Speech Communication 143 (2022) 33–45

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

A method for improving bot effectiveness by recognising implicit customer intent in contact centre conversations

Łukasz Pawlik [a], Mirosław Płaza [b,*], Stanisław Deniziak [b], Ewa Boksa [c]

[a] Altar Sp. z o.o., ul. Zagnańska 61, 25-528 Kielce, Poland
[b] Kielce University of Technology, Al. Tysiąclecia P.P. 7, 25-314 Kielce, Poland
[c] Jan Kochanowski University, ul. Żeromskiego 5, 25-369 Kielce, Poland

A R T I C L E I N F O

Keywords:
Contact centre
Intent recognition methods
Emotions in a conversation
Virtual assistant
Voicebot
Chatbot

A B S T R A C T

Contact centre systems are increasingly using intelligent voicebots and chatbots. These solutions are constantly evolving and improving. One of the main tasks of a virtual assistant is to recognise customers' intentions. The accuracy of intent recognition should reach the highest possible level, because higher effectiveness of intent recognition translates into higher effectiveness of customer service. The purpose of this paper is to propose a new method for recognising customer intent. The functionality of this method takes into account emotions that accompany particular conversations, which is a new approach to the problem of recognising intent and determining the target flow of a conversation by bots. The method developed by the authors is intended for direct application in the customer service industry. The present paper analyses currently known solutions in terms of the intent recognition methods used. A linguistic analysis of actual conversations from contact centre hotlines was conducted for both the voice channel and the text channel. This analysis shows the clear impact of emotions that accompany conversations on the correctness of operation of virtual assistants. The results presented further in the paper show that it is possible to improve the correctness of the actions taken by bots by as much as above 50%. The solution presented herein is expected to improve customer service and optimise many KPIs.

1. Introduction

ICT companies that implement call/contact centre (CC) systems are experiencing rapid growth. This is due, amongst other things, to the outbreak of the COVID-19 pandemic, which definitely transformed the model of interpersonal communication. A significant portion of both personal and business relationships is now conducted remotely. In all likelihood, we can assume that this forced trend, especially in business relations, will continue. As a result, CC hotline system solutions are increasingly being used in many industries. These systems are currently being developed with increasing use of artificial intelligence (AI) algorithms, including mainly machine learning (ML) and natural language processing (NLP) methods, as well as Big Data analytics. Implementation of solutions supported by AI algorithms is particularly important in the design of virtual assistants taking the form of chatbots and voicebots (Boonstra, 2021).

The basic challenge in chatbot/voicebot design is to correctly identify the intent expressed in customers' utterances made in contact centre conversations. Currently, there are many different methods on the market for recognising customer intent, which constitute excellent support platforms for designing virtual assistants that operate in both the voice channel and the text channel (Singh et al., 2019). They are available free of charge or on a commercial basis, most often in the SaaS (Software as a Service) model and sometimes in the On-premises model. These solutions offer functionalities that provide intent recognition capabilities for different language models. Some of the solutions (IBM, Watson Assistant: Visual Chatbot Builder, 2020; Google Cloud, Dialogflow CX, 2021) also enable designing complex conversation scenarios and provide programming APIs and mechanisms for creating custom plug-ins in different programming languages. Table 1 summarises the most well-known intent recognition systems.

As indicated in Table 1, some intent recognition methods also have integrated sentiment evaluation mechanisms (Google Cloud. Natural Language API Basics [WWW Document], 2021). However, in the existing solutions, sentiment determination is only possible for English and does not cover other, more complex languages, such as Polish (Google Cloud, Language reference [WWW Document], 2021). Additionally, sentiment is only recognised based on the text (Google Cloud. Natural

* Corresponding author.
E-mail address: m.plaza@tu.kielce.pl (M. Płaza).

https://doi.org/10.1016/j.specom.2022.07.003
Received 12 October 2021; Received in revised form 17 March 2022; Accepted 14 July 2022
Available online 16 July 2022

Language API Basics [WWW Document], 2021) and does not take into account the parameters of the audio signal. In this case, the conversation sentiment recognition scores are normalised and range from −1.0 to 1.0, allowing identification of negative, positive, and neutral sentiment. From the point of view of building intelligent bots that carry out conversations in CC hotlines automatically, the approach that takes sentiment into account is not fully sufficient. It would be much more beneficial to be able to identify the specific types of emotions present in CC conversations.

To the best of our knowledge regarding the methods of intent recognition in CC helplines, there are no solutions that take into account specific types of emotions for the Polish language that accompany a customer during a conversation with an agent or a virtual assistant. From a semantic point of view, the presence of a particular emotion is, in many cases, extremely important in the proper recognition of actual intent (Sehgal et al., 2018). For example, identical phrases uttered with *happiness* often have a completely different meaning than those uttered with *anger*, which also means that the actual intent of the caller is completely different. This observation was the main motivation for the authors to conduct detailed research with the aim of determining the impact of emotions present in CC conversations on the correctness of automatic conduct of conversations by virtual assistant systems. As a result of cooperation with psychologists and linguists specialising in CCs, five main types of emotions were identified, which are of the greatest importance in customer-agent conversations. These emotions are *happiness, anger, sadness, fear,* and *neutrality* (Płaza et al., 2022).

The main research problem of this paper is to show that by applying appropriate inference rules, taking into account the impact of recognised emotions on the actual intentions of the customer, it is possible to significantly increase the effectiveness of voicebots and chatbots used in CCs. To this end, a new method was developed to improve the recognition of CC customers' intentions when the recognition error is caused by the emotions that alter the bot's understanding of the utterances. The paper will demonstrate that this problem concerns a significant number of CC conversations.

The contribution to the body of knowledge of this paper is as follows:

- determination of the impact and importance of specific emotion types in intention recognition using examples from various datasets created from real CC conversations;
- development of universal inference rules that can be used with different conversational AI platforms based on customer emotions identified as entities;
- development of a comprehensive method intended specifically for the CC industry, providing the possibility of correcting wrongly recognised intentions based on recognised emotions;
- determination of the impact of transcription the effectiveness of intention recognition.

During the development of the proposed method, it was assumed that one of the components would be the existing intent recognition engines, which were selected based on the parameters summarised in Table 1. The main criteria that were taken into account were the need for native, high-level support for the Polish language model and the need to provide the possibility to design custom conversation scenarios. The language criterion was not sufficiently met by the Microsoft LUIS and IBM Watson Assistant solutions. ON the other hand, the Facebook Wit.ai and VoiceLab Intent Recognition solutions do not provide the ability to create custom conversation scenarios. Therefore, the proposed method uses the Google Dialogflow and RASA platforms. Dialogflow is a commercial platform operating in a cloud model, while RASA is an open-source platform providing software in the on-premises model. Due to the planned commercial implementation of the solution described herein, the research was conducted on the Polish language model; however, the method described in the paper can be adapted and generalised to other language models as well.

The rest of the paper is organised as follows: Section 2 explains in detail the impact of emotions present in customers' utterances on the correctness of intent recognition from the linguistic and psychological perspective. A number of examples of real-life dialogues between customers and agents, obtained directly from a large CC helpline, were collated. The collected examples are meant to indicate the direct impact of emotions on the correctness of customer intent recognition, which is the motivation for the research described herein. Section 3 presents the research methodology and includes the definitions of the basic concepts and components of an intent recognition system. Section 4 describes the new method proposed by the authors for recognising customer intent, which is designed for direct application in CC systems. The proposed method takes into account in the actions taken by bots the emotional states that accompany the conversations. Section 5 describes the research results. Section 6 presents the conclusions and the directions for further research.

2. Intent and emotions in the utterances of cc customers in the linguistic approach

Interpersonal communication is about revealing intent (Sperber and Wilson, 1986; Taylor, 1995; Zhezhnych et al., 2019). The sender of a message firstly informs the recipient about something and secondly consciously or unconsciously reveals his or her goals and needs. Thus, the following pattern of the communication process can be presented: X (the sender), by saying p (content), has q (meaning, sense) in mind; that is, X, by saying p, reveals his or her intent to communicate q. Understanding of the implicit (inferred from the context) meanings and interaction between the literal meaning and the relevant information contained in the situational context are possible not so much through the understanding of the literal meaning in a communication, but through its coincidence with the emotional layer (Batliner and Schuller, 2013). In humans, the above process, where one arrives at the recognition of a new statement on the basis of sentences that have already been accepted as true, is perfectly natural. Natural language is characterised by great flexibility, allowing the same content to be expressed in many ways. However, this assumption is not the basis for language learning by automated intelligent bots. In order for a bot to correctly recognise the

Table 1
Selected intent recognition systems that can be implemented in CC helplines.

No.	Name of conversational AI platform	Polish language	Building scenario	License	API	Software model On premises	Software model SaaS	Sentiment	References
1.	Microsoft LUIS	No	Yes	Commercial	Yes	No	Yes	No	(Aahill, 2021)
2.	Google Dialogflow	Yes	Yes	Commercial	Yes	No	Yes	Yes	(Sabharwal and Agrawal, 2020)
3.	Facebook Wit.ai	Yes	No	Commercial	Yes	No	Yes	No	(Wit.ai [WWW Document], 2021)
4.	IBM Watson Assistant	No	Yes	Commercial	Yes	Yes	Yes	No	(IBM, Watson Assistant: Intelligent Virtual Agent, 2021)
5.	VoiceLab Intent Recognition	Yes	No	Commercial	Yes	Yes	Yes	No	(VoiceLab.ai, 2021)
6.	Rasa	Yes	Yes	Open source	Yes	Yes	No	No	(Rasa, Training Data Format [WWW Document], 2021)

meaning of utterances addressed to it by a customer, it must learn the meanings of certain words or phrases used in interpersonal communication in specific situations. Consequently, the starting point for the development of algorithms for communication between customers and virtual agents is the recognition of the language processing mechanism in the customer-agent diade.

This paper focuses on agent-customer interactions in a real contact centre. The goal of our research is to develop a method that allows bots to correctly conduct conversations based on recognised intent in dialogues between a customer and an agent, while taking into account the emotions recognised in the customer's utterances. The motivation for our research was a linguistic analysis of the structures of selected acts of speech in coincidence with the customer's emotional factor. The material for the analysis consisted of actual contact centre conversations between agents and customers. Table 2 shows examples where understanding a customer's utterance requires some knowledge of the world, the current situation, and the ability to recognise other nonverbal communication signs. The table also denotes the types of emotions identified by psychologists in the CC dialogues presented.

As an analysis of the examples given in Table 2 shows, the interpretation of words or phrases in conjunction with a particular type of emotion can completely change the direction of a conversation. This is related to the understanding of the meanings conveyed by the customer verbally, but also emotionally. The same phrases uttered with a specific emotion have a meaning that is different from the meaning of the same phrases devoid of an emotional background (Roche et al., 2015). The task of virtual assistants should therefore be to be able to carry out conversations in such a way that dialogues can be made dependant on both recognised intent and recognised emotions, if any. The bot will then be able to treat a phrase with a negative emotion differently from the same phrase with a positive or neutral emotion. Therefore, in communication between a human and a bot, the term intent should be clarified. In this approach, intent is a communicative value with an emotional background, as shown in Fig. 1.

Recognition of emotions contained in an utterance can be based on an analysis of the content of the utterance, as well as on an analysis of prosodic or cepstral features of the voice (Koolagudi and Rao, 2012). This is often done using machine learning methods, mainly methods for classification of selected voice features, and deep learning (Khalil et al., 2019; Płaza et al., 2018). Other methods are based on Fourier parameters (Nwe et al., 2003) or hidden Markov chains (Wang et al., 2015). The existing methods can achieve emotion recognition accuracy of more than 70% (Bota et al., 2019). There are also attempts to build empathy into virtual assistant mechanisms (Ghandeharioun et al., 2019; (Kyo-Joong et al., 2017) and to improve the human-computer interface (Ramakrishnan and El Emary, 2011). However, there are no known methods that make it possible to correct improperly recognised CC customers' intentions when the recognition error is caused by the

Table 2
Selected examples of information intent correlated with emotions of anger, happiness, fear, and sadness.

No.	C/A	Conversation excerpt	Literal meaning (no emotional background)	Recognised emotion	Implied meaning with emotional background
EXAMPLES OF REAL CONVERSATIONS FROM THE VOICE CHANNEL					
1.	C:	*I'm sorry, but it's wrong to pay for something I didn't receive.*	I am grateful to be able to use the company's services	Anger	I am sick of dealing with this company.
	A:	*I see.*			
	C:	***Thank you for being such a company.***			
		Conversation analysis: A statement, an affirmative sentence.			
2.	A:	*Have you read the terms and conditions of the selected service and do you agree with them?*	I have read the terms and conditions of the service.	Fear	I have not read the terms and conditions, but in order for my case to be considered I confirmed this fact.
	C:	*Okay,* ***I accept and confirm***			
		Conversation analysis: An affirmative sentence; the use of the word *okay* reinforces the fact that the terms of service have been formally approved, but its use before the verb *confirm* indicates that the customer is unsure whether he knows the contents of the terms and conditions.			
3.	A:	*Did the contact with the hotline help resolve the issue? After the call you will receive a text message.*	Thank you, I will definitely respond.	Happiness	Thank you I will definitely respond, the review will be positive.
	C:	***Thank you, I will definitely respond.***			
		Conversation analysis: An affirmative sentence; the use of the word *definitely* indicates that the customer, whose case was resolved as expected, will praise the company for its effectiveness.			
4.	A:	*Do you want to cancel the number you are calling from ?*	I want to cancel all numbers.	Sadness	I am disappointed and if you don't resolve the issue I will cancel all numbers.
	C:	**I will cancel all numbers.**			
		Conversation analysis: An affirmative sentence. The use of the universalising pronoun in the "**all** numbers" phrase indicates the customer's great disappointment and serves to generalise and emphasise the negative emotions the customer has towards the company. This kind of generalisation is one of the communication barriers that consists in exaggeration in attributing to the company even those mistakes it did not make.			
EXAMPLES OF REAL CONVERSATIONS FROM THE TEXT CHANNEL					
1.	A:	*How can I help you?*	Oh, we are meeting again.	Sadness	I'm not happy that you are going to serve me again.
	C:	***Oh, it's you again!:(***			
	A:	*I will try to help you as best I can.*			
		An exclamation, using a sad face emoticon, indicates the customer's disappointment that he or she will be served by someone whose services the customer was not satisfied with.			
2.	C:	*Please activate the subscription.*	Thank you for your help, the subscription has been activated.	Happiness	You helped me. That's what I wanted.
	A:	*The subscription has been activated. Is there anything else I can do to help?*			
	C:	***Thanks for your help for now.*** *Bye:)*			
		Conversation analysis: The use in the utterance of colloquial words, such as *thanks* and *for now*, emphasises the completion of an activity.			
3.	C:	*Then who is able to answer my question?*	I'm very satisfied.	Anger	I am unhappy, even angry, that no one has answered my question.
	A:	*Unfortunately, no one.*			
	C:	***I'm very satisfied!!!:(****Good bye!!!*			
		Conversation analysis: An exclamation sentence; the customer's dissatisfaction is reinforced by exclamation marks at the end of the phrase and a sad face emoticon.			
4.	A:	*Your request has been sent, your request number is 123–456.*	I'll be without a phone number for a long time	Fear	My phone will not be available for a long time and I'm concerned about it.
	C:	*The request number is of no use to me, they will work on the repair for a week and* ***I will be without a phone during this whole time*** ☹			
	A:	*Service provision may commence within 24 h of the Activation.*			
		Conversation analysis: An affirmative sentence, the content of which is reinforced by a sad face emoticon, reinforces the customer's concern that he or she will not be able to use the phone and will suffer some inconvenience.			

C – Customer, A - Agent.

emotions that alter the actual bot's understanding of the utterance.

Automatic caller intent recognition is the subject of research related to the development of bots, recommender systems, and customer service systems. In the AntProphet system (Chen et al., 2019) customer intent is determined using a machine learning method based on the customer's past behaviour. The SBotScope system (Zhang et al., 2013) was developed to identify the intent in queries entered in web search engines. The main goal was to identify malicious behaviour, such as detecting vulnerabilities in systems or collecting email addresses. Also, the smart Search Bot (Gupta et al., 2012) was developed to recognise the intent in queries. In this case, however, the goal is to learn to recognise user intent in order to produce the most relevant search results. Recognition of customer intents based on machine learning methods was applied in Tian et al. (2020). For this purpose, a Knowledge Graph was used to support inference of fuzzy user requirements. The above-mentioned methods of intent recognition are based on an analysis of the content of utterances and the history of the caller's behaviour. As shown above, this is not always sufficient and bots often do not correctly recognise the user's true intents.

The authors emphasise that, in the context of CCs, they have no knowledge of any empirical studies on natural language processing that would take into account the impact of emotional states present in utterances communicated by senders on their comprehension by recipients. However, the feature of intent related to the expression of emotional attitudes seems to be particularly relevant from the point of view of communicative interaction. Taking into account the role of emotions, in particular emotional attitudes which constitute the core of intent, in processing phrases expressed literally or non-literally in a system for processing new data by a bot, can significantly increase the level of understanding of customers' utterances by the bot, which results in an increase in the number of solved cases.

3. Research methodology

The process used so far for bots to determine responses is shown in Fig. 2. In order for a virtual assistant to start a conversation, it is necessary to provide text to the input of the conversational platform in the form of a text phrase. Based on the utterance provided, the conversational platform recognises the actual intent of the customer. After correct recognition of the intent, the response to be given to the customer by the bot (bot utter action) is sent to the output (Ohta et al., 2021).

In order to use a given conversational platform for a specific application, it is necessary to first perform a learning process using the machine learning mechanism built into the conversational platform. Teaching the conversational system used by the bot requires defining the following sets:

- TRP: a training phrase set for teaching the platform;
- INT: an intent set that specifies the possible intent of a caller;
- TEP: a set of testing phrases used by customers, on the basis of which later functional tests are performed, leading to the determination of the correctness of bot actions;
- BUA: a set of possible bot responses (bot utter actions) for particular intent and input phrases

As shown in Section 2, due to the unique characteristics of CC conversations, there are many instances when responses to automatically recognised customer intent should be different depending on the emotions that are present in each conversation. Therefore, our approach uses a proprietary method to recognise customer emotions in CC audio channels and text channels (Płaza et al., 2022). In the proposed method (Fig. 3), an input phrase and the value of the recognised emotion are delivered to the input of the conversational platform. The input phrase undergoes standard processing by the selected conversational platform, as a result of which the intent is recognised. The recognised intent is then adjusted according to the recognised emotions in the customer's utterance; thus the conversational dialogue is switched according to the defined inference rules for the specific types of emotions recognised for the previously recognised intent. The output receives a response to the customer, generated by the bot on the basis of the prediction of the customer's intent taking into account the customer's emotional state. In the Fig. 3, the additional colour indicates new elements in the proposed approach to the subject matter described in the paper.

Table 3 presents excerpts of customers' utterances selected directly from actual CC conversations, for which, depending on the emotional state of the customers, the inferred actions performed by the bot should be completely different. The examples shown concern both the voice channel and the text channel.

An entity mechanism was chosen to convey information about recognised emotions (Google Cloud, Entities [WWW Document], 2021; Rasa, Open source conversational, 2020).

This mechanism is a standard for various conversational AI platforms (Deloitte Digital, 2019). An entity defines the type of specific information that can be extracted from the input data. It corresponds to many different common data types, such as entities matching dates, email addresses, or times. It is also possible to create custom entities to match custom data. In order to implement an intent recognition method that takes into account customer emotions, a custom entity named emotionEntity was defined for the purpose of transferring data about recognised emotions. In the method in question, the emotion types specified above are transferred as entities. The use of entity mechanisms allows arbitrary values to be transferred in phrases and does not adversely affect the learning and testing processes.

In order to take into account the emotions associated with customer utterances, our approach further defines a set of possible emotions, defined as EM={*ANGER, SADNESS, FEAR, HAPPINESS, NEUTRAL*}. This set also specifies all possible values of the emotionEntity added to the input phrases by the emotion recognition system. If the emotion recognition method does not indicate any emotion, then the value of the entity remains empty, indicating the neutral emotional state of the caller {*NEUTRAL*}. According to the proposed approach, developing a bot for a specific CC consists of the following steps:

1. Identification of the INT and BUA sets: these sets are determined on the basis of an analysis of the cases being discussed in a given CC campaign.
2. Teaching the conversation platform:
   - Developing the TRP and TEP sets: the phrase sets should be developed based on actual conversations between agents and customers conducted in a given CC.

Fig. 1. Informative intent in communication between a human and a virtual assistant.

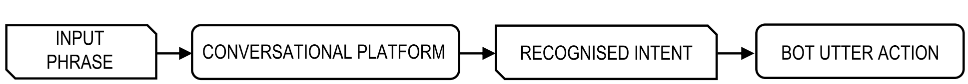


Fig. 2. Process for determining the responses given by bots.

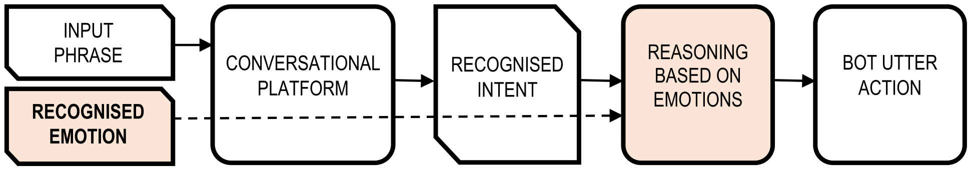


Fig. 3. Bot response process that takes into account the emotional state of the customer.

Table 3
Examples of building a conversation flow depending on the identified emotions.

CHANNEL	CUSTOMER PHRASE	INTENT	EMOTION		BOT UTTER ACTION
Voice	*Thank you for being such a company.*	Satisfaction with customer service	Anger, sadness, fear	→	The customer is dissatisfied, ask why.
			Happiness	→	The customer is satisfied, thank him or herand send a survey.
			Neutral	→	The customer is satisfied, thank him or her.
Voice	*Okay, I accept and confirm*	Approving anything	Sadness, fear	→	Confirms, but make sure
			Happiness, anger, neutral	→	Receive the confirmation.
Voice	*Well, I didn't get it.*	Not approving anything	Anger, sadness, fear	→	Denies, but make sure.
			Happiness, neutral	→	Accept the refusal.
Voice	*I will cancel all numbers*	Termination of the contract.	Sadness, fear	→	Make sure there is an intent to cancel.
			Anger	→	Cancels, make a better offer.
			Neutral	→	Accept the cancelling.
Text	*I can't reset my password.*	Password recovery	Anger, sadness	→	Password reset by an agent.
			Happiness, fear, neutral	→	Password reset via email.
Text	*Reinstalling the application had no effect.*	Application problems.	Anger	→	Application reset by an agent.
			Sadness, fear, happiness, neutral	→	Application reinstalled by the customer.
Text	*Please connect me with an agent.*	Connect with an agent.	Anger	→	Connect with an agent immediately.
			Sadness, fear	→	Put in a waiting queue for an agent.
			Happiness, neutral	→	Try to solve the problem automatically.

- Platform training: using the TRP set, the system must be taught to recognise intent from the INT set.
- Verification of the correctness of the identified intent.

3 Developing inference rules based on recognised emotions: the set of rules forms an inference mechanism that selects an appropriate action from the BUA set.

4 Bot validation: testing and evaluating bot actions using the TEP set. In case of unsatisfactory results, return to step 2.

The effectiveness of the proposed method in recognising customer intent and making bots take appropriate actions was demonstrated by developing a complete flow of customer-agent conversation processing for a sample CC campaign. Next, experiments were performed to compare the results obtained using this method with the results obtained with the solutions used so far.

4. Intentions recognition environment for CC applications

The proposed intent recognition method was integrated with other conversation analysis methods to form a complete environment for developing chatbots/voicebots (Fig. 4). The method uses recognised emotions in voice channels and text channels. The environment consists of the following components:

- a module for transcription of conversations originating from CC audio channels;
- a module for recognition of emotions in the utterances of CC customers;
- a module for preparation of training and test data; and
- an intent recognition module that takes into account recognised emotions.

As a result of the cooperation between the individual components, in the first phase the emotion recognition system is taught and the rules of inference are created, and in the second phase of operation of the method, a specific action is inferred by the virtual assistant on the basis of the learned rules.

In the Fig. 4, the additional colour indicates new elements in the proposed approach to the subject matter described in the paper.

4.1. Voice call transcription module

The voice call transcription method used (Płaza et al., 2021) is intended directly for use in CC systems. This method makes it possible to increase the efficiency of automatic transcription of calls by using selected algorithms for pre-processing of call recordings and post-processing of the original automatic transcription. The pre-processing part consists of three modules: channels separation module, training of the ASR (Automatic Speech Recognition) systems module, and audio signal correction module. As far as postprocessing is concerned, depending on the needs, the following modules may be used: text correction module, close-sounding and foreign words module, and lemmatisation module. The developed pre-processing and postprocessing methods have been shown to achieve transcription quality of 90–92% compared to 86% (Narayanan et al., 2018) achieved by currently used approaches. Some of the preprocessing and post-processing methods are used directly during intention recognition processes.

4.2. Emotion recognition module

The emotion recognition method (Płaza et al., 2022) uses artificial intelligence and machine learning mechanisms, as well as natural language processing methods to recognise emotions in caller utterances. The emotion recognition tool includes a voice analyser and a text analyser. With their help, features are extracted that contain the

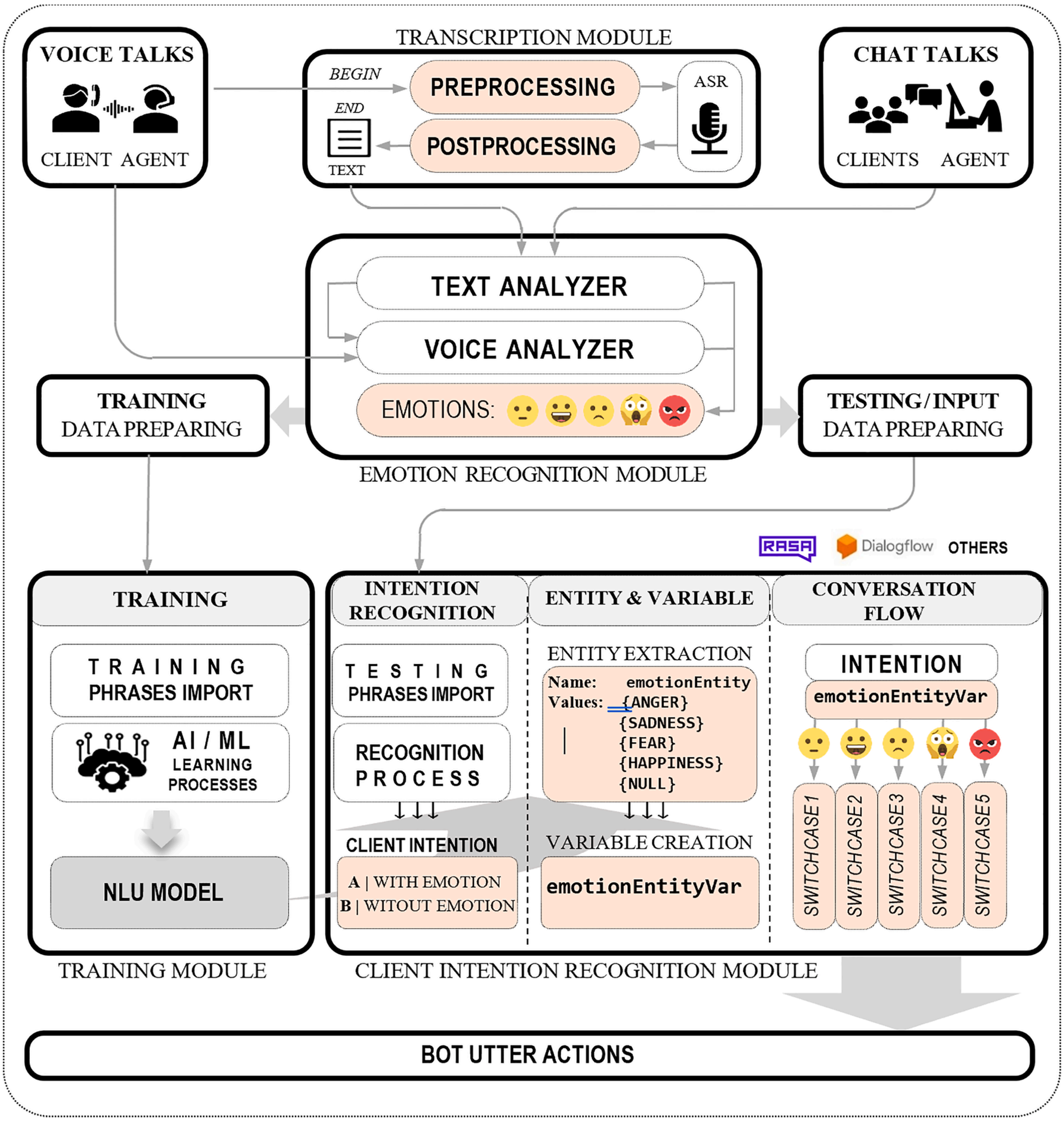


Fig. 4. Intention recognition environment for a contact centre application.

information necessary to determine the relevant emotional state of the customer (Cheang and Pell, 2008; Akçay and Oğuz, 2020). These features are the input of the classifier responsible for determination of the emotion type. The voice analysis process is carried out based on the speech signal parameters selected in the course of research, which resulted in the best convergence in determination of emotions. The following speech signal descriptors were chosen: basic frequency f_0, melcepstral coefficients *MFCC* and deviations described by the *jitter* and *shimmer* parameters. The method uses the first thirteen *MFCC* coefficients describing the frequency parameters of speech, as they contain most of the information regarding the recognised emotions. The basic frequency f_0 contains pitch information and, therefore, enables taking gender into account. These features are extended to include voice frequency changes during speech, i.e., the *jitter* parameter and amplitude changes described by the *shimmer* parameter. The process of emotion analysis in text channels, on the other hand, is based on property vectors. This technique enabled a semantic analysis of words, taking into account the context of their occurrence. In addition, the TF (Term Frequency) technique was used, which is based on counting the occurrences in the text of key words that have emotional connotations. In order to improve the performance of this technique, keywords that occur in all texts regardless of the emotion present were also considered, for which the IDV (Inverse Document Frequency) method was used.

4.3. *Creating the training and testing data sets*

TRAINING DATA (TRP set) is prepared in the form of training phrases using developed algorithms that analyse data in the form of chat conversations (for the text channel) or in the form of transcriptions of voice conversations (for the audio channel). The texts of the conversations are supplemented with information about the recognised emotional states of the caller. Consequently, training phrases ultimately contain the text of the selected utterance, which may be supplemented with a value describing the recognised emotion. TRAINING DATA is transferred to the learning module of the conversational system.

The module for the preparation of test and input data is designed for testing the solution in a production environment during face-to-face conversations on CC hotlines. TESTING DATA (TEP set) is used during the verification of the system operation in laboratory conditions, while INPUT DATA is the data used during the direct use of the developed solution. This module analyses conversations together with the emotions defined for them and transfers selected phrases to the conversational platform. Like the TESTING PHRASES, TRAINING PHRASES contain the text of the utterance and possibly also added information about the emotion present in the form of the emotionEntityValue entity.

4.4. Teaching the conversational platform

The TRAINING module is responsible for the process of import of the training phrases and for the training process performed using the imported phrases, which should also contain values from the emotionEntity set. The key result of the training process is the creation of the NLU (Natural Language Understanding) model, which is the basic element used by conversational platforms in recognising intent from test phrases and extracting entity values from the recognised intent. In the proposed solution, this model is important in the operation of the INTENT RECOGNITION module and in the ENTITY EXTRACTION process. To activate the entity mechanisms, each determined intent is learned, as schematically shown in Fig. 5. Training is done:

- with phrases containing relevant entities that define emotion types (A),
- with phrases that do not contain entities (B).

B phrases provide the ability to define a neutral emotional state.

4.5. Customer intent recognition module

The most significant functional component is the developed customer intent recognition module, with which Dialogflow and RASA platforms are integrated. From the point of view of operation of the proposed method, the above two platforms are not the only ones that can be used. The module shown in Fig. 4 is composed of the following blocks: INTENT RECOGNITION, ENTITY & VARIABLE, and CONVERSATION FLOW.

For the model trained according to the assumptions described herein, test phrases containing a specific emotion or indicating a neutral state of the caller may appear as the test data imported into the INTENT RECOGNITION block. In case A, when an emotion is recognised, then information about the emotion is appended to the end of the test phrase. Later, this information is separated as an entity and stored in a variable named emotionEntityVar, which takes place in the ENTITY & VARIABLE block. When information about a particular emotion is reduced to the form of a variable, it is possible to create inference rules that determine the responses used by bots. In the next step, the CONVERSATION FLOW block implements the target conversation flow using the new approach. It is not only based on the recognised intent, but also enables making the response given by the bot dependant on the recognised customer emotions stored in the emotionEntityVar variable. The inference rules developed, which are built on the emotions contained in the utterance, are responsible for the selection of the appropriate answer. In this way, the BOT UTTER ACTION response takes into account both the text of the utterance and the emotions it contains. All this results in a response that is appropriate for the actual situation based on which the customer's actual intent (A) is determined. When emotions are not recognised, bots do not have to choose alternative scenarios for the conversation (B). Fig. 6 shows a diagram of the process of development of the conversation flow depending on the received parameters.

The implementation of the way in which a conversation is conducted depending on the value of emotionEntity is different for the two platforms studied. In the case of the Dialogflow platform, the parameters embedded in individual intents are used. These parameters are automatically populated with the value of the entity, and the language of the conditions available on this platform is used to determine the further course of the conversation. Based on those conditions, the next steps of the conversation are determined. The RASA platform uses slots that act as variables that store entity values. Based on the filled slots, depending on their value, a specific action (bot utter action) is determined, i.e. the answer given by the virtual assistant.

5. Results and discussion

In order to evaluate the effectiveness of the method proposed by the authors, experiments were carried out to assess the impact of customers' emotions present during conversations on CC hotlines on the correctness of intent recognition. Since the performance of the Dialogflow and RASA platforms may depend on the quality of the input data, three groups of experiments were performed. In the first step, the impact of transcription quality on the intent recognition performance of the Dialogflow and RASA platforms was investigated, as shown in Section 5.1. In the second step, described in Section 5.2, the impact of the emotions recognised in the conversations on the correctness of the recognised intents was tested. Section 5.3, on the other hand, describes the results obtained during tests of the complete solution, which takes into account both the recognised emotions and the pre-processing and postprocessing elements in the automatic transcriptions performed.

For the purpose of the research, two databases were used, which contained a total of:

- 1054 actual calls made in the audio channels of a CC system and
- 545 actual chat conversations conducted in text channels.

The collection of voice calls used comprises a total of 100 h 53 min and 41 s of recordings. All recordings were transcribed manually (by a human) and automatically (by the system). An important element in the process of intent recognition is the quality of the automatic transcriptions; therefore, a proprietary transcription method designed for the CC industry was used in this study (Płaza et al., 2021). Based on the

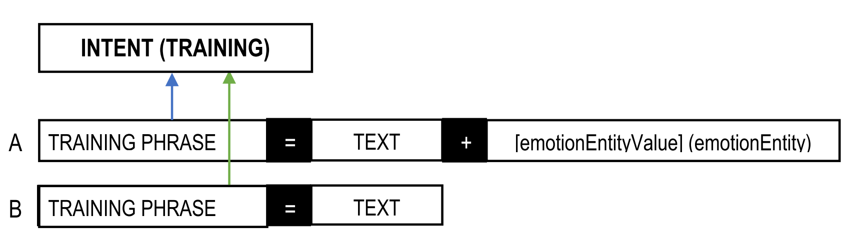


Fig. 5. Diagram of the training process used in the intent recognition method.

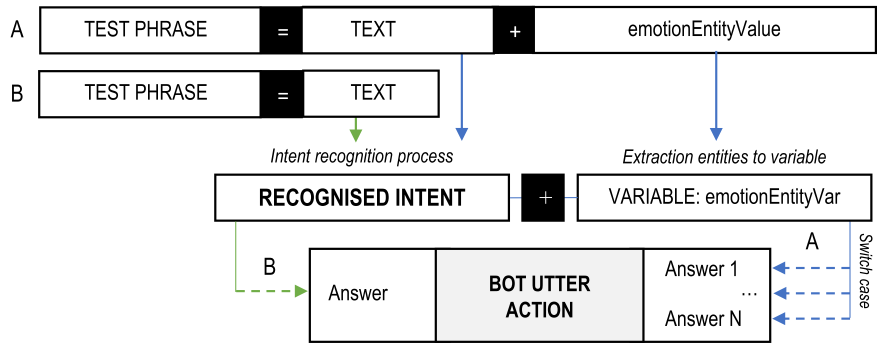


Fig. 6. Illustrative diagram of the conversation flow development process used in the intent recognition method.

conversations collected in the databases, the 15 most frequent intents were selected and divided into three groups with a breakdown into both communication channels (Table 4). The first group comprises intents for which customer emotions matter the most; the second group comprises intents with medium importance of emotions, and the third group comprises a set of intents where the customer's emotions matter the least.

For each of the selected intents, 25 training phrases (for the audio channel, these phrases were selected from the manual transcriptions) and 15 test phrases (in the audio channel, these phrases were from the automatic transcriptions) were prepared.

5.1. The impact of transcription quality on intent recognition performance

In CC systems, there are two basic communication channels: the voice channel and the text channel. Since customer intent recognition methods work using only textual data, voice calls must undergo a prior transcription process. Depending on the ASR system used, the quality of this transcription usually varies greatly. It was therefore necessary to check how the quality of the transcription of the calls at a CC, obtained with the method used in our approach, affects the process of intent recognition. For this purpose, a comparative study was performed using automatic transcriptions generated by Microsoft's basic ASR system (Microsoft Speech to Text) and transcriptions made using a proprietary method designed specifically for the CC industry (Płaza et al., 2021). Table 5 summarises the results of the study showing the effectiveness of customer intent recognition for the aforementioned automatic transcription methods.

For the data in Table 5 which is shaded and marked with symbols (a) to (c), the corresponding confusion matrices are summarised in Fig. 7. They illustrate the detailed distributions of the results obtained, which show the correctness of the recognition of particular intents for the test phrases. The first row in all columns shows the data obtained using the basic ASR system, while the second row contains the distributions for the data obtained by using the proprietary solution in the functionality of the ASR system.

On the basis of the tests performed, the observation was made for the defined individual groups (1, 2, 3) of intent (V1-V5) that the applied automatic transcription method designed for the CC industry clearly improves the effectiveness of correct intent recognition. Depending on the tested sample, for the Dialogflow platform, the improvement ranges from 1.33% to as much as 18.67%, while for the RASA platform it ranges from 2.66% to 6.67%. The described results translate directly into improved bot utter action. Depending on the tested sample, for the Dialogflow platform, the improvement related to the bot utter action ranges from 1.33% to as much as 18.67%, while for the RASA platform it ranges from 4% to 8%. As the NLU model used was trained with training phrases from manual (human-made) transcriptions, these transcriptions were improved so as to be error-free. However, errors did occur in the automatic transcriptions that were drawn randomly as the test phrases.

Table 4
Individual groups of selected intents.

Group	SELECTED INTENTS ID	Voice channel	ID	Text channel	Potential influence of customer emotions
1	V1:	Approving anything	T1:	Approving anything	Large
	V2:	Not approving anything	T2:	Not approving anything	
	V3:	Authorisation	T3:	Authorisation	
	V4:	Termination of the contract	T4:	Termination of the contract	
	V5:	Satisfaction with customer service	T5:	Satisfaction with customer service	
2	V1:	Service status	T1:	Password recovery	Medium
	V2:	Contract information	T2:	Subscription activation	
	V3:	Internet service activation	T3:	Application problems	
	V4:	Device repair	T4:	Internet service activation	
	V5:	Contract extension	T5:	Issue status	
3	V1:	Invoice information	T1:	Service status	Small
	V2:	Internet connection problem	T2:	Connect with agent	
	V3:	Connect with agent	T3:	SIM activation	
	V4:	Damage description	T4:	Internet connection problem	
	V5:	Password recovery	T5:	Solution deadline	

Legend: V1-V5 - designation of an individual intent in a given group for the voice channel; T1-T5 - designation of an individual intent in a given group for the text channel.

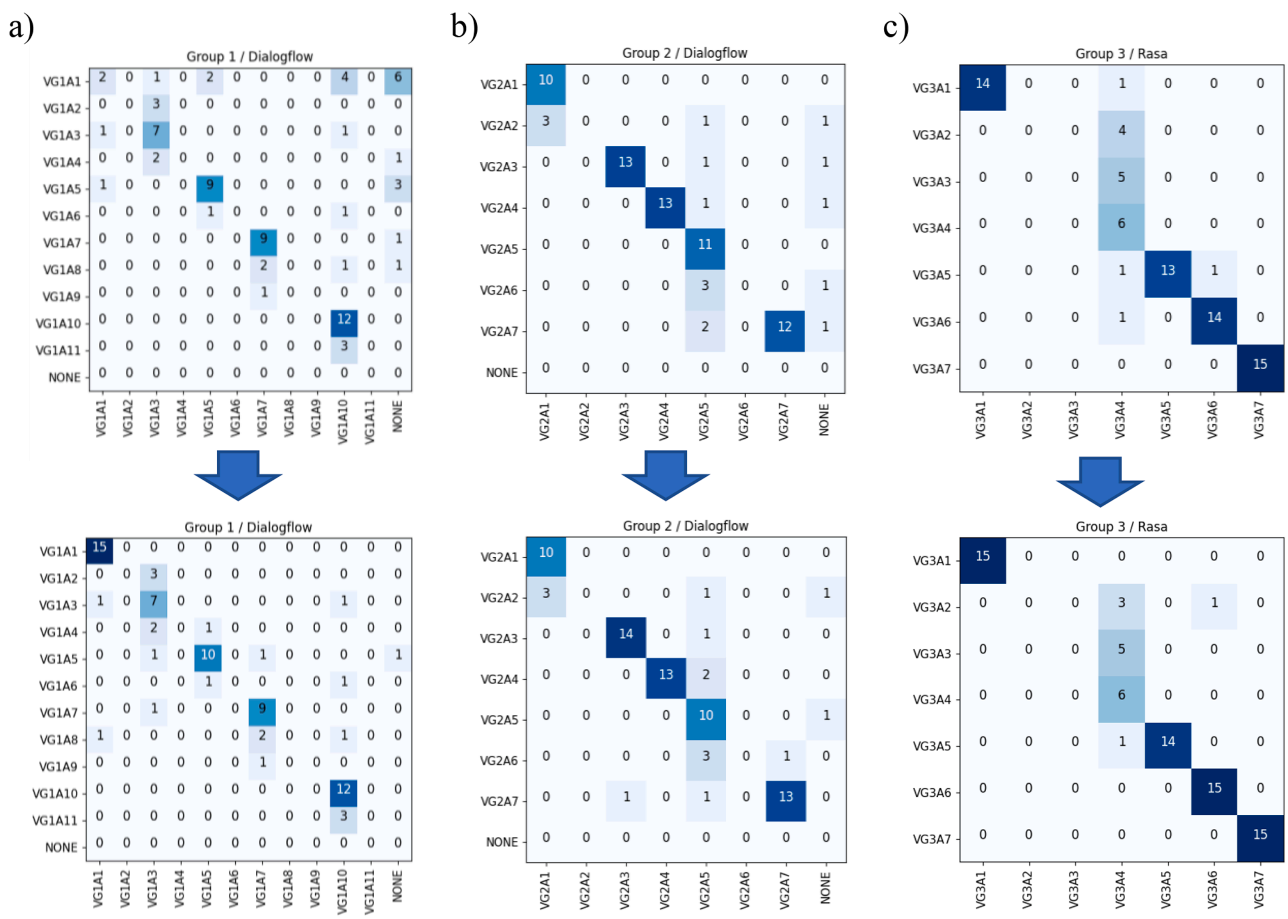
Nevertheless, far fewer errors occur when using the transcription method proposed by the authors, which includes pre-processing and postprocessing elements.

This is clearly visible in the selected confusion matrices shown in Fig. 7, especially for the data in group 1 (Fig. 7a). In group 2, the number of errors in the automatic transcriptions was small, hence the results did not improve very much (Fig. 7b). In the set of test phrases contained in group 3, the improvement was quite pronounced (Fig. 7c) and was due to the moderate number of errors in the automatic transcriptions. The above discussion confirms the desirability of implementation of an efficient automatic transcription method in the proposed solution. By improving the quality of the transcription, the performance of the bot utter actions was improved. However, as these results were not fully

Table 5
The effect of transcription quality on the accuracy of customer intent recognition.

Group	Conversational AI platform	RECOGNISING INTENT Basic ASR method voice [%]	pcs/75	Our ASR method voice [%]	pcs/75	BOT UTTER ACTION Basic ASR method voice [%]	pcs/75	Our ASR method Voice [%]	pcs/75
1	Dialogflow	68	51	86.67	65	52 [(a)]	39	70.67 [(a)]	53
	Rasa	81.33	61	88	66	61.33	46	69.33	52
2	Dialogflow	86.67	65	88	66	78,67 [(b)]	59	80 [(b)]	60
	Rasa	92	69	96	72	81.33	61	85.33	64
3	Dialogflow	80	60	84	63	70,67	53	74.67	56
	Rasa	94.67	71	97.33	73	82.67 [(c)]	62	86.67 [(c)]	65

[(a), (b), (c)] - data shown in Figure 7.



Legend: **VG1-3A1-11** - Bot utter action in each group (1, 2, or 3) for the voice channel; **NONE** - No action was recognized; The ordinate axes (y) describe the **TRUE** values; The abscissa axes (x) describe the **PREDICTED** values

Fig. 7. Confusion matrices for selected data shown in Table 5.

optimal, customer emotions were included as another element in the proposed method.

5.2. The impact of emotions on the correctness of bot utter action selection

In the proposed method, emotions in utterances are used to identify the hidden actual intent of the customer. On this basis, the inference mechanism selects the correct action to be performed by the bot. Creating an inference mechanism involves defining inference rules for all possible emotions that may be relevant to each intent. Table 6 shows the defined inference rules associated with the intents listed in Table 4. As you can see, in some cases emotions don't matter (e.g. V3). This means that regardless of the emotions in an utterance, there is no hidden intent of the customer. In other cases, the presence of selected emotions may mean that the actual intent is the result of the emotion and a willingness to change one's decision may be considered as the hidden intent (e.g. V4). An interesting case is the V5 intent. In this case, the emotions may signify a hidden intent that is completely different from the intent detected in the course of the analysis of the utterance text. Failure to take such cases into account leads to a completely wrong operation of the bot.

The results of the experiments showing the impact of the recognised customer emotions on the conversation in a CC hotline are shown in Table 7. To include only the impact of our proposed inference mechanism in the experiments, the standard Microsoft Speech to Text transcription system was used for the voice channels. In this way, the differences in the results obtained are only due to the fact that emotions were taken into account.

When analysing the results in Table 7, one can see a clear impact of customer emotions on the correctness of a virtual assistant's conversation. Theoretically, tests using identical test phrases for cases when emotions were recognised and when emotions were not present should produce the same result for intent recognition. In reality, however, the results of intent recognition may vary slightly. This is because new NLU models are created with each test and the conversational platforms used are not deterministic in this respect. This means that development of two NLU models from the same training phrases does not guarantee the same results. However, it should be noted that the differences in intent recognition are small enough to be ignored in further evaluation of the results regarding bot utter action analysis.

Depending on the platform studied and the selected set of intents, the correctness of bot utter action selection for the Dialogflow platform increases, respectively: in the voice channel from 2.66% to 5.33% and in the text channel from 4% to 17.33%. On the other hand, for the RASA platform, in the voice channel it increases from 5.33% to 9.34% and in the text channel - from 2.67% to 17.33%. For the data in Table 7 which is shaded and marked with symbols (a) to (c), the corresponding confusion matrices are summarised in Fig. 8. The first row in all columns shows the data obtained by testing the solution when the emotion recognition method was not used, while the second row contains the distributions for the method that takes into account inference rules based on recognised emotions.

The data shown above refers to the text channel; in the voice channel there are analogous relationships. It is natural to vary the number of possible actions performed by the bot depending on the emotional characteristics of each set. Therefore, for group 1 there are 10 different possible dialogues (Fig. 8a), for group 2 there are 7 (Fig. 8b), while for group 3, where there are the fewest emotion, there are only 6 possible actions (Fig. 8c). For the 1st group of intents, the conversational flow that was developed is largely determined by the recognised emotions (SWITCH CASES), and due to the high rate of correct emotion recognition (84%), the results for the bot utter action parameter also improved significantly. For the second group, emotions were recognised correctly at the rate of 92%, so that bot utter action scores also improve significantly. The smallest improvement in terms of bot utter action was observed for the third group, which is also obvious because, as mentioned before, it is the set least characterised by emotional utterances. The results clearly confirm the significant importance of the emotion recognition method in enhancing the bot's performance.

5.3. *The impact of transcription and emotion on the effectiveness of bot utter action in the voice channel*

This section presents the results of the study on the solution proposed in this paper, which integrates both a dedicated proprietary transcription method and a customer emotion recognition method. The test results obtained during the comprehensive testing of the solution developed in this study are presented in Table 8.

The value of the IMP factor, which determines the relative improvement of the fraction of proper bot utter actions, is defined as follows:

$$IMP = \frac{M2 - M1}{M1} \cdot 100\% \quad (1)$$

An analysis of the data in Table 8 clearly shows that the correctness of the answer given to the customer by the virtual assistant increases very significantly when our solution is usedFor the data from the first group for the Dialogflow platform, the improvement is 51.29%, while for the RASA platform, the improvement is 23.92%. For the second group, the results for the Dialogflow platform are 6.77% better, while for the RASA platform they are 14.75% better. A similar trend is observed for the third group of intents, where the data for the Dialogflow platform improved by 11.32%, while the data for the RASA platform improved by 9.67%. The results shown in the table confirm the selected confusion matrices shown in Fig. 9. The matrices refer to selected instances in the voice channel, which are shaded in Table 8. The first row in all columns in Fig. 9 contains the data obtained with the known solutions, while the second row contains the distributions for the solution proposed in this paper.

Table 6
Bot utter actions resulting from inference rules.

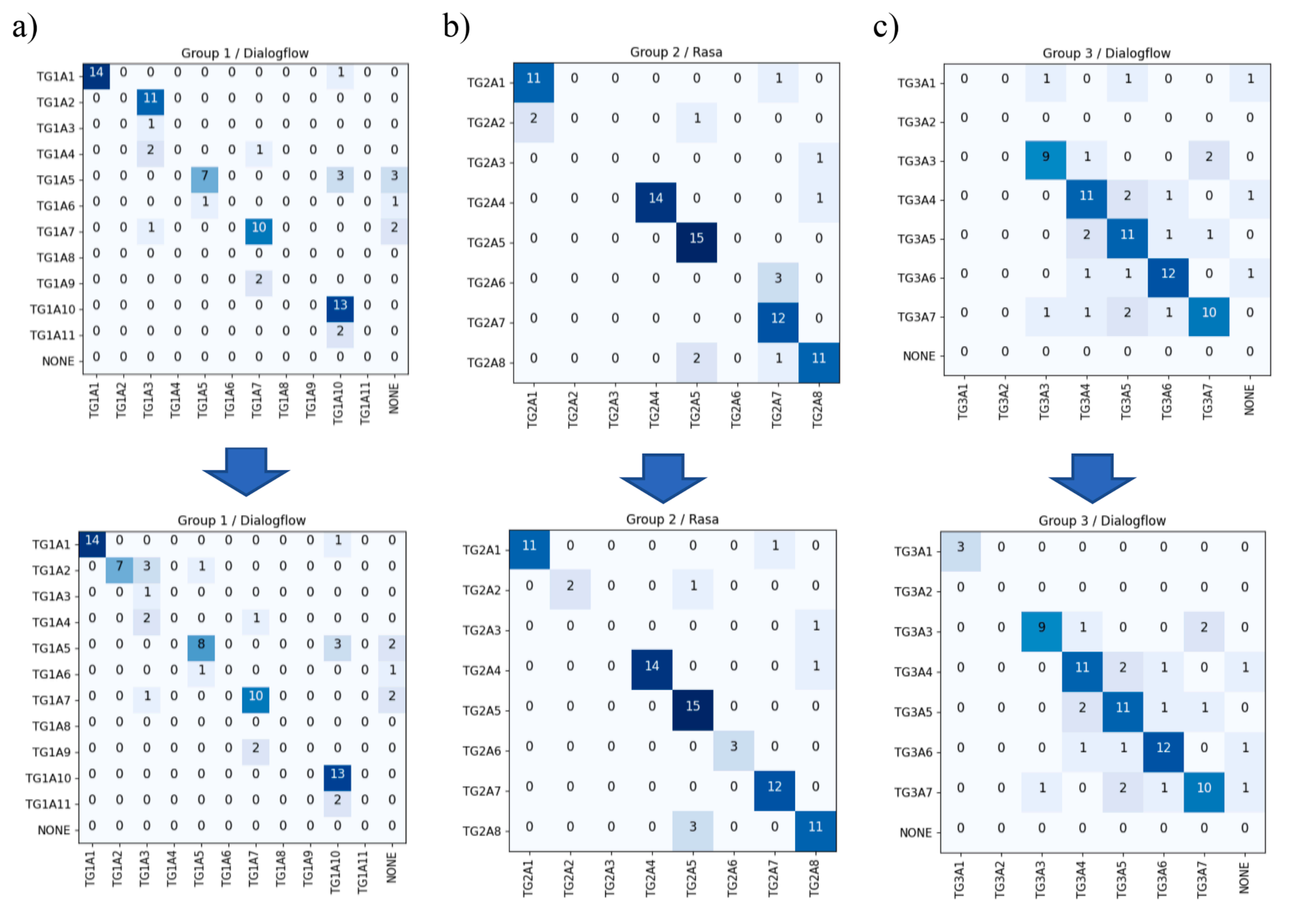
Channel	ID	INTENT	Intent inference rule	Action ID	BOT UTTER ACTION
Voice	V 1	Approving anything	IF (emotionEntityVar = Sadness OR emotionEntityVar = Fear) EXECUTE	VG1A6	Confirms, but make sure
			EXECUTE	VG1A5	Receive the confirmation
	V 2	Not approving anything	IF (emotionEntityVar = Anger OR emotionEntityVar = Sadness OR emotionEntityVar = Fear) EXECUTE	VG1A11	Denies, but make sure.
			EXECUTE	VG1A10	Accept the refusal.
	V 3	Authorisation	EXECUTE	VG1A1	Authorise
	V 4	Termination of the contract	IF (emotionEntityVar = Sadness OR emotionEntityVar = Fear) EXECUTE	VG1A9	Make sure there is an intent to cancel
			IF (emotionEntityVar = Anger) EXECUTE	VG1A8	Cancels, make a better offer.
			EXECUTE	VG1A7	Cancels
	V 5	Satisfaction with customer service	IF (emotionEntityVar = Anger OR emotionEntityVar = Sadness OR emotionEntityVar = Fear) EXECUTE	VG1A4	The customer is dissatisfied, ask why
			IF (emotionEntityVar = Satisfaction) EXECUTE	VG1A2	Satisfied, thank him or her, send a survey.
			EXECUTE	VG1A3	The customer is very satisfied, send a survey.
Text	T 1	Password recovery	IF (emotionEntityVar = Sadness OR emotionEntityVar = Anger) EXECUTE	TG2A6	Password reset by an agent.
			EXECUTE	TG2A7	Password reset via email.
	T 2	Subscription activation	EXECUTE	TG2A8	Enable subscription
	T 3	Application problems	IF (emotionEntityVar = Sadness OR emotionEntityVar = Anger) EXECUTE	TG2A2	Application reset by an agent
			EXECUTE	TG2A1	Application reinstalled by the customer
	T 4	Internet service activation	EXECUTE	TG2A5	Enable Internet service
	T 5	Issue status	EXECUTE	TG2A4	Give the status of the case

Legend: V1-V5 - intents in a given group in the voice channel; T1-T5 - intents in a given group in the text channel; VG - individual actions performed by the bot in the voice channel; TG - individual actions performed by the bot in the text channel.

Table 7
The impact of emotions present on the correctness of bot utter action.

Group	Conversational AI platform	RECOGNISING EMOTIONS		INTENT				BOT UTTER ACTION			
				Without emotion		With emotion		Without emotion		With emotion	
		voice [%]	text [%]	voice [%]	text [%]	voice [%]	text [%]	voice [%]	text [%]	voice [%]	text [%]
1	Dialogflow	80	84	68	84	68	84	52	60[(a)]	57.33	77.33[(a)]
	Rasa			81.33	93.33	81.33	93.33	61.33	70.67	69.33	88
2	Dialogflow	78.67	92	86.67	77.33	84	78.67	78.67	69.33	81.33	77.33
	Rasa			92	92	94.67	92	81.33	84[(b)]	90.67	90.67[(b)]
3	Dialogflow	65.33	89.33	80	70.67	81.33	74.67	70.67	70.67[(c)]	76	74.67[(c)]
	Rasa			94.67	76	94.67	74.67	82.67	72	88	74.67

(a), (b), (c) – data shown in Fig. 8.



Legend: **TG1-3A1-11** - Bot utter action in each group (1, 2, or 3) for the text channel; **NONE** - No action was recognized; The ordinate axes (y) describe the **TRUE** values; The abscissa axes (x) describe the **PREDICTED** values

Fig. 8. Confusion matrices for selected data shown in Table 7.

6. Conclusion

This paper proposes a new method for recognising the intent of a customer contacting a CC hotline, which is designed specifically for this industry. The main new functional feature of the method is the ability to take into account the emotional states of customers, which are recognised in the course of the conversation. Emotions such as happiness, anger, sadness, fear, and a neutral state are recognised. As demonstrated in the article, this enables the voicebot or chatbot to lead the conversation appropriately and far more effectively. For the results presented here, the improvement in bot utter action for the set in which emotions in utterances mattered most was above 50% for the Dialogflow platform. The results obtained for the other cases also showed a very high potential of the proposed solution. The solution presented herein is expected to improve customer service and optimise the relevant KPIs (Płaza and Pawlik, 2021). The results presented herein clearly show that in many cases the presence of emotions in conversations requires developing an appropriate conversation flow, which significantly improves the effectiveness of correct answers given by bots. These conclusions apply to both the voice channel and the text channel. It is also very important to note that the proposed method performed well for both selected conversational platforms, which are available both commercially and in an open-source licensing model. This shows the potential for integration of other platforms as well, and also offers great

Table 8
The effect of transcription quality and the emotions present on the correctness of bot utter action recognition.

Group	Conversational AI platform	Recognising emotions	Bot utter action		
			M1	M2	IMP
		voice [%]	voice [%]	voice [%]	voice [%]
1	Dialogflow	80	52[(a)]	78,67[(a)]	**51,29**
	Rasa		61,33	76	**23,92**
2	Dialogflow	78,67	78,67	84	**6,77**
	Rasa		81,33[(b)]	93,33[(b)]	**14,75**
3	Dialogflow	65,33	70,67	78,67	**11,32**
	Rasa		82,67[(c)]	90,67[(c)]	**9,68**

[(a), (b), (c)] - data shown in Fig. 9; M1 – Base method without emotions recognition and with standard ASR; M2 – Our method with emotions recognition and with our ASR system; IMP – Relative value of the improvement coefficient.

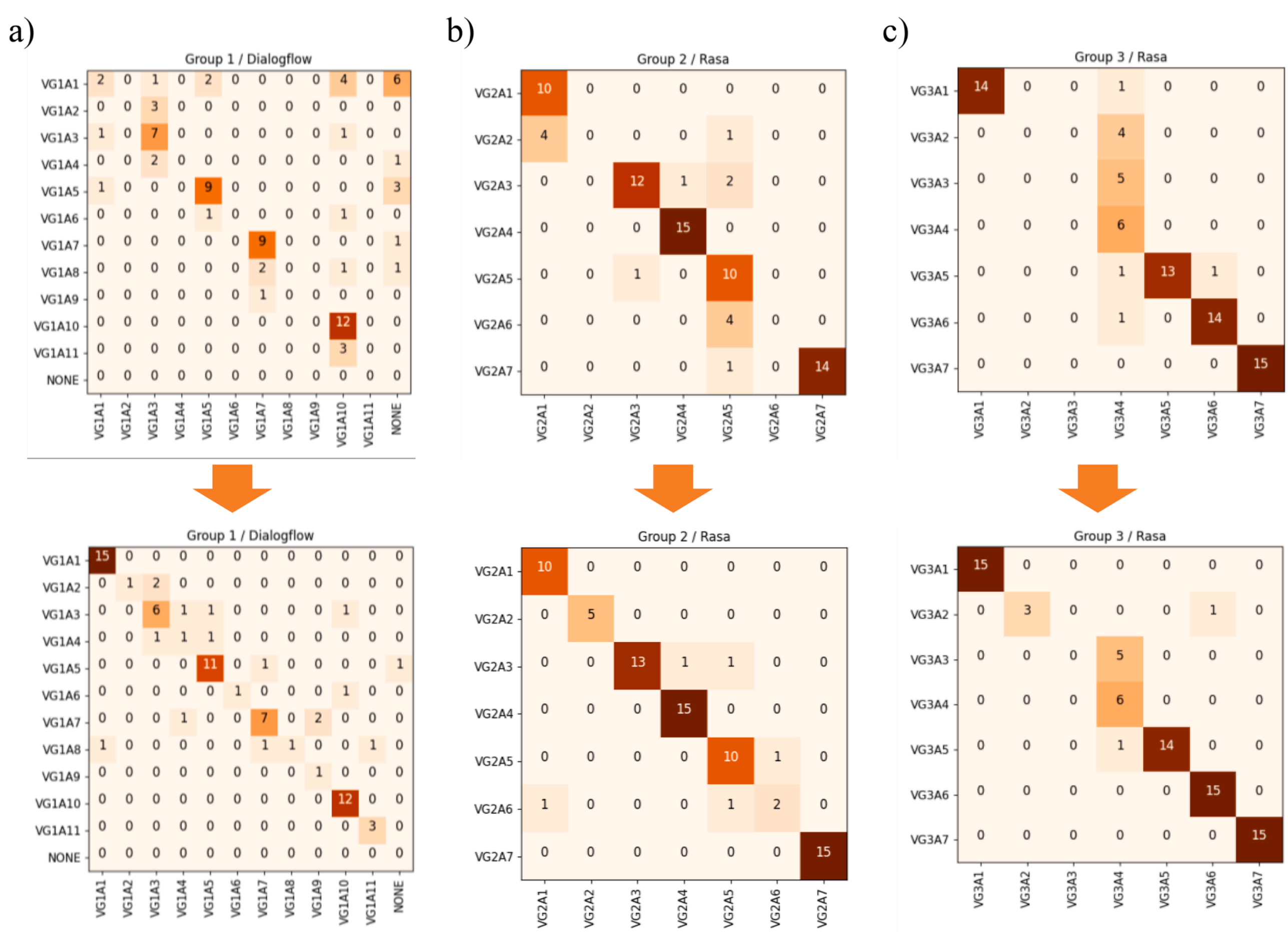
potential for widespread use of this method.

A high level of customer service is increasingly important in the CC industry. In recent years, intelligent voicebots and chatbots have been increasingly used for direct service, largely eliminating the human factor. However, this has some limitations, which are explained in detail in the paper. One of the basic tasks of a bot is to correctly identify the customer's intent which, as has been mentioned, in many cases also depends on the emotions of the caller. In many CC campaigns, proper determination of the emotion can even be crucial to a high-quality service. These campaigns include debt collection, insurance claim settlement, and all kinds of satisfaction surveys. These types of campaigns are very effective when they are handled by people for whom proper understanding of emotionally uttered content is a natural process. For voicebots and chatbots, proper handling of campaigns in which conversations are usually characterised by many extreme emotions is a big challenge. The method proposed by the authors is intended to help solve these problems.

Further work on the development of the proposed subject includes research related to the development and implementation of voicebots and chatbots prepared for different subject-specific CC campaigns. It is also expected that the method proposed in this paper can be used in CC systems of the future which, according to a literature review (Płaza and Pawlik, 2021), can integrate Internet of Things (IoT) technologies, which are currently popular and are being implemented in many technology sectors (Płaza et al., 2019; Pięta et al., 2018; Belka et al., 2018). In addition, it is predicted that video technologies (Królikowski et al., 2017), which could also support emotion recognition, will become much more important in customer service.

CRediT authorship contribution statement

Łukasz Pawlik: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Visualization. **Mirosław Płaza:** Conceptualization, Methodology, Investigation, Validation, Writing – original draft, Supervision, Project administration, Funding acquisition. **Stanisław Deniziak:** Conceptualization, Writing – review & editing, Supervision. **Ewa Boksa:** Investigation.



Legend: **VG1-3A1-11** - Bot utter action in each group (1, 2, or 3) for the voice channel; **NONE** - No action was recognized; The ordinate axes (y) describe the **TRUE** values; The abscissa axes (x) describe the **PREDICTED** values

Fig. 9. Confusion matrices for selected data shown in Table 8.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by European Union's Smart Growth Operational Programme 2014–2020, under grant agreement no POIR.04.01.04–00–0079/19.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.specom.2022.07.003.

References

Aahill, 2021. Language Understanding (LUIS) overview - azure cognitive services [WWW Document]. Microsoft.com. URL https://docs.microsoft.com/en-us/azure/cognitive-services/luis/what-is-luis (accessed 10.7.21).

Akçay, M.B., Oğuz, K., 2020. Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Commun. 116, 56–76. https://doi.org/10.1016/j.specom.2019.12.001.

Batliner, A., Schuller, B., 2013. Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. John Wiley & Sons, New York.

Belka, R., Deniziak, S.R., Płaza, M., Hejduk, M., Pięta, P., Płaza, M., Czekaj, P., Wołowiec, P., Ludwinek, K., 2018. Integrated visitor support system for tourism industry based on IoT technologies. Proceedings of the Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments. https://doi.org/10.1117/12.2326403, 2018.

Boonstra, L., 2021. Introduction to conversational AI. Definit. Guide Conversat. AI Dialogflow Google Cloud 1–27. https://doi.org/10.1007/978-1-4842-7014-1_1.

Bota, P.J., Wang, C., Fred, A.L.N., Placido Da Silva, H., 2019. A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. IEEE Access 7, 140990–141020. https://doi.org/10.1109/access.2019.2944001.

Cheang, H.S., Pell, M.D., 2008. The sound of sarcasm. Speech Commun. 50, 366–381. https://doi.org/10.1016/j.specom.2007.11.003.

Chen, C., Zhang, X., Ju, S., Fu, C., Tang, C., Zhou, J., Li, X., 2019. AntProphet: an intention mining system behind Alipay's intelligent customer service bot. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence.

Deloitte Digital, Conversational AI - the next wave of customer and employee experiences [WWW Document], 2019. URL https://www2.deloitte.com/content/dam/Deloitte/au/Documents/strategy/au-deloitte-conversational-ai.pdf (accessed 8.10.21).

Ghandeharioun, A., McDuff, D., Czerwinski, M., Rowan, K., 2019. EMMA: an emotion-aware wellbeing Chatbot. In: Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction. ACII.

Google Cloud, Entities [WWW Document], 2021. URL https://cloud.google.com/dialogflow/cx/docs/concept/entity (accessed 10.7.21).

Google Cloud, Language reference [WWW Document], 2021. URL https://cloud.google.com/dialogflow/es/docs/reference/language (accessed 10.7.21).

Google Cloud. Natural Language API Basics [WWW Document], 2021. URL https://cloud.google.com/natural-language/docs/basics#sentiment_analysis (accessed 10.7.21).

Gupta, V., Garg, N., Gupta, T., 2012. Search bot: search intention based filtering using decision tree based technique. In: 2012 Third International Conference on Intelligent Systems Modelling and Simulation. https://doi.org/10.1109/isms.2012.78.

IBM, Watson Assistant: Intelligent Virtual Agent [WWW Document], 2021. URL https://www.ibm.com/cloud/watson-assistant (accessed 10.7.21).

IBM, Watson Assistant: Visual Chatbot Builder [WWW Document], 2020. URL https://www.ibm.com/cloud/watson-assistant/visual-builder (accessed 10.7.21).

Khalil, R.A., Jones, E., Babar, M.I., Jan, T., Zafar, M.H., Alhussain, T., 2019. Speech emotion recognition using deep learning techniques: a Review. IEEE Access 7, 117327–117345. https://doi.org/10.1109/access.2019.2936124.

Koolagudi, S.G., Rao, K.S., 2012. Emotion recognition from speech: a review. Int. J. Speech Technol. 15, 99–117. https://doi.org/10.1007/s10772-011-9125-1.

Królikowski, M., Płaza, M., Szcześniak, Z., 2017. Chosen sources of signal interference in HD-TVI technology. Proceedings of the Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments. https://doi.org/10.1117/12.2280534, 2017.

Kyo-Joong, O., Lee, D., Byungsoo K., Ho-Jin Ch., 2017. Empathy bot: conversational service for psychiatric counseling with chat assistant [WWW Document]. ResearchGate. URL https://www.researchgate.net/publication/322252913_Empathy_Bot_Conversational_Service_for_Psychiatric_Counseling_with_Chat_Assistant (accessed 10.7.21).

Narayanan, A., Misra, A., Sim, K., Pundak, G., Tripathi, A., Elfeky, M., Haghani, P., Strohman, T., Bacchiani, M., 2018. Toward Domain-Invariant Speech Recognition via Large Scale Training. Google, USA. arXiv:1808.05312.

Nwe, T.L., Foo, S.W., De Silva, L.C., 2003. Speech emotion recognition using hidden Markov models. Speech Commun. 41, 603–623. https://doi.org/10.1016/s0167-6393(03)00099-2.

Ohta, K., Nishimura, R., Kitaoka, N., 2021. Response type selection for chat-like spoken dialog systems based on LSTM and multi-task learning. Speech Commun. 133, 23–30. https://doi.org/10.1016/j.specom.2021.07.003.

Pięta, P., Deniziak, S., Belka, R., Płaza, M., Płaza, M., 2018. Multi-domain model for simulating smart IoT-based theme parks. Proceedings of the Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments. https://doi.org/10.1117/12.2501659, 2018.

Płaza, M., Pawlik, L., 2021. Influence of the contact center systems development on key performance indicators. IEEE Access 9, 44580–44591. https://doi.org/10.1109/access.2021.3066801.

Płaza, M., Pawlik, L., Deniziak, S., 2021a. Call transcription methodology for contact center systems. IEEE Access 9, 110975–110988. https://doi.org/10.1109/access.2021.3102502.

Płaza, M., Belka, R., Szcześniak, Z., 2019. Towards a different world – on the potential of the internet of everything. Inf. Control Meas. Econ. Environ. Prot. 9, 8–11. https://doi.org/10.5604/01.3001.0013.2539.

Płaza, M., Deniziak, S., Płaza, M., Belka, R., Pięta, P., 2018. Analysis of parallel computational models for clustering. Photon. Appl. Astron. Commun. Ind. High-Energy Phys. Exp. https://doi.org/10.1117/12.2500795, 2018.

Płaza, M., Kazała, R., Krechowicz, A., Koruba, Z., Kozłowski, M., Lucińska, M., Sitek, K., Spyrka, J., 2022. Emotions Recognition Method for Call/Contact Center Systems, Unpublished results.

Płaza, M., Trusz, S., Kęczkowska, J., Boksa, E., Sadowski S., Koruba, Z., 2022. Machine Learning Algorithms for Detection and Classifications of Emotions in Contact Center Applications. Sensors 22, no 14: 5311. 10.3390/s22145311.

Ramakrishnan, S., El Emary, I.M.M., 2011. Speech emotion recognition approaches in human computer interaction. Telecommun. Syst. 52, 1467–1478. https://doi.org/10.1007/s11235-011-9624-z.

Rasa, Training Data Format [WWW Document], 2021. URL https://rasa.com/docs/rasa/training-data-format#entities (accessed 10.7.21).

Rasa, Open source conversational AI [WWW Document], 2020. URL https://rasa.com/ (accessed 10.7.21).

Roche, J.M., Peters, B., Dale, R., 2015. Your Tone Says It All: the processing and interpretation of affective language. Speech Commun. 66, 47–64, 2015.

Sabharwal, N., Agrawal, A., 2020. Introduction to Google Dialogflow In book: Cognitive Virtual Assistants Using Google Dialogflow. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-5741-8_2.

Sehgal, R.R., Agarwal, S., Raj, G., 2018. Interactive voice response using sentiment analysis in automatic speech recognition systems. In: 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE). https://doi.org/10.1109/icacce.2018.8441741.

Singh, A., Ramasubramanian, K., Shivam, S., 2019. Introduction to Microsoft Bot, RASA, and Google Dialogflow. Build. Enterprise Chatbot 281–302. https://doi.org/10.1007/978-1-4842-5034-1_7.

Sperber, D., Wilson, D., 1986. Relevance. Communication and Cognition. Blackwell, Oxford.

Taylor, J.R., 1995. Linguistic Categorization Prototypes in Linguistic Theory. Oxford University Press, Oxford.

Tian, J., Tu, Z., Wang, Z., Xu, X., Liu, M., 2020. User intention recognition and requirement elicitation method for conversational AI services [WWW Document]. arXiv.org. URL https://arxiv.org/abs/2009.01509 (accessed 10.7.21).

VoiceLab.ai [WWW Document], 2021. URL https://voicelab.ai/ (accessed 10.7.21).

Wang, K., An, N., Nan Li, B., Zhang, Y., Li, L., 2015. Speech emotion recognition using fourier parameters. IEEE Trans. Affect. Comput. 6, 69–75. https://doi.org/10.1109/taffc.2015.2392101.

Wit.ai [WWW Document], 2021. URL https://wit.ai/ (accessed 10.7.21).

Zhang, J., Xie, Y., Yu, F., Soukal, D., Lee, W., 2013. Intention and origination: an inside look at large-scale bot queries [WWW Document]. In: NDSS Symposium. URL https://www.ndss-symposium.org/ndss2013/ndss-2013-programme/intention-and-origination-inside-look-large-scale-bot-queries/ (accessed 10.7.21).

Zhezhnych, P., Shilinh, A., Melnyk, V., 2019. Linguistic analysis of user motivations of information content for university Entrant's web-forum. Int. J. Comput. 18, 67–74.