

Influence of Emotion Distribution and Classification on a Call Processing for an Emergency Call Center

Milana Bojanić, Vlado Delić, and Alexey Karpov

Abstract — The article addresses the influence of two aspects on speech emotion recognition utilization for an emergency call center: a frequency of a caller experiencing certain emotional state and classification methods used for speech emotion recognition. In situations when more simultaneous calls in an emergency call center are received, the aim is to detect more urgent callers, e.g. in a life threatening situation, and give them priority in a callers' queue. Three different emotion distributions based on the corpora from real-world emergency call centers are considered. The influence of those emotion distributions on the proposed call redistribution and subsequent time savings are reported and discussed. Regarding speech emotion classification, two approaches are presented, namely the linear Bayes classifier and a multilayer perceptron-based neural network. Their recognition results on the corpus of acted emotional Serbian speech are presented and potential application in an emergency call center is discussed.

Keywords — affective computing, call center, speech emotion recognition.

I. INTRODUCTION

ALTHOUGH almost 25 years of research have passed, speech emotion recognition (SER) is still an increasingly interesting topic due to its possible usage in various systems based on human-machine interaction and speech analysis for health, security, robotics and many other applications [1]. SER is a part of spoken dialogue systems whose task is automatic recognition of speaker's emotional state based upon the analysis of his/her speech.

Paper received May 26, 2021; accepted November 11, 2021. Date of publication December 30, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dragana Šumarac Pavlović.

This paper is revised and expanded version of the paper presented at the 28th Telecommunications Forum TELFOR 2020 [16].

This research was supported by the Science Fund of the Republic of Serbia, (grant #6524560, AI – S-ADAPT), and by the Serbian Ministry of Education, Science and Technological Development (project no. 451-03-68/2020-14/200156: "Innovative scientific and artistic research from the Faculty of Technical Sciences activity domain").

Milana Bojanić (corresponding author), Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: milana.bojanic@uns.ac.rs).

Vlado Delić, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: vlado.delic@uns.ac.rs).

Alexey Karpov, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, SPC RAS, 14th Line 39, 199178, St. Petersburg, Russia; (e-mail: karpov@iias.spb.su).

Regarding application in a call center, one of the first attempts to employ SER was a study for prioritizing telephone voice message in a call center [2]. Application of SER for a customer service call-center, where customers seek information or have a dispute to resolve, and employment of ensemble methods for speech emotion classification was proposed in [3]. The study of Devillers et al. provided annotation and analysis of emotional speech recorded in a real medical emergency call center comprised in the CEMO corpus [4]. In our previous study, the algorithm of call redistribution for a call center has been proposed in order to recognize more urgent callers and give them priority while waiting in a caller's queue [5]. The proposed call redistribution is based on speech emotion recognition with an assumption that all the emotions have a uniform distribution. The focus of this research was to examine how different emotion distributions influenced the proposed call redistribution regarding a possible application in an emergency call center. The reported statistics for fine grained and coarse labeled emotions in the CEMO corpus provided an insight into emotion distribution in real-world emergency call center recordings [6]. It was a motivation and basis to examine our proposed algorithm, tested in a simulated call center, in a more realistic scenario of basic emotion distribution which is likely expected in an emergency call center.

The article is organized as follows. Section II gives an overview of the proposed call redistribution algorithm. Modeling of the emotion distribution in a caller's queue for an emergency call center is proposed in Section III. Specification of implemented SER system is given in Section IV. The experimental results are reported and discussed in Section V. Conclusion and direction for future research are given in Section VI.

II. PROPOSED CALL REDISTRIBUTION ALGORITHM

The call redistribution algorithm used in the article was proposed in [5]. In this section we will briefly present its main parts. The proposed procedure has four main steps: recording the initial part of the caller's speech, recognition of caller's emotional state (SER module), monitoring next incoming calls, call redistribution according to an adopted priority vector, and final call processing in a new order. The flow of call processing is shown in Fig.1.

First, if all operators in a call center are busy, the system automatically asks for the reason of the call and records the initial part of the caller's speech containing 1-2 sentences.

It is sufficient for the SER analysis in the next step. In the SER part, the feature vector is extracted for each utterance and forwarded to the classifier which classifies a given utterance into five emotion categories (fear, anger, neutral, sadness, and joy). More details about SER module are given in the next section. In the monitoring phase, if a new call is received within 30 s from the last call then they are grouped together as calls that are received simultaneously. Monitoring is continued until a predefined number of simultaneously received calls is achieved. In the experiments the maximum number of calls in a group was seven. All calls grouped together go through the first and second step i.e. for each call the initial part of the call is recorded and then processed by the SER module. SER module provides an emotion label for every call in a group of the simultaneously received calls. Based on those emotion labels, calls are redistributed according to an adopted priority vector. In the priority vector introduced in [5] fear is put first, afterwards anger, sadness, neutral and joy in the last place. The proposed algorithm allows utilization of different priority vectors. The proposed priority vector used in the paper is considered suitable for application in an emergency call center. After redistribution, calls are processed in a newly obtained order determined by the used priority vector.

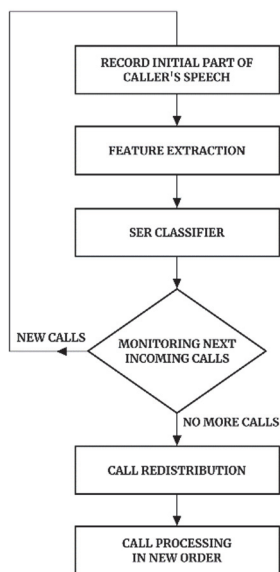


Fig. 1. Block diagram of the proposed algorithm for call redistribution.

III. IMPACT OF EMOTION DISTRIBUTIONS IN AN EMERGENCY CALL CENTER

In our previous experiments reported in [5], each emotion was chosen with an equal probability, i.e. all emotions in a caller's queue were equally distributed. Those experiments are denoted as baseline experiments in the given paper and they encompassed applications both for an emergency call center and a health care center for elderly people. In this research, the experiments were conducted regarding an application of call redistribution for an emergency call center only. Thus there was a possibility to adjust the probability of observed five emotional states which

normally are not equal in such a call center. The CEMO corpus was recorded in a real-world emergency call center, and based on emotion annotations reported in [4] almost 65% of client dialogs were labeled with the neutral emotional state, 20% with anxiety, 8.7% stress, 3.3% relief, 0.6% annoyance and 3.2% with other fine-grained emotion labels from the list of 21 emotion labels. Taking into account the analysis of emotional classes present in the CEMO corpus and reported in [4], [6], three different emotion distributions are proposed and their effect on the proposed call redistribution for an emergency call center is studied.

Emotion probabilities in a baseline experiment and three different emotion distributions used in the current research are shown in Table 1. In the first emotion distribution (ED1) the probability (i.e. frequency of occurrence) of fear is increased at the expense of probability of joy compared to the baseline distribution. Reason for this change is grounded in an expected trend for an emergency call center that there are more callers experiencing some kind of fear than callers expressing joy and happiness. The second distribution (ED2) and the third distribution (ED3) are motivated by the representation of emotional classes in the CEMO corpus. Thus, emotion probabilities in ED2 and ED3 are chosen to approximately match reported emotion distributions in the CEMO corpus. Regarding that the neutral state is dominant in the CEMO corpus, the probability of neutral state was significantly increased: 50% in ED2, and 66% in ED3, followed by fear (25%) in both ED2 and ED3. The probabilities of anger and sadness were reduced to 10% and 3% for ED2 and ED3, respectively. The probability of joy was set to the lowest rate among five emotions, 5% and 3% in ED2 and ED3, respectively.

TABLE 1: EMOTION PROBABILITIES IN BASELINE AND THREE STUDIED EMOTION DISTRIBUTIONS (ED1, ED2, ED3)

| <i>Emotion</i> | <i>Baseline</i> | <i>ED1</i> | <i>ED2</i> | <i>ED3</i> |
|----------------|-----------------|------------|------------|------------|
| fear | 0.2 | 0.25 | 0.25 | 0.25 |
| anger | 0.2 | 0.2 | 0.1 | 0.03 |
| sadness | 0.2 | 0.2 | 0.1 | 0.03 |
| neutral | 0.2 | 0.2 | 0.5 | 0.66 |
| joy | 0.2 | 0.15 | 0.05 | 0.03 |

IV. SER MODULE

SER module typically has three main parts on which research is focused: a) choice/design of an appropriate emotional speech corpus, b) extraction and selection of a suitable feature vector, and c) design/choice of classification technique that will be exploited. We will briefly give a description of our SER module.

A. Emotional Speech Corpus

As to emotional speech corpus, our research was restricted to Serbian and GEES corpus which is the only one available in Serbian. The GEES corpus is the corpus of acted emotional speech and it contains recordings of semantically neutral sentences uttered by six actors in five basic emotional states: anger, joy, fear, sadness, and neutral [7]. In human listening tests, average human recognition accuracy for the GEES was 94.7% which is a sign of expressive and clearly distinctive emotions in the corpus

[7]. Although there are objections to spontaneity of emotions present in acted speech corpora, they are still widely used due to their public availability and well annotated data [8], [9]. As in our case a goal was to build SER intended for an emergency call center, the presence of more intensive emotion expressions in acted speech corpus was an additional advantage.

B. Feature Extraction

For the task of acoustic modeling of emotional speech there is a traditional approach, so-called hand-crafted features, or recently introduced learned representations where an end-to-end system learns data representation directly from the raw speech signal [9]. Hand-crafted features are based on the extraction of various prosodic, spectral, cepstral and voice quality features, usually analyzed on a supra-segmental level with statistical functionals (e.g. mean, median, variance, minimum, maximum etc.) [10]. Hand-crafted features are grounded on expert knowledge about acoustic and prosodic variations present in a speech signal during the vocal expression of emotions. On the other hand, development of Deep Neural Networks (DNN) enables that the classifier discovers a suitable feature representation by itself [11] using as input raw speech signal or its spectral representation.

In the research, the set of hand-crafted features is employed. Low level descriptors (LLD) including pitch, energy and Mel Frequency Cepstral Coefficients (MFCC) are extracted for each processed utterance. Then a set of 12 statistical functionals is applied on those LLDs and their first derivative. The final vector comprises 384 features for each utterance. More details about extracted feature vector can be found in [5].

C. Classification Techniques

In the field of SER many classification techniques have been used, and recently the research has shifted towards DNN, specifically Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [12]. Still, there is no consensus on the best classification technique for the SER task.

In our study, focus was on simple and efficient classification methods. In our previous study [13] the recognition accuracies of the Linear Bayes (LB) classifier, perceptron rule and kNN were compared. The linear Bayes classifier achieved the best accuracy of 91.5% in a five-class emotion recognition task for the GEES corpus. The linear Bayes classifier is considered to be efficient for both acted and spontaneous emotional speech [10], at the same time it is a simple classification technique with low computational demands. As the second classifier, an artificial neural network (ANN) was employed. ANNs are capable of learning complex nonlinear relations between inputs and desirable outputs and their training is faster than the rest of the neural network based methods [14]. Our previous study presented the results of the neural network application in SER on the example of the GEES corpus [15]. The best results were obtained with the multilayer perceptron (MLP) composed of 3 layers (the input layer, the hidden layer with 15 neurons, and the output layer). During

training, the log sigmoid activation function and standard backpropagation (BP) as a learning algorithm were used. The MLP was chosen as a representative of ANNs, which together with the deep learning techniques, have attained the research focus in the field of SER [14].

The results presented in the paper are obtained with LB and MLP classifier. In experiments, the recognition accuracy was estimated after 10-fold cross-validation including all speakers in the corpus. Thus, the presented results belong to speaker-dependent experiments.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental results are obtained in experiments conducted in a simulated emergency call center. In the simulation process, the cases of three, five and seven simultaneously received calls are considered. These numbers of simultaneously received calls are chosen regarding they reflect situations of moderate to higher call center load. During experiments there was a need to simulate the order of incoming calls in a call queue regarding caller's emotional state and call duration. For this task, a pseudo-random number generator was used in the simulations. Unlike the baseline experiments reported in [5], where each emotion had an equal probability of occurrence, in experiments with a simulated emergency call center the probability of particular emotional state is modeled by three introduced distributions, namely ED1, ED2 and ED3. In such a way, the influence of emotion probability distribution on a used call redistribution is studied. A call duration for each call was also generated as a random value from 30 s to 3 min 50 s. To the best of our knowledge, there is no available statistics of call duration reported for real call centers, so this range is chosen to cover both shorter and longer phone calls. Every incoming call in simulation is represented with an utterance from the GEES corpus.

Let us give a simulation example of three simultaneously received calls. The order of emotions in those three calls is determined by a pseudo-random number and accordingly three utterances belonging to those emotion classes are selected from the GEES. A call duration, also determined by a random number, is appended to every utterance. The initial waiting time of some caller equals the sum of call durations for all callers which are preceding him/her in the queue. Those three utterances are processed by SER module and as a result every utterance is labeled with a recognized emotional state. Using those emotion labels, calls are redistributed according to the priority vector. A new waiting time after redistribution is evaluated for each caller, representing some emotional state, taking into account his/her new position in the caller's queue. So, a waiting time that a caller will spend in a caller's queue is evaluated twice: without the proposed algorithm (i.e., an initial waiting time) and after application of SER and the proposed algorithm (i.e., a new waiting time after redistribution). An average waiting time for each emotion is estimated as the mean value of waiting time obtained in 1000 iterations of simulation.

Experimental results are presented for two research directions: 1) SER classification results based on two

classification methods, and 2) results of average waiting time without and after the application of SER and the proposed call redistribution algorithm.

A. SER classification results

The classification results are obtained using the linear Bayes classifier and MLP with one hidden layer. In order to get more information about recognition accuracies, the normalized confusion matrix for LB as a classifier is presented in Table 2 [5]. The confusion matrix provides the insight into performance of the classifier. Each row in the matrix represents the instances of an actual emotion class, while each column represents the instances in a predicted emotion class. The diagonal elements in the matrix correspond to the percentage of correctly recognized instances in an emotion class. The emotion classes with the highest recognition rate are: neutral state (97.1%), sadness (94.8%) and fear (92.5%). It could be noted that anger and joy have lower recognition rates due to mutual misclassification (11.2% of angry callers are misrecognized as joyful ones, and 14.9% of joyful callers are misrecognized as angry).

TABLE 2: NORMALIZED CONFUSION MATRIX FOR LINEAR BAYES CLASSIFIER

| <i>True Emotion Class</i> | <i>Recognized Emotion Class [%]</i> | | | | |
|---------------------------|-------------------------------------|-------------|-------------|-------------|-------------|
| | anger | fear | joy | neutral | sadness |
| anger | 88.8 | 0 | 11.2 | 0 | 0 |
| fear | 1.4 | 92.5 | 1.4 | 3.2 | 1.4 |
| joy | 14.9 | 0.6 | 84.2 | 0.3 | 0 |
| neutral | 0 | 1.1 | 1.2 | 97.1 | 0.6 |
| sadness | 0 | 1.2 | 0 | 4 | 94.8 |

Three connected classification measures, namely precision, recall and F_1 score, are calculated based on the recognition results of the linear Bayes classifier and they are presented in Table 3. Precision reflects the number of relevant instances among all instances classified in some class. Sadness, fear and neutral emotion have the highest precision of 0.979, 0.969, and 0.927, respectively.

TABLE 3: PRECISION, RECALL AND F_1 SCORE FOR LINEAR BAYES CLASSIFIER

| <i>Measure</i> | <i>Emotion class</i> | | | | |
|------------------|----------------------|-------|--------------|--------------|-------|
| | fear | anger | sadness | neutral | joy |
| <i>Precision</i> | 0.969 | 0.845 | 0.979 | 0.927 | 0.859 |
| <i>Recall</i> | 0.925 | 0.888 | 0.948 | 0.971 | 0.842 |
| <i>F1 score</i> | 0.947 | 0.866 | 0.963 | 0.949 | 0.851 |

Joy and anger have lower precision due to mutual misclassification (this property of the GEES corpus is also observed in the human listening tests [7]). Recall as a measure reflects the fraction of relevant instances retrieved in some class. Regarding the recall, neutral, sadness and fear are emotions with the highest recall rates. An equal number of emotion instances from 5-classes were included in our test experiments. F_1 score is calculated as the harmonic mean of precision and recall, and it is a measure of a model's accuracy on a dataset. It can be observed from Table 3, that sadness, neutral and fear emotions have rather

high values of F_1 score. Compared to those three emotions, anger and joy have lower values of F_1 score.

The normalized confusion matrix in the case of MLP, which was used in experimental tests, is shown in Table 4. As compared to LB classifier, it can be noted that MLP achieved slightly lower results in the case of anger, fear and joy. The recognition rate of the neutral state is decreased by 3.4%, but the recognition rate of sadness is raised by 2.9%.

TABLE 4: NORMALIZED CONFUSION MATRIX FOR THREE LAYER MLP CLASSIFIER

| <i>True Emotion Class</i> | <i>Recognized Emotion Class [%]</i> | | | | |
|---------------------------|-------------------------------------|-------------|-------------|-------------|-------------|
| | anger | fear | joy | neutral | sadness |
| anger | 86.5 | 1.1 | 11.8 | 0.6 | 0 |
| fear | 0.8 | 91.7 | 2.3 | 2 | 3.2 |
| joy | 13.2 | 2.3 | 82.5 | 2 | 0 |
| neutral | 0 | 1.4 | 2.6 | 93.7 | 2.3 |
| sadness | 0 | 1.4 | 0 | 0.9 | 97.7 |

Based on the results presented in Table 4, precision, recall and F_1 score are calculated for every emotion class regarding three-layer MLP as a classifier and are shown in Table 5. As to precision, sadness, neutral and fear are emotions with a highest precision rate in experiments with MLP. Comparing F_1 score results from Table 3 and Table 5, it can be noted that both classification methods achieved a high F_1 score among all emotion classes. It indicates that these two classification methods have a certain trade-off between precision and recall rates for some emotion classes, which finally results in high F_1 scores for every class.

TABLE 5: PRECISION, RECALL AND F_1 SCORE FOR THREE LAYER MLP CLASSIFIER

| <i>Measure</i> | <i>Emotion class</i> | | | | |
|------------------|----------------------|-------|--------------|---------|-------|
| | fear | anger | sadness | neutral | joy |
| <i>Precision</i> | 0.937 | 0.861 | 0.947 | 0.945 | 0.832 |
| <i>Recall</i> | 0.917 | 0.865 | 0.977 | 0.937 | 0.825 |
| <i>F1 score</i> | 0.927 | 0.863 | 0.962 | 0.941 | 0.828 |

Given the emotion probabilities in studied emotion distributions (given in Table 1), and given the measures of recall for emotions (as shown in Table 3 and Table 5), the probability of correctly recognized caller can be calculated using the equation:

$$P(TR) = \sum_i P(TR|E_i) \cdot P(E_i) \quad , i = 1, \dots, 5, \quad (1)$$

where $P(E_i)$ denotes probability of an emotion occurrence in a caller's queue, and $P(TR|E_i)$ denotes the probability that a caller expressing a certain emotion will be correctly recognized by the classifier. The conditional probability $P(TR|E_i)$ is equal to the recall rate for a specific emotion i (5 emotions). The average probability to correctly recognize a caller, in observed five emotional states using LB and MLP as classifiers, is presented in Table 6. It can be noted that the probability to correctly recognize the emotional state of a caller increases as emotion distribution is more adapted for an emergency call center (ED2 and ED3), especially in the case of LB classifier.

TABLE 6: PROBABILITY [%] OF CORRECTLY RECOGNIZED EMOTION OF A CALLER

| Classifier | Emotion distribution | | | |
|------------|----------------------|-------|-------|-------|
| | Baseline | ED1 | ED2 | ED3 |
| LB | 91.48 | 91.89 | 94.25 | 95.24 |
| MLP | 90.42 | 90.88 | 92.32 | 92.77 |

B. Estimated caller waiting time

We recall that experiments integrated two research directions: a) three emotion distributions modeling the emotional state of a caller, b) two classification methods for SER. In all experimental settings, an average waiting time is estimated for cases of 3, 5, and 7 simultaneously received calls, without and after application of SER and the proposed call redistribution algorithm.

Table 7 shows an average waiting time estimated in a simulated call center using ED1 as a basis for emotion distribution and the LB as a classifier. It can be noticed that reduction in a waiting time is achieved for the callers who are recognized as being in a state of fear and anger. A neutral, sad, and joyful caller are evaluated to have an increased waiting time in all observed cases.

TABLE 7: AVERAGE WAITING TIME OBTAINED FOR THE FIRST EMOTION DISTRIBUTION ED1 AND LB BASED CLASSIFIER

| Emotion/Calls | Without the Proposed Algorithm [min]:[s] | | | After Application of SER and the Proposed Algorithm [min]:[s] | | |
|---------------|--|------|------|---|------|-------|
| | 3 | 5 | 7 | 3 | 5 | 7 |
| | fear | 2:05 | 4:12 | 6:15 | 0:19 | 0:40 |
| anger | 2:10 | 4:15 | 6:23 | 1:29 | 2:54 | 4:27 |
| sadness | 2:13 | 4:23 | 6:20 | 2:32 | 4:58 | 7:08 |
| neutral | 2:14 | 4:17 | 6:13 | 3:22 | 6:50 | 10:02 |
| joy | 2:05 | 4:17 | 6:26 | 4:05 | 8:11 | 12:12 |

Comparison of the results obtained with a baseline emotion distribution and with ED1, shows that fear, anger, sadness, and joy have an increased waiting time if emotion distribution is modeled by ED1. This is expected because in ED1 there is by 5% more callers featuring fear than in baseline distribution. After redistribution, those callers recognized as afraid are put first so they cause an additional waiting time for all other callers in a queue.

Table 8 shows an average waiting time estimated in a simulated call center using ED2 as a basis for emotion distribution and LB as a classifier. Callers featuring three emotions, namely fear, anger and sadness, have a decreased waiting time after redistribution. On the other hand, joyful and neutral callers have an increased waiting time after redistribution. The results obtained using ED2 for emotion distribution modeling, show that fear, neutral and joy have an increased waiting time compared with a baseline distribution. Angry callers are estimated to have slight changes of waiting time compared to a baseline distribution. Only callers detected as sad are estimated to have a noticeable decrease in a waiting time. It can be explained by the fact that for sad callers in ED2 there is a decreased probability of callers with a higher priority (fear and anger) compared to the baseline distribution. A high percentage of neutral callers in ED2 (50%) caused a slight increase in a

waiting time for this emotional state as well as for joyful callers.

TABLE 8: AVERAGE WAITING TIME OBTAINED FOR THE SECOND EMOTION DISTRIBUTION ED2 AND LB BASED CLASSIFIER

| Emotion/Calls | Without the Proposed Algorithm [min]:[s] | | | After Application of SER and the Proposed Algorithm [min]:[s] | | |
|---------------|--|------|------|---|------|-------|
| | 3 | 5 | 7 | 3 | 5 | 7 |
| | fear | 2:08 | 4:15 | 6:28 | 0:24 | 0:41 |
| anger | 2:03 | 4:15 | 6:30 | 1:11 | 2:19 | 4:05 |
| sadness | 2:06 | 4:24 | 6:15 | 1:43 | 3:14 | 5:21 |
| neutral | 2:13 | 4:15 | 6:33 | 3:17 | 6:27 | 9:42 |
| joy | 2:11 | 4:27 | 6:13 | 4:25 | 8:23 | 12:27 |

Average waiting times, before and after redistribution, assessed in experiments using LB based classification method and ED3 for emotion distribution are given in Table 9. Similar to experiment using ED2, calls featuring three emotions, namely fear, anger and sadness, have a decreased waiting time after redistribution. On the other hand, joyful and neutral callers have an increased waiting time after redistribution.

TABLE 9: AVERAGE WAITING TIME OBTAINED FOR THE THIRD EMOTION DISTRIBUTION ED3 AND LB BASED CLASSIFIER

| Emotion/Calls | Without the Proposed Algorithm [min]:[s] | | | After Application of SER and the Proposed Algorithm [min]:[s] | | |
|---------------|--|------|------|---|------|-------|
| | 3 | 5 | 7 | 3 | 5 | 7 |
| | fear | 2:12 | 4:13 | 6:32 | 0:21 | 0:41 |
| anger | 2:13 | 4:13 | 6:38 | 1:03 | 2:22 | 3:18 |
| sadness | 2:12 | 4:09 | 6:55 | 1:29 | 2:44 | 3:52 |
| neutral | 2:06 | 4:15 | 6:17 | 3:01 | 5:55 | 8:30 |
| joy | 2:30 | 4:02 | 6:27 | 4:33 | 8:13 | 12:34 |

Comparing the results evaluated with a baseline distribution and with ED3, it can be observed that fear has an increased waiting time due to a higher probability of fear in ED3. Callers in angry, sad and neutral states have a decreased waiting time compared to the baseline results. Reasons for this are a decreased probability of appearance of callers with a higher priority and a reduced number of callers in aforementioned emotion classes. Calls featuring joy are evaluated to have a slightly increased waiting time.

Table 10 shows an average waiting time estimated in a simulated call center using ED1 as a basis for emotion distribution and MLP as a classifier. Compared to the results shown in Table 7, a similar tendency of increase/decrease in waiting times can be observed considering all emotions.

Average waiting times, before and after redistribution, estimated in experiments using MLP based classification method in the case of ED2 and ED3 emotion distribution are given in Tables 11 and 12, respectively. The comparative results of LB and MLP in the case of ED2 and ED3 distribution show that there is a slight increase in a waiting time for the neutral callers using MLP classifier. This is due to the high probability of occurrence of the neutral caller in these two emotion distributions and the fact that MLP achieved a slightly lower recognition rate for a neutral state.

TABLE 10: AVERAGE WAITING TIME OBTAINED FOR THE FIRST EMOTION DISTRIBUTION ED1 AND MLP BASED CLASSIFIER

| Emotion/Calls | Without the Proposed Algorithm [min]:[s] | | | After Application of SER and the Proposed Algorithm [min]:[s] | | |
|---------------|---|------|------|--|------|-------|
| | 3 | 5 | 7 | 3 | 5 | 7 |
| fear | 2:11 | 4:13 | 6:24 | 0:20 | 0:41 | 1:09 |
| anger | 2:08 | 4:22 | 6:34 | 1:25 | 2:54 | 4:27 |
| sadness | 2:19 | 4:21 | 6:31 | 2:26 | 4:45 | 7:11 |
| neutral | 2:10 | 4:26 | 6:19 | 3:20 | 6:44 | 10:01 |
| joy | 1:56 | 4:18 | 6:29 | 4:00 | 8:20 | 12:19 |

TABLE 11: AVERAGE WAITING TIME OBTAINED FOR THE SECOND EMOTION DISTRIBUTION ED2 AND MLP BASED CLASSIFIER

| Emotion/Calls | Without the Proposed Algorithm [min]:[s] | | | After Application of SER and the Proposed Algorithm [min]:[s] | | |
|---------------|---|------|------|--|------|-------|
| | 3 | 5 | 7 | 3 | 5 | 7 |
| fear | 2:15 | 4:09 | 6:27 | 0:18 | 0:42 | 1:11 |
| anger | 2:04 | 4:36 | 6:10 | 1:11 | 2:34 | 3:53 |
| sadness | 2:15 | 4:00 | 6:25 | 1:37 | 3:20 | 5:16 |
| neutral | 2:08 | 4:06 | 6:35 | 3:24 | 6:47 | 9:55 |
| joy | 2:05 | 4:19 | 6:12 | 4:09 | 8:30 | 12:48 |

TABLE 12: AVERAGE WAITING TIME OBTAINED FOR THE THIRD EMOTION DISTRIBUTION ED3 AND MLP BASED CLASSIFIER

| Emotion/Calls | Without the Proposed Algorithm [min]:[s] | | | After Application of SER and the Proposed Algorithm [min]:[s] | | |
|---------------|---|------|------|--|------|-------|
| | 3 | 5 | 7 | 3 | 5 | 7 |
| fear | 2:08 | 4:11 | 6:30 | 0:20 | 0:43 | 1:13 |
| anger | 1:48 | 4:26 | 6:32 | 0:55 | 2:21 | 3:28 |
| sadness | 2:06 | 4:19 | 6:48 | 1:30 | 2:36 | 3:51 |
| neutral | 2:08 | 4:17 | 6:24 | 3:17 | 6:23 | 9:43 |
| joy | 2:16 | 4:34 | 6:51 | 4:09 | 8:30 | 12:52 |

VI. CONCLUSIONS

In the article, the influence of three different emotion distributions on the algorithm for call redistribution for an emergency call center is examined. While the first distribution has slightly changed probabilities compared to the baseline experiments, the second and the third distributions reflect a realistic emotion distribution that can be expected in an emergency call center. The results show that for the second and third emotion distributions, the proposed call redistribution algorithm results in a reduced waiting time for calls featuring emotions of fear, anger, and sadness. These results confirm the utility of the proposed call redistribution algorithm and give a better estimation of possible time savings for emergency call centers. In the case of calls featuring joy or neutral states, the proposed redistribution results in an increased waiting time for all three observed distributions. As to a neutral state, the

increase is lower in the case of ED3 than in the case of the baseline distribution. The paper also presented the results of two speech emotion classifications based on the linear Bayes and MLP methods. The experimental results show that the utilization of adequate emotion distribution has increased the average recognition rate for both classification methods. In future work, the proposed algorithm and its validity should be verified on data from a real-world emergency call center.

REFERENCES

- [1] B. W. Schuller, "Speech emotion recognition: two decades in a nutshell, benchmarks, and ongoing trends," *Commun. ACM*, vol. 61, no. 5, pp. 90–99, 2018.
- [2] V. Petrushin, "Emotion in speech: recognition and application to call centers," in *Proc. of Conf. on Artificial Neural Networks in Engineering (ANNIE)*, USA, 1999, pp. 7–10.
- [3] D. Morrison, R. Wang, L.C. De Silva, "Ensemble methods for spoken emotion recognition in call-centres," *Speech Commun.*, vol. 49(2), pp. 98–112, 2007. doi:10.1016/j.specom.2006.11.004
- [4] L. Devillers, L. Vidrascu, and L. Lamel, "Challenges in real-life emotion annotation and machine learning based detection," *Neural Networks*, vol. 18, pp. 407–422, 2005.
- [5] M. Bojanić, V. Delić, and A. Karpov, "Call redistribution for a call center based on speech emotion recognition," *Appl. Sci.*, vol. 10(13), 4653, 2020. doi:10.3390/app10134653
- [6] L. Devillers, C. Vaudable, and C. Chastagnol, "Real-life emotion-related states detection in call centers: a cross-corpora study," in *Proc. of the INTERSPEECH 2010*, Japan, 2010, pp. 2350–2353.
- [7] S.T. Jovičić, Z. Kašić, M. Djordjević, and M. Rajković, "Serbian emotional speech database: Design, processing and evaluation," in *Proc. 9th Int. Conf. SPECOM'2004*, Russia, 2004, pp. 77–81.
- [8] C. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011," *Artif Intell Rev*, vol. 43, pp. 155–177, 2015.
- [9] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: a review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019, doi: 10.1109/ACCESS.2019.2936124
- [10] B. Schüller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Speech Commun.*, vol. 53, pp. 1062–1087, 2011. doi:10.1016/j.specom.2011.01.011
- [11] J. Wagner, D. Schiller, A. Seiderer, and E. André, "Deep learning in paralinguistic recognition tasks: are hand-crafted features still relevant?" in *Proc. of the INTERSPEECH 2018*, pp. 147–151.
- [12] M.B. Akçay, and K. Oğuz, "Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Commun.*, vol. 116, pp. 56–76, 2020. doi:10.1016/j.specom.2019.12.001
- [13] V. Delić, M. Bojanić, M. Gnjatović, M. Sečujski, and S. T. Jovičić, "Discrimination capability of prosodic and spectral features for emotional speech recognition," *Elektron. ir Elektrotehnika*, vol. 18, pp. 51–54, 2012. doi:10.5755/j01.eee.18.9.2806
- [14] B. J. Abbaschian, D. Sierra-Sosa, and A. Elmaghraby, "Deep learning techniques for speech emotion recognition, from databases to models," *Sensors*, vol. 21, 1249, 2021. doi:10.3390/s21041249
- [15] M. Bojanić, V. Crnojević, and V. Delić, "Application of neural networks in emotional speech recognition," in *Proc. 11th Symposium on Neural Network Applications in Electrical Engineering*, Belgrade, Serbia, 2012, pp. 223–226, doi:10.1109/NEUREL.2012.6420016
- [16] M. Bojanić, V. Delić, and A. Karpov, "Effect of Emotion Distribution on a Call Processing for an Emergency Call Center," *28th Telecommunications Forum (TELFOR)*, Belgrade, 2020, pp. 1–4, doi:10.1109/TELFOR51502.2020.9306564.