A deep learning approach to prepare participants for negotiations by recognizing emotions with voice analysis

Jonas Maier, Daniel Schlechte, Marc Fernandes,
Andreas Theissler [0000-0003-0746-0424]

Aalen University of Applied Sciences, Beethovenstr. 1,
73430 Aalen, Germany

jonas.maier711@googlemail.com
danielschlechte@googlemail.com
marc.fernandes@hs-aalen.de

Abstract: Emotional intelligence plays an essential role in negotiations. In this work we present an approach for emotion recognition (sentiment detection) based on the characteristics of the human voice. It is shown how such an approach can support negotiators to train their emotional intelligence prior to negotiations. Using methods from the field of deep learning, in specific convolutional neural networks, a prototype was developed. The approach is tested and thus validated by various experiments on voice recordings, showing that it is possible to determine emotions to a certain accuracy in real time. This can help, especially young and inexperienced negotiators, to prepare for important negotiations. It is shown that the prototype can make a valuable contribution in the run-up to negotiations through constructive projects.

Keywords: Emotions, Emotional Intelligence, Sentiment Detection, Negotiation, Artificial Neural Network

Introduction

Many people are not aware that they are unconsciously aggressive or even fearful. With these patterns of behavior, for example, one party at the negotiation can be suppressed and thus (significantly) reduce the success for their own company. However, the conflicting interests of the various parties can also give rise to these patterns of behavior. For some people they become noticeable through their behavior, for others through the way they express themselves, or even through their voice. This shows that emotions appear in different ways and can occur especially in the undesired moments. Particularly in distributive negotiations, the focus is on maximizing your own benefit. Nevertheless, it is neglected that the relationship with the other parties is an important element to bringing a negotiation situation to a successful conclusion. The described technique is used in integrative negotiations [1]. Therefore, it is not far off that emotions play an essential role here, or would you want to give your counterpart a piece of the cake, although he just verbally hit you in a tour? - Probably not.

So, wouldn't it make sense to be able to recognize and train emotions in real time by means of the voice in order to make negotiations influenceable in advance?

To make the responsible emotions recognizable on both sides, a technical application is able to help. Results have already been achieved in this area. In particular that very strong emotions have a strong influence on negotiations. [2] [3] [11] However, in the most papers only the physical characteristics of the people were considered. [11] [12] [13] In order to extend the approach from the physical characteristics or even to replace it, an application like this prototype can help. This is primarily intended to recognize emotions by voice and help inexperienced negotiators in training scenarios as follows. Such a training scenario could be conducted by negotiation trainers or psychologists. The prototype shows simultaneously the hidden emotional states. This makes the negotiation training more efficient and the trainers can directly point out the weaknesses of the person to be trained. Subsequently, the emotional intelligence can be trained directly at the weak point. For this reason, the following two research questions attempt to clarify how helpful a developed application that classifies emotions in real time is:

(1) Is it possible to measure and classify emotions with the help of a deep neural network using the characteristics of the voice.

(2) Is a software able to support negotiations in advance by classifying emotions (e.g. during trainings or educations)?

A first prototype was developed in the context of this work and validated by a test. In the further process the prototype will be improved by additional experiments.

Theoretical background and related work

In psychology, emotions are described by emotional states. These usually consist of a chain of feelings such as joy, sadness, anger, fear, disgust or surprise. Emotions can affect feelings, the physical state and behavior. Through personal experience, the intensity of the perception of emotions can vary. This makes it almost impossible to trigger and measure emotions for everyone in the same way. Furthermore, studies prove that emotions can have an impact on a person's health. However, the expression of emotions can also be lived out in different ways depending on the culture. In context of this emotions can also have a very deep impact in negotiations [4] [5] [7] [8] [10]. Nevertheless, the different emotions can be simplified in a two-dimensional model. This concept has already been described by Russel [6] in 1980 and by Sharma et al. in 2008 [12]. With the help of this model, emotions can be classified according to their intensity and their negative or positive rating. It is important to know what physical effects can be triggered by emotions. For example, the heart rate of a person will increase in stressful situations [11]. For others, skin conductance decreases when they feel angry [12]. In addition to the effects already mentioned, there are other effects, such as changes in brain waves or blood pressure [11] [12] [13]. Furthermore, people's facial expressions and gestures could change. Thus, a camera can be combined with face tracking to confirm whether a person is happy or angry [13] [14]. These findings can be used to generate new training data or to check the current emotion using medical parameters. By triggering emotions, the physical effects on the body described above can be triggered. Emotions can be evoked naturally and artificially. Triggers can be pictures and music- or videoclips. These triggers can consciously and unconsciously trigger various emotions. However, the culture or personal taste plays an important role too [10] [11] [12] [15]. Angry and loud people and their posture can also be triggered. Depending on the level of emotion and how it appears, a disinterest can be triggered in conversation partners [15] [16] [17] [18]. This can massively affect negotiations. That is why we have to bring the emotions closer in the process of negotiations. At first, we can mention Bercovitch [19]. He discussed the problems in negotiations in 1984. In his scientific work he described the problems of a conflict as well as the topic of conflict management in negotiations. Both parts are mostly accompanied with emotions. He also dealt with the psychological approaches, as well as the personal abilities of the negotiating partners. This topic was also addressed by Druckman et al. [17]. The authors showed in general how emotions influence negotiations. They described that i. e. angry can influence the outcome of a negotiation on a positive and negative way. It depends on the objective of the negotiators. In order to make the results of Druckman et al. even clearer, the strong emotion of fear can also be considered. Once the negotiators are in an anxious state, there is a risk that they may react intimidated, dejected or even shaken. Positive emotions, such as happy, can also influence the outcome of negotiations through mutual trust and good cooperation [10] [5] [9]. These relationships can be used to recognize emotions and to train the emotional intelligence. In order to be able to use emotions consciously, Bosley et al. investigated, among other things, the training of emotions [20]. Holt et al. also examined how emotional intelligence can affect negotiations and other professional situations [21]. Finally, we want to establish a connection between those issues.

The human language contains essential information about their emotions. For example, emotions can be recognized through semantics. In this case people can fake certain emotions by choosing their words. However, it is also possible to recognize emotions by the pitch of the sound of voice. This can be achieved with software and hardware. Especially the author Bernd Schuller plays a major role in the area of emotions and the ways to recognize them [22] [23] [24]. In his first relevant article to this work, Schuller et al. [24] presented a prototype that runs as an application on Android smartphones and recognizes emotions by means of a neural network. This application was developed for a company, called audEERING, and is now actively marketed to support call centers [34]. Another article by Schuller et al. [23] describes the use of a new innovative neural network (Long Short-Term Memory Recurrent Neural Network - LSTM RNN, proposed by Hochreiter and Schmidhuber back in 1997 [34]). This network is designed to detect speech overlaps and filters out non-speech. It also recognizes the gender of the speaker. Furthermore, it was also used for the application already presented above. In addition to Schuller, other scientists also investigated the problem of sentiment detection. Based on these sources, it can be said that negotiations can be specifically influenced by training of the voice. This can be realized by a software that recognizes emotions in the voice in real time. It will be explained in the next section.

In order to be able to measure emotions via the voice, we have used the technology of deep neural networks. Artificial neural networks (ANN) are comparable to the human nervous system. These nervous systems consist of innumerable neurons that can simultaneously receive, process and output information. Neural networks belong to the subfield of artificial intelligence (AI) [22] [25] [26] [27]. The components of neural networks can be used and configured individually. We have selected so called Convolutional Neural Networks (CNN). These will be trained with labelled data and can then independently and reliably divide patterns and relationships into predefined classes. They are a widespread form of deep neural networks and are often used in speech and image recognition. Like a conventional neural network, CNNs consist of different layers. In convolution layers, each neuron is connected to each neuron of the previous layer. For example, CNNs exploit certain properties of input data from neighborhoods between pixels in an image. All of these properties can be seen as differentiating them from other machine learning methods, such as ensembler and clustering. However, the theory of depth is not part of this work and will not be discussed further. The developed prototype is explained in the next section.

The approach: Sentiment detection using deep learning

The problem of sentiment detection from audio signals was addressed with methods of machine learning, in specific classification, which refers to the assignment of data vectors to one of N classes. In a first step a machine learning model is trained on data vectors with known classes. Following that, unknown data is classified using the fitted model. The input data consists of voice recordings of sentences spoken by various actors in different moods: “neutral”, “happy”, “angry” and “fearful”. As training data, we used the "RAVDESS Dataset" [30] and the "European Stimulus Dataset" [31]. The audio signals correspond to one univariate time series per test person and sentence. The goal is to build a model that can classify new voice recordings as any of these four classes, i.e. sentiment detection on voice recordings. This could be achieved using feature extraction methods from the speech recognition domain, e.g. the extraction of features from the frequency domain. Alternatively, machine learning models like recurrent neural networks could be applied to the raw time series. In this paper, the knowledge about relationships between ascendant datapoints within the time series is exploited: each voice recording is transformed to a two-dimensional image showing the signal plot (wave plot). The images are used to train a convolutional neural network – which is a specific type of artificial neural network from the field of deep learning.

Tensorflow and Keras, which are special software libraries for machine learning, were used as a basis [28] [29]. By the interplay of these two frameworks an artificial neural network can be developed relatively easily. For the implementation we used a so-called 2-dimensional Convolutional Neural Network. However, before the network can be trained, the voice recordings in the training data were transformed to waveplots (= signal plots).

In the process of hyperparameter tuning (which means testing of different configurations of the neural network), a Convolutional Neural Network with three convolutional layers, three dense layers and the available training data was used. Additionally, the function to reduce the learning rate on a plateau was used. This function looks for certain conditions and adjusts the speed of the learning process. This can improve the accuracy of the neural network. The optimal configuration of the neural network can only be determined by multiple tests with different hyperparameters. Finally, the configuration of the network which has achieved the highest accuracy and lowest loss will be selected.

The trained model was embedded into a program with a graphical user interface (GUI). This GUI takes contains a start and stop button. By pressing the "start button" the voice recording is started, recording 2.5 seconds of audio. Subsequently, a waveplot is generated from this recording and transferred to the neural network. This creates a prediction of the emotion and writes it into a text file. To terminate the voice recording, the "stop button" must be pressed. Only a standard microphone is required for the voice recordings.

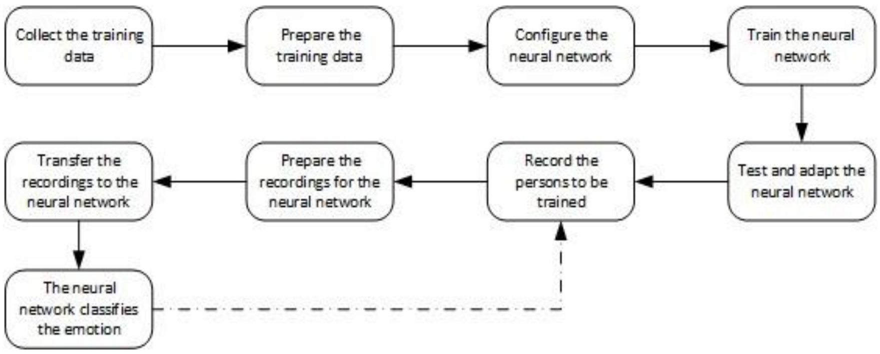

Figure 1: Data flow model

Scientific Experiments

Essentially, three experiments were planned in order to present the individual results in a structured manner. These should validate our prototype and show whether a neural network can recognize emotions. In the further course of the conference it will also be clarified whether the created prototype can be used in negotiation situations or for training scenarios. So far, the first experiment has been conducted. The second and third experiments will be prepared soon. In the following, the individual scientific experiments will be explained in more detail.

Scientific Experiment I: hold-out test set

The first experiment evaluates the accuracy of the neural network with a hold-out test set of preselected voice recordings. For this purpose, about 80 speech recordings of each emotion were separated from the complete data set prior to training and not used to train the neural network. This is a common approach in machine learning to estimate the error on new data. It might however underestimate the effect of overfitting, since this test set contains recordings of the test persons included in the training set.

Scientific Experiment II: test on new test persons

In the second experiment, the neural network classifies voice recordings that were recorded from randomly selected persons not included in the training set. The goal is to test how the neural network can deal with unknown voices in potentially different environments.

Scientific Experiment III: test data from real negotiations

The last scientific experiment is carried out with voice recordings from test persons that were not included in the training set, and in addition to experiment II, were conducting real negotiations. This scientific experiment is based on the so-called dictator game which is a modification of the ultimatum game. Basically, two partners (A1 and A2) negotiate with each other. For example, A1 has 20€ and has to decide how much of his money he/she wants to give to A2. In a standard ultimatum game, A2 can accept or reject the offer. If A2 rejects the offer, neither of the two parties get any money. However, if A2 accepts the offer, the money is distributed as agreed. In the dictator game, party A2 cannot reject A1's offer. In addition, in our variant, A2 should trade for the highest possible amount and convince A1 to give away as much of his/her share as possible. Furthermore, A2 has only a certain amount of time to convince A1 to accept the offer. In this scientific experiment party A1 is the role of the test persons. Time pressure and the need for increasing the own outcome should trigger various emotions in A2. For this purpose, we used the developed prototype, which records and evaluates the speech in real time and saves the determined emotion in a text file. This text file will be evaluated to validate the neural network in a real-time environment [32] [33].

Experimental Results

Results on validation sets during training

2272 recordings were used by the neural network for training. However, 15% must still be deducted for the validation set. With these recordings we achieved a val_accuracy of almost 70%. The val_accuracy indicates the accuracy of the predictions of a randomly separated validation set after each training period. However, the general accuracy continued to increase (see Figure 2). It is interesting to note that the val_accuracy does not decrease and remains constant at about 70%, which may be due to the overfitting described below.

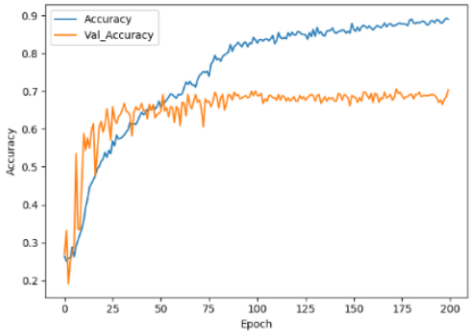

Figure 2: Accuracy vs. val_accuracy

The values for loss are similar (see Figure 3). The general loss decreases after almost every epoch and approaches the value 0, whereas the val_loss stagnates like the val_accuracy. The loss shows how close the neural network is to the optimum. At about 0.85 it stops decreasing and gets a very slightly increasing tendency.

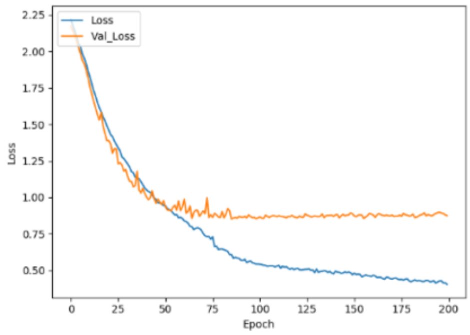

Figure 3: Loss vs. val_loss

Both the accuracy and the loss indicate an overfitting of the neural network. Overfitting is the memorization or over adaptation of training data and the resulting lack of generalization. This can cause the neural network to classify test data incorrectly because it cannot handle the generalization of the problem. This can possibly be solved with a few improvements (see discussion and conclusion).

Scientific experiment I: results on hold-out test set

For the first scientific experiment, we selected 275 recordings in advance, which were not used to train the ANN. This enabled us to test the neural network with the same subjects with whom it was trained. So, we had 80 recordings for each emotion, with the exception of the emotion "fear". In this case we just used 35 recordings. In this experiment we reached a general accuracy of 66.91%. The exact results are shown in the Confusion Matrix (Figure 4). Especially the emotion "neutral" (accuracy of 81.25%) could be predicted very well by the neural network. Likewise, the neural network was able to detect "fear" quite well with an accuracy of 77.14%. The emotions "happy" and "angry" (accuracy of 58.75% and 56.25%) were not recognized very well. We noticed that the neural network recognized "neutral" for the emotion "happy" and the emotion "happy" for "angry" more often than the other emotions.

	Neutral	Happy	Angry	Fear
Neutral	65	14	11	2
Happy	4	47	13	5
Angry	2	10	45	1
Fear	9	9	11	27

Accuracy: 66.91 %

Figure 4: confusion matrix for the scientific experiment 1

Discussion and Conclusion

There is still some work left. The prototype and the experiment will be evaluated in more detail. Furthermore, the research questions posed in the introduction are to be answered. As this is a prototype, the main focus of this scientific work was to answer the questions of (1) how emotions can be recognized and classified with the help of software and (2) whether this software can support negotiations in advance.

(1) Is it possible to measure and classify emotions with the help of a deep neural network using the characteristics of the voice.

During the individual experiments it was noticed that the neural network still has problems with different voices. In order to recognize individual emotions with poor performance better, it is possible to train them with more training data. However, due to the limited amount of training and test data the number could not be increased individually. Therefore, the neural network should be trained with the trainees prior to each training session. By this way the neural network is prepared for the new voices and can predict the trainees' emotions more exactly. During the additional training, the emotions of the speakers should also be measured, e.g. by using clinical measurements, in order to avoid incorrect inputs. Furthermore, a function should be included which can recognize the different negotiators by their voice. In addition, the neural network can be continuously improved using this approach by obtaining more and more training data. Over this way the current deficit of training data can be eliminated in the long-term. Nevertheless, with the help of scientific experiments and literature research we can assume that the prototype can be improved to generate better results.

(2) Is a software able to support negotiations in advance by classifying emotions (e.g. during trainings or educations)?

This research question is intended to clarify whether the prototype can help to train emotional intelligence. The prototype presented in this article is based on a person's voice and classifies emotions using waveplots in real time. Thus, the neural network does not exclusively accept words and cannot be manipulated by the choice of words. Due to this characteristic, a training of emotional intelligence can be performed prior to a negotiation. This can be used to prepare future participants of negotiations. Thus, a subsequent analysis and final training of the emotional intelligence should be possible. The goal should be to train participants to get deeper insights into their own feelings and to better control them.

After addressing the two research questions, it can be said, that an application that recognizes emotions by voice and therefore helps to train emotional intelligence can provide great value. Especially young and inexperienced people, who may be attending their first negotiation, can benefit from such an application and learn a lot about their own emotions.

Outlook

In the following, the authors' suggestions for further work are explained. The legal, technical and medical areas can be considered, which depend very much on regional characteristics.

The current state of the art allows applications to be created that penetrate into the world of human emotions and emotional intelligence. With these novel applications, human behavior can be observed and analyzed much more closely. The presented prototype of a fully functional neural network, which is used for the recognition and training of emotional intelligence, can become a pioneering project through further research. Above all, more training data must be made available. Here it would make sense to evoke the emotions with the help of different triggers and to validate them by clinical measurement methods. In addition, market studies should be conducted on the acceptance of such a technique, because negotiations as well as preparations and training still take place in the traditional way, without any technical support. The introduction of modern technologies into traditional procedures is often delayed and this work should accelerate this process. In this way, negotiations can be optimized in advance in a more fundamental way than before.

References

[1] Stoshikj, M.: "Integrative and distributive negotiations and negotiation behavior", Journal of Service Science Research, pp. 29 – 69 (2014).

[2] Van Kleef, A., G.; De Dreu, K., W., C.; Manstead, R., S., A.: "Supplication and Appeasement in Conflict and Negotiation: The Interpersonal Effects of Disappointment, Worry, Guilt, and Regret", Journal of Personality and Social Psychology, Volume 91, pp. 124 – 142 (2006).

[3] Polzin, T. S., Waibel, A. H.: "Detecting Emotions in Speech", Proceedings of the CMC (1998).

[4] Fisher, R.; Ury, W.; Patton, B.: "Das Harvard-Konzept – Die unschlagbare Methode für beste Verhandlungsergebnisse", Volume 25, Campus Verlag, Frankfurt / New York (2015).

[5] Shapiro, D. L.: "Emotions in negotiation: peril or promise?", Marquette law review, Volume 87 (2004).

[6] Russel, A. J.: "A Circumplex Model of Affect", Journal of Personality and Social Psychology, Volume 39, Issue 6, pp. 1161 – 1178 (1980).

[7] Helfrich, H.: "Kulturvergleichende Psychologie", Basiswissen Psychologie, Springer VS (2013).

[8] Smith, H. M.: "Emotions", Encyclopedia of Aging and Public Health, Springer, Boston (2008).

[9] Adler, S., R.; Rosen, B.; Silverstein, M. E.: "Emotions in Negotiations: How to Manage Fear and Anger", Negotiation Journal, Volume 14, Issue 2, pp. 161 – 179 (1998).

[10] Yoon, K.; Lee, J.; Kim, M.: "Music Recommendation System Using Emotion Triggering Low-level Features", IEEE Transactions on Consumer Electronics, Volume 58, Issue 2, pp. 612 – 618 (2012).

[11] Canento, F.; Fred, A.; Silva, H.; Gamboa, H.; Lourenço, A.: "Multimodal Biosignal Sensor Data Handling for Emotion Recognition", Sensors (2011 IEEE), Limerick (2011).

[12] Sharma, T.; Bhardwaj, S.; Maringanti, B. H.: "Emotion estimation using physiological signals", TENCON 2008 – 2008 IEEE Region 10 Conference, Hyderabad (2008).

[13] Haag, A.; Goronzy, S.; Schaich, P.; Williams, J.: "Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System", Tutorial and Research Workshop on Affective Dialogue System Systems, pp. 36 – 48 (2004).

[14] Bailenson, N. J.; Pontikakis, D. E.; Mauss, B. I.; Gross, J. J.; Jabon, E. M.; Hutcherson, C., A., C.; Nass, C.; John, O.: "Real-time classification of evoked emotions using facial feature tracking and physiological responses“, International Journal of Human-Computer Studies, Volume 66, Issue 5, pp. 303 – 317 (2008).

[15] Adler, S., R.; Rosen, B.; Silverstein, M. E.: "Emotions in Negotiations: How to Manage Fear and Anger", Negotiation Journal, Volume 14, Issue 2, pp. 161 – 179 (1998).

[16] Allred, G., K.; Mallozzi, S., J.; Matsui, F.; Raia, P., C.: "The Influence of Anger and Compassion on Negotiation Performance", Organizational Behavior and Human Decision Processes, Volume 70, Issue 3, pp. 175 – 187 (1997).

[17] Druckman, D.; Olekalns, M.: "Emotions in negotiation", Group Decision and Negotiation, Volume 17, Issue 1, pp. 1 – 11 (2007).

[18] Santos, R.; Marreiros, G.; Ramos, C.; Neves, J.; Bulas-Cruz, J.: “Personality, Emotion, and Mood in Agent- Based Group Decision Making”, IEEE Intelligent Systems, Volume 26, pp. 58 – 66 (2011).
[19] Bercovitch, J.: “Problems and Approaches in The Study of Bargaining and Negotiation”, Political Science, Volume 36, Issue 2, pp. 125 – 144 (1984).
[20] Bosley, I.; Kasten, E.: “Emotionale Intelligenz - Ein Ratgeber mit Übungsaufgaben für Kinder, Jugendliche und Erwachsene”, Springer Verlag (2018).
[21] Holt S.; Wood, A.: “Leadership and Emotional Intelligence”, In: Marques J., Dhiman S. (eds) Leadership Today. Springer Texts in Business and Economics. Springer, pp. 111 – 138 (2017).
[22] Deng, J.; Eyben, F.; Schuller, B.; Burkhard, F.: “Deep Neural Networks for Anger Detection from Real Life Speech Data”, 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), San Antnio (2017).
[23] Hagerer, G.; Pandit, V.; Eyben, F.; Schuller, B.: “Enhancing LSTM RNN-based Speech Overlap Detection by Artificially Mixed Data”, Semantic Audio, Erlangen (2017).
[24] Marchi, E.; Eyben, F.; Hagerer, G.; Schuller, B.: “Real-time Tracking of Speakers’ Emotions, States, and Traits on Mobile Platforms”, Interspeech 2016: Show & Tell Contribution, pp. 1182 – 1183, San Franzisco (2016).
[25] McCulloch, S., W.; Pitts, W.: “A logical calculus of the ideas immanent in nervous activity”, Bulletin of Mathematical Biophysics, Volume 5, pp. 115 – 133 (1943).
[26] Rumelhart, E. D.; Widrow, B.; Lehr, A., M.: “The Basic Ideas in Neural Networks”, Communications of the ACM, Volume 37, Issue 3, pp. 87 – 92 (1994).
[27] Ciresan, D.; Meier, U.; Masci, J.; Gambardella, M., L.; Schmidhuber, J.: “Flexible, High Performance Convolutional Neural Networks for Image Classification”, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Spain (2011).
[28] Tensorflow Documentation: https://www.tensorflow.org/.
[29] Keras Documentation: https://keras.io.
[30] SMART Lab, Ryerson University: https://smartlaboratory.org/ravdess/.
[31] University of Cambridge: https://www.autismresearchcentre.com/projects/Emoticons.aspx.
[32] Güth, W.; Schmittberger, R.; Schwarze, B.: “An experimental analysis of ultimatum bargaining”, Journal of Economic Behavior and Organization, pp. 367 – 388, North-Holland (1983).
[33] Nelissen, R.; Leliveld, M.; Van Dijk, E.; Zeelenberg, M.: “Fear and guilt in proposers: Using emotions to explain offers in ultimatum bargaining”, European Journal of Social Psychology, pp. 78 – 85 (2011).
[34] AudEERING GmbH, https://www.audeering.com/.
[35] Hochreiter, S; Schmidhuber, J.: “Long short-term memory In”, Neural Computation, Volume 9, issue 8, pp. 1735 – 1780 (1997).