Online Music Production and Collaboration Method based on Internet of Things and Computer Assistance

Meng Xiao[1]

[1]College of Music, Jiaozuo Normal College, Jiaozuo, Henan 454001, China, meng2019@jzsz.edu.cn

Corresponding author: Meng Xiao, meng2019@jzsz.edu.cn

Abstract. In response to the large amount of audio and video data that needs to be processed in social production and life, an embedded intelligent Internet of Things processing system has been developed. Established a DRTU data terminal system with Raspberry Pi as the core and STM32 auxiliary control. The control system can realize the functions of data collection, transmission and display. Use the Raspberry Pi interface to connect to the wireless technology module to transmit a large amount of audio and video data, and transmit it to the Mongodb cloud database through the mobile base station. A series of cloud services are applied to meet the functional requirements of the Internet of Things. These services are all running in virtual machines in the cloud, so they can be expanded at any time according to traffic needs. Through the main interface test of the engineer station, the system has realized the designed audio and video data processing function. In addition, this article is based on the audio and video online production and collaboration system and method assisted by the Internet of Things equipment, which involves the technical field of the Internet of Things. The research in this paper specifically includes IoT hardware devices, cameras, storage cloud servers, and application terminals; IoT hardware devices are used to receive the recording trigger signal and metadata information sent by the camera, and synchronize the video signal with the camera in the IoT hardware device. The collection of recording and metadata information, the acquisition of "proxy" video files consistent with the video file length/picture/metadata information recorded in the camera, and transfer to the storage cloud server. The research content of this article can be controlled by the audio and video production team members and collaborators without being limited to the actual address, in any scene where people can be connected to the Internet; "Material files are subject to collaborative review and post-production, and the review, post-production and distribution of shooting materials are completed in a short period of time.

Keywords: Internet of things; computer aided; online music production and collaboration
DOI: https://doi.org/10.14733/cadaps.2022.S6.102-112

1 INTRODUCTION

Sabrina and Ramet [1] consider the Internet of Things technology is a comprehensive technology. The key to the planning and design of the Internet of Things lies in the research in the fields of RFID, sensors, embedded software, and data transmission. Traditional Internet of Things technology is widely used in the field of smart home, and many smart home technologies are combined with access control systems. Through the research of traditional smart home technology, combined with Internet technology and cloud platform technology, this subject has developed a cloud platform technology for intelligent Internet of Things audio and video information processing. In order to improve the efficiency of user management, this paper designs a cloud platform system for information processing such as intelligent Internet of things audio and video. The cloud platform system provides users with services such as remote data collection, remote automatic control, immediate decision-making, command scheduling, and data analysis through remote audio and video data collection, analysis, and remote security control. Make its Internet of Things technology no longer stay in the concept, but widely used in various fields. Mastnak and Neuwirthová. [2] think the online real-time collaboration platform based on the Internet can greatly improve the work efficiency of the team, but the premise is that the work done by everyone can be completed on the network. Fowdur et al. [3] think in audio and video production, there are only a handful of networked equipment and equipment in the early video shooting scene, especially on the camera, the core of the set. Cheng and Jin [4] think although camera manufacturers provide some networked functions on their cameras, usually simple on-site control functions are really difficult to produce practical value. If you want the pre-shooting scene to be connected with the Internet online collaboration platform, you need to synchronize the core asset of the scene-the video material under shooting to the cloud in real time, so that the production team can access/process and distribute it online. In the 5G era, Baltasar and Ortin [5] think low-latency and high-bandwidth network transmission makes it more realistic to synchronize digital assets such as video materials to the cloud in real time; how to achieve this is a problem that the audio and video production industry is looking forward to solving.

With the rapid development of the computer Internet, the previously unimaginable interactive transmission of Internet voice and even video has been realized, and the high Internet tariffs are becoming more and more accessible to the people. Mobile Internet voice and video have also brought huge convenience to countless people [6,7]. At present, the portable music production and audio processing equipment has made more recording studios tend to be miniaturized, and more musicians, audio workers, and producers have become accustomed to recording in small recording studios or even home studios. Instrument, music vocal track, dubbing, and then send the recorded sound track to the customer for live or to the mixer to continue to make the mix, or wait for the producer's feedback. In recent years, with the development of cloud storage technology, many people no longer have to regret that they forgot to copy the project files that need to be taken home to continue processing after get off work, nor do they need to worry about the urgent need to view or modify documents outside and unable to use the USB flash drive. Distressed, cloud storage can easily solve these problems. Recently, new versions and new plug-ins of mainstream workstation software including Cubase/Nuendo and Pro Tools have added cloud storage functions to meet these needs, and complete music production and audio processing through cloud collaboration. How to achieve cloud collaboration? For example, a Beijing record company needs to produce a song. According to the production requirements, it finds an excellent drummer in Los Angeles to complete the recording of the drum part. Due to time constraints, it is impossible to wait for the drummer to return to Beijing to record; besides, drummers often cooperate [8,9]. The bass player is participating in the music festival in Shanghai. In this situation, it is entirely possible to use Internet cloud collaboration to allow producers/boss, performers/musicians to complete music production, audio processing and other tasks simultaneously in their respective places [10].

2 AUDIO AND VIDEO PROCESSING OVERALL ARCHITECTURE

This system designs the IoT audio and video information processing cloud platform into the following 5 layers: perception layer, transmission layer, convergence layer, distribution layer and application layer. The perception layer mainly includes on-site various industrial instruments, analyzers, measurement and control boards, various sensors, PLC/DCS, cameras and GPS positioning devices. The transport layer is responsible for audio and video data transmission and control, and transmits the data to DRTU (Data Transfer Unit and Remote Terminal Unit) in real time through 3G wireless transmission technology. The convergence layer is the convergence point of audio and video information. DRTU performs various transformations on the received data and transmits it to the cloud center through the industrial communication protocol Modbus. Then, the cloud computing analyzes and processes various field data, mainly including the generation of historical data, Check alarm data, generate summary data, and perform various mining analysis and predictions on big data. The distribution layer is a convergence point for users to access network resources. The PC client obtains the results of cloud processing through a unified distribution layer interface to ensure data security. The application layer is directly connected with the application program interface and provides network application services for data retrieval. Due to the relatively large amount of audio and video data, a separate cloud service is used for processing to ensure the fluency of audio and video, and at the same time, the audio and video data is stored for further identification and analysis. Among them, information cloud service platforms such as Raspberry Pi, STM32, Internet of Things audio and video, and PC clients are combined to form a complete smart industry, smart logistics, smart grid, smart transportation, smart agriculture, etc. The overall functional framework of the IoT audio information processing cloud platform is shown in Figure 1.

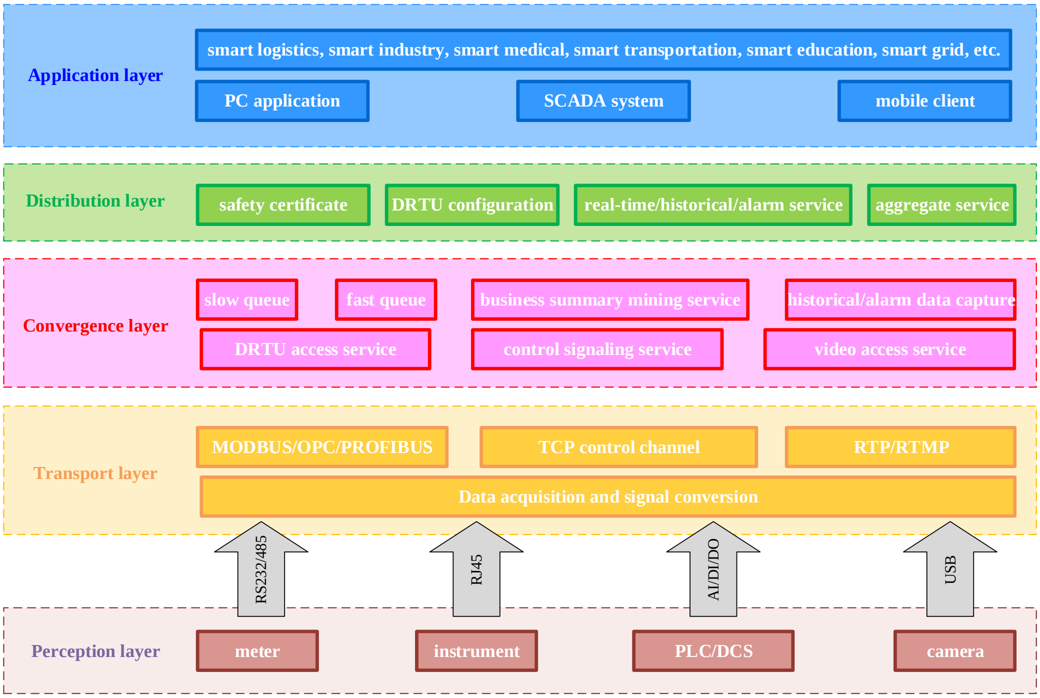

Figure 1: Overall frame diagram.

The online audio and video production and collaboration system proposed in this paper is shown in Figure 2. It includes IoT hardware devices, cameras connected to IoT hardware devices using signal cables or wireless networks, and IoT hardware A storage cloud server connected in real time by a wired network or a wireless network or a mobile Internet adapter between the devices, and an application terminal connected to the storage cloud server through a wireless network or a wired network or mobile Internet of Things; the Internet of Things hardware device is used to receive the camera The recording trigger signal is synchronized with the camera in the IoT hardware device to record the video signal, obtain the "proxy" video file that is the same as the video file length/picture recorded in the camera, and transfer it to the storage cloud server, and read the camera at the same time The metadata (Metadata) information transmitted through the video signal cable or wireless network is recorded in the "agent" video file; the Internet of Things hardware device includes an FPGA chip for analyzing camera shooting data, and is connected to the FPGA chip for The encoding SOC chip for encoding and recording video signals, the network module connected to the FPGA chip for data transmission to the storage cloud server, two lithium batteries connected to the FPGA chip for providing redundant uninterrupted power supply, and battery voltage alarm Display LED light; application terminal adopts smart phone/tablet computer/PC.

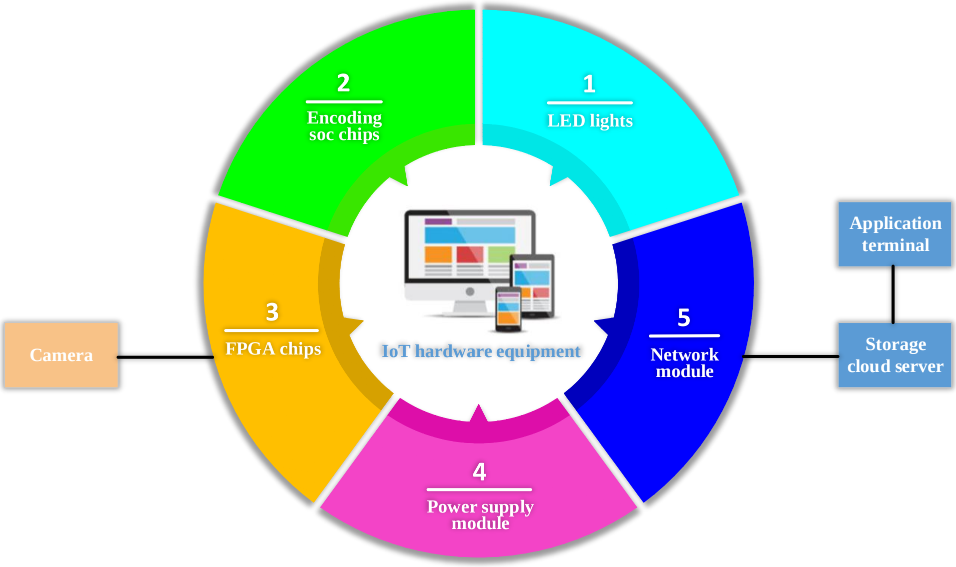

Figure 2: Online production and collaboration system.

The method of online audio and video production and collaboration based on IoT device assistance includes the following steps: 1) System installation and IoT hardware device initialization: the user installs the battery on the IoT hardware device at the shooting site and connects the camera; application The terminal connects to the Internet of Things hardware equipment through wireless network/wired network or mobile Internet of Things; connects to the Internet of Things hardware equipment through the application terminal to complete the initialization of the Internet of Things hardware equipment; 2) Use the application terminal to connect to the storage through the application The cloud server account is bound, the shooting project is established, and the shooting team is invited to shoot to obtain the shooting material in the form of video stream; 3) The camera shoots the material and triggers the simultaneous recording of the IoT hardware. When a piece of material is taken, the video Upload to the storage cloud server; 4) After the upload is completed, the project members can access/comment/approve or download, edit, and produce the synchronized video material on the storage cloud server in real time through the application terminal. Among them, in step 1, after the lithium battery is installed, the FPGA chip judges the two output voltages of the two lithium batteries, and preferentially uses the lower voltage battery power supply as the load power supply; when the voltage of any battery is lower than the threshold, then Prompt by LED lights, and automatically switch to another battery power supply as the load power supply for power supply. When the load power supply is switched, it will not cause the IoT hardware device to power off and restart. Among them, in step 2, the application terminal is bound to the storage cloud server account through the application, and the method of obtaining the AccessToken and the storage server address is adopted. The hardware device uses the AccessToken to obtain the write permission to the storage server, and the recorded video is recorded in real time. The file is uploaded to the storage cloud server, and the IoT hardware device receives the callback parameters of the storage cloud server and feeds back the upload result to the user. Among them, step 3 is specifically implemented by the following method: parse the Auxiliary Video information InfoFrame part of the HDMI signal of the camera or the ANC part of the SDI signal through the FPGA chip to obtain the resolution and frame rate of the video stream transmitted by the camera, and automatically Adjust the frame rate of the recording code to be consistent with it, and at the same time look for the Rec Trigger sent by the camera. When receiving the Rec Trigger, send a signal to the coding SoC chip to start encoding and recording the video signal, and after receiving the Rec Trigger from the camera At the same time StopTrigger sends a signal to the encoding SoC chip to stop the encoding and recording of the video. Among them, the FPGA chip parses the Auxiliary Video information InfoFrame part of the HDMI signal of the camera or the ANC part of the SDI signal, and also obtains the time code (TimeCode)/volume name (ReeIName)/file name (FileName) transmitted by the camera, and sends it to the encoding SoC, it is written into the metadata of the video file recorded under the encoding.

The advantages of the method proposed in this article over the prior art include: 1. With this system and method, members of the audio and video production team and coordinators are not limited to their actual addresses, and can perform matching in any scenario where the Internet can be accessed. Control of the shooting progress and shooting quality in the early stage. 2. Using this system and method, audio and video production team members and collaborators can use the "agents" on the cloud platform in real time. Collaborative review and post-production of the material files, complete the review of the shooting materials in a short period of time, post-production and distribution 3. The application terminal of the method proposed in this article configures the hardware equipment and the method of binding platform accounts and projects is not limited the actual address, in any scene where people can be connected to the Internet, can conduct the preliminary shooting progress, shooting quality, and video production team members and collaborators can use the "agent" material files on the cloud platform to conduct collaborative review and post-production in real time. Complete the review, post-production and distribution of shooting materials within a short period of time.

3 MODULE FUNCTION

As shown in Figure 3, the data terminal takes the Raspberry Pi as the core, and the Raspberry Pi combines with the STM32 controller to perform efficient work. The data terminal connects the camera with the CSI interface of the Raspberry Pi, collects audio and video data and transmits it to the mobile base station through wireless technology, and uses the STM32 serial port and other sensors to expand its system functions. At the same time, STM32 controls the human-computer interaction function of the Diwen screen, and the PC client can call the database data by calling MOGonDB.

The function of the wireless data control terminal is to collect various field data, convert them into digital signals, and upload them to the access service through the field general industrial bus Modbus technology. Wifi, 3G and other wireless technologies are used for transmission. After the data is processed by the access service, the data is saved in a cloud database based on Mongodb.

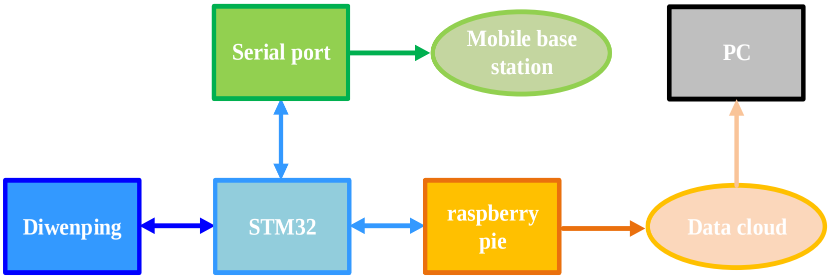

Figure 3: Schematic diagram of the data terminal.

The control terminal starts audio and video collection, compression, transmission, GPS positioning and other functions, which can realize the monitoring of audio and video information and map positioning, etc. The system uses a 32-bit processor and an embedded real-time multitasking operating system (LINUX). STM32 itself can also work independently, and is equipped with a 5-inch display. In the display, various parameters of the device can be set, various variable values can be input, and historical curves of various variables can be viewed on site.

DWINOS based on the DGUS screen is designed for parameter setting and data display applications. The data terminal display module realizes human-computer interaction process control, communication, data acquisition and processing, database operation, mathematical calculations and data analysis through DWINOS embedded in the DGUS screen. Function to meet the requirements of secondary development. The data terminal uses STM32 to control the Diwen screen, manually operate the data terminal's display screen, input corresponding parameters and other operations. The parameters of the Diwen screen display interface are set to the following 6 items: analog input channel, switch input channel, switch output channel, curve display, slave station data monitoring, and remote transmission equipment link.

Audio and video cloud services mainly include storage data services, access to DRTU services, control signaling services, access to audio and video services, capture historical data and generate alarm data, summary services and data mining, and unified data query access interface services. These services are all running in virtual machines in the cloud, so they can be expanded at any time according to traffic needs, and data can be backed up regularly to avoid data loss. The storage data service uses the cloud-based NoSql MongoDB database. MongoDB has the advantages of supporting large amounts of data and high concurrency applications. The built-in horizontal expansion mechanism of MongoDB provides data processing capabilities ranging from one million to one billion, and it is simple to use and has the characteristics of out-of-the-box use [7]. Since MongoDB adopts a bson-based data storage structure, there is no hard requirement for the number and types of "fields", which improves the scalability of the system. DRTU sends a link request to the cloud service, establishes two TCP-based links, and then sends a login command in the following format to the cloud service for login permission authentication. The content sent by DRTU is as follows: loginuuid: xxx; user: admin; pwd: xxx; the value of the carriage return and line feed protocol can be modbus, OPC, profibus, etc.; the cloud service response is as follows: replyresult: ok; or replyresult: err; access to the DRTU service If the corresponding record is retrieved from the database, it will reply ok, login is successful, and it will be displayed online after refreshing the DRTU. At the same time, the link session will be transmitted to the background service, and the real-time data access service of the sensor will be completed according to the configuration parameters. If the record is missing, the missing information is returned to the DRTU device, prompting a login error.

The control signaling service link is initiated by the DRTU and maintains a long connection. When the PC client initiates a video request to a DRTU, it needs to send the UUID of the DRTU, the device number, and an on or off command. Figure 4 shows the signaling working process: the PC client sends an audio and video switch command to the control signaling service. After the control signaling service receives the command, it retrieves the number of consumers of the corresponding device in the database. If the client sends an open command and the consumer value is 0, it sends an open device command to DRTU, otherwise no command is sent; if the client sends an off command and the consumer value is 1, then it sends an off command to DRTU At the same time, the consumer value is set to 0; otherwise, no command is sent, only the consumer value is reduced. Error handling mechanism: The server is down and all audio and video consumer values are cleared.

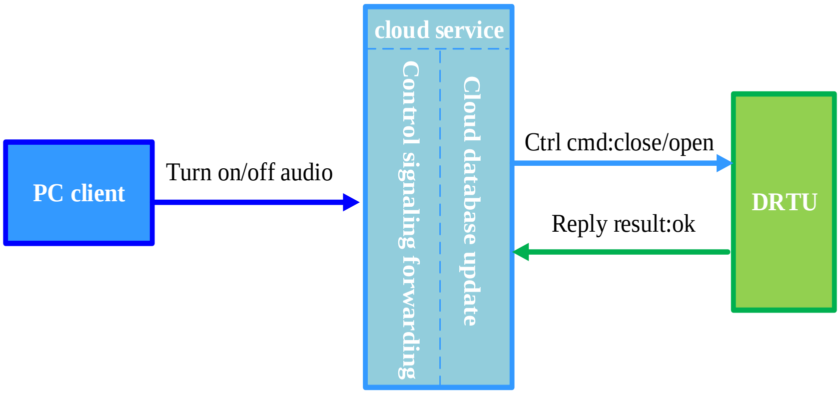

Figure 4: Signaling working process.

When the client opens the audio and video, the number of consumers of the device increases by one. When the consumer is equal to 1, it sends an open command to the DRTU. When the client closes the audio and video interface, the number of consumers is reduced by one. When the number of consumers is 0, it sends a close command to the DRTU through control signaling. DRTU collects and compresses the audio and video and forwards it to the audio and video server. The audio and video server maps different DRTUs to different channels. If a DRTU has multiple cameras, each camera is distinguished by a number, and the number is placed in the RUL when accessing.

The audio and video cloud includes a service for capturing historical data at regular intervals, and the service cycle is 5000ms (can be set as required). Copy the DRTU online data to the historical database for future summary analysis and data mining forecast applications. In the process of generating historical data, the data is verified by alarm rules. Currently, there are two types of alarms: one is orange alarm, and the other is red alarm. Because DI/DO/PI/PO only has two values of true and false, the orange alarm means the value changes from true to false, and the red alarm means the value changes from false to true. The specific data is shown in Figure 5.

Business summary mainly includes generating different reports for historical data according to user needs, in order to prepare users for various queries and printing. The summary information includes data of different devices, different regions, different types, and different times. Data mining includes classifying business data, clustering analysis, predicting and analyzing business trends, generating decision trees, and helping business experts and company leaders make production decisions and operation management.

Business summary mainly includes generating different reports for historical data according to user needs, in order to prepare users for various queries and printing. The summary information includes data of different devices, different regions, different types, and different times. Data mining includes classifying business data, clustering analysis, predicting and analyzing business trends, generating decision trees, and helping business experts and company leaders make production decisions and operation management.

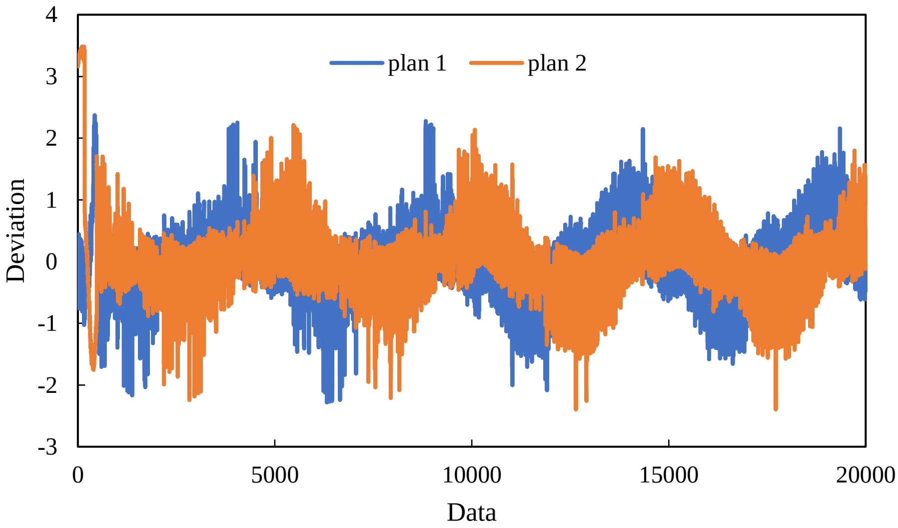

Figure 5: Deviation vs. Data.

The data storage service develops a unified interface for the client to prevent users from deliberately performing destructive operations on the database, such as script injection, destroying the table structure and other users' data. The unified service interface includes the following types: 1) Real-time data query interface. Query the value of each point of the current DRTU. Since the points of different DRTUs are different, and the names of the points are different, and the data types of the point values are also different, the service system transmits all the data in json format. The customer After receiving the data, it needs to be analyzed according to specific requirements. 2) Historical data query. Query certain DRTU points according to the time period, and the result of the query is for the user to make a historical curve. 3) Alarm data query. Query the alarm log according to the user's area. The query conditions include time period, alarm level, alarm status, whether to confirm, etc. The unconfirmed alarm can be confirmed. 4) User authority authentication. It is mainly used to authenticate DRTU and user login. 5) DRTU configuration. Configure the basic information of DRTU and the address value of each point, data type, address classification, alarm requirements, etc. In order to increase the speed of engineers' configuration, DRTUs of the same structure can be entered in batches to improve the efficiency of engineers. 6) Summarize and query the results of mining. The user queries various data aggregated in different time periods through the client, and displays the results of data mining to the manager in a graphical way. The results are shown in Figure 6, Figure 7 and Figure 8.

4 CONCLUSIONS

The cloud platform based on the efficient embedded operating system designed by the author of this article can complete the processing of data. The DRTU data terminal efficiently completes the tasks of data transmission, control, analysis and calculation of the Internet of Things, and realizes the transmission of audio and video cloud data. Its system can handle the collection, transmission, storage and display of audio and video information. The intelligent Internet of Things audio and video information processing cloud service receiving and dispatching processing system, CLIENT end data monitoring, configuration, management, dispatch command, security monitoring provide meticulous display. A single system is equipped with MDPA500DRTU data terminal, which can accept 1 million point data to work at the same time, and control at the same time. The joint system can easily expand to tens of millions to billions of data processing capacity, and can form a provincial, national, and global The cloud computing system of the Internet of Things audio and video is of great significance for improving the standard of production and living.

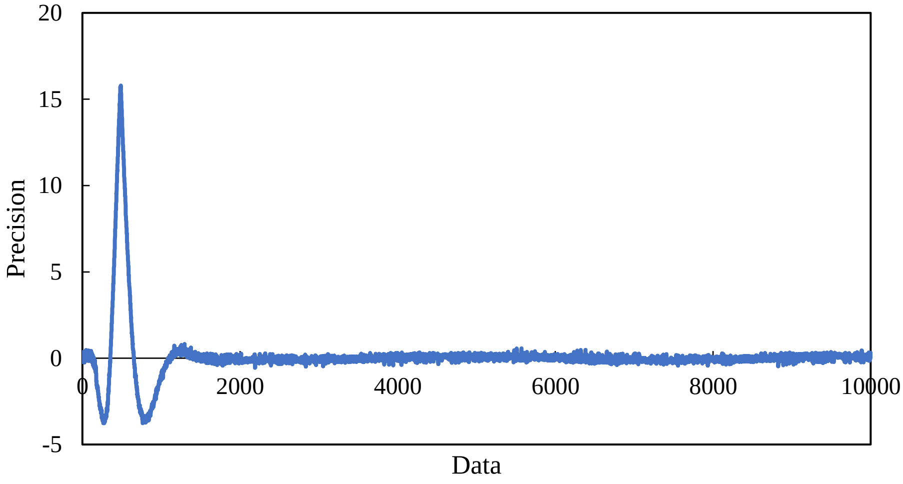

Figure 6: Precision vs. Data.

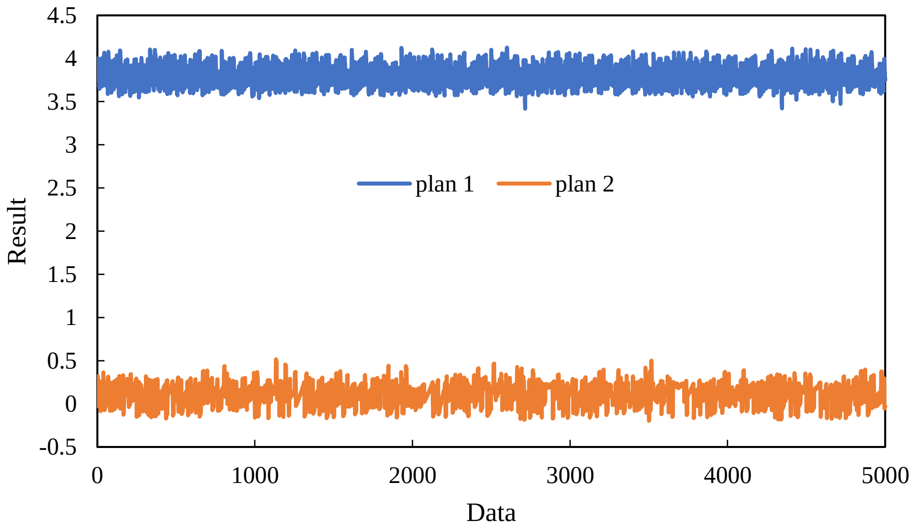

Figure 7: Predicted result.

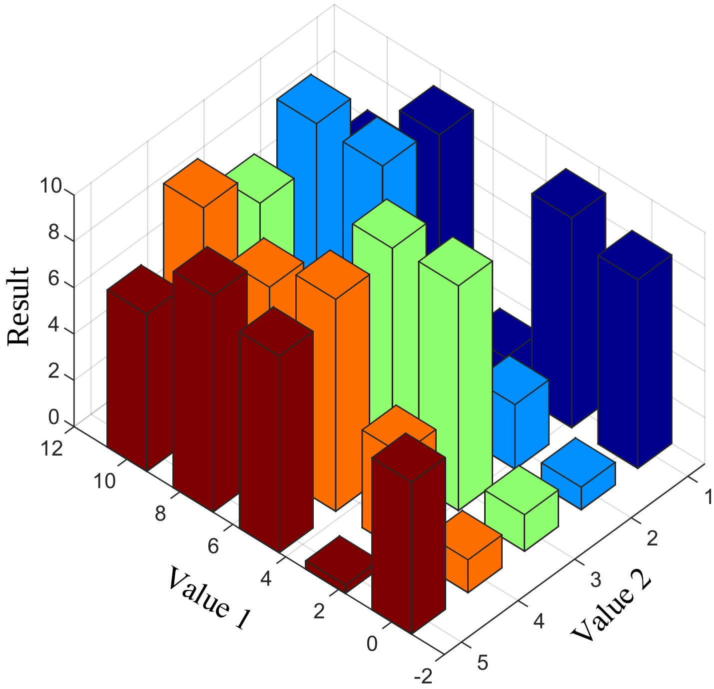

Figure 8: Predicted result with different values.

Meng Xiao, https://orcid.org/0000-0001-5087-8506

REFERENCES

[1] Sabrina, P.; Ramet, V.: The three phases of rock music in the Czech lands, Communist and Post-Communist Studies, 52(1), 2019, 59-70. http://doi.org/10.1016/j.postcomstud.2019.02.002

[2] Mastnak, L.; Neuwirthová, A.: Music Therapeutic Practice & Research at the Social Care Centre Tloskov, Czech Republic. A Short Report. Journal of Russian & East European Psychology, 55(1), 2018, 42-52. https://doi.org/10.1080/10610405.2018.1491239

[3] Fowdur, T.-P.; Ramkorun, N.; Chiniah, P.-K: Performance Analysis of WebRTC and SIP-based Audio and Video Communication Systems, SN Computer Science, 1(6), 2020, 1-22. https://doi.org/10.1007/s42979-020-00380-z

[4] Cheng, M.; Jin, X.: What do Airbnb users care about? An analysis of online review comments, International Journal of Hospitality Management, 76, 2019, 58-70. https://doi.org/10.1016/j.ijhm.2018.04.004

[5] Baltasar, G.; Ortin, F.: A didactic object-oriented, prototype-based visual programming environment, Science of Computer Programming, 176, 2019, 1-13. https://doi.org/10.1016/j.scico.2019.02.004

[6] Huang Y.: Research on Interactive Creation of Computer-Aided Music Teaching, Solid State Technology, 64(01), 2021, 577-587. https://doi.org/10.1109/ICMT.2011.6001675

[7] Kaleli Y.-S.: The Effect of Computer-Assisted Instruction on Piano Education: An Experimental Study with Pre-Service Music Teachers, International Journal of Technology in Education and Science, 4(3), 2020, 235-246. https://doi.org/ 10.46328/ijtes.v4i3.115

[8] Farha, F.; Sugandhi, K.; Raju, G.; Debabrata, S.; Biswaranjan, A.; Manas, Ranjan P.: A Novel Threshold based Method for Vessel Intensity Detection and Extraction from Retinal Images,

International Journal of Advanced Computer Science and Applications, 12(6), 2021, 1-9. https://doi.org/10.14569/IJACSA.2021.0120663
[9] Le, D.; Biswaranjan, A.; Ajaya, K.; Jyotir, M.; Raghvendra, K.: NoSQL Database Classification: New Era of Databases for Big Data, International Journal of Knowledge - Based Organizations, 9(1), 2019, 50-65. https://doi.org/10.4018/IJKBO.2019010105
[10] Pérez,G.; Manuel, T.-J.; Morant, R.: Cantus: Construction and evaluation of a software solution for real-time vocal music training and musical intonation assessment, Journal of Music Technology and Education, 9(2), 2016, 125-144. https://doi.org/ 10.1386/jmte.9.2.125_1