Journal of Advanced Research in Intelligence Systems and Robotics
Volume 5, Issue 2 - 2023, Pg. No. 27-33
Peer Reviewed Journal

Research Article

Psychological Detection Using Artificial Intelligence

Gourav Singh

UG Student, CSE, Anurag University, Hyderabad.

I N F O

E-mail Id:
gouravgs24@gmail.com
Orcid Id:
https://orcid.org/0009-0007-8815-3798

How to cite this article:
Singh G. Psychological Detection Using Artificial Intelligence. *J Adv Res Intel Sys Robot 2023*; 5(2): 27-33.

Date of Submission: 2023-09-18
Date of Acceptance: 2023-11-05

A B S T R A C T

In recent years, the attention to the study of emotional content of speech signal has been increased for the interaction of human with machines. New paradigm of human-computer interaction like emotion, gesture and force-feedback are emerging. In spite of that, one imperative component for natural interaction is still missing i.e., voice. This paper describes the classification of different emotional states by using voice quality information only. Emotions that are being observed in this paper are six basic emotions which includes happy, sad, angry, surprise, disgust and neutral. Different researchers inspected dataset of various voice parameters by doing multiple experiments for detecting psychological emotions by using different voice signals and achieved a reliable recognition rate. The best accuracy that is observed after analyzing all the experiments thoroughly is around 80% which is a comparable accuracy. Though the exact accuracy for voice analyzing is not still found and human-computer voice recognition can be improved, the authors are seeking to increase the dataset by adding more voice parameters.

Keywords: Emotion recognition, Classifiers, Feature extraction, Continuous emotions, SVM (Support Vector Machine), CNN (Convolutional neural network), ERIC (Education Resources Information Centre), IEEE (Institute of Electrical and Electronics Engineers) DNN (Deep Neural Networks), AI (Artificial Intelligence), ML (Machine Learning)

Introduction

Voice is a complex process that presents bio markers. It's a complex physical and cognitive process pumping air through our lungs, varying the tension in our local folds, changing the shape of our local cavity with our tongue, with our jaw.[1,6] Expelling a time varying pressure wave between 50 and 7000 hertz and every step along the way your voice production is being affected by factors like how awake are you, our energy level, hormones, our stress level etc and so we can analyse all these bio markers. Emotion perception from speech is not always that simple this is because in lot of cases you actually rely on context. So, it's very natural to represent using categories. As mentioned in Fig 1 Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral and many more. So, listed above seven, but it's supposed to be big six. But of course, it's neutral which means there's no emotion expressed and a lot of times it's actually really challenging to decide the number of categories that's suitable because certainly human emotion is more than this right, we can have contempt and can have excitement. Some features used for emotion classification are Mel Frequency Cepstral Coefficient (MFCC), Fundamental frequencies, and Linear Prediction Cepstral Coefficient (LPCC).

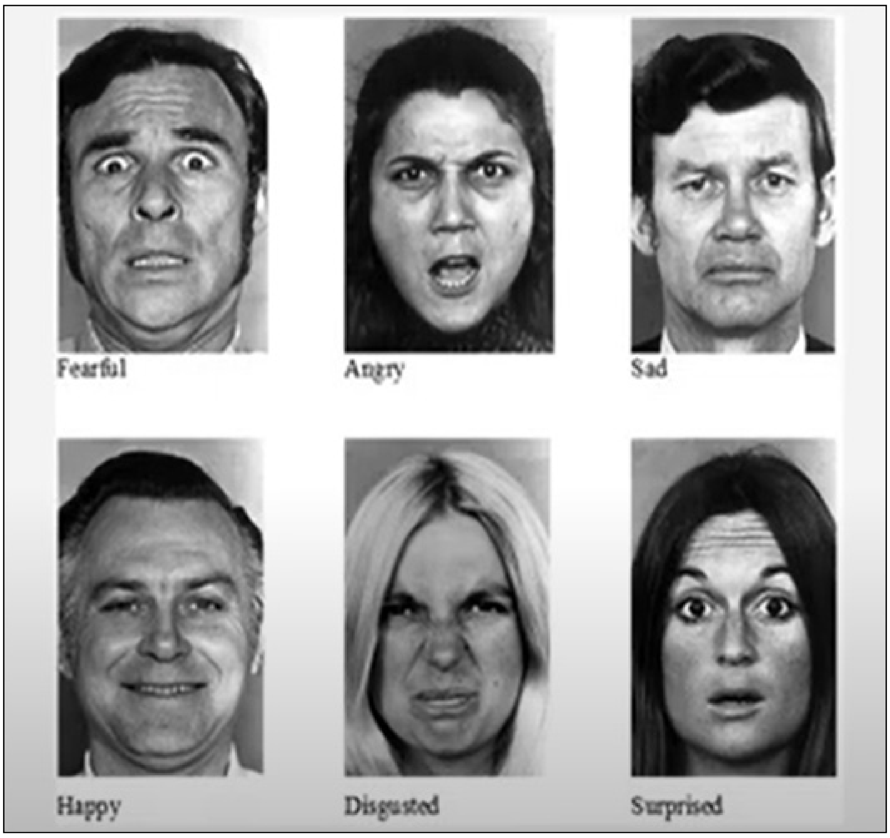

Figure 1.Discrete Emotions

According to the author human's ability to perceive different emotions is extremely sophisticated. The study of how humans perceive emotions perceive different sounds is what we call auditory perception.[14] It affects the production of the sound a speaker's emotional state effects such as the tension in the muscles or the rapid breathing. This in turn affects the five level details in the voice. Such as the voice quality or the pitch. In the upcoming 20 years humans would be more dependent on machines. So, the need is to make the communication natural between human and machine. An overview of how the emotions is detected from speech can be seen in Figure 2.

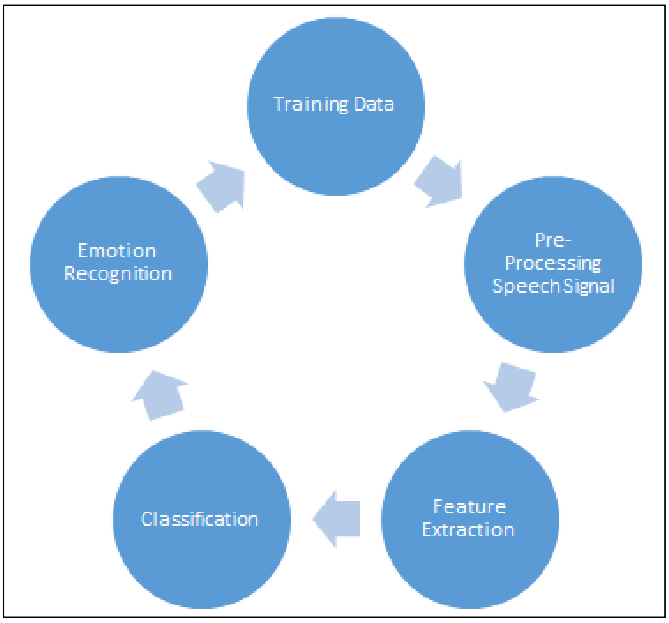

Figure 2.An overview of Emotion Detection System

These details can be characterized in terms of feature representations which can be deduced to figure out emotions such as if the person is angry or the person is sad or any of the six universal emotions by Paul Ekman but why stop at six emotions only ,why not eight emotions, why not a whole spectrum of emotions.[6] The thing is humans demonstrate complex emotional states and so if you're sort of discretizing emotions such as this you're sort of limiting the power of expression as well as perception and so there's a very famous model in the speech community which is called emotion primitives or the valence activation dominance which is what we use in our study, where we basically do is we project emotions on a 3-D continuous space such that you can capture the nuances that are in the emotional states .

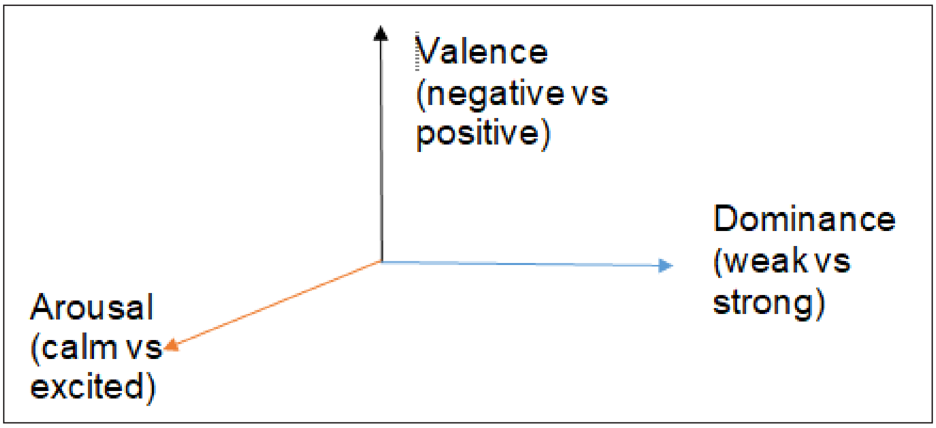

Figure 3.Continuous emotions

So, some of the applications of this technology are obvious things like security checkpoints or intelligence, human machine team. This is the idea of humans and AI systems working closely together just as it's important for humans to be able to trust and understand the AI systems that we're working with it's going to be important for those systems those machines to understand to perceive to reach to their human counterparts.

Related Work

In this field of psychology detection, author have come across a few researches which are very much related to psychological detection using voice. The conclusion from previous research papers shows that various voice parameters can be used to detect the mental health of a person. The audio-visual dataset consists of more than 2000 video clips that describe the most significant emotion states which are happiness, sadness, anger, surprise, disgust, and neutral. This particular research is based on 'The audio-visual dataset for natural emotion'.[21] The paper is organized in various sections where the first two sections display an overview of existing databases and the type of database available. And further in the next two sections, the database collection process was discussed. In the next section, databases structures and contents are present. Overall audio based and video experiments are performed for the collection of datasets. This dataset ended up having 82 subjects of different ages and genders covering six basic emotions. With this exclusive dataset, they ended up obtaining an accuracy of 54.5% and 57.9% using audio and video data respectively.[21] These six parameters can be proved helpful for our aim to collect dataset of various parameters for psychological detection by voice only.

In addition to this, coming across another research which is performed for the detection of deception using the voice. The methodology for this particular research is the level of stress of an individual which would offer a number of advantages over current psychological monitoring. But the results of a low and high stress study show that reliable changes in the voice are evidenced by the changes in the autonomic nervous system only which is unlike our project. Recent unpublished work shows that the level of stress experienced by the subject is an important factor affecting the accuracy of voice analysis.

One more research is there which has purposed for emotion recognition using music and lyrics only. But the performance attained by the employed lyrical dataset is significantly worse as compared to previous models which are created from standard audio, melodic and midi features. In another research, which is the cascaded emotion classification via psychological emotion detection using a large set of voice quality parameters, improved the speaker independent emotion recognition by two measures only. Firstly, there is an extension in the voice quality parameter set. Secondly, used a psychological motivated cascaded classification strategy. The psychological motivated cascaded classification is done by 2 stage approach. In this, 6 basic emotions are divided into high and low. The author has purposed of large set of 67 voice quality parameters and a new feature set with 2 stage classification led to recognition rate of 83.5%. With this approach, there could be a further increase overall recognition rate to 88.8%.[25] With this particular research, more than 60 voice quality parameters along with 6 emotions as well which has already succeeded for 80% of recognition rate, have concluded.

In this field of psychology detection, author compares different approaches for emotions recognition task and we propose an efficient solution based on combination of these approaches. The goal of the psychology detection using voice is to recognize the user's emotional state precisely. One of the sticking points is what features should be used. In this paper, a large set of 67 voice quality parameters along with 6 emotions has purposed as well which has already succeeded for 80% of recognition rate.[25]

In the future, the author is aiming to increase the size of the database for covering more emotions using voice only.

Methodology

Finding reliable and valid sources of information on psychological detection was a challenge for us. The researcher used emotion detection using keyword-searching methods to locate traditional and online sources on the topic. The primary database used to locate sources was ERIC. The ERIC database was useful in locating full-text articles from well-known research journals and publications. ERIC is a database that focuses on retrieving sources related to education and new technology. The researcher was most successful in finding pertinent information when using the following descriptors: emotions, psychology, technology use, and DNN. Another database used to locate sources was the IEEE library. The researcher used IEEE primarily to locate traditional sources such as books and some research articles. The third source of locating traditional sources was the faculty within the Chandigarh University. The researcher was able to review books from respected researchers in the fields of AI, emotion detection, and CNN. The final source of locating information was the use of World Wide Web search engines. The researcher was able to locate full-text research articles from online journals using search engines. The challenge of citing resources from the World Wide Web is that the researcher had to check for the credibility of the information found. This was an important issue because the researcher recognizes that anyone can publish information online that is false or misleading. To check for credibility, the researcher found background information on the authors of the online sources and determined if the information was credible. To determine further credibility, the researcher entered the authors' names into the ERIC database and found that many of the authors have several publications in the fields of emotion detection, psychology, AI and/or instructional technology. The primary rationale for selecting the sources described above was reliability. The second rationale for selecting the above sources is the researcher's own interest in the topic of emotion detection and use of technology.

Table 1. Comparative analysis of emotion recognition systems and algorithms

S.no	Year	Author Name	Dataset	Classifier	Accuracy	Result
1.	2009	Shrikanth Narayan et al.[24]	Usc-Iemocap	SVC GMM	70%	Emotions Based
2.	2018	Zhang et al.[1]	Emo-Db Rml Enterface05 Baum-1s	SVM SVM SVM SVM	86.30% 75.20% 79.40% 44.03%	Emotions Based
3.	2015	Shivaji Chaudhari et al.[2]	Recorded audio clip	SVM AND GMM	81%	Polarity Based
4.	2010	Ezzat et al.[3]	Call centre database of 36 audio files	HC KNN SVM	50% 66% 94.4%	Polarity Based
5.	2010	Hassan et al.[4]	500 Utterances spoken by actors	J48 CART K MLP	79% 64.58% 67.58% 64.50%	Polarity Based
6.	2017	Ahmad et al.[5]	Berlin Dataset	CNN	84.3%	Emotions Based
7.	2017	Pascale Fung et al.[6]	Data from ongoing annotation project	CNN	66.1%	Emotions Based

8.	2016	Fayek et al.[7]	Enterface05 Savee	DNN DNN	60.53% 59.7%	Emotions Based
9.	2005	Chul Min Lee et al.[8]	7200 Utterances of calls	LDC KNN	Not Available	Polarity Based
10.	2014	Huang et al.[10]	Buaa-Db	SVM	86.5%	Emotions Based
11.	2016	Bertero et al.[11]	Ted-Lium-Db	CNN	65.7%	Emotions Based
12.	2010	Chavhan et al.[12]	Berlin-Db	SVM	93.75%	Emotions Based
13.	2001	Feng Yu et al.[9]	Chinese Teleplays-Db	SVM	77.16%	Emotion Based
14.	2012	Gayar et al.[13]	36 recorded audio files	Decision Tree SVM K-nearest	86.1% 94.4% 91.7%	Polarity Based
15.	2015	Chen et al.[14]	Usc-Iemocap	SVM	69.2%	Emotions Based
16.	2015	Zheng et al.[15]	Iemocap	PCA-DCNNs-SER	40.02%	Emotions Based
17	2015	Knuxia Wang et al.[16]	Emodb Casia Eesdb	SVM	79.51%	Emotions Based
18	2009	Tania Banziger et al.[17]	Danva	MERT ERI JACFEE PONS DANVA	65% 63% 52% 39% 55%	Emotions Based
19.	2007	Mark Lugger et al.[18]	Prosodic	GMM	80.2%	Emotions Based
20.	2006	Nicu Sebe et al.[19]	Bimodal Testing	HMM	90%	Emotions Based
21.	2010	Moataz Ei et al.[20]	Front end processing unit and classifier	HMM GMM ANN SVN	75-78% 74-81% 51-52% 75-81%	Emotions Based
22.	2019	Ftoon Abu Shagra et al.[21]	3000 clip of audio and video	SVM MLP K-NN Decision Tree	54.5%	Emotions Based
23.	2014	Fabien Ringeval et al.[22]	Emotiw14 dataset	SVM	Improved by 9.80%	Emotions Based
24.	2019	Muhammad Fahreza Alghifari et al.[23]	Emo-Db Low Quality	VAD	Improved by 3.5%	Emotions Based
25.	2018	Huiyi Wu et al..	Dataset of Emotion	CNN	77.2%	Emotions Based

26.	2016	Savita Sondhi et al.	RML	PRAAT	87.58	Emotions Based
27.	2008	Marko Lugger et al.	Dataset of 6 basic emotions	CNN	14% -88.8%	Emotions Based
28.	2019	Maryam Imani et al.	Emotional database	CNN	95% -63%	Emotions Based
29.	2020	Zhou Qing et al.	Audio clip RML	KNN	78.0%	Emotions Based
30.	2019	Tanya Keshari et al.	Functions Using facial expressions	CNN	90%	Emotions Based
31.	2019	Kartic Hausal et al.	Audio Clips	CNN	78%	Emotions Based
32.	2004	K.Takahashi	EMO-DB	SVM	41.01%	Emotions Based
33.	2021	Kunyu Peng	Emotional Database	SVM	50.0%	Emotions Based
34.	2020	Wei sun et al.	RML	SVM	86.3%	Emotions Based
35.	2018	Fen Xu;	Dataset of emotion recognition	CNN	38.41%	Emotions Based
36.	2021	Fangyu and Li	Usc-Iemocap	CNN	73.6%	Emotions Based
37.	2018	Subhas Mita Sahoo	38 recorded files	SVM	93%	Emotions Based
38.	2021	Sofia de La	EMO-DB	CNN	73%	Emotions Based

Result

According to the study of perception, researcher said that yes expressions are expressions over people demonstrate expressions differently but then people perceive emotions differently as well, different genders have different ways of perceiving emotions and so, we use the data that we have which is the data in the wild to study these emotions and come up with novel analysis where we can make use of this noisy data per session.

Conclusion

This paper gives us a summarized overview of the ongoing updates on emotion detection using AI. In the field of emotion identification, database selection plays an important role in perfect accuracy of result. The next important step is featuring extraction from the speech signal in this field. From our examination, we find that the SVM algorithm is widely used in most of the paper for speech signals as well feature extraction because it performs better than the other approach in the case of noise. [15,16] Next major part of sentiment analysis is the use of Classifier. For solving Sentiment analysis issues authors have used the Support Vector Machines ML algorithms. [1,13].

Still, there are many open issues that need to be solved like diversity in emotion, recognize spontaneous emotion and speaker recognition in case of simultaneous conversation. Further research is expected to improve accuracy and increment the number of emotions in an input speech signal.

References

1. Zhang, Shiqing, Shiliang Zhang, Tiejun Huang. et al. "Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching." IEEE Transactions on Multimedia 20, no. 6 (2017).
2. Chaudhari, Shivaji J, Ramesh M. et al. "Automatic speaker age estimation and gender dependent emotion recognition." *International Journal of Computer Applications* 117, no. 17 (2015).
3. Souraya Ezzat, Neamat El Ghayar, Moustafa M. Ghanem. "Investigating analysis of speech content through Text Classification". International Conference of Soft Computing and Pattern Recognition (2010).

4. Esraa Ali Hassan, Neamat El Gayar, Moustafa M. Ghanem. "Emotions analysis of speech for call classification". 10th International Conference on Intelligent Systems Design and Applications (2010).
5. Abdul Malik Babshah, Jamil Ahmad, Nasir Rahim and Sung Wook Baik. "Speech emotion recognition from spectrograms with deep convolutional neural network". IEEE (2017).
6. Bertero, Dario, and Pascale Fung. "A first look into a convolutional neural network for speech emotion detection." In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE (2017).
7. Fayek, Haytham M, Margaret Lech, et al. "Towards realtime speech emotion recognition using deep neural networks." In 2015 9th international conference on signal processing and communication systems (ICSPCS), IEEE (2015).
8. Chul Min Lee, Shrikanth S, Narayanan. "Towards detecting emotions in spoken dialogs". IEEE Transactions on speech and audio processing, vol. 13, no. 2, (2005).
9. Yu, Feng, Eric Chang, Ying-Qing Xu, et al. "Emotion detection from speech to enrich multimedia content." In Pacific-Rim Conference on Multimedia, Springer, Berlin, Heidelberg, (2001).
10. Chavhan, Yashpalsing, M L Dhore. et al. "Speech emotion recognition using support vector machine." *International Journal of Computer Applications* 1, no. 20 (2010).
11. Souraya Ezzat, Neamat El Gayar, and Moustafa M. Ghanem. "Sentiment Analysis of Call Centre Audio Conversations using Text Classification". International *Jornal of Computer Information Systems and Industrial Management Applications*, volume 4 (2012).
12. Jin, Qin, Chengxin Li. et al. "Speech emotion recognition with acoustic and lexical features." In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE (2015).
13. Zheng, W Q, J S Yu. et al. "An experimental study of speech emotion recognition based on deep convolutional neural networks." In 2015 international conference on affective computing and intelligent interaction (ACII), IEEE (2015).
14. Sun, Chi, Luyao Huang. et al. "Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence." arXiv preprint arXiv:1903.09588 (2019).
15. Gan, Chenquan, Lu Wang, et al. "Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis." Knowledge-Based Systems 188 (2020).
16. Kunxia Wang, Ning An, Bing Nan Li, et al. "Speech Emotion Recognition Using Fourier Parameters." In IEEE Transactions on Affective Computing (Volume: 6, Issue: 1, Jan.-March 1 2015).
17. Tanja Ba¨nziger, Didier Grandjean, Klaus R Scherer. "Emotion Recognition from Expressions in Face, Voice, and Body: The Multimodal Emotion Recognition Test (MERT)." In American Psychological Association (2009).
18. Marko Lugger and Bin Yang. "The Relevance of Voice Quality Features In Speaker Independent Emotion Recognition". In IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07 (2007).
19. Nicu Sebe1, Ira Cohen2, Theo Gevers1. et al. "Emotion Recognition Based on Joint Visual and Audio Cues." In 18th International Conference on Pattern Recognition (ICPR'06) (2006).
20. Moataz El Ayadi a, Mohamed S, Kamel b , et al. "Survey on speech emotion recognition: Features, classification schemes, and database." In Pattern Recognition (2009).
21. Ftoon Abu shaqra, Rehab Duwairi, Mahmoud Al-Ayyoub. "The Audio-Visual Arabic Dataset for Natural Emotions." In 7th International Conference on Future Internet of Things and Cloud (FiCloud) (2019).
22. Fabien Ringeval, Shahin Amiriparian, Florian Eyben. et al. "Emotion Recognition in the Wild: Incorporating Voice and Lip Activity in Multimodal Decision-Level Fusion." In 16th International Conference on Multimodal Interaction (2014).
23. Muhammad Fahreza Alghifari1 , Teddy Surya Gunawan2 , Mimi Aminah binti Wan Nordin3. et al. "On the use of voice activity detection in speech emotion recognition." In Bulletin of Electrical Engineering and Informatics (2019).
24. Angeliki Metallinou, Sungbok Lee and Shrikanth Narayanan. "Audio-Visual Emotion Recognition using Gaussian Mixture Models for Face and Voice." In Tenth IEEE International Symposium on Multimedia (2009).
25. Dimensions Using A Large Set of Voice Quality Parameters." In IEEE International Conference on Acoustics, Speech and Signal Processing (2008).