3G
A 3G Video Phone
Solution of Improving
Synchronization between
Audio and Video
A 3G Video Phone Solution of Improving Synchronization between Audio and Video

Gao Yongqing , Shang Dan
Department of Electronic Information Engineering
North China Institute of Science and Technology
Sanhe, China
e-mail: gaoyq2008@163.com

Ping Yucai
R&D Center
LG Electrics(China) Ltd.
Beijing, China
e-mail: yucai.ping@163.com

Abstract-This paper analyzes the reasons of asynchronous between audio and video image for current 3G video phones. A solution for improving the synchronization between audio and video data is proposed. The solution proposed in this work mainly consists of software components, according to which hardware needs to be selected. It also gives the video phone terminal structure, block diagram for H.324M video phone terminal, the main software functional blocks, the implementation flow chart of each module and the whole implementation flow chart of the designed solution. The synchronization performance of audio and video can be improved by adding user interface module, timestamp monitoring module and audio decoder controlling module, by the use of software to control the decoding rate of the sound decoder and video decoder, improving the effects of 3G video phone communications.

Keywords-Third Generation; Video Phone; Audio; Video; Synchronous

I. INTRODUCTION

Mobile video phone is a video and voice point-to-point communication service, which enables real-time audio and video exchange between two mobile terminals, mobile terminal and fixed video phone or PC[1-2]. It not only brings many benefits to our daily life, but also provides a cutting-edge for business.

In the current 3G video phone, we always first heard the voice, then see the image. Video image shows a clear lag in the sound playback, lag time is normally about 1~2 seconds, there are slightly different in terminals of different manufacturers. At present, the standard of China Mobile UE-SEV-Video-001 and 3GPP standards require images and sounds of delay no more than 1.2s in both sides point to point video calls[3].

II. REASONS OF ASYNCHRONIZATION BETWEEN AUDIO AND VIDEO

The current 3G terminal, the responsibility of the video phone protocol stack H.324M is to receive the data from other side, and is to separate the audio data and video data as frame in a unit. Audio data is to be sent to the voice decoder driver through a functional interface, there is AMR format of audio data at current home. It is to send the video image data to the video decoder drivers at same time, data format is H.263 or MPEG4. For example, the MUSE decoder chip of CORELOGI Inc. or the MV decoder chip of MV Inc. The current mechanism is H.324M protocol stack to send out data which is no longer concerned by H.324M protocol stack, when can decode them in the end depending on the use of the speed of the codec chip.

Generally speaking, a small amount of audio data is about 12kbits/second. The audio data sent out by the protocol stack soon be decoded, the delay is less than 0.2 seconds. However the video data is 48kbits/second, whose amount of data is about 4 times higher than the audio data. And the video data also need to queue in the cache, the video data waiting in front of queue is to be decoded after which has been processed, which will result in the delay time that is normally ranging from 1 to 2 seconds. For example, a word said by A side, will be first heard by side B, B side is to see his expression about 1 to 2 seconds later. At this time, side A may continue to speak another word, which will result in a mismatch. Icon is shown in Figure 1[4]:

III. DESIGN THINKING

The designed solution can improve the synchronization performance of audio and video by detecting the current decoding states of both audio decoder chip and video decoder chip, as well as by the use of software to control the decoding rate of the sound decoder and video decoder.

The solution proposed in this work mainly consists of software components, according to which hardware needs to be selected.

a) Hardware: Mobile phones need to be equipped with the camera for video phone.
b) Software: Supporting functionality of video telephone.

It can be chose whether to open the function which can improve the synchronization performance between sound and video images by the menu function setting.

IV. VIDEO PHONE TERMINAL STRUCTURE

Video phone call service consists of MMI(Man-Machine Interface)/APP(Application), Video Telephony Engine, CS64Kbps Modem, H.324M Protocol Stack, Video Codec and Audio Codec. These components are independent of each other, except that Video Telephony Engine coordinates their work together. H.324M Protocol Stack implements video phone interaction of both signaling and data sending or receiving. Block diagram for H.324M video phone terminal is shown in Figure 2:

V. THE OVERVIEW OF DESIGNED SOLUTION

The designed solution is taken to solve the problems asynchronous between the audio and video data in video phone as the following program:

a) According to the data carrying the timestamp attributes of data, it make audio data and video image data separate after video phone protocol stack has received audio and video data from the network side. Of course, each separated audio data and video data has itself time stamp.

b) Since the decoding of video data is always slower than the decoding of audio data, so it can be synchronized according to the same timestamp of audio and video data before audio and video data is to be decoded.

c) There is need to add a module to monitor the timestamp when video decoder is decoding video images data, and send messages to the audio decoder before it is to be decoded.

d) Audio decoder is to decode after it has received the message.

So, it basically make the audio and video decode at the same time, which improve the synchronization performance of audio and video.

VI. MAIN SOFTWARE FUNCTION MODULES

It is necessary to add three modules in the existing 3G mobile phone software modules. These three modules are user interface module, timestamp monitoring module and audio decoder controlling module.

A. User Interface Module

The module has a visual UI module, through which user can enter the UI module, and set it the function which can improve the synchronization performance between sound and video images in video phone through menu.

The function of the user interface module of video phone is to write the setting results into the user database. When the video phone is switched on, the initialization program will access the database and check the settings before the video decoder is to decode, and turn on this function according to settings. It will then implement the corresponding program process depending on whether or not to open this function. The work flow of the user interface module is shown in Figure 3:

B. Timestamp Monitoring Module

The timestamp monitoring module is the background module running when the mobile video phone is on a call. When the video decoding program has obtained the video data from H.324M protocol stack, firstly, it will load these data into the buffer zone, and then to be decoded according to first come, first decode. The module can retrieve the current timestamp of the video data frame before each video frame is to be decoded by H.263 standard protocol packet format, and sent it to the audio decoder driver in a message, which can provide the audio decoder with decoding synchronized. The work flow of the timestamp monitoring module is shown in Figure 4:

![Figure 2: Block diagram for H.324M video phone terminal](image-url)
Audio Decoder Controlling Module

Firstly, the module is to set up a buffer for storing audio data received from H.324M protocol stack, and obtain the timestamp of audio data. Then the module poll and receive the messages sent from the timestamp monitoring module of the video decoder, according to the message carrying the timestamp information, is to retrieve the timestamp of the local audio data. Because the audio and video data which have the same timestamp is synchronous data, therefore, the module need to control the audio decoder decode the audio data timestamp which has the same timestamp, to reach the purpose of synchronization between audio and video.

In summary, the whole implementation flow chart of the designed solution shown in Figure 5:

As message passing is very fast within the mobile terminal, the time consumed by timestamp information is generally 10 to 50 milliseconds which can be ignored. So, the designed solution basically can realize the synchronization between voice and video data in video phone.

VII. CONCLUSION

The designed solution has been applied to a well-known international mobile phone manufacturers, and has been tested and compared with the former production, which has achieved benefits for its use. It make the delay time which video image display lags behind the sound playback reduce to 0.4 seconds from 1 to 2 seconds, Which make 3G video mobile audio and video synchronization performance has been significantly improved. It brings great commercial return and technical value to the development and application of 3G video phone in future, as well as far-reaching impact to improving the quality of 3G services.

ACKNOWLEDGMENT

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project.

REFERENCES

[1] Dou Zhongzhao. CDMA Wireless Communication Theory [M]. Beijing:Tsinghua University Press. 2003.
[2] Zhao Jianyong. Applied Televisions Technology[M]. Xian:Xidian University Press. 2003.
[3] Ministry of Industry and Information Technology of the People’s Republic of China. The service specification of the third generation mobile communication(Trial). 2009.
[4] ITU-T Recommendation H.234. Encryption key management and authentication systey for audiovisual services. 2002.
[5] ITU-T Recommendation H.261. Video codec for audiovisual services at p=64 kbit/s. 1993.
[6] ITU-T Recommendation H.263. Video coding for low bitrate communication. 1996.
[7] ITU-T Recommendation H.324. Terminal for low bitrate multimedia communication. 1998.
[8] ITU-T Recommendation H.245. Control Protocol for multimedia communication. 1997.
[9] ITU-T Recommendation V.34. A modem operating at data signalling rates of up to 33600 bit/s for use on the general switched telephone network and on leased point-to-point 2-wire telephone-type circuits. 1998.
[10] ITU-T Recommendation V.14. Transmission of start-stop characters over synchronous bearer channels. 1993.
[11] ITU-T Recommendation T.120. Data protocols for multimedia data conferencing. 1996.
[12] ITU-T Recommendation H.239. Role management and additional media channels for H.300-series terminals. 2005.