

Speech Recognition as One of the Methods for Determining the Mental State of a Person

B Zandan^{1*}, AI Baskakov² and B Odsuren^{1,2}

¹Laboratory of Radio electronics of the Institute of Physics and Technology of the Academy of Sciences of Mongolia, Mongolia

²Department of Radio Engineering Devices and Antenna Systems of the Moscow Power Engineering Institute, Russia

*Corresponding author: B Zandan, Laboratory of Radio electronics of the Institute of Physics and Technology of the Academy of Sciences of Mongolia, Mongolia



ARTICLE INFO

Received:  March 22, 2021

Published:  April 01, 2021

Citation: B Zandan, AI Baskakov, B Odsuren. Speech Recognition as One of the Methods for Determining the Mental State of a Person. Biomed J Sci & Tech Res 34(5)-2021. BJSTR. MS.ID.005625.

ABSTRACT

In this work, we investigated the work of a neural network for recognizing speech in the Mongolian language and the emotional state in it. For the experiment, we used four commonly used words, which are composed of frequently used consonants and vowels. Emotional states in the speech were selected taking into account further research on changes in human mental disorders. For the analysis, we used a database of speeches of 12 men and women with eight types of emotions. Neural net recognizes words by syllables with an efficiency of up to 96 percent, and emotions - up to 80 percent. These results show that for further analysis of human mental disorders, it remains only to collect a database from the speeches of patients of the relevant medical institutions.

Keywords: Speech Recognition; Mental Disorders; Word; Syllable; Neural Network

Mini Review

Speech recognition as one of the methods of automatic speech analysis allows analyzing a person's mental state, such as loneliness [1], depression [2] and predicts the development of psychosis that accompanies many mental disorders, including, for example, schizophrenia and bipolar disorder [3,4]. A change in the psychological and mental state of a person is found in the violation of both semantic and syntactic components: for example, poverty (using simple constructions and understanding only closed types of questions) and confusion (impaired speech coherence) of speech [5]. For the analysis of human speech, the choice of parameters by which we determine the mental state of a person also depends on whether it is written or oral. If audio recordings are examined, then the spatial and temporal parameters of speech should be taken into account: pauses, voice pitch, and the average time spent on pronouncing one word. Also, to increase the reliability of the result, the most frequent words and phrases used by people with different diagnoses should be calculated. Based on these parameters, using them to distinguish features, the neural network issues a "diagnosis" about the mental state of a person [2]. Therefore, to master the methods of automatic analysis of a person's mental state based on

oral speech in Mongolian, we needed to develop a neural network that recognizes not only words, but also syllables in a word, and also distinguishes such parameters of a speech signal as pauses, voice pitch and average time spent pronouncing one word.

Materials and Methods

To recognize words and a syllable in a word, we used a neural network based on the Levenberg-Marquardt method [6]; after training, the neural network recognizes short words or speech syllables in Mongolian with an accuracy of 96.5 percent. (Table 1) shows the words and syllables used, and the numbering of the columns in the table corresponds to the numbering on the error matrix [7] used to assess the accuracy (Figure 1). The error matrix allows you to evaluate the efficiency of the algorithm used by comparing the prediction and the real value. These words were chosen for the reason that they are often used in colloquial speech and they contain the basic sounds of speech in the Mongolian language. Besides, noise is recognized to better highlight non-voiced vowels and consonants when splitting words into syllables. In addition to recognizing words and syllables, for the analysis of a person's mental state, such parameters of a speech signal as pauses,

voice pitch, and the average time spent on pronouncing one word are distinguished. Calculating such parameters of a speech signal as speech rate, speech signal energy, pitch frequency, I, II, and III formants, as well as I, II, and III coefficients of the chalk filter bank [8,9], we recognized eight types of emotions in speech human [10,11]:

- a. Neutral speech
- b. Calm speech

- c. Joyful speech
- d. Sad speech
- e. Angry speech
- f. Frightened speech
- g. Speech with disgust
- h. Surprised speech

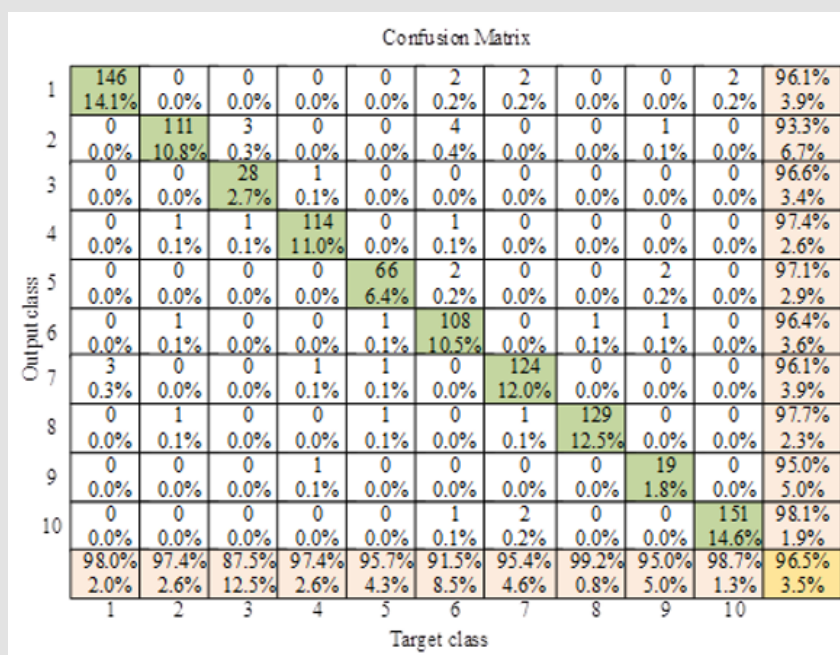


Figure 1: Confusion matrix in word recognition and syllable in a speech word in Mongolian.

Table 1: Words and syllables of words in the Mongolian language for training a neural network.

Columns Correspond to the Numbering in the Error Matrix	1	2	3	4	5	6	7	8	9	10
Words and syllables in Mongolian	ac	буд	будэг	дэг	г	шуугиан	тар	ун	унтар	тод
Transcription	As	Bud	Budge	Deg	G	Noise	Tar	Un	Untar	Tod
Number of samples for training	149	114	32	117	69	118	130	130	20	153

To analyze the emotional state of a person with the help of speech signals and train the neural network, a database consisting of the speech of 12 men and 12 women was used. And the result of training and recognition of the emotional state of a person from his speech is also shown in the form of error matrices (Figure 2). The error matrix in (Figure 2) shows that the neural network recognizes emotion in human speech with a probability of 80.5 percent. The results obtained enable us to conclude that we correctly calculate

the speech parameters we need to analyze and determine the mental state of a person. Speech recognition using a neural network not only allows you to analyze the mental state of a person but also allows you to detect and classify speech pathologies [12], helps in the detection of Parkinson's disease [13], and to identify loneliness, which becomes a factor in the development of type II diabetes [14] and ischemic heart disease [15].

Output class	1	2	3	4	5	6	7	8	
1	69 4.8%	10 0.7%	7 0.5%	6 0.4%	1 0.1%	1 0.1%	2 0.2%	0 0.0%	66.3% 33.7%
2	15 1.0%	171 11.9%	4 0.3%	10 0.7%	0 0.0%	1 0.1%	7 0.5%	2 0.1%	81.4% 18.6%
3	2 0.1%	0 0.0%	149 10.3%	5 0.3%	0 0.0%	4 0.3%	2 0.1%	11 0.8%	86.1% 13.9%
4	8 0.6%	8 0.6%	8 0.6%	155 10.8%	3 0.2%	10 0.7%	15 1.0%	6 0.4%	72.8% 27.2%
5	0 0.0%	0 0.0%	1 0.1%	4 0.3%	170 11.8%	10 0.7%	8 0.6%	3 0.2%	86.7% 13.3%
6	0 0.0%	1 0.1%	8 0.6%	3 0.2%	3 0.2%	144 10.0%	3 0.2%	5 0.3%	86.2% 13.8%
7	1 0.1%	1 0.1%	3 0.2%	3 0.2%	10 0.7%	8 0.6%	143 9.9%	4 0.3%	82.7% 17.3%
8	1 0.1%	1 0.1%	12 0.8%	6 0.4%	5 0.3%	14 1.0%	7 0.5%	158 11.0%	77.5% 22.5%
	71.9% 28.1%	89.1% 10.9%	77.6% 22.4%	80.7% 19.3%	88.5% 11.5%	75.0% 25.0%	74.5% 25.5%	82.3% 17.7%	80.5% 19.5%
	1	2	3	4	5	6	7	8	
	Target class								

Figure 2: Confusion matrix in recognizing emotions in a person's oral speech: the numbering corresponds to emotions in a person's speech.

Conclusion and Further Work

It is obvious that the introduction and development of modern artificial intelligence technology in various fields of medicine is a very important task, but most of the research and experiments were carried out in a native speaker of English or another popular language. Therefore, we began our research on a neural network for speech recognition in the Mongolian language and highlighting the parameters necessary for analyzing the speech signal. The neural networks that we trained for the Mongolian language are good at recognizing both words and emotions with efficiency up to 96 and 80 percent, respectively. In the future, we plan to contact medical centers in the country and work with patients who need one or another support, both mental and pathological treatment.

References

1. Varsha D Badal, Sarah A Graham, Colin A Depp, Kaoru Shinkawa, Yasunori Yamada, et al. (2020) Prediction of Loneliness in Older Adults Using Natural Language Processing: Exploring Sex Differences in Speech. *The American Journal of Geriatric Psychiatry*.
2. Al Hanai T, Ghassemi M, Glass J (2018) Detecting Depression with Audio/Text Sequence Modeling of Interviews. *Proc Interspeech*, pp. 1716-1720.
3. Corcoran CM, Carrillo F, Fernández Slezak D, Bedi G, Klim C, et al. (2018) Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry* 17(1): 67-75.
4. Rezaii N, Walker E, Wolff P (2019) A machine learning approach to predicting psychosis using semantic density and latent content analysis. *npj Schizophr* 5: 9.
5. Andreasen NC, Grove WM (1986) Thought, language, and communication in schizophrenia: diagnosis and prognosis. *Schizophr Bull* 12(3): 348-359

6. M Sanaullah, MH Chowdhury (2014) Neural network based classification of stressed speech using nonlinear spectral and cepstral features. 2014 IEEE 12th International New Circuits and Systems Conference (NEWCAS): 33-36.
7. Ting KM (2011) Confusion Matrix. In: Sammut C, Webb GI (Eds.), Encyclopedia of Machine Learning. Springer, Boston, MA.
8. Akash Shaw, Rohan Kumar Vardhan, Siddharth Saxena (2016) Emotion Recognition and Classification in Speech using Artificial Neural Networks. International Journal of Computer Applications 145(8): 5-9.
9. Gharavian D, Sheikhan M (2010) Emotion Recognition and Emotion Spotting Improvement Using Formant-Related Features. Majlesi Journal of Electrical Engineering 4(4).
10. Banse R, Scherer KR (1996) Acoustic profiles in vocal emotion expression. J Pers Soc Psychol 70(3): 614-636.
11. Javidi MM, Roshan Ebrahim (2013) Speech Emotion Recognition by Using Combinations of C5.0, Neural Network (NN), and Support Vector Machines (SVM) Classification Methods. Journal of mathematics and computer Science 6: 191-200.
12. S Santhana Megala, R Padmapriya, B Jayanthi, M Suganya (2019) Detection and Classification Of Speech Pathology Using Deep Learning. International Journal Of Scientific & Technology Research 8(12): 3045-3051.
13. Noor Afza, Manoj Challa, Jitendranath Mungara (2013) Speech Processing Algorithm for Detection Of Parkinson's Disease. International Journal of Engineering Research & Technology (IJERT) 2(4): 1798-1803.
14. Hackett RA, Hudson JL, Chilcot J (2020) Loneliness and type 2 diabetes incidence: findings from the English Longitudinal Study of Ageing. Diabetologia 63(11): 2329-2338.
15. Valtorta NK, Kanaan M, Gilbody S, Sara Ronzi, Barbara Hanratty (2016) Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies. Heart 102: 1009-1016.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2021.34.005625

B Zandan. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>