services*. [Online]. Available: https://docs.microsoft.com/en-US/azure/cognitive-services/speech-service/custom-speech-overview

[26] Google Cloud. (2021). *Improve Transcription Results With Speech Adaptation*. [Online]. Available: https://cloud.google.com/speech-to-text/docs/speech-adaptation

[27] I. Ali, M. Asif, M. Shahbaz, A. Khalid, M. Rehman, and A. Guergachi, ''Text categorization approach for secure design pattern selection using software requirement specification,'' *IEEE Access*, vol. 6, pp. 73928–73939, 2018, doi: 10.1109/access.2018.2883077.

[28] M. Woliński, M. Miłkowski, M. Ogrodniczuk, A. Przepiórkowski, and Ł. Szałkiewicz, ''PoliMorf: A (not so) new open morphological dictionary for Polish,'' in *Proc. LREC*, 2012, pp. 860–864. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2012/pdf/263_Paper.pdf

[29] GitHub. (2020). *Github Repository—Morfologik-Stemming*. [Online]. Available: https://github.com/morfologik/morfologik-stemming

[30] T. Cottam. (2016). Contact centre 2025: Trends, opportunities, strategies. Telesperience. [Online]. Available: https://www.nice.com/optimizing-customer-engagements/Lists/WhitePapers/Contact_centre_2025.pdf

[31] M. Płaza, R. Belka, and Z. Szcześniak, ''Towards a different world—On the potential of the internet of everything,'' *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, vol. 9, no. 2, pp. 8–11, 2019.

[32] P. Pięta, S. Deniziak, R. Belka, M. Płaza, and M. Płaza, ''Multi-domain model for simulating smart IoT-based theme parks,'' *Proc. SPIE*, vol. 10808, pp. 867–878, Oct. 2018.

[33] R. Belka, S. R. Deniziak, M. Płaza, M. Hejduk, P. Pięta, M. Płaza, P. Czekaj, P. Wołowiec, and K. Ludwinek, ''Integrated visitor support system for tourism industry based on IoT technologies,'' *Proc. SPIE*, vol. 10808, pp. 447–454, Oct. 2018.

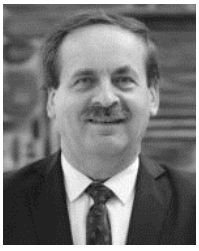
[34] A. Melnyk and V. Melnyk, ''Remote synthesis of computer devices for FPGA-based IoT nodes,'' in *Proc. 10th Int. Conf. Adv. Comput. Inf. Technol. (ACIT)*, Sep. 2020, pp. 254–259, doi: 10.1109/ACIT49673.2020.9208882.

[35] M. Królikowski, M. Płaza, and Z. Szcześniak, ''Chosen sources of signal interference in HD-TVI technology,'' *Proc. SPIE*, vol. 10445, Aug. 2017, Art. no. 104455M, doi: 10.1117/12.2280534.

MIROSŁAW PŁAZA (Member, IEEE) received the Doctor of Technical Sciences degree from Kielce University of Technology, in 2005. Since 2002, he has been a member of the Academic Staff at Kielce University of Technology. He is currently a Graduate Student with Kielce University of Technology, with a focus on telecommunications and computer engineering. He is also the Head of ICT and IoT CyberLAB and the Head and Instructor of the Cisco Networking Academy. He has conducted numerous research and development studies for EU-financed projects. At present, he is the Head of the Project titled ''EMOTICA AI–The Intelligent Contact Centre System.'' The results of his research include two patents as the main author (both received awards in the Świętokrzyski Racjonalizator Contest), one monograph, and over 50 scientific publications. His certificates include CCNA, CCNP, CCNA security/cybersecurity, the IoT, and big data and analytics. Since 2007, he has been a member of the Management Board of Kielce Branch of the Polish Society of Theoretical and Applied Electrical Engineering (PTETiS). From 2003 to 2009, he was the Vice President of AMICUS Foundation operating for Kielce University of Technology. As a result of his long cooperation with Cisco, Kielce University of Technology opened a joined specialization and major named ICT. The curriculum for this specialization was recognized at the national level and won the title of ''Leader of Cooperation with the IT Business'' in 2018. He was awarded (2015, 2017, and 2020) the status of Expert Instructor, the Above and Beyond Award (2018), and the Letter of Recommendation from the C++ Institute (2018) in recognition of increasing the quality of the C++ course offered to more than 9500 instructors around the world. He is a winner of the International European Kangaroo Mathematics Contest, where he won the main prize.

ŁUKASZ PAWLIK (Member, IEEE) received the M.S. degree in computer science from Kielce University of Technology, Kielce, Poland, in 2002.

Since 2000, he has been working at Altar Sp. z o.o., as a Programmer and a Technical Leader. During his work, he has been involved in the design and programming of contact center, workflow, and billing systems. He is currently leading the first stage of a research project on speech transcription and intention recognition in the field of contact center systems. The project is carried out jointly by Altar Sp. z o.o. and Kielce University of Technology. He is also involved in research on the impact of new technologies on the optimization of key performance indicators in contact centers. He has participated in many conferences in Poland, where he promoted new technologies in the call center industry, including ''Mobile applications in Contact Center services,'' ''Visual IVR in Contact Centers,'' and ''New possibilities in customer communication with a company.'' His interests include development of voicebots and chatbots using artificial intelligence.



STANISŁAW DENIZIAK (Member, IEEE) graduated at the Faculty of Electronics and Information Technologies, Warsaw University of Technology, in 1988. He received the Ph.D. degree from Gdańsk University of Technology, in 1994, and the D.Sc. degree from Warsaw University of Technology, in 2006. From 2012 to 2019, he was the Vice Dean of research and promotion. Since 2020, he has been the Dean of the Faculty of Electrical Engineering, Automatics and Computer Science, Kielce University of Technology. He is currently a Professor of computer science with the Department of Information Systems, Kielce University of Technology, Poland, where he is also the Head of the Division of Computer Science. He has published more than 130 research papers in various journals, books, and conferences. His research interests include design of embedded systems, the Internet of Things, logic synthesis for FPGAs, big data, and machine learning. He is a member of IEEE Computer Society. He is also a member of Program/Scientific Committee and a Reviewer of many scientific conferences, such as IEEE DDECS, RUC, SETIT, IEEE DAC, and EPMCCS. He is also an Active Reviewer and Editorial Board Member of many international journals, such as *Computer Networks*, *Microprocessors and Microsystems*, *Sensors*, *Remote Sensing*, IEEE Access, *Applied Sciences*, *Multimedia Tools and Applications*, *Journal of Systems and Software*, *Computing*, and *Computers&Security*.