AIRO JOURNALS
Peer Reviewed
Multidisciplinary
International

ISSN:2320-3714
Volume:2 Issue:2
May 2024
Impact Factor: 10.2
Subject: Psychology

Emotion Recognition in Human Speech - A Comprehensive Review of Techniques and Applications

Anusha Shinde
Research Scholar
High School Senior, Gokuldham High School and Junior College, Mumbai

DECLARATION: I AS AN AUTHOR OF THIS PAPER /ARTICLE, HERE BY DECLARE THAT THE PAPER SUBMITTED BY ME FOR PUBLICATION IN THE JOURNAL IS COMPLETELY MY OWN GENUINE PAPER. IF ANY ISSUE REGARDING COPYRIGHT/PATENT/ OTHER REAL AUTHOR ARISES, THE PUBLISHER WILL NOT BE LEGALLY RESPONSIBLE. IF ANY OF SUCH MATTERS OCCUR PUBLISHER MAY REMOVE MY CONTENT FROM THE JOURNAL WEBSITE. FOR THE REASON OF CONTENT AMENDMENT /OR ANY TECHNICAL ISSUE WITHNO VISIBILITY ON WEBSITE /UPDATES, IHAVE RESUBMITTED THIS PAPER FOR THE PUBLICATION.FOR ANY PUBLICATION MATTERS ORANY INFORMATION INTENTIONALLY HIDDEN BY ME OR OTHERWISE, I SHALL BE LEGALLY RESPONSIBLE.(COMPLETE DECLARATION OF THE AUTHOR AT THE LAST PAGE OF THIS PAPER/ARTICLE

ABSTRACT

The field of Affective computing is rapidly emerging as a game-changer within computer science, striving to enhance the interaction between humans and machines. One of its focal points lies in its ability to make machines cognizant of human emotions, garnering significant attention across various disciplines such as Information Sciences and Human-Computer Interaction (HCI). Human speech is a fundamental means of communication, replete with emotional cues. Consequently, this has piqued interest in affecting computer research into emotion recognition speech.

This research paper aims to highlight the numerous strategies applied to identify emotional states in vocal expressions, by providing a short review of the ongoing research in this domain. Since it is critical to extract useful features from speech signals, the article also analyzes machine learning approaches, focussing on classifiers and datasets. Machine learning can evaluate retrieved speech characteristics and categorize emotional states, thereby precisely identifying emotions. Emotion recognition can have many uses, including the medical field, field of Psychology, Cognitive sciences and even marketing. In the medical field, it can gauge patients' emotional distress and track mental health. In psychology and cognitive sciences, it can decipher emotional and behavioral patterns. Tailored marketing and sentiment analysis of the customer are critical in the field of marketing.

***Keywords**: Affective computing, Information Sciences, Human-Computer Interaction (HCI)*

1. INTRODUCTION

Speech and language allow humans to communicate vocally, facilitating the quicker dissemination of ideas, inventions and messages. Human communication entails not only what humans say, but also the manner of saying it. Also, among non-verbal cues, facial expressions account for around 55%, vocal intonation for approximately 38%, and words themselves for about 7% of the message perception.

Language is the medium that fundamentally separates the human species from the rest of existing ones - state P. Boxer and V. Kenny. By origin, human beings are a language-based community, thus human nature is not a 'natural nature' but a nature created by humans via language standards. This nature makes us unique and grants a defining threshold apart from anything else that lacks the adjective 'human'.

The basic command "go home" can be interpreted either as an extremely positive or highly negative remark, just by writing it down without using any tone of voice or facial expressions.

For instance, if a student is ill and the instructor sends him home, or if the student is misbehaving in class and the instructor sends him home. Humans convey emotions - possibly associated with evaluation and subjective perception - through intonation and facial expressions. The rational model of emotion theory describes the process by

which humans produce emotions; it is based on the complex process that embodies the mind-body inter-relation.

Emotions impact judgement depending upon the object that influences judgement. N. Kamaruddin and Wahab defined emotion as a cognitive process occurring randomly due to confounding triggers relayed by physical articulations, shaped via cultural norms and used in specific contexts. In other words, people's emotional perceptions impact the decisions they make.

AIRO JOURNALS
Peer Reviewed
Multidisciplinary
International
ISSN:2320-3714
Volume:2 Issue:2
May 2024
Impact Factor: 10.2
Subject: Psychology

Computer-assisted emotion detection could be useful in certain contexts, like high-stakes aviation or professional driving, where it could improve perception, lead to more informed decisions, and so on. Machine emotion recognition has so far out-performed human emotion recognition. Austermann and Kleinjohann cited research using a fuzzy-rule based system to identify emotions in ordinary speech and independent of a speaker; the former received recognition varying between 55-95 percent, whereas humans only managed about 60 percent when presented with unknown speakers. An online implementation of the method was also created, although it performed 8% worse than the offline implementation.

Emotion recognition technology might aid humans in making accurate decisions, particularly in cases where thinking rationally would take too long, and prompt action is required.

When people's mental and emotional states can be purposefully manipulated, it can be useful for preventing mental diseases from occurring and improving performance in stressful situations.

The manner in which one's brain behaves can be changed by asserting certain resources and taking away others, resulting in each of our major emotional states. The way humans think is structured on a framework based on the following ways to think :

• automatic responses at the most fundamental level,
• responses by learning,
• thorough deliberation,
• contemplation,
• contemplative self-analysis,
• at the most advanced level, there are self-awareness emotions.

Each feeling activates a particular 'way to think' while simultaneously deactivating another.

2. BRIEF OVERVIEW OF AFFECTIVE COMPUTING

R. Picard states that experts of human-computer interaction (HCI) and affective computing focus on four main areas : 1) making users feel more at ease, 2) development of socio-emotional abilities, 3) infrastructure and apps for processing affective input, and 4) conveyance of user emotion. Knowledge of feelings helps to overcome hurdles pertaining to favorable and natural man-machine interaction for applications in patient care, geriatric nursing, call centers, psychological consultation, and human communication. Thus Affective Computing measures observations of the motor system behavior that correlate with high accuracy to an underlying emotion or combination of emotions. Accordingly, it is possible to discern the observer's emotional condition by examining their facial expressions. An exposure to a particular event is necessary for the subject to reach a given state. Such a user experience can also be shaped by machines.

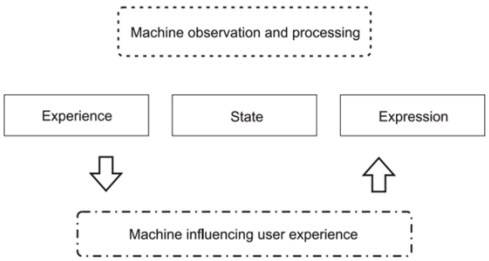

Figure-1. presents a general method within the emotional computing framework (framework of Affective Computing).

This framework emphasizes the multi-disciplinary character of emotional computing and gives a high-level outline of the steps involved, from data acquisition to application integration, computer vision, machine learning, signal processing, and human-computer interface approaches.

The subject's expression may serve as a proxy for the new condition that has been induced by the encounter. We could watch this process unfold while also manipulating it in an artificial way.

The broader societal impact of this new product must be taken into account when studying the human-machine interaction as a single integrated system. New socio-technical theories and models may emerge as a result of this change in focus, with potential applications in the fields of social science, health care, management, and the economy.

3. CONSIDERATIONS FOR VOICE RECOGNITION

A successful machine learning system will be able to deduce human emotions from spoken language by applying various classifiers to the data derived from audio recordings.

The results of tests with similar approaches used to recognise emotions in speech are presented in the next paragraphs. Collecting data from which a machine can learn is the initial stage in teaching it to

ISSN:2320-3714
Volume:2 Issue:2
May 2024
Impact Factor: 10.2
Subject: Psychology

detect emotions. Using both main and secondary sources could accomplish this. Actor recordings might serve as the main input. It can convey many emotions through the reading of the same text, with secondary input potentially drawing from databases that have already been created by other academics. It is also feasible to employ some hybrid strategies.

The most useful data comes from real-world examples; for instance, S. Yamamoto examined a baby's voice in one study. This helped to enhance affective computing technologies that are specifically designed for the needs of new-borns and young children and provided important insights into the early phases of emotional development. Using R. Sharma's input learning sets derived from movie databases is another option. U. Neumann gathered speech samples from different online sources and then assessed emotions by humans; utilizing a machine to generate voices. Danish Emotional Speech Database (DES) by I. Engberg, the Berlin Emotional Speech Database (BES) created by F. Burkhardt and the SUSAS Database created by J. Burkhardt are some valuable resources for studying the effects of stress on speech.

What computers can detect and interpret are the acoustic features of a sound. Some examples of these characteristics are intonation, the degree to which a voice sounds high or low, intensity, and the pattern of frequencies that make up a sound. Speech evolution (TF transformation), linear prediction (based on past sounds to anticipate future speech), and wave properties (cepstral coefficients, formants, harmonicity, and perturbation) are further aspects.

The term 'linguistic features' refers to several aspects of language, including phonemes, words, and paralinguistics, which includes things like pauses and laughter. Rules and symbols (phonetic symbols, words) can represent these traits, and can disclose a great deal of the speaker's attitude, emotional condition, and purpose.

Considerations such as the speaker's age, gender, and mental or physical health are examples of non-linguistic factors. When trying to decipher non-verbal signs of emotion, these elements can have a role in shaping the way speech sounds. Academics can analyze human speech with the use of Praat Software. It can monitor things like voice onset time, provide visual representations of sound, process and filter speech signals, and segment and label speech elements like words and syllables. Many fields rely on it for phonetics and speech recognition.

After features have been retrieved, machine learning algorithms can examine them for statistical correlations or patterns that may indicate an emotional state. The classifier is an integral component of the system that aids in the automatic recognition of emotions using speech data.

The most popular technique for classifying data is Artificial Neural Network (ANN), a computer based system that imitates the way the human brain operates. Emotion recognition and other data-intensive jobs become much easier for them as they learn to spot patterns and correlations.

The k-Nearest Neighbour (k-NN) algorithm sorts data items according to what their closest neighbors think. It works well for emotion recognition tasks since it is simple and effective. As effective classifiers, Support Vector Machines (SVMs) determine the best decision boundary for data classification. Emotion identification is one area where they shine because of how well they handle complicated data.

In order to categorize various emotional states based on information derived from speech, these algorithms and models are frequently employed in emotion recognition tasks. A person's emotional state is a dependent variable that arises from their intrinsic qualities, changes in their environment, and interactions with that environment. By watching them, we can gain a profound understanding of the complexities of actual circumstances. Machines' processing capability, rather than the number of human agents involved, determines the maximum output of such analyses. The quantity of human agents accessible to analyze employee emotional states is now limiting management's understanding of these emotions. On the other hand, if we could automate this process, we could save money and have a better grasp on how employees are feeling.

The substance of communication as well as people's emotional responses can be understood through the analysis of emotions in speech. Using non-invasive methods such as emotion

speech recognition, we can install tiny devices in boardrooms to learn the emotional reactions and responses of managers to various subjects. Because of this, we are able to record not just what people say but also how they feel when conversing.

4. DISCUSSION

By identifying emotion in speech, one can follow the emotional state and behavior pattern of various

ISSN:2320-3714
Volume:2 Issue:2
May 2024
Impact Factor: 10.2
Subject: Psychology

social groups, which could potentially assist public services. Additionally, such methods could be applied in the academic world to attain finer-grained results, particularly in the realm of social science research. One could, for instance, identify feelings while interviewing people verbally. The same data gathering processes might be used to examine an additional signal that could be produced by such an approach.

One promising area for the advancement and use of AI robots in the future is the ability to recognise emotions in human speech. When interacting with humans, robots that can detect and respond to emotions are much more effective. As a result, it is possible to strengthen bonds between humans and robots through more organic and compassionate interactions. When it comes to people's emotional needs, emotion-aware Al robots are able to tailor their assistance and support accordingly. As an illustration, in healthcare contexts, robots can provide emotional support to patients by picking up on their emotions during talks. Similarly, AI-powered tutoring or teaching tools can adjust their approach depending on how students react emotionally. Robots can greatly improve learning outcomes by sensing when pupils are interested, frustrated, or confused, and then adapting their instructional tactics accordingly. They can be companions with emotional intelligence for people, especially to support mental health or situations involving the care of the elderly, where interpersonal communication is vital. They are capable of carrying on natural-sounding dialogues, offering emotional support and providing companionship, based on their knowledge of human emotions.

The development of Al robots with ever-more-advanced emotional intelligence will be propelled by innovations in emotion recognition technology. To make emotion-aware robots even more effective, researchers in this field will look into new algorithms, methods, and applications.

All things considered, Al robots may play a game-changing role in fields as diverse as medicine, teaching, customer service, and companionship if they could detect and respond to human emotions through speech. Contributing to the growth of Al technology and the improvement of human well-being, these robots can perceive and respond to human emotions, allowing them to provide users with more personalized, compassionate, and successful interactions.

5. CONCLUSION

Emotion recognition in speech and emotional computing are the subjects of this paper's intended literature review. This is an effort to bring together other areas of study, such as the social sciences and humanities with this technology. Systems that individuals engage with, can be made smarter by incorporating emotional understanding into speech recognition. The advantages surpass the disadvantages because this has the potential to enhance perception and comprehension, and machines are capable of doing this efficiently. But there are moral concerns to think about, like who pays for what, how data is shared, who uses it to make judgements, and what counts as success. These obstacles aside, the next few decades should see significant advancements in this field.

6. REFERENCES

[1] C. C. Chibelushi and F. Bourel, “Facial expression recognition: A brief tutorial overview,” CVonline: On-Line Compendium of Computer Vision, vol. 9, 2003.

[2] P. Boxer and V. Kenny, “The economy of discourses: a third order cybernetics?,” Human Systems Management, vol. 9(4), pp. 205–224, 1990.

[3] G. L. Clore and J. R. Huntsinger, “How emotions inform judgment and regulate thought,” Trends in cognitive sciences, vol. 11(9), pp. 393–399, 2007.

[4] R. S. Lazarus and B. N. Lazarus, “Passion and reason: Making sense of our emotions,” Oxford University Press, New York, 1994.

[5] N. Kamaruddin and A. Wahab, “Heterogeneous driver behavior state recognition using speech signal,” in Proceedings of The 10th WSEAS International Conference on System Science and Simulation in Engineering, pp. 207–212, 2011.

[6] A. Austermann, N. Esau, L. Kleinjohann, and B. Kleinjohann, “Fuzzy emotion recognition in natural speech dialogue,” IEEE International Workshop on Robot and Human Interactive Communication 2005, ROMAN 2005, IEEE, pp. 317–322, 2005.

[7] E. Vyzas and R. W. Picard, “Offline and Online Recognition of Emotion Expression from Physiological Data,” in Proceedings of the Workshop on Emotion-Based Agent Architectures at the Third International Conference on Autonomous Agents, pp. 135–142, 1999.

[8] R. W. Picard, “Affective Computing for HCI,” In HCI (1), pp. 829–833, 1999.

ISSN:2320-3714
Volume:2 Issue:2
May 2024
Impact Factor: 10.2
Subject: Psychology

[9] K. Ćosić, S. Popović, M. Horvat, D. Kukolja, B. Dropuljić, B. Kovač, and M. Jakovljević, "Computer-aided psychotherapy based on multimodal elicitation, estimation and regulation of emotion," Psychiatr. Danub., vol. 25, pp. 340–346, 2013.

[10] K. Ćosić, S. Popović, M. Horvat, D. Kukolja, B. Dropuljić, I. Kostović, M. Judaš, M. Radoš, M. Radoš, L. Vasung, B. B. Spajić, S. Doričić, and D. Mesić, "Virtual reality adaptive stimulation in stress resistance training," In HFM-205 Symposium on "Mental Health and Well-Being across the Military Spectrum", 2011.

[11] D. Kukolja, S. Popović, M. Horvat, B. Kovač, and K. Ćosić, "Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications," International journal of human-computer studies, vol. 72(10), pp. 717–727, 2014.

[12] N. Wiener, "The Human Use of Human Beings," Houghton Mifflin, 1954.

[13] M. Minsky, "The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind," Simon and Schuster, 2007.

[14] F. Ren, "From cloud computing to language engineering, affective computing and advanced intelligence," International Journal of Advanced Intelligence, vol. 2(1), pp. 1–14, 2010.

[15] R. W. Picard, "Affective Computing," M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321, 1995.

[16] B. C. Stahl, "Virtual suicide and other ethical issues of emerging information technologies," Futures 50, pp. 35–43, 2013.

[17] E. Hollnagel, "From human factors to cognitive systems engineering: human machine interactions in the 21st century," InAnzen-no-Tankyu (Researches on Safety), M. Kitamura, I. Kimura (Eds.), ERC Publishing, Tokyo, 2001.

[18] S. Yamamoto, Y. Yoshitomi, M. Tabuse, K. Kushida, and T. Asada, "Detection of baby voice and its application using speech recognition system and fundamental frequency analysis," in Proceedings of 10th WSEAS International Conference on Applied Computer Science, pp. 341–345, 2010.

[19] R. Sharma, U. Neumann, and C. Kim, "Emotion Recognition In Spontaneous Emotional Utterances From Movie Sequences," in Proceedings of the WSEAS International Conference on Electronics, Control & Signal Processing, 2002.

[20] M. Kurematsu, H, Jun, and F. Hamido, "An extraction of emotion in human speech using speech synthesize and classifiers for each emotion," in Proceedings of the WSEAS Transaction on MIPRO 2016/CIS 1555 Information Science and Applications, vol. 5(3), pp. 246–251, 2008.

[21] B. Dropuljić, S. Popović, D. Petrinović, and K. Ćosić, "Estimation of emotional states enhanced by a priori knowledge," in Proceedings of the 4th International Conference on Cognitive Infocommunications (CogInfoCom), IEEE, pp. 481–486, 2013.

[22] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. Mcrorie, J. C. Martin, L. Devillers, S. Abrilian, A. Batliner, and N. Amir, "The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data," Affective computing and intelligent interaction, Springer Berlin Heidelberg, pp. 488–500, 2007.

[23] I. S. Engberg, A. V. Hansen, O. Andersen, and P. Dalsgaard, "Design, recording and verification of a danish emotional speech database," Eurospeech, 1997.

[24] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of German emotional speech," Interspeech,vol. 5, pp. 1517–1520, 2005.

[25] J. H. Hansen, S. E. Bou-Ghazale, R. Sarikaya, and B. Pellom, "Getting started with SUSAS: a speech under simulated and actual stress database," EUROSPEECH, vol. 97(4), 1997.

[26] J. Kleckova and B. Mahdian, "Nonverbal Communication in Spontaneous Speech Recognition", WSEAS Transactions on Electronics, vol. 1(3), pp. 531–536, 2004.

[27] V. Petrushin, "Emotion in speech: Recognition and application to call centers," in Proceedings of Artificial Neural Networks in Engineering, 1999.

[28] M. Khan, T. Goskula, M. Nasiruddin, and R. Quazi, "Comparison between k-nn and svm method for speech emotion recognition," International

ISSN:2320-3714
Volume:2 Issue:2
May 2024
Impact Factor: 10.2
Subject: Psychology

Journal on Computer Science & Engineering, vol. 3(2), 2011.

[29] C.-C. Lee, E. Mower, C. Busso, S. Lee, S. Narayanan, "Emotion recognition using a hierarchical binary decision tree approach," Speech Communication, vol. 53(9), pp. 1162–1171, 2011.

[30] X. Cheng and Q. Duan, "Speech Emotion Recognition Using Gaussian Mixture Model," Proceedings of The 2nd International Conference on Computer Application and System Modeling,. 2012.

[31] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden Markov models," Speech communication, vol. 41(4), pp. 603–623, 2003.

[32] J. Hakura, M. Kurematsu, and H. Fujita, "An exploration toward emotion estimation from facial expressions for systems with quasi-personality," International Journal of Circuits, Systems and Signal Processing, vol 1(2), pp. 137–144, 2008.

[33] S. Barsade and O. A. O'Neill, "Manage your Emotional culture," Harvard Business Review, vol. 94(1–2), pp. 58–66, 2016

Author's Declaration

I as an author of the above research paper/article, here by, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website /amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct, I shall always be legally responsible. With my whole responsibility legally and formally have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and hentriacontane is genuinely mine. If any issue arises related to Plagiarism/ Guide Name/ Educational Qualification /Designation /Address of my university/ college/institution/ Structure or Formatting/ Resubmission /Submission /Copyright /Patent /Submission for any higher degree or Job/Primary Data/Secondary Data Issues. I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible or shuffled or vanished from the database due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who finds trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide or do not submit the copy of my original documents(Andhra/Driving License/Any Identity Proof and Photo) in spite of demand from the publisher then my paper maybe rejected or removed from the website anytime and may not be consider for verification. I accept the fact that As the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds Any complication or error or anything hidden or implemented otherwise, my paper maybe removed from the website or the watermark of remark/actuality maybe mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me

Anusha Shinde

******