Anti-school Bullying System Based on Voiceprint Recognition

Hao Yu
School of software, Zhengzhou University, Zhengzhou, China
2604962074@qq.com

Abstract. The article is devoted to the study of an Anti-school Bullying system based on Voiceprint Recognition. Through the introduction of the social background of the phenomenon of bullying and the existing relevant technological background, the design of an Anti-school Bullying system based on Voiceprint Recognition technology is introduced, and the analysis of the system is made.

1. Research background

1.1. Social background
School bullying is a deviant behavior in sociology. The theory of cultural transmission focuses on the influence of deviant students' compliance and deviant social facts on them in the process of socialization. The structural stress theory holds that deviant students' academic success is affected by their position in the disadvantaged group. Social control theory holds that the link between the perpetrator and the society is weakened. Label theory explains how deviant labels are "branded". Broken window theory believes that the long-term existence of school bullying is caused by the absence of effective restrictive measures by the education department and the judicial department[1]. In a broad sense, school bullying is defined as a violation of social ethics imposed on students by teachers, students, or personnel outside the school. In a narrow sense, bullying is defined as strong physical or mental conflict among students on campus. School bullying is a common concern of the hot topics, in view of the campus bullying, China education steering committee of the state council's office in April 2016 to all parts of the country issued “Notice on Launching a Special Campaign against School Bullying”. in November 2016, the Chinese ministry of education in combination with the comprehensive management of public, the supreme people's court, the supreme people's procurator ate, ministry of public security, the ministry of civil affairs, justice, the communist youth league central committee and the all-china women's federation issued “Guidelines On the Prevention and Treatment of Bullying and Violence Among Primary and Secondary School Students” request to strengthen the prevention and punishment according to law education and comprehensive treatment, and to control students bullying and violence [2]. In The United States, in order to prevent school bullying incidents, The federal government has issued a series of ACTS to take "zero tolerance" measures against bullying. As early as 1990, the federal government passed “The Clery Act” and congress passed “The Better Education for Students and Teachers Act” in 2001.

Countries to prevent and stop the campus bullying put in very big, means of countries to prevent and stop bullying campus basic it is, by using the method of introduced legislation and regulations from the years of practice, the effectiveness of the solution is not ideal, countries invested a lot of human resources, and the school campus bullying is ordinary, visible to prevent and stop the campus
bullying on manpower and act alone is not enough. Today's society is a society where science and technology have entered into life and developed rapidly. When it cannot be solved well by relying on human resources, it is advisable to start with science and technology and seek solutions using science and technology.

1.2. Relevant technical background

1.2.1. Background of Voiceprint Recognition technology. Voiceprint Recognition technology is a kind of biometric identification technology, which originated in the United States. The first person who studied this technology was Lawrence kerster of bell LABS. He analysed and identified tens of thousands of voice print images of more than 100 healthy people, and the recognition accuracy reached 99.65%. Scholars and experts in the field of information technology have great expectations for the application of this technology. This technology has been applied in many fields. For example, in the financial field, there are online payment based on voice print recognition and auxiliary verification of bank account identity. In the field of e-commerce, there is an e-commerce platform login system based on voice print recognition; in the field of reconnaissance and case solving, there is a suspect identification system based on voice print recognition, etc. With the wide application of this technology, the technology itself has been developed and improved.

1.2.2. Background of Speech Synthesis technology. Speech Synthesis technology research has a history of nearly two hundred years, but the real practical significance of modern speech synthesis technology is with the development of computer technology and digital signal processing technology and development, main is to enable the computer to produce high definition, high degree of natural continuous speech[3], it is a very mature technology, the applications of this technology is more extensive, for example, the voice on the audio mailbox, finance reimbursement, calls inquiry of speech, etc.

1.2.3. Background of Fog Computing technology. The concept of Fog Computing was proposed in 2011, and was defined in detail in 2012. The name of Fog Computing was proposed by professor Stelfer from Columbia University in New York in the United States, whose purpose at that time was to use "fog" to ward off hackers. Fog Computing is a very new technology. Fog Computing is mainly a technology serving the Internet of things. With the arrival of the 5th generation of communication technology serving the Internet of things, the combination of Fog Computing and 5G is an inevitable trend, so is the perfection and promotion of fog computing.

2. Overview of related technical principles

2.1. Voiceprint Recognition

Voiceprint Recognition is broadly divided into speaker identification and speaker confirmed two kinds, the former is used to identify in more than one person the identity of the speaker, the latter is used to determine whether a voice and data specified said, two different way of working is suitable for different areas, this paper puts forward the system use to identify this way to work.

Voiceprint Recognition is the extraction of speech acoustic spectrum by electro-acoustic instruments. Human speech acoustic spectrum is different like fingerprint, and voiceprint recognition use this difference to identify and confirm the owner of speech data.

Voiceprint Recognition process can be divided into: sound data extraction, feature extraction, feature database established, pattern matching, four parts, feature extraction and pattern matching is the most critical part, the main technology used for feature extraction are: VQ Vectorization model, stochastic model and vector quantization neural network model, related to the pattern matching technology are: probability statistics method, dynamic time neat method, vector quantization, Hidden Markov Model, neural network method.
2.2. Speech Synthesis

Speech Synthesis technology is the technology that produces artificial speech through mechanical and electronic methods. It converts the text data generated by the computer or the text data input from the outside into natural speech data. The goal of speech synthesis technology is to synthesize electronic sounds with a high degree of naturalness. Naturalness refers to the similarity between synthesized speech and human speech. To improve naturalness of synthesized speech, text regularity, lexical segmentation, grammatical analysis and semantic analysis are required. Through nearly two hundred years of research, the naturalness of speech synthesis has reached a high level.

Speech Synthesis techniques can be subdivided into three parts of text analysis, modelling and rhythm, Speech Synthesis. Speech Synthesis is the main working process according to the result of rhythm model, from the original library to retrieve the corresponding element of voice, using specific speech synthesis technology for voice primitive rhythm features, adjust and modify the final synthetic natural degree high voice[4]. The related Speech Synthesis technologies include Formant Synthesis, Linear Prediction Synthesis, PSOLA Synthesis and LMA Channel Simulation Synthesis.

2.3. Fog Computing

Fog computing is an extension of cloud computing. Fog Computing is proposed for the marginalization of data computing, mainly to deal with the huge amount of data in the development process of the Internet of things and the high requirements on servers. Fog Computing marginalizes data computing, adopts the distributed principle, and replaces large cloud servers of large enterprises with personal cloud, private cloud and small cloud, which has the characteristics of low delay, good mobility and low requirements on equipment performance. The working process is that the data detected by the terminal devices on the edge of the Internet of things are processed in the terminal itself, and the interactive processing of data can be conducted directly between the terminal devices. Meanwhile, the data processed by the terminal devices can be sent to the small cloud, and the data can also be transmitted between the small clouds. Based on the marginalization of the processing of this data, the pressure on the server is improved, and the corresponding services can be completed without the support of high-performance server.

3. System introduction

Against the background of high attention paid by various countries to the phenomenon of school bullying, this paper proposes a kind of Anti-school bullying system based on Voiceprint Recognition technology, Voice Recognition technology, Speech Synthesis technology and Fog Computing technology. It is mainly composed of detection terminal, cloud server and mobile terminal APP. The functions and structures of each part are described as follows:

(1) Detection terminal: the main control chip of the detection terminal is based on ARM architecture, which is mainly composed of voiceprint recognition module, voice recognition module, speech synthesis module, camera module and NB-IOT module integrated with GPS. Voiceprint recognition module is used to extract voice feature information and identify the identity of the speaker, voice recognition module is used to identify sensitive words, voice synthesis module is used to synthesize electronic sounds for warning, camera module is used to extract scene image information, and NB-IOT module integrated with GPS is used for wireless transmission and positioning of data. The main control chip analyses and synthesizes the information generated by the three modules and conducts data communication with the small cloud server.

(2) Cloud server: the server records and statistics the data sent by the detection terminal and the data sent by the mobile terminal, and timely feeds the information back to the mobile terminal.

(3) Mobile terminal: mobile terminal is convenient to view data and submit data feedback to the server.
4. System operation mode introduction
The school needs to do a good job in the early stage, extract students' voiceprint information to establish a feature database, and place the detection terminal of the system in the school without any blind side. The detection terminal usually only has voiceprint recognition module, voice recognition module and camera module in working state. Meanwhile, all teachers and school security personnel need to install the corresponding mobile terminal APP.

The work basis of the system is the setting of sensitive words. The sensitive words defined here are the set of words that are used most frequently by some students when they are playing with each other. For the scheme of establishing sensitive words, please refer to the following section, "building sensitive words". Detection terminal real-time voice data, when detect the voice and data, voice and data with vocabulary and other sensitive data matching, matching failure, discarding the voice and data, the match is successful, the use of voiceprint recognition module for voiceprint recognition of voice and data model match, identify the voice the identity of the sender, camera module can record the time of the scene information, speech synthesis module synthesis electronic sound to warn of relevant personnel. At the same time, system will use detection terminal NB-IOT module into the surrounding mobile terminal APP to send packets, packet content to detect the location of the detection terminal information and detect the scene information, The mobile terminal closest to the detection terminal APP will be the first to receive packets, with the corresponding mobile terminal APP teacher can through the APP test response is connected to the detection terminal, and on the mobile terminal APP to check the detection terminal camera module to extract information of scene, and then to respond in a timely manner, when the detection terminal access to specify a value when the number of mobile terminal APP stop sending packets to the surrounding.

Mentioned voiceprint recognition module will use words people's identity, this will appear two situations, first, the user is already recorded voiceprint information of students, the detection terminal not only can send packets to the surrounding, also through the cloud servers for the students of the teacher in charge the use of mobile terminal APP account, and orientation to the students of the teacher in charge to send packets, can the teacher in charge according to the detection terminal of scene information, fill in the corresponding students in the mobile terminal APP punishment; Second, if the identified voice print information is not entered, it can be determined that the sender of the sensitive words is not a student of our school. In this case, besides sending data packets to the surrounding, the detection terminal will also search the mobile terminal APP account of the personnel in the school security office through the cloud server, and send data packets to the school security office directly.

5. building sensitive words
(1) Crawl video or audio data related to violent images. For video with subtitles, use the Corner Detection Principle or the Binarization Method of Interpolation, Amplification and Mixed Threshold to extract subtitle information and generate text data; for video or audio information without subtitles, voice is converted into text data using Acoustic Model Library and Speech Model Library.

(2) The processing of the obtained text data mainly use Regular Expressions to find and delete the Stop Words in the text data, and compares and corrects the text data obtained through video subtitles and the text data obtained through audio with the Text Model Library, so as to optimize the output of the text data with high accuracy.

(3) Through the method of probability statistics, the modified text data can be used to obtain the words with the frequency of over 90%. The processed text data is input into the neural network for learning and generating sensitive word bank.

6. System analysis
The system applies science and technology to prevent and stop the phenomenon of bullying on campus, making campus management scientific, intelligent and efficient. As the system itself has camera module, the system is an intelligent version of campus monitoring system and a better
alternative to campus monitoring system. The existing campus monitoring system need to by security personnel 24 hours of uninterrupted corresponding terminal monitoring the situation in the school, and need security guards to patrol the campus, the system can mobilize all the teachers and school security personnel to monitor whether the school campus bullying incident, the system can effectively reduce countries and schools to prevent and stop the campus bullying on the investment of manpower and material resources. To provide more effective and feasible solutions to prevent and stop school bullying.

There are still some shortcomings in this system. The key deficiency lies in the establishment of the sensitive word bank. In order to accurately identify bullying incidents, the accuracy of the sensitive word bank is highly required, and there is still a lot of room for improvement in the establishment of the sensitive word bank. And the system can't tell on its own whether students intentionally use sensitive words against the system, and there's actually no bullying, and teachers and security personnel have to use the information recorded on the camera to figure it out. Therefore, it is also necessary for the school to make some rules according to the system to restrain students' behaviour that is deliberately targeted at the system.

7. Conclusion
Many countries have introduced a lot of laws and measures on bullying, and schools have also invested a lot of manpower and material resources to prevent and stop bullying. It can be seen that human resources alone are not enough for bullying. The campus anti-bullying system based on voice print recognition proposed in this paper combines voice print recognition technology, voice recognition technology, voice synthesis technology, fog calculation, machine learning, etc., to make the school security management intelligent, automatic and scientific. It also forms an alternative to the campus monitoring system and puts forward a feasible scientific and technological plan for preventing and stopping bullying incidents.

Acknowledgments
The Self-service Project of the Innovation and Entrepreneurship Training Program of the University Students in Zhengzhou University.

References
[1] Zhang guoping, Sociological Analysis of School Bullying, Contemporary Youth Research, 2011. 08. 26.
[2] Gong qingqiang, Special Protection For Healthy Growth, Examination Questions and Research (senior high school entrance examination), 2017. 01. 01
[3] Zhang shili, prediction and automatic labeling of prosodic structure in Chinese speech synthesis, Peking University, 2006. 06. 01.
[4] Tian weidong, application of sound channel frequency response synthesis method in military name corpus, Northeastern University, 2008. 12. 29.