



(11) **EP 2 073 543 B9**

(12) **CORRECTED EUROPEAN PATENT SPECIFICATION**

(15) Correction information:
Corrected version no 1 (W1 B1)
Corrections, see
Description Paragraph(s) 29

(48) Corrigendum issued on:
29.02.2012 Bulletin 2012/09

(45) Date of publication and mention
of the grant of the patent:
06.07.2011 Bulletin 2011/27

(21) Application number: **07817026.3**

(22) Date of filing: **29.09.2007**

(51) Int Cl.:
H04N 7/15 (2006.01) H04M 3/56 (2006.01)

(86) International application number:
PCT/CN2007/070835

(87) International publication number:
WO 2008/040258 (10.04.2008 Gazette 2008/15)

(54) **SYSTEM AND METHOD FOR REALIZING MULTI-LANGUAGE CONFERENCE**

SYSTEM UND VERFAHREN ZUM REALISIEREN EINER MEHRSPRACHIGEN KONFERENZ
SYSTÈME ET PROCÉDÉ PERMETTANT DE TENIR UNE CONFÉRENCE MULTILINGUE

(84) Designated Contracting States:
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IS IT LI LT LU LV MC MT NL PL PT RO SE
SI SK TR**

(30) Priority: **30.09.2006 CN 200610063017**

(43) Date of publication of application:
24.06.2009 Bulletin 2009/26

(73) Proprietor: **Huawei Technologies Co., Ltd.**
Longgang District, Shenzhen
Guangdong 518129 (CN)

(72) Inventors:
• **LIU, Zhihui**
Shenzhen Guangdong 518129 (CN)

• **YUE, Zhonghui**
Shenzhen Guangdong 518129 (CN)

(74) Representative: **Körber, Martin Hans et al**
Mitscherlich & Partner
Patent- und Rechtsanwälte
Sonnenstraße 33
80331 München (DE)

(56) References cited:
EP-A- 1 357 493 WO-A-99/63756
WO-A2-2006/023961 CN-A- 1 685 697
CN-A- 1 685 698 CN-A- 1 845 573
CN-A- 1 937 664 JP-A- 2006 268 561
US-A- 1 685 696 US-A- 5 815 196
US-A1- 2006 120 307

EP 2 073 543 B9

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description

FIELD OF THE TECHNOLOGY

[0001] The present invention relates to the field of communication technology, and more particularly, to a system and method for realizing multi-language conference.

BACKGROUND

[0002] In recent years, with the development of communication technologies, video conference has found broad applications. The video conference may be understood as TV conference service in common sense. Through multimedia communication means, a conference is held by using television equipment and communication networks, so as to realize an interaction of images, voices and data simultaneously between two or more places. As shown in Figure 1, a video conference is generally composed of video terminals (i.e. conference terminals), a transmission network and a multipoint control unit (MCU).

[0003] Video terminal equipment mainly includes video input/output equipment, audio input/output equipment, a video codec, an audio codec, information communication equipment and multiplexing/signal distributing equipment etc. The video terminals have a basic function of performing compression coding on the image signals shot by local cameras, sound signals captured by microphones, and transmitting the signals to a remote conference site through the transmission network. At the same time, the video terminals receive signals from the remote conference site, reduce the signals to analog images and sound signals after decoding. The processing of the audio signals is shown in Figure 2 and the description thereof is given below.

[0004] In order to form a complete TV conference system, the video terminal equipments and the MCU have to be connected together through the communication network, and transmitting channels may be in the form of optical fibers, electric cables, microwaves or satellites.

[0005] The MCU is a control core of the video conference. When the number of the conference terminals participating in the conference is more than two, a control through the MCU is necessary. All conference terminals need to be connected to the MCU through standard interfaces. The MCU is realized according to protocols such as international standard H.221, and H.245. The MCU mainly functions to realize the mixing and exchange of images and voices and realize the control of all conference sites.

[0006] The MCU processes the audio data to realize sound mixing of multipoint conference sites, and the conference sites participating in the sound mixing are the conference sites with higher volumes among the multipoint conference sites. If a three-point sound mixing is to be realized, the conference sites participating in the sound mixing are three conference sites with largest vol-

umes among the multipoint conference sites. A sound mixing policy is introduced as follows.

[0007] 1) When a speech is given on one conference site, the speaker on its conference site may not hear its own voice and participants on all other conference sites may hear the voice of the speaking conference site.

[0008] 2) When a speech is given on two conference sites, both of the speakers on the speaking conference sites may hear the voice of each other, but may not hear its own voice, and participants on all other conference sites may simultaneously hear the voice of the two speaking conference sites.

[0009] 3) When a speech is given on three or more conference sites, three conference sites with largest volumes participate in the sound mixing. As shown in Figure 3, T1, T2, T3 are the three conference sites with largest sound volumes among the current conference sites, speaker on any one of the three conference sites may hear the voice of other two conference sites, for example, the speaker on T1 conference site may hear the voice of T2 and T3 conference sites, and the speakers on all the other conference sites may simultaneously hear the voice of the three conference sites.

[0010] When a conference is held employing current video conference system, the processing of sound by the conference system is as shown in Figure 2. The data of speaking conference site is sent to the MCU after being encoded, the MCU performs sound mixing process on the sound data of the speaking conference site and sends the processed sound data to other conference terminals, and the participants on the other terminals may hear the sound of the speaking conference site after decoding the sound data. In the current video conference system, the MCU and the video terminal process the sound data of a certain conference site as one flow of data. When only one language is spoken in the entire conference system, the communication among multiple conference sites may be performed smoothly. But when two or more languages are spoken in the entire conference system, obstacles in language communication among participants may occur. Directed to this problem, the conventional art has the following two solutions. The following description takes a conference mixing Chinese and English as an example, and the case in which the participants on one of multiple conference sites speak in is considered.

[0011] The first method is that all other Chinese conference sites are allocated with their own interpreters respectively to interpret English into Chinese for their own conference sites in order to understand the speaking contents of the above English conference site. As a result, if the conference scale is relative large, many interpreters are required, thus leading to waste of personnel. Moreover, when the interpreter on each Chinese conference site is interpreting, the microphone delivers the interpreted data to other conference sites, so that the scene may be chaotic and is not feasible in reality. If the microphone is set not to deliver the interpreted data to other conference sites, the speaking conference site does not know

whether the interpretation on Chinese conference sites is completed or not, and may not control its own speaking speed, thus resulting in a poor quality of the entire conference.

[0012] The second method is that a conference site is assigned as a dedicated interpreting terminal to interpret the speech of the conference sites participating in the sound mixing. But the solution also has disadvantages. If English is spoken on a conference site, and the interpreting terminal interprets English into Chinese, participants on each of the conference sites may hear English first, and then hear Chinese. But in fact, participants on the Chinese conference sites do not need to hear the English speech, and participants on the English conference sites also do not need to hear the interpreted Chinese speech. This causes the participants to have to hear much useless information. Chinese and English are mixed, thus causing conference disorders, and the participants are easy to get weary. In addition, the interpretation may slow down the conference rhythm, and reduce the efficiency.

[0013] Considering the case in which three or more languages are spoken in a conference, and the languages are spoken concurrently on multiple conference sites, the above two solutions lead to poor conference quality and are not applicable considering the effect of sound mixing.

US 2006/0120307 A1 discloses a videophone interpretation system. The videophone interpretation system accepts a call from a caller terminal and refers an interpreter registration table to extract the terminal number of an interpreter capable of interpreting between the language of a caller and the language of a callee and connects the caller terminal, a callee terminal and an interpreter terminal. The videophone interpretation system includes a function to communicate video and audio necessary for interpretation between the terminals. The audio of an interpreter is transmitted either to the caller or callee, which is specified by the interpreter terminal. The audio of the conversation partner is suppressed or interrupted when the audio of the interpreter is detected by an audio synthesizer, thereby providing a quick and precise interpretation service.

SUMMARY

[0014] In view of this, embodiments of the present invention provides a system, method and multipoint control unit for realizing a multi-language conference to make each of the conference sites only use a selected language to participate in the conference with a few interpretation resources in a multi-language multipoint conference system.

[0015] The embodiment of the present invention provides a multipoint control unit for realizing a multi-language conference.

[0016] The multipoint control unit is adapted to receive audio data from conference terminals each being as-

signed a language by the MCU, at least one of which is an interpreting terminal, which is adapted to interpret a pair of languages of speech from the conference terminals; decode the audio data received from the conference terminals; for each language perform a sound mixing process of the decoded audio data from the conference terminals of the respective language and from at least one interpreting terminal of a same corresponding language; and after the multipoint control unit performs said sound mixing process, so that a sound mixing result is obtained for each language, send the sound mixing result comprising mixed audio data of the respective language to the conference terminals with the corresponding assigned language respectively.

[0017] The embodiment of the present invention provides a method for realizing a multi-language conference, which includes the following steps.

[0018] Transforming, by each of conference terminals, a speech of the conference into audio data for sending to a multipoint control unit, MCU, and sending, by conference terminals, the audio data to the multipoint control unit. Then, a multipoint control unit receives audio data transmitted from each of conferences terminals each being assigned a language by the MCU wherein at least one of the conference terminals is an interpreting terminal, which is adapted to interpret a pair of languages of speech from the conference terminals, decodes the audio data received from conference terminals; for each language performs a sound mixing process of the decoded audio data from the conference terminals and from at least one interpreting terminal of a same corresponding language, and after the MCU performs sound mixing process, so that a sound mixing result is obtained for each language, sends a sound mixing result comprising mixed audio data of the respective language to the conference terminals with the corresponding assigned language respectively.

[0019] The audio data received by the multipoint control unit includes the audio data sent to the multipoint control unit by the at least one interpreting terminal after the interpreting terminal interprets the speech of the conference terminal according to the audio data transmitted from the multipoint control unit and processes the interpreted audio data.

[0020] As is known from the above technical scheme, in the embodiments of the present invention, because multi-channel technology is used, participants on each of the conference sites may hear the speech in a single language, and may also hear the speech in needed language as desired, thereby successfully avoiding useless language information. Because multi-channel technology is used, speaking in multiple languages in a conference system is realized successfully, in the mean time the multiple languages do not disturb each other, and different languages are combined seamlessly, thereby satisfying the demand of languages for different persons.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] Figure 1 is a schematic view of a system for holding a multipoint video conference of the conventional art;

[0022] Figure 2 is a schematic view of a processing scheme of an audio data by conference terminals in the conventional art;

[0023] Figure 3 is a schematic view of a sound mixing policy of a multipoint control unit in the conventional art;

[0024] Figure 4 is a schematic view of an application scene of an embodiment of the present invention;

[0025] Figure 5 is a schematic view of a system according to an embodiment of the present invention;

[0026] Figure 6 is a schematic view of a processing scheme of an audio data by conference terminals in the system according to a first embodiment of the present invention;

[0027] Figure 7 is a schematic view of a processing scheme of an audio data by a multipoint control unit in the system according to the first embodiment the present invention;

[0028] Figure 8 is a schematic view of a sound mixing scheme when three languages including Chinese, English and Russian exist in the system according to the first embodiment of the present invention;

[0029] Figure 9 is a schematic view of a sound mixing scheme when three languages including Chinese, English and Russian exist in the system according to the first embodiment of the present invention;

[0030] Figure 10 is a schematic view of the relationship between conference terminals and a multipoint control unit in the system according to a second or a third embodiment of the present invention;

[0031] Figure 11 is a schematic view of an implementing scheme of a multipoint control unit in the system according to the third embodiment of the present invention;

[0032] Figure 12 is a schematic view of a system according to a fourth embodiment of the present invention;

[0033] Figure 13 is a schematic view of a system according to a fifth embodiment of the present invention;

[0034] Figure 14 is a schematic view of an implementing scheme of a multipoint control unit in the system according to the fifth embodiment of the present invention;

[0035] Figure 15 is a first schematic view of a sound mixing solution when three languages including Chinese, English and Russian exist in the system according to an embodiment of the present invention; and

[0036] Figure 16 is a second schematic view of a sound mixing solution when three languages including Chinese, English and Russian exist in the system according to an embodiment of the present invention.

DETAILED DESCRIPTION

[0037] In the system and method described according to the embodiment of the present invention, by using a multi-channel processing technology in an MCU of a

multipoint conference system, different sound channels are set according to different languages, and the audio data in different languages is processed according to different sound channels. The process according to different sound channels refers to processing audio data in various languages through the respective corresponding sound channels. For example, Chinese audio data is processed in sound channel 1, English audio data is processed in sound channel 2 and the processing of audio data in various languages does not disturb each other. The multipoint conference system described in the embodiment of the present invention includes not only a conventional pure-voice multipoint conference system, but also a TV conference system or a video conference system.

[0038] The embodiment of the present invention takes a Chinese-English multipoint conference as an example and other conferences using two or more languages are processed in a similar way.

[0039] Referring to Figure 4, an application scene of an embodiment of the present invention is described. Each of the conference sites corresponds to a conference terminal, and any one of the conference sites in the multipoint conference system is assigned as an interpreting conference site responsible for interpreting Chinese into English or interpreting English into Chinese for the speech of each of the conference sites. Referring to Figure 5, the system according to the embodiment of the present invention includes each of the conference terminals and an MCU.

[0040] A first embodiment of the system of the present invention is described below.

[0041] The conference terminal and an interpreting terminal (the conference terminal of interpretation conference site is referred to as the "interpreting terminal") only use one input and output interface, the MCU adopts mixing of sounds from multiple sound channels, and the MCU assigns languages to which the conference terminals correspond and the conference terminal acting as the interpreting terminal. In this scheme, there are no special requirements for the conference terminals and the interpreting terminal, and the majority of work is completed by the MCU. Various parts are described in detail as follows.

[0042] As in Figure 6, the conference terminal gathers locally input audio signals, encode the signals and send the encoded signals to the MCU. The conference terminal also receives an audio code stream sent from the MCU and plays the audio code stream after decoding. As such, the conference terminal does not need to differentiate input and output languages.

[0043] The interpreting terminal is responsible for interpreting the language of the speech in each of the conference sites, for example, interpreting Chinese speech into English, or interpreting English speech into Chinese. In this embodiment, simultaneous interpretation by interpreters may be performed on the interpreting conference site, or intelligent simultaneous machine interpretation

may also be performed by using interpreting machines. Because simultaneous interpretation is introduced, the conference delay caused by interpretation may be ignored essentially, thereby effectively ensuring the smooth completion of the conference.

[0044] As shown in Figure 7, when dispatching the conference, the MCU assigns language information for each of the conference terminals first and assigns a conference terminal having simultaneous interpretation function as the interpreting terminal. For example, the language of a conference terminal 1 is assigned to be Chinese, the language of a conference terminal 2 is assigned to be English and a conference terminal 3 is assigned to be the interpreting terminal in the mean time. When the MCU receives audio data from a conference terminal, decoding is performed first, and then the decoded voice data is sent to a corresponding language mixer for sound mixing according to the language information assigned for the conference terminal. For example, for a conference terminal being assigned Chinese, the MCU sends the received audio data of this conference terminal to a Chinese mixer for sound mixing, and sends the audio data of the conference terminal assigned English to an English mixer for sound mixing. The audio data sent by the interpreting terminal to the MCU always participates in the sound mixing, and the language of the sound mixing in which it participates is introduced below. After the MCU performs sound mixing individually according to different languages, the sound mixing policy for various languages may be obtained with reference to the conventional art, the data after the sound mixing is sent to the corresponding conference terminals being assigned languages respectively. For example, conference terminals participating in Chinese sound mixing receive the corresponding data subject to Chinese sound mixing, and conference terminals participating in English sound mixing receive corresponding data subject to English sound mixing.

[0045] The MCU sound mixing has two steps. Firstly, the volume size of each of the conference sites is compared, and secondly, the conference sites whose sound is to be broadcasted are selected according to the volume size thereof, and the sound is sent to receiving conference sites. The sound mixing may adopt one-party sound mixing, two-party sound mixing, three-party sound mixing, even four or more party sound mixing. In one-party sound mixing, only the party with the highest sound is broadcasted to each of the conference sites. In two-party sound mixing, only two parties with the highest sound are broadcasted after being mingled, and the participants on the two conference sites with the highest sound being broadcasted may not hear their own sound, but may hear the sound of the other party. One-party sound mixing and two-party sound mixing tend to result in the disadvantage of word-cut in an interactive conference involving more than three conference sites. Therefore, three-party sound mixing is put forward, and a schematic view of three-party sound mixing is shown in Figure 3.

[0046] The volumes of different conference sites are compared to determine three conference sites with the largest sounds, namely, T1, T2, and T3, which participating in the sound mixing. The data of the T1, T2 and T3 is sent to other conference sites which do not participate in the sound mixing after data adding. The participants on the T1, T2, and T3 hear respectively the mingled sound of other two conference sites, i.e. the participants on the T1 hear the mingled sound of the T2 and T3 conference sites, the participants on the T2 hear the mingled sound of the T1 and T3 conference sites, and the participants on the T3 hear the mingled sound of the T1 and T2 conference sites. The one-party sound mixing is easy to be realized, and is relatively fit for the case in which only the chairman speaks, but the interactive performance thereof is poor. The two-party sound mixing may have certain level of interactivity, but when a third party inserts a word, the case in which the sound of one party is cut tends to occur. The interactive effect of three-party sound mixing is relatively good.

[0047] After the MCU has mixed the sound independently according to different languages, the sound mixing result is sent respectively to corresponding conference terminals being assigned languages. For example, conference terminals participating in Chinese sound mixing receive corresponding data subject to Chinese sound mixing, and conference terminals participating in English sound mixing receive corresponding data subject to English sound mixing.

[0048] In Figure 7, each pair of languages use only one interpreting terminal, for example, a Chinese-English interpreting terminal may interpret Chinese into English, and may also interpret English into Chinese. Because the interpreting work of speeches of various conference sites is processed densely in a few conference sites, the interpretation resources are saved greatly while guaranteeing the quality of interpretation, thus reducing the conference cost. When one-party sound mixing is adopted, sound of a party in the Chinese and English languages with the highest sound is sent to the interpreting terminal. The interpreting terminal, which does not participate in the volume comparison, sends the audio data with the highest volume to the interpreting conference site. If Chinese is spoken at the highest volume, then the audio data language sent to the interpreting terminal is Chinese. In the next sound mixing, the audio data of the interpreting terminal participates in the English language sound mixing. If English is spoken at the highest volume, the audio data language sent to the interpreting terminal is English. In the next sound mixing, the audio data of the interpreting terminal participates in Chinese language sound mixing. Although only the louder one of Chinese and English languages is sent to the interpreting terminal, the sound mixing of Chinese language or English language itself may adopt one-party, two-party, three-party, even four or more party sound mixing. The interpreting terminal always participates in the sound mixing of the corresponding language with the highest

volume or one of the highest volumes.

[0049] When two-party sound mixing is used, the audio data with the highest volumes of two parties is selected to be sent to the interpreting terminal for interpretation, and thus two interpreting terminals are required. One interpreting terminal interprets the sound of the conference site with the first highest volume, and the other interpreting terminal interprets the sound of the conference site with the second highest volume. The volume of the interpreting conference site does not participate in the volume comparison. The interpreted sound is sent to the mixer of a corresponding language for sound mixing. If both parties with the highest volumes are Chinese or English, the data interpreted by the interpreting terminals participates in the sound mixing of English or Chinese in the next sound mixing as two parties with the highest volumes or two of several parties with the highest volumes. If one of the two parties with the highest volumes is Chinese and the other is English, then the two parties participate in English and Chinese sound mixing respectively after being interpreted by corresponding interpreting terminal. The interpreting terminal always participates in the sound mixing, and the volume thereof may be equal to that of the conference site for which it interprets, or may be one of the highest volumes. Likewise, Chinese language mixer or English language mixer may adopt two-party, three-party or multi-party sound mixing.

[0050] When three-party or multi-party sound mixing is used, the processing of sound mixing is similar to that of two-party sound mixing, and three or more interpreting terminals are required to interpret the volumes of several parties with the highest volumes.

[0051] A multi-language conference in two languages of Chinese and English is described above for the ease of illustration, and any multi-language conference in two languages may be implemented according to the above method. The number of languages holding a conference simultaneously may be greater than two. When a language is added, the processing of terminals does not change, and the MCU adds a mixer accordingly. As shown in Figure 8, taking a conference in three languages including Chinese, English and Russian as an example, three interpreting terminals are provided, one is used for realizing Chinese-English interpretation, one is used for realizing Chinese-Russian interpretation, and one is used for realizing English-Russian interpretation. The sound mixing result of Chinese, English and Russian is sent to the respective listening conference site. One-party sound mixing is adopted in Figure 8, data of one party with the highest volume among Chinese, English and Russian is sent to the corresponding interpreting conference terminal after being encoded, and the volume of the interpreting conference terminal does not participate in the comparison. If Chinese is spoken at the highest volume, the Chinese data is sent to Chinese-English and Chinese-Russian interpreting conference sites. If English or Russian is spoken at highest volume, the processing method is similar. In this method, only one of Chinese,

English and Russian with the highest volume is interpreted. The situation for four or more languages may be deduced by analog.

[0052] The processing of two or more parties sound mixing is similar to the above description, but several more interpreting terminals are needed to interpret the sound of the conference site with the second highest volume.

[0053] Figure 9 shows a more flexible way of implementing sound mixing. In the sound mixing, the volumes are compared for Chinese and English, for Chinese and Russian, and for English and Russian respectively, and data of one party with the highest volume is sent to the corresponding interpreting terminal for interpretation. For example, if Chinese and English are compared in volume size, one party with the higher volume is sent to the Chinese-English interpreting conference terminal. If Chinese and Russian are compared in volume size, the one party with the higher volume is sent to the Chinese-Russian interpreting conference terminal. If English and Russian are compared in volume size, the one party with the higher volume is sent to the English-Russian interpreting conference terminal. The interpreted audio data participates in the sound mixing of the corresponding language.

For example, the data participating in Chinese sound mixing contains the data of Chinese, Chinese interpreted from English, and Chinese interpreted from Russian. Chinese sound mixing includes one-party sound mixing, two-party sound mixing, three-party sound mixing, and even more party sound mixing. In Chinese sound mixing, the volume of Chinese interpreted from English participates in the sound mixing according to the volume of English, and the volume of Chinese interpreted from Russian participates in the sound mixing according to the volume of Russian. The sound mixing processing of English and Russian are similar to that of Chinese, and the situation for four or more languages may be deduced by analog.

[0054] In Figure 9, two or more party sound mixing may also be implemented between every two languages, but interpreting terminals are to be increased accordingly.

[0055] A second embodiment of the system of the present invention is described below.

[0056] The conference terminals and the interpreting terminal use only one input and output interface. The conference terminals need to inform the MCU its language or inform the MCU that it is an interpreting terminal in a manner of signaling. The MCU adopts the mixing of sounds from multiple sound channels. Codecs used by the conference terminals, the interpreting terminal, and the MCU are single-channel codecs. Various parts are described in detail as follows.

[0057] The implementation of the conference terminals is substantially the same as that in the first embodiment. What is different is that as shown in Figure 10, the conference terminals support language assignment for the conference sites, and confirm the MCU the languages of the conference sites through signaling such as H.245. For example, a conference terminal selects Chinese as

the language, and the MCU is informed that the conference terminal belongs to a Chinese conference site..

[0058] The implementation of the interpreting terminal is the same as that in the first embodiment, what is different is that the conference terminal is assigned as the interpreting terminal, and the MCU is informed that the conference terminal is the interpreting terminal through signaling such as H.245.

[0059] The implementation of the MCU is the same as that in the first embodiment, what is different is that the MCU knows the language assigned for the conference terminals or whether or not a conference terminal serves as the interpreting terminal through signaling such as H.245. That is, the conference site attribute assignment information in Figure 7 is obtained from the conference terminals through signaling such as H.245.

[0060] A third embodiment of the system of the present invention is described below.

[0061] The conference terminals and the interpreting terminal use only one input and output interface. The conference terminal needs to inform the MCU its language in an in-band manner, or inform the MCU that it is an interpreting terminal. The MCU adopts the mixing of sounds from multiple sound channels. Codecs used by the conference terminals, the interpreting terminal, and the MCU are single-channel codecs. Various parts are described in detail as follows.

[0062] The implementation of the conference terminal is the same as that in the second embodiment, what is different is that the conference terminal receives and encodes the input audio data, and packs the encoded code stream and the language information corresponding to the code stream for sending to the MCU. See Figure 10, for example, the Chinese conference terminal adds Chinese identifiers in the code stream to be sent to the MCU, and the English conference terminal adds English identifiers in the code stream to be sent to the MCU, rather than indicating the MCU its language through signaling.

[0063] The implementation of the interpreting terminal is the same as the second embodiment, what is different is that the conference terminal is assigned as an interpreting terminal, the interpreting terminal adds an identifier of the interpreting terminal in the code stream to be sent to the MCU for informing the MCU that it is the interpreting terminal, rather than indicating the MCU its identity as the interpreting terminal through signaling, see Figure 10.

[0064] The implementation of the MCU is the same as that in the second embodiment, what is different is that, see Figure 11, the MCU determines the language of the code stream and whether or not the conference terminal is an interpreting terminal by analyzing the received audio code stream, rather than through signaling.

[0065] A fourth embodiment of the system of the present invention is described below.

[0066] Sometimes, on one conference site, there may be speeches given or listened to in multiple languages. If the multiple languages on the conference site are spoken in speeches alternately, the language of the confer-

ence terminal needs to be updated continuously according to the foregoing method, which is quite inconvenient in operation. In this implementation, the conference terminal uses multiple input-output interfaces.

[0067] The conference terminal has two input channels, a channel language allocation module of the conference terminal assigns the language corresponding to each input channel, for example, Chinese corresponds to channel 1, and English corresponds to channel 2. The data of channel 1 and channel 2 is compared first in volume before being sent to an encoder, the party with the higher volume is sent to the encoder for encoding, as shown in Figure 12. The language information corresponding to the channel is added to the encoded code stream before the encoded code stream is sent to the MCU. For example, when only Chinese language is spoken on a conference site, the data in the channel 1 is sent to the encoder for encoding, and the encoded code stream is sent to the MCU after the identifier of Chinese is marked in the code stream. When only English language is spoken on the conference site, the data in the channel 2 is sent to the encoder for encoding, and the encoded code stream is sent to the MCU after the identifier of English is marked in the code stream. If Chinese and English languages are spoken at the same time, the party with the higher volume is sent to the MCU after being encoded and added with corresponding language information.

[0068] The conference terminal being assigned as the interpreting terminal adds the interpreting terminal identifier in the code stream for sending to the MCU. Of course, the conference terminal being assigned as the interpreting terminal may also indicate the MCU its identity through signaling such as H.245.

[0069] As shown in Figure 11, after receiving the audio code stream from the conference terminal, the MCU identifies the language corresponding to the code stream and whether or not the code stream is from the interpreting terminal through the terminal language and the identifier of the interpreting terminal. After being decoded, the code stream is sent to mixers for different languages for sound mixing according to the language information after being identified. For example, Chinese language is sent to a Chinese mixer for sound mixing, and English language is sent to an English mixer for sound mixing. After the sound mixing, the sound mixing results of Chinese and English are paired one by one according to volume size, encoded and marked with language identifiers respectively, and then sent to corresponding receiving conference terminals. For example, the pair with the highest volume including Chinese sound mixing output and English sound mixing output is sent to the conference sites that receives the highest volume of Chinese or English.

[0070] As shown in Figure 12, in this scheme, the conference terminals may also select languages to receive. The conference terminals inform the MCU the language they select to receive through signaling such as H.245. After sound mixing, the MCU sends the mixed data to

the conference terminals according to the language information that the conference terminals select to listen to. If the conference terminal selects to listen to the language information of Chinese and English, the data sent to the conference terminal by the MCU is processed as described above. If the conference terminal only selects to listen to Chinese, the result after Chinese sound mixing is sent to the corresponding receiving conference terminal, and English is processed similarly. The interpreting terminal needs to be set so as to receive both Chinese and English languages.

[0071] The conference terminal decodes selectively or all of the received MCU code streams according to assignment of the language to be listened to, and then output the decoded code streams to corresponding output channels. For example, if the output channel 1 is assigned to output Chinese, the output channel 2 is assigned to output English, when the conference terminal receives Chinese and English code streams, the conference terminal decodes the code streams respectively, and outputs the decoded code streams to corresponding output channels. If the output channel 1 is only assigned to output Chinese, the data of English code stream is discarded, and only the data of Chinese code stream is decoded and sent to the output channel 1 for outputting. If the output channel 2 is only assigned to output English, the data of Chinese code stream is discarded, only the data of English code stream is decoded and sent to the output channel 2 for outputting.

[0072] The audio data sent to the interpreting terminal and the language sound mixing in which the interpreting terminal participates is processed as follows. The MCU compares the volume size of the received Chinese and English audio data (the interpreting terminal does not participate in the volume comparison), sends the audio data with the highest volume to the interpreting conference site. If Chinese is spoken at the highest volume, the language of audio data sent to the interpreting terminal is Chinese, and in the next sound mixing, the audio data of the interpreting terminal received by the MCU participates in the English language sound mixing. If English is spoken at the highest volume, the language of audio data sent to the interpreting terminal is English, and after completion of the interpretation, the interpreting terminal sends the interpreted Chinese data to the MCU for participating in the Chinese language sound mixing.

[0073] A fifth embodiment of the system of the present invention is described below.

[0074] This scheme is similar to scheme 4, the differences include that multi-channel encoders are adopted.

[0075] As is shown in Figure 13, the difference between the processing of the conference terminal in this scheme and that in scheme 4 lies in that, the encoding of different languages corresponds to encoding sound channels of the multi-channel encoder. For example, Chinese is encoded correspondingly in a left channel, English is encoded in a right channel, and the encoded data is sent to the MCU after being packed.

[0076] The conference terminal being assigned as the interpreting terminal adds interpreting terminal identifier in the code stream to be sent to the MCU. It may also indicate the MCU its identity through signaling such as H.245.

[0077] As is shown in Figure 14, the MCU receives and decodes the audio code stream sent from the conference terminal, and then delivers sound channel data to corresponding mixers for sound mixing. For example, the data of the left channel is delivered to the Chinese mixer for sound mixing, and the data of the right channel is delivered to the English mixer for sound mixing. After the sound mixing, the sound mixing results of Chinese and English are paired one by one according to volume size, encoded in the corresponding sound channels of the multi-channel encoder, and then sent to the conference terminals.

[0078] The conference terminal receives and decodes the code stream sent by the MCU, and sends the decoded data of corresponding language to the output channel according to the selected output language. The interpreting terminal is set to receive both Chinese and English languages.

[0079] The audio data sent to the interpreting terminal and the language sound mixing in which the interpreting terminal participates are processed as follows. The MCU compares the volume size of the received Chinese and English audio data (the interpreting terminal does not participate in the volume comparison), sends the audio data with the highest volume to the interpreting conference site. If Chinese language is spoken at the highest volume, the language of audio data sent to the interpreting terminal is Chinese, and in the next sound mixing, the audio data of the interpreting terminal received by the MCU participates in the English language sound mixing. If English is spoken at the highest volume, the language of audio data sent to the interpreting terminal is English, and after completion of the interpretation, the interpreting terminal sends the interpreted Chinese data to the MCU for participating in the Chinese language sound mixing.

[0080] A multi-language conference in two languages of Chinese and English is described above for the ease of illustration, and the multi-language conference in any two languages may be implemented according to the above-mentioned method. The number of languages simultaneously used when holding a conference may be greater than two. When a language is added, the processing of terminals does not change, but the MCU adds a mixer accordingly. As shown in Figure 15, taking a conference in three languages of Chinese, English and Russian as an example, the sound mixing result of Chinese, English and Russian is sent to their respective listening conference site. In addition, one party among Chinese, English and Russian with the highest volume is sent to the corresponding interpreting conference terminal after being encoded. It should be noted that the volume from the interpreting terminal does not participate

in the comparison. If Chinese is spoken at the highest volume, the Chinese data is sent to Chinese-English and Chinese-Russian interpreting conference sites. If English or Russian is spoken at the highest volume, the processing method is similar. In this method, only one party of Chinese, English and Russian with the highest volume is interpreted. The situation for four or more languages may be deduced by analog.

[0081] Figure 16 is a more flexible way for implementing sound mixing. After the sound mixing, the volumes are compared for Chinese and English, for Chinese and Russian, and for English and Russian respectively, and then the languages are sent to corresponding interpreting terminals for interpretation. In this way, participants on the conference sites of Chinese, English and Russia may be spoken at the same time. The situation for four or more languages may be deduced by analog.

[0082] Implementation embodiments of the method of the present invention are described below.

[0083] Being the same as the embodiment of the above system embodiments, various conference sites are first assigned with the language information and the input-output interfaces connecting conference terminals correctly for respective conference sites, a conference site is assigned as an interpreting conference site responsible for interpreting the speeches of the conference sites. When only two languages exist in a multipoint conference system, it is enough to assign one conference site as the interpreting conference site. When more than three languages exist, multiple interpreting conference sites should be assigned. Each conference terminal transforms the speech of the conference into audio data for sending to the multipoint control unit. The method further includes the following steps.

[0084] The interpreting conference site performs simultaneous interpretation on the speeches of the conference sites, and sends the interpreted speeches to the multipoint control unit according to language types.

[0085] The multipoint control unit mixes the audio data from the conference terminals in different sound channels according to language types, and sends the processed audio data to the conference terminals according to the language preset by the multipoint control unit or the language selection of the conference terminals.

[0086] The conference terminals process the audio data from the multipoint control unit and output the processed audio data.

[0087] The implementations of the conference terminals and the multipoint control unit may be obtained with reference to the first to fifth embodiments of the above system, which are not described in detail here.

[0088] When using the method, no matter for the case in which only participants on the interpreting conference site speak, or the case in which only participants on other conference sites speak, or the case in which both the interpreting conference site and other conference sites participate in the speaking, processing in different sound channels according to language types may be realized

successfully, the speech contents in various languages do not disturb each other, and various conference sites may select the language they need to participate in the conference.

[0089] The above embodiments of the present invention take the Chinese and English conference as an example, but in practical application process of the embodiment of the present invention, the languages used in the conference are not limited to Chinese and English, and may be a combination of any two languages. The languages used in the conference are not limited to two kinds, and may be three kinds or more. The sound mixing policies used by the multipoint control unit are not limited to the method used by the conventional art, and other sound mixing policies are applicable as well. When the system and method according to the embodiment of the present invention are used, each of the conference sites may participate in the conference according to the language it selects. For example, when three languages of Chinese, English and Russian exist in the conference system, it is enough to provide three sound channels to process the audio data of corresponding languages. Three interpreting conference sites are assigned at the same time, namely Chinese-English interpreting conference site, Chinese-Russian interpreting conference site, and English-Russian interpreting conference site. Each conference site decides to output only the language it needs according to assignment information, and shields useless speech information.

[0090] Finally, it should be understood that the above embodiments are only used to explain, but not to limit the technical solution of the present invention. In despite of the detailed description of the present invention with referring to above preferred embodiments, it should be understood that various modifications, changes or equivalent replacements can be made by those skilled in the art without departing from the spirit and scope of the present invention and covered in the claims of the present invention.

Claims

1. A multipoint control unit, MCU, for realizing multi-language conference, wherein the MCU is adapted to:

receive audio data from conference terminals each being assigned a language by the MCU, at least one of which is an interpreting terminal, which is adapted to interpret a pair of languages of speech from the conference terminals;
decode the audio data received from the conference terminals;
for each language perform a sound mixing process of the decoded audio data from the conference terminals of the respective language and from at least one interpreting terminal of a same

- respective language, and
 after the MCU performs said sound mixing process, so that a sound mixing result is obtained for each language, send the sound mixing result comprising mixed audio data of the respective language to the conference terminals with the corresponding assigned language respectively.
2. The MCU according to claim 1, wherein the MCU is adopting one-party sound mixing, two-party sound mixing, three-party sound mixing, or more-party sound mixing policy, whereby in one-party sound mixing, only the conference terminal with the highest sound is broadcasted to each of the other conference terminals, in two-party sound mixing, the two conference terminals with the highest sound are broadcasted after being mixed and accordingly for further multiple parties analogous sound mixing policies are employed.
 3. A system for realizing multi-language conference, comprising conference terminals and the MCU according to claim 1 or 2, wherein the conference terminals are adapted to send a speech after being processed to an MCU, process an audio data received from the MCU and output the processed audio data;
 wherein the at least one interpreting terminal is adapted to interpret the speech according to the audio data transmitted from the MCU, process the interpreted audio data and send the processed audio data to the MCU.
 4. The system according to claim 3, wherein each of the conference terminals is a conference terminal having one input interface and one output interface; or
 each of the conference terminals is a conference terminal having multiple input interfaces and multiple output interfaces.
 5. The system according to claim 4, wherein each of the conference terminals is a conference terminal adopting a single-channel codec to perform encoding and decoding, or
 each of the conference terminals is the conference terminal adopting a multi-channel codec to perform encoding and decoding.
 6. The system according to claim 3, wherein the MCU is a MCU adopting one-party sound mixing, two-party sound mixing, three-party sound mixing, or more-party sound mixing policy.
 7. A method for realizing multi-language conference, comprising:
 transforming, by each conference terminal, a speech of the conference into audio data for sending to a multipoint control unit, MCU, and sending, by conference terminals, the audio data to the MCU
 receiving, by the MCU, audio data transmitted from each of the conference terminals each being assigned a language by the MCU wherein at least one of the conference terminals is an interpreting terminal, which is adapted to interpret a pair of languages of speech from the conference terminals;
 decode the audio data received from conference terminals;
 for each language performing a sound mixing process of the decoded audio data from the conference terminals of the respective language and from at least one interpreting terminal of a same respective language; and
 after the MCU performs said sound mixing process, so that a sound mixing result is obtained for each language, sending the sound mixing result comprising mixed audio data of the respective language to the conference terminals with the corresponding assigned language respectively,
 wherein the at least one interpreting terminal sends the audio data to the MCU after the interpreting terminal interprets the speech of the conference terminal according to the audio data transmitted from the MCU and processes the interpreted audio data.
 8. The method according to claim 7, wherein the conference terminal processes the audio data through one input interface and one output interface; or
 the conference terminal processes the audio data through multiple input interfaces and multiple output interfaces.
 9. The method according to claim 8, wherein the conference terminals process the audio data by adopting a single-channel codec; or
 the conference terminals process the audio data by adopting a multi-channel codec.
 10. The method according to claim 8, wherein the sound mixing process is: one-party sound mixing, two-party sound mixing, three-party sound mixing or more-party sound mixing process, whereby in one-party sound mixing, only the conference terminal with the highest sound is broadcasted to each of the other conference terminals, in two-party sound mixing, the two conference terminals with the highest sound are broadcasted after being mixed and accordingly for further multiple parties analogous sound mixing policies are employed.

transforming, by each conference terminal, a

11. The method according to claim 7, wherein the sending the processed audio data to the corresponding conference terminals being assigned languages respectively comprises:

5 sending, by the MCU, the audio data after the sound mixing process of a corresponding sound channel to the conference terminals according to language information corresponding to the conference terminals and identifier information of the interpreting terminal. 10

12. The method according to claim 11, further comprising:

15 presetting, by the MCU, the language information corresponding to the conference terminals and/or the identifier information of the interpreting terminal; and/or receiving and storing, by the MCU, the language information and/or the identifier information of the interpreting terminal transmitted from the conference terminals. 20

13. The method according to any one of claims 8 to 12, further comprising:

25 selecting, by the MCU, conference sites according to volume size of the speech of each of the conference sites, and sending the audio data of the selected conference sites to the interpreting terminal. 30

14. The method according to claim 13, wherein the selected conference sites comprise one or more conference sites, and the audio data transmitted from the interpreting terminal always participates in the sound mixing processing in different sound channels. 35

Patentansprüche

1. Mehrpunkt-Steuereinheit MCU zum Realisieren einer mehrsprachigen Konferenz, wobei die MCU für Folgendes ausgelegt ist: 45

Empfangen von Audiodaten von Konferenzendgeräten, denen jeweils durch die MCU eine Sprache zugewiesen wird, wobei mindestens eines ein dolmetschendes Endgerät ist, das dafür ausgelegt ist, ein Paar von Sprachen der Sprache von den Konferenzendgeräten zu dolmetschen; 50
Decodieren der von den Konferenzendgeräten empfangenen Audiodaten; 55
Durchführen eines Tonmischprozesses der decodierten Audiodaten von den Konferenzend-

geräten der jeweiligen Sprache und von mindestens einem dolmetschenden Endgerät einer selben jeweiligen Sprache für jede Sprache und nachdem die MCU den Tonmischprozess durchführt, sodass ein Tonmischergebnis für jede Sprache erhalten wird, Senden des Tonmischergebnisses, umfassend gemischte Audiodaten der jeweiligen Sprache, zu den Konferenzendgeräten jeweils mit der entsprechenden zugewiesenen Sprache.

2. MCU nach Anspruch 1, wobei die MCU eine Einteilnehmer-Tonmisch-, Zweiteilnehmer-Tonmisch-, Dreiteilnehmer-Tonmisch- oder Mehrteilnehmer-Tonmischrichtlinie verwendet, wodurch beim Einteilnehmer-Tonmischen nur das Konferenzendgerät mit dem stärksten Ton zu jedem der anderen Konferenzendgeräte rundgesendet wird, beim Zweiteilnehmer-Tonmischen die zwei Konferenzendgeräte mit dem stärksten Ton rundgesendet werden, nachdem sie gemischt wurden, und entsprechend für weitere mehrere Teilnehmer analoge Tonmischrichtlinien verwendet werden.

3. System zum Realisieren einer mehrsprachigen Konferenz, das Konferenzendgeräte und die MCU nach Anspruch 1 oder 2 umfasst, wobei die Konferenzendgeräte dafür ausgelegt sind, eine Sprache nach Bearbeitung zu einer MCU zu senden, von der MCU empfangene Audiodaten zu verarbeiten und die verarbeiteten Audiodaten auszugeben; wobei das mindestens eine dolmetschende Endgerät dafür ausgelegt ist, die Sprache gemäß den von der MCU gesendeten Audiodaten zu dolmetschen, die gedolmetschten Audiodaten zu verarbeiten und die verarbeiteten Audiodaten zu der MCU zu senden. 40

4. System nach Anspruch 3, wobei jedes der Konferenzendgeräte ein Konferenzendgerät mit einer Eingangsschnittstelle und einer Ausgangsschnittstelle ist; oder jedes der Konferenzendgeräte ein Konferenzendgerät mit mehreren Eingangsschnittstellen und mehreren Ausgangsschnittstellen ist. 45

5. System nach Anspruch 4, wobei jedes der Konferenzendgeräte ein Konferenzendgerät ist, das einen Einkanal-Codec zum Durchführen von Codierung und Decodierung verwendet, oder jedes der Konferenzendgeräte das Konferenzendgerät ist, das einen Mehrkanal-Codec zum Durchführen von Codierung und Decodierung verwendet. 50

6. System nach Anspruch 3, wobei die MCU eine MCU ist, die eine Einteilnehmer-Tonmisch-, Zweiteilnehmer-Tonmisch-, Dreiteilnehmer-Tonmisch- oder Mehrteilnehmer-Tonmischrichtlinie 55

verwendet.

7. Verfahren zum Realisieren einer mehrsprachigen Konferenz, mit den folgenden Schritten:

Transformieren einer Sprache der Konferenz durch jedes Konferenzendgerät zu Audiodaten zum Senden zu einer Mehrpunkt-Steuereinheit MCU, und Senden der Audiodaten durch Konferenzendgeräte zu der MCU;

Empfangen von Audiodaten, die von jedem der Konferenzendgeräte gesendet werden, durch die MCU, denen jeweils durch die MCU eine Sprache zugewiesen wird, wobei mindestens eines der Konferenzendgeräte ein dolmetschendes Endgerät ist, das dafür ausgelegt ist, ein Paar von Sprachen der Sprache von den Konferenzendgeräten zu dolmetschen;

Decodieren der von Konferenzendgeräten empfangenen Audiodaten;

Durchführen eines Tonmischprozesses der decodierten Audiodaten von den Konferenzendgeräten der jeweiligen Sprache und von mindestens einem dolmetschenden Endgerät einer selben jeweiligen Sprache für jede Sprache; und nachdem die MCU den Tonmischprozess durchführt, sodass ein Tonmischergebnis für jede Sprache erhalten wird, Senden des Tonmischergebnisses, umfassend gemischte Audiodaten der jeweiligen Sprache, zu den Konferenzendgeräten jeweils mit der entsprechenden zugewiesenen Sprache, wobei das mindestens eine dolmetschende Endgerät die Audiodaten zu der MCU sendet, nachdem das dolmetschende Endgerät die Sprache des Konferenzendgeräts gemäß den von der MCU gesendeten Audiodaten dolmetscht und die gedolmetschten Audiodaten verarbeitet.

8. Verfahren nach Anspruch 7, wobei das Konferenzendgerät die Audiodaten durch eine Eingangsschnittstelle und eine Ausgangsschnittstelle verarbeitet; oder das Konferenzendgerät die Audiodaten durch mehrere Eingangsschnittstellen und mehrere Ausgangsschnittstellen verarbeitet.

9. Verfahren nach Anspruch 8, wobei die Konferenzendgeräte die Audiodaten durch Verwendung eines Einkanal-Codecs verarbeiten; oder die Konferenzendgeräte die Audiodaten durch Verwendung eines Mehrkanal-Codecs verarbeiten.

10. Verfahren nach Anspruch 8, wobei der Tonmischprozess Folgendes ist: ein Einteilnehmer-Tonmisch-, Zweiteilnehmer-Tonmisch-, Dreiteilnehmer-Tonmisch- oder Mehrteilnehmer-Tonmischprozess,

wodurch bei dem Einteilnehmer-Tonmischen nur das Konferenzendgerät mit dem stärksten Ton zu jedem der anderen Konferenzendgeräte rundgesendet wird, bei dem Zweiteilnehmer-Tonmischen die zwei Konferenzendgeräte mit dem stärksten Ton rundgesendet werden, nachdem sie gemischt wurden, und entsprechend für weitere mehrere Teilnehmer analoge Tonmischrichtlinien verwendet werden.

11. Verfahren nach Anspruch 7, wobei das Senden der verarbeiteten Audiodaten zu den entsprechenden Konferenzendgeräten, denen jeweils Sprachen zugewiesen sind, Folgendes umfasst:

Senden der Audiodaten durch die MCU nach dem Tonmischprozess eines entsprechenden Tonkanals zu den Konferenzendgeräten gemäß Spracheninformationen, die den Konferenzendgeräten entsprechen, und Kennungsinformationen des dolmetschenden Endgeräts.

12. Verfahren nach Anspruch 11, ferner mit den folgenden Schritten:

Voreinstellen der Spracheninformationen, die den Konferenzendgeräten entsprechen, und/oder der Kennungsinformationen des dolmetschenden Endgeräts durch die MCU; und/oder Empfangen und Speichern der Spracheninformationen und/oder Kennungsinformationen des dolmetschenden Endgeräts, die von den Konferenzendgeräten gesendet werden, durch die MCU.

13. Verfahren nach einem der Ansprüche 8 bis 12, ferner mit den folgenden Schritten:

Auswählen von Konferenzorten durch die MCU gemäß Lautstärke der Sprache jedes der Konferenzorte und Senden der Audiodaten der gewählten Konferenzorte zu dem dolmetschenden Endgerät.

14. Verfahren nach Anspruch 13, wobei die gewählten Konferenzorte einen oder mehrere Konferenzorte umfassen und die von dem dolmetschenden Endgerät gesendeten Audiodaten immer an der Tonmischverarbeitung in verschiedenen Tonkanälen teilnehmen.

Revendications

1. Unité de commande multipoint, MCU, pour réaliser une conférence multilingue, la MCU étant adaptée pour :

- recevoir des données audio depuis des terminaux de conférence à chacun desquels la MCU a assigné une langue, dont au moins l'un est un terminal interprète, lequel est adapté pour interpréter une paire de langues de conversation des terminaux de conférence; 5
- décoder les données audio reçues des terminaux de conférence; 5
- pour chaque langue effectuer un processus de mixage sonore des données audio décodées des terminaux de conférence de la langue respective et d'au moins un terminal interprète d'une même langue respective; et 10
- quand la MCU a terminé ledit processus de mixage sonore, afin d'obtenir un résultat de mixage sonore pour chaque langue, envoyer le résultat de mixage sonore comprenant les données audio mixtes de la langue respective aux terminaux de conférence de la langue assignée correspondante respectivement. 15 20
2. MCU selon la revendication 1, la MCU adoptant une règle de mixage sonore pour un correspondant, deux correspondants, trois correspondants, ou plus, selon laquelle dans le mixage sonore pour un correspondant, seul le terminal de conférence ayant le son le plus haut est diffusé à chacun des autres terminaux de conférence, dans le mixage sonore pour deux correspondants, les deux terminaux de conférence ayant le son le plus haut sont diffusés après avoir été mixés et pour de multiples correspondants des règles de mixage sonore analogues sont employées comme il convient. 25 30
3. Système de réalisation d'une conférence multilingue, comprenant des terminaux de conférence et la MCU selon la revendication 1 ou 2, dans lequel les terminaux de conférence sont adaptés pour envoyer à une MCU une conversation après son traitement, traiter des données audio reçues depuis la MCU et produire en sortie les données audio traitées; 35 40
- dans lequel l'au moins un terminal interprète est adapté pour interpréter la conversation en fonction des données audio transmises depuis la MCU, traiter les données audio interprétées et envoyer les données audio traitées à la MCU. 45
4. Système selon la revendication 3, dans lequel chacun des terminaux de conférence est un terminal de conférence ayant une seule interface d'entrée et une seule interface de sortie; ou 50
- chacun des terminaux de conférence est un terminal de conférence ayant de multiples interfaces d'entrée et de multiples interfaces de sortie. 55
5. Système selon la revendication 4, dans lequel chacun des terminaux de conférence est un terminal
- de conférence adoptant un codec à canal unique pour effectuer un codage et décodage, ou chacun des terminaux de conférence est un terminal de conférence adoptant un codec à canaux multiples pour effectuer un codage et décodage.
6. Système selon la revendication 3, dans lequel la MCU est une MCU adoptant une règle de mixage sonore pour un correspondant, deux correspondants, trois correspondants, ou plus.
7. Procédé de réalisation d'une conférence multilingue, comprenant:
- la transformation, par chaque terminal de conférence, d'une conversation de la conférence en données audio à envoyer à une unité de commande multipoint, MCU, et l'envoi, par les terminaux de conférence, des données audio à la MCU;
- la réception, par la MCU, de données audio transmises par chacun des terminaux de conférence auxquels la MCU a assigné une langue, au moins l'un des terminaux de conférence étant un terminal interprète, lequel est adapté pour interpréter une paire de langues de conversation des terminaux de conférence;
- le décodage des données audio reçues des terminaux de conférence;
- pour chaque langue l'exécution d'un processus de mixage sonore des données audio décodées des terminaux de conférence de la langue respective et d'au moins un terminal interprète d'une même langue respective; et
- quand la MCU a terminé ledit processus de mixage sonore, afin d'obtenir un résultat de mixage sonore pour chaque langue, l'envoi du résultat de mixage sonore comprenant les données audio mixtes de la langue respective aux terminaux de conférence de la langue assignée correspondante respectivement;
- dans lequel l'au moins un terminal interprète envoie les données audio à la MCU après que le terminal interprète interprète la conversation du terminal de conférence en fonction des données audio transmises par la MCU et traite les données audio interprétées.
8. Procédé selon la revendication 7, dans lequel le terminal de conférence traite les données audio par le biais d'une seule interface d'entrée et d'une seule interface de sortie; ou
- le terminal de conférence traite les données audio par le biais de multiples interfaces d'entrée et de multiples interfaces de sortie.
9. Procédé selon la revendication 8, dans lequel les terminaux de conférence traitent les données

audio en adoptant un codec à canal unique ; ou les terminaux de conférence traitent les données audio en adoptant un codec à canaux multiples.

10. Procédé selon la revendication 8, dans lequel le processus de mixage sonore est : un processus de mixage sonore pour un correspondant, deux correspondants, trois correspondants, ou plus, selon lequel dans le mixage sonore pour un correspondant, seul le terminal de conférence ayant le son le plus haut est diffusé à chacun des autres terminaux de conférence, dans le mixage sonore pour deux correspondants, les deux terminaux de conférence ayant le son le plus haut sont diffusés après avoir été mixés et pour de multiples correspondants des règles de mixage sonore analogues sont employées comme il convient. 5
10
15

11. Procédé selon la revendication 7, dans lequel l'envoi des données audio traitées aux terminaux de conférence correspondants auxquels sont respectivement assignées des langues comprend :
 - l'envoi, par la MCU, des données audio après le processus de mixage sonore d'un canal sonore correspondant aux terminaux de conférence en fonction d'informations de langue correspondant aux terminaux de conférence et d'informations d'identifiant du terminal interprète. 20
25
30

12. Procédé selon la revendication 11, comprenant en outre:
 - le prééglage, par la MCU, des informations de langue correspondant aux terminaux de conférence et/ou des informations d'identifiant du terminal interprète ; et/ou la réception et la mémorisation, par la MCU, des informations de langue et/ou des informations d'identifiant du terminal interprète transmises par les terminaux de conférence. 35
40

13. Procédé selon l'une quelconque des revendications 8 à 12, comprenant en outre: 45
 - la sélection, par la MCU, de sites de conférence en fonction du volume de conversation de chacun des sites de conférence, et l'envoi des données audio des sites de conférence sélectionnés au terminal interprète. 50

14. Procédé selon la revendication 13, dans lequel les sites de conférence sélectionnés comprennent un ou plusieurs sites de conférence, et les données audio transmises par le terminal interprète participent toujours au traitement de mixage sonore dans différents canaux sonores. 55

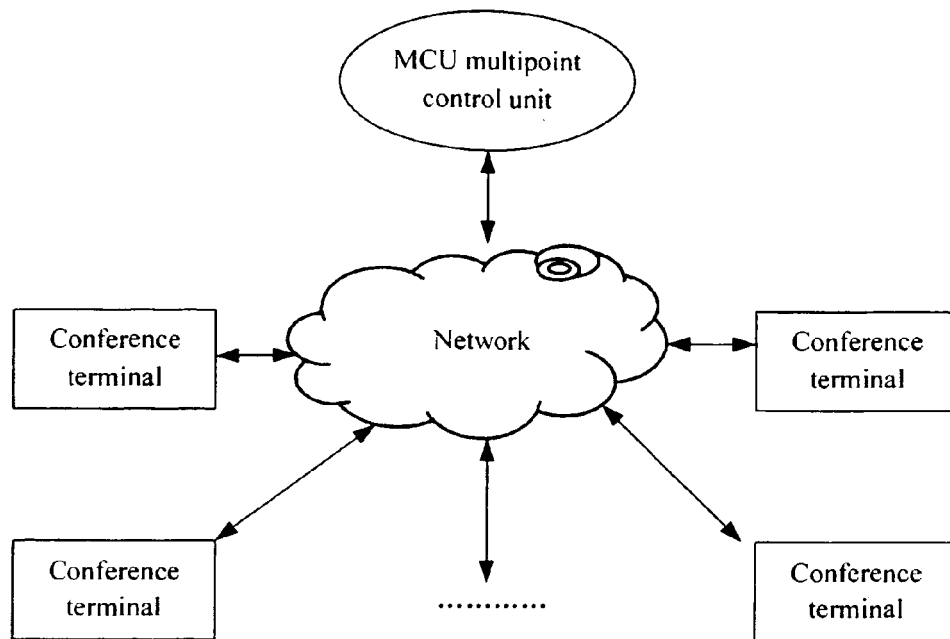


FIG. 1

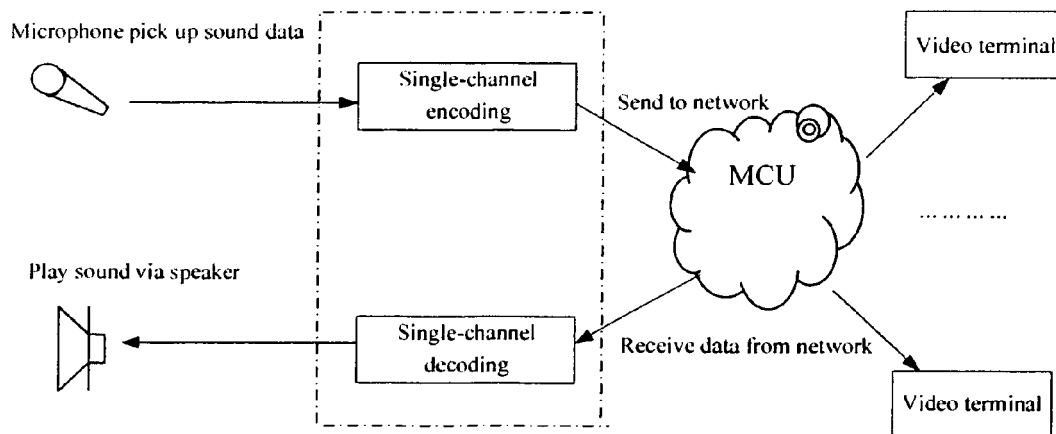


FIG. 2

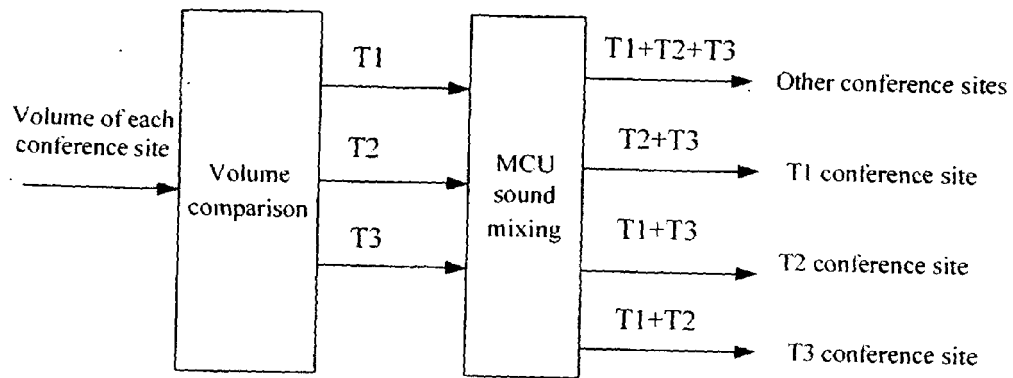


FIG. 3

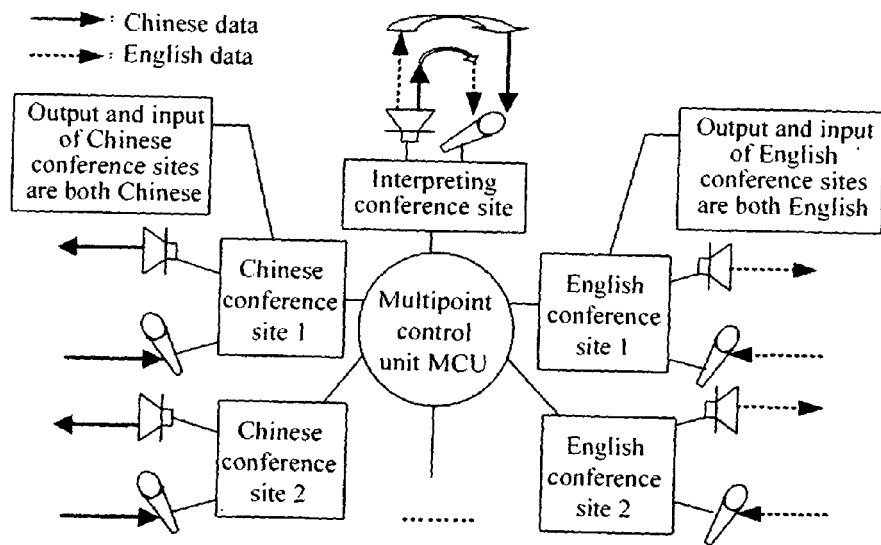


FIG. 4

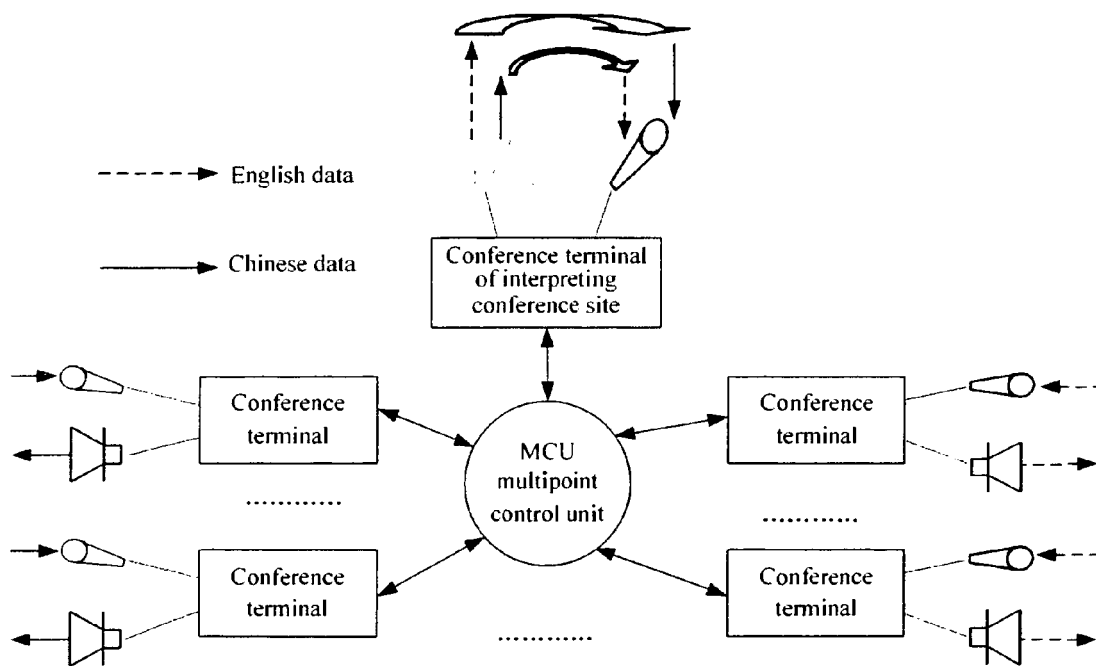


FIG. 5

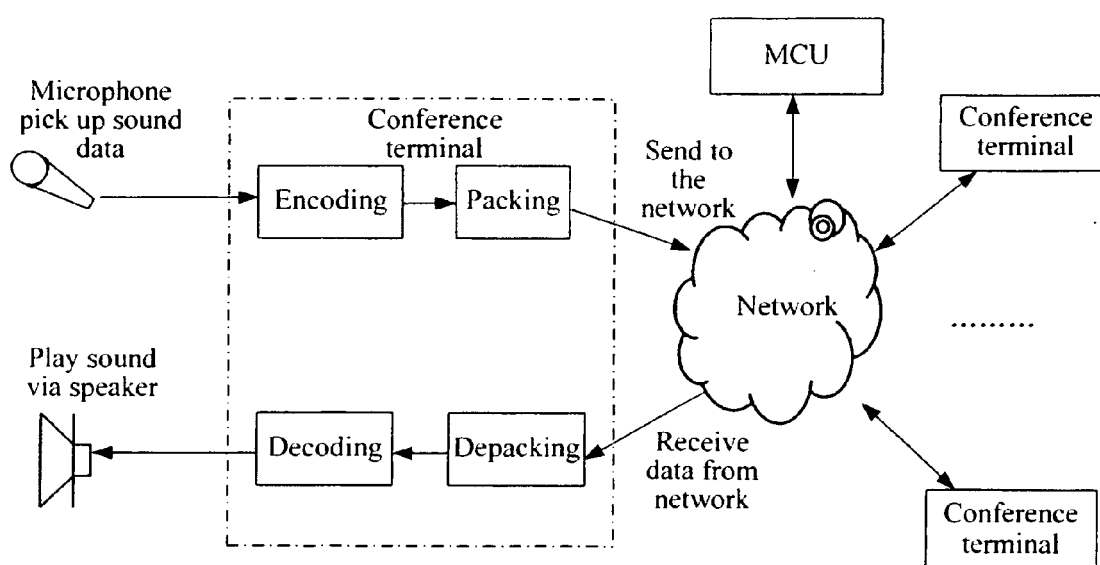


FIG. 6

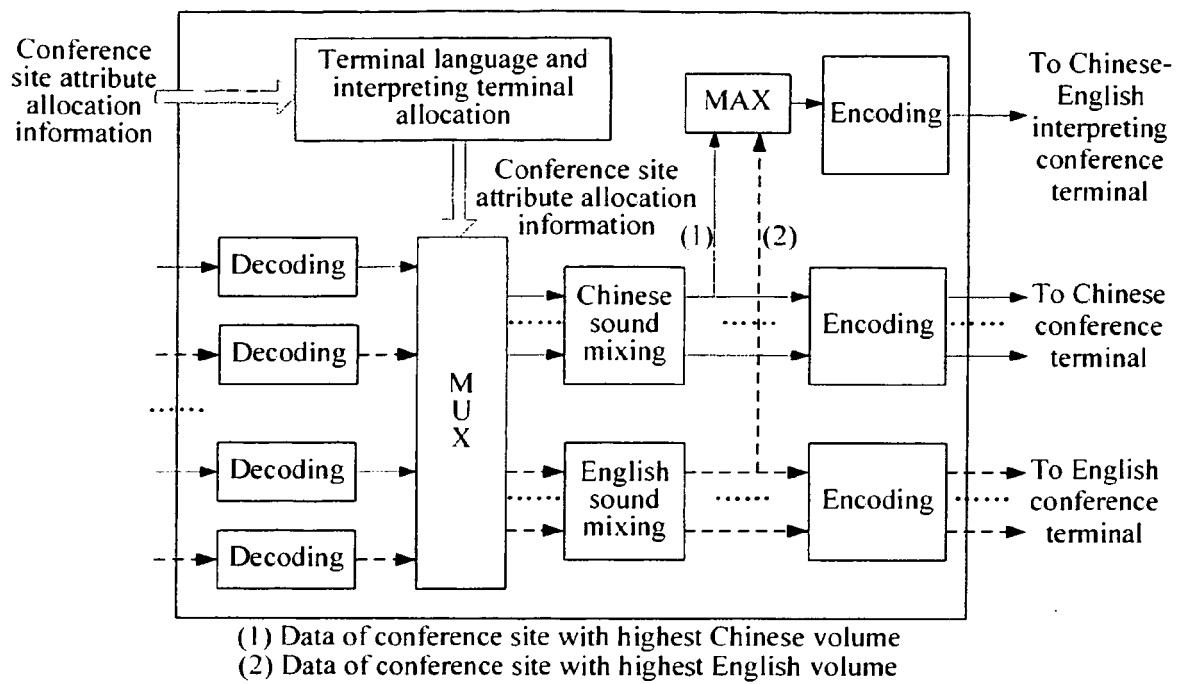


FIG. 7

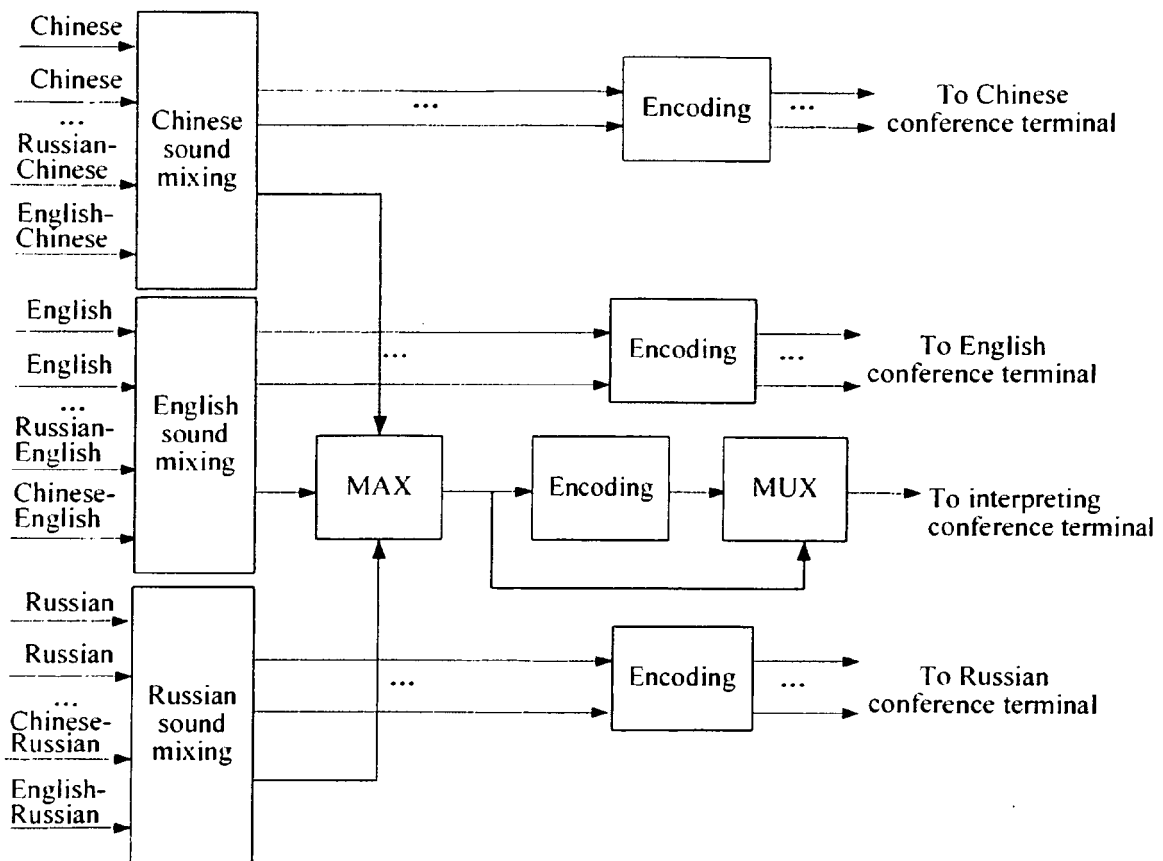


FIG. 8

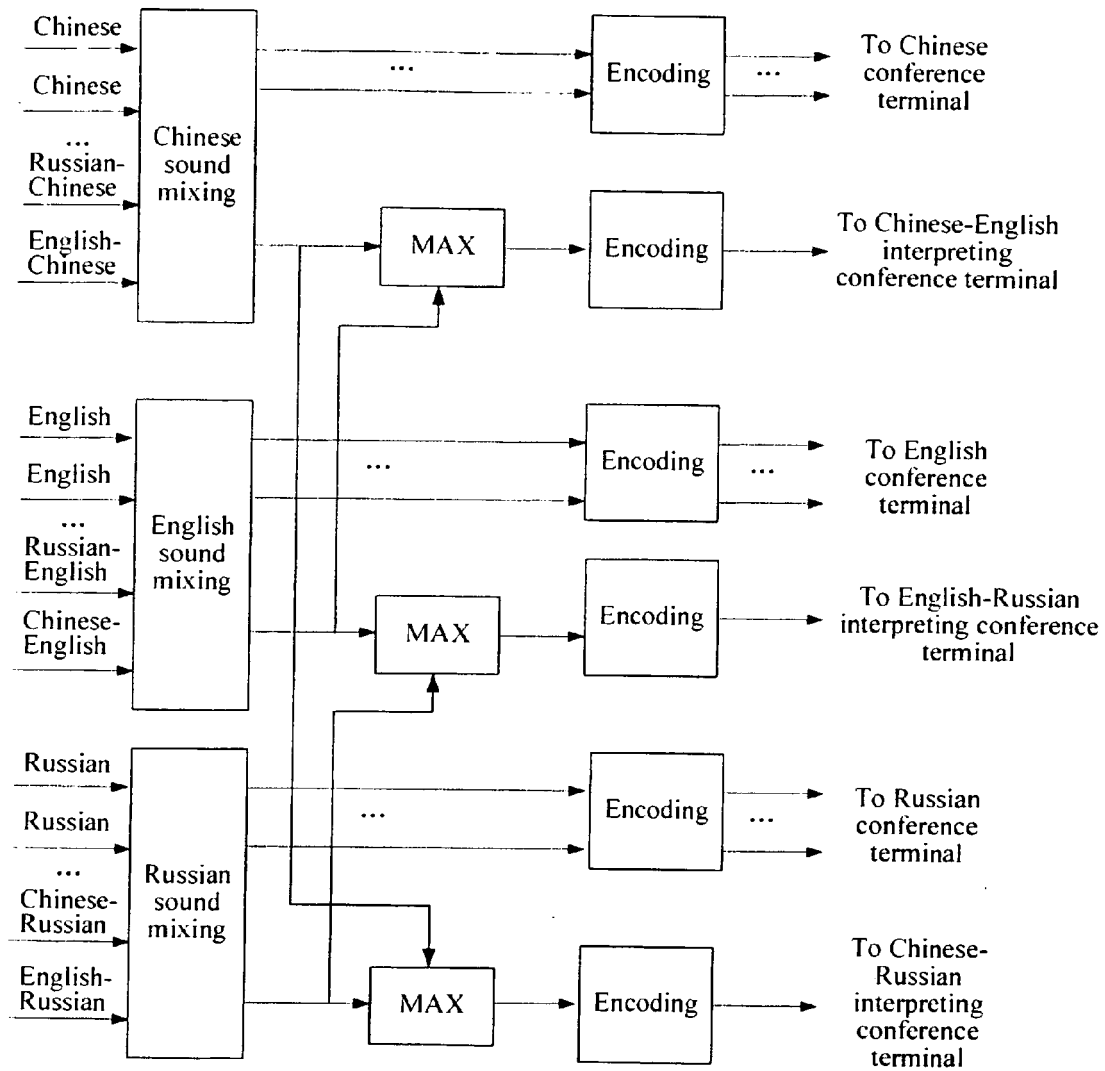


FIG. 9

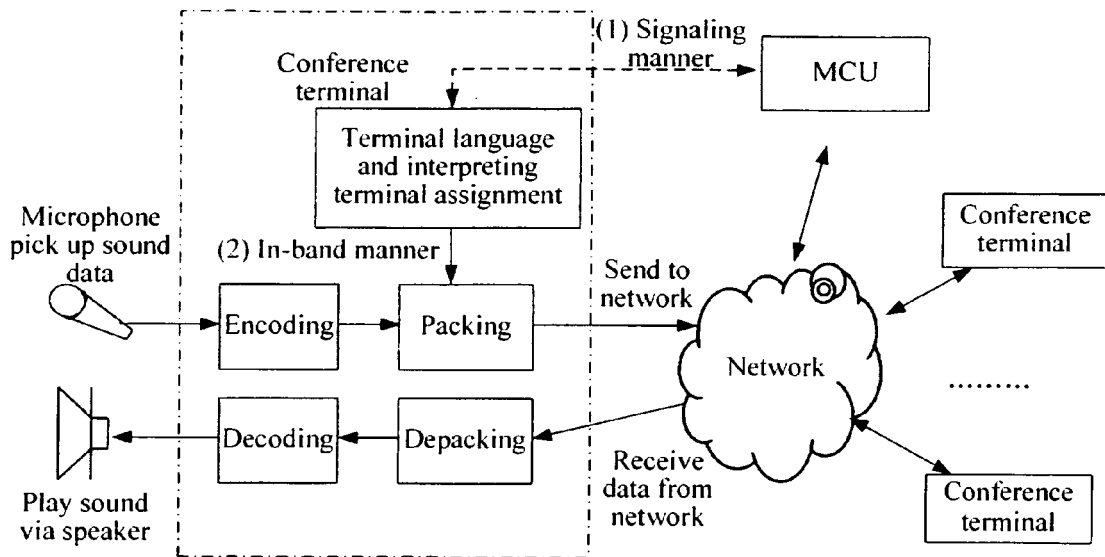
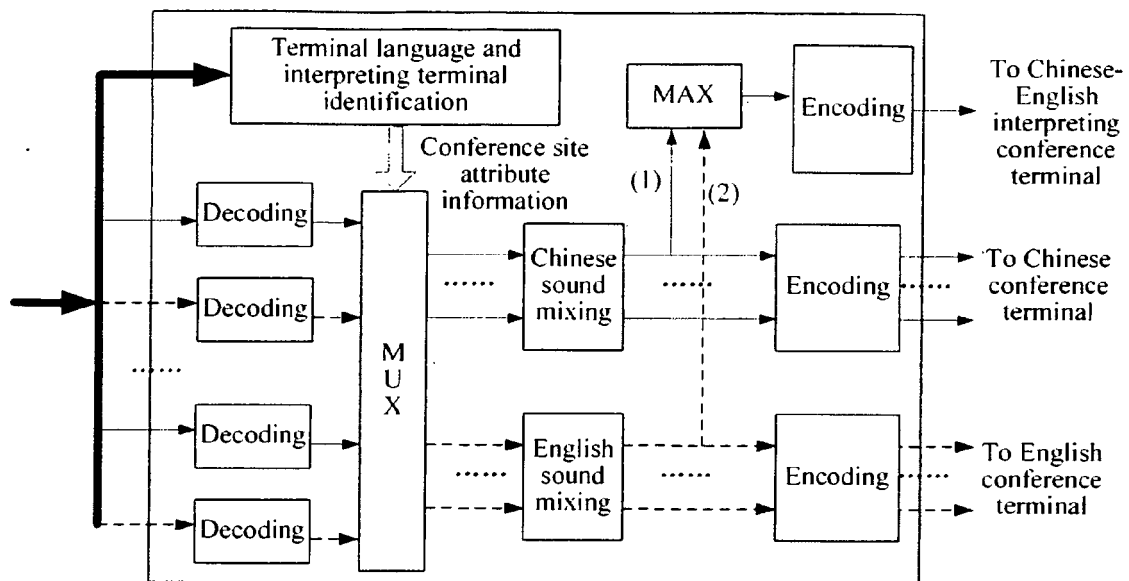


FIG. 10



(1) Data of conference site with highest Chinese volume

(2) Data of conference site with highest English volume

FIG. 11

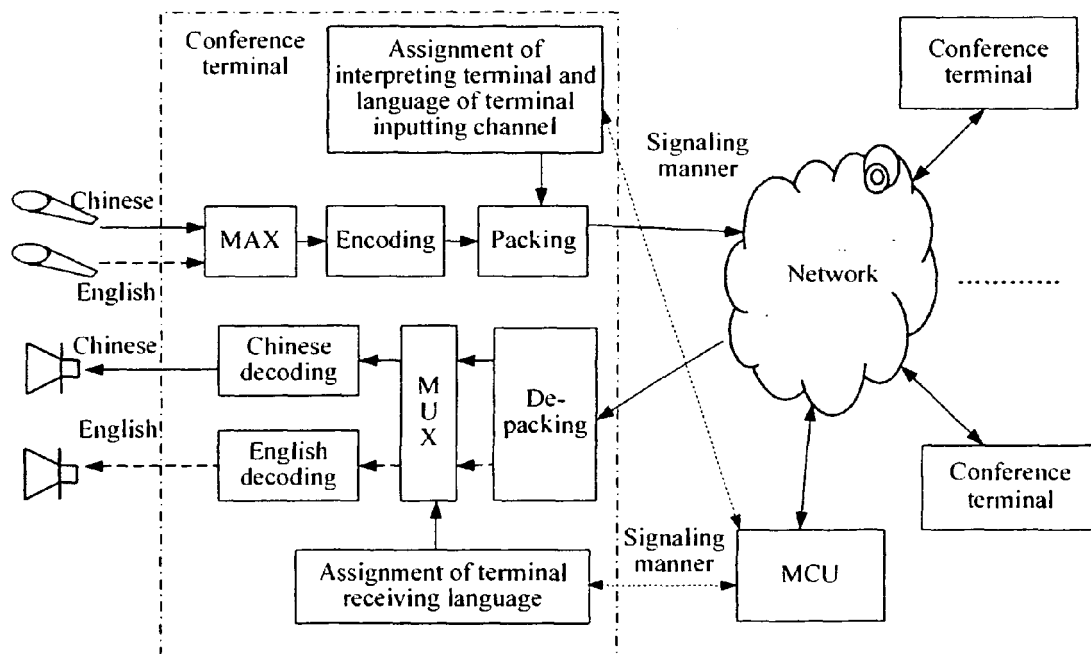


FIG. 12

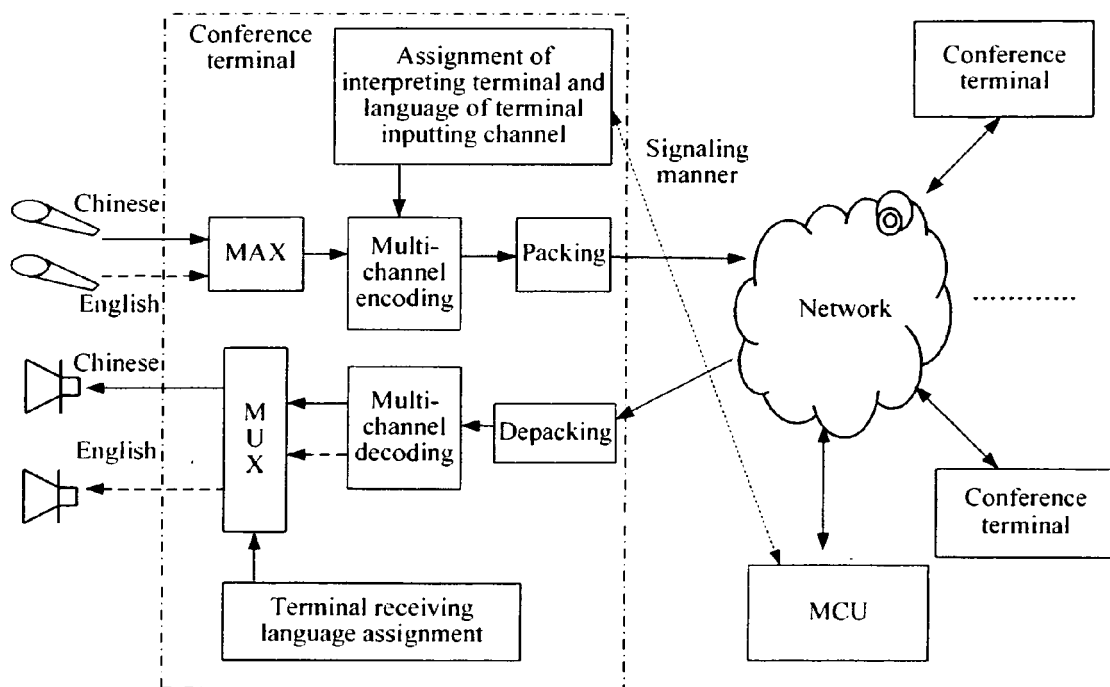


FIG. 13

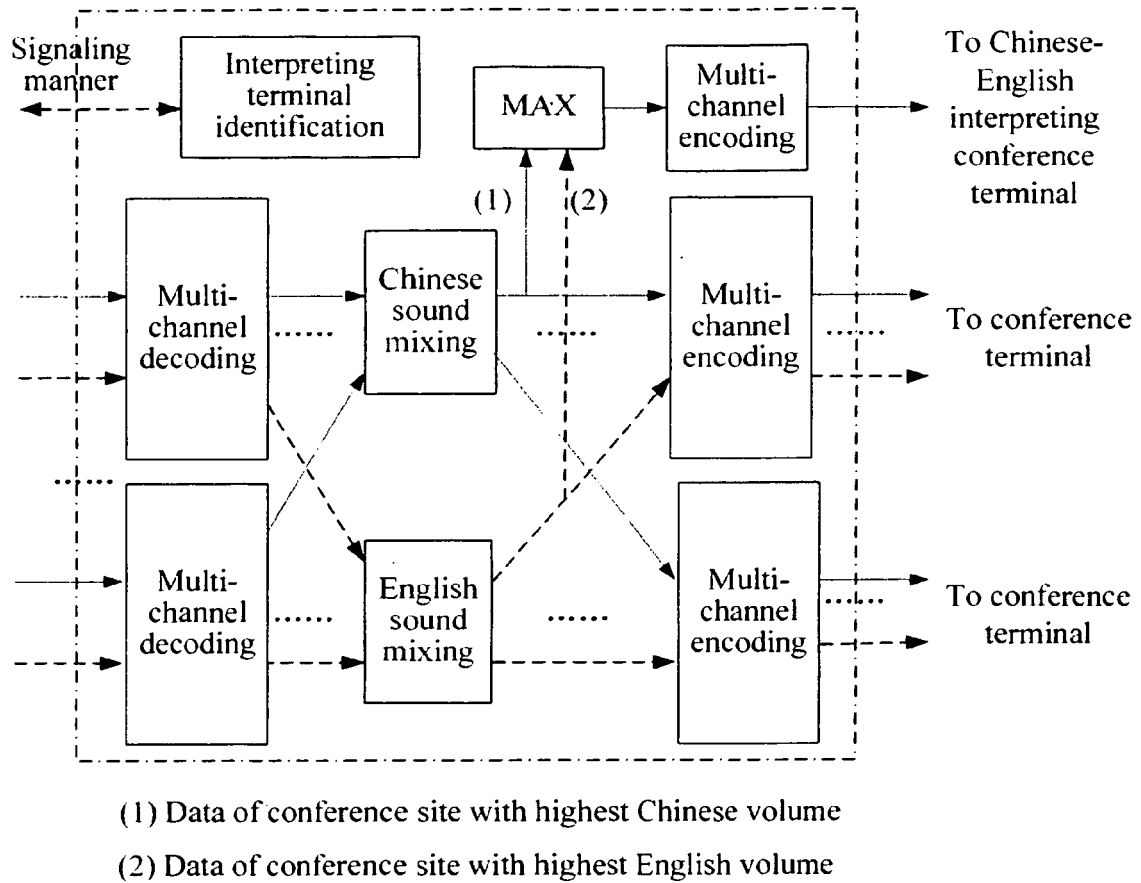


FIG. 14

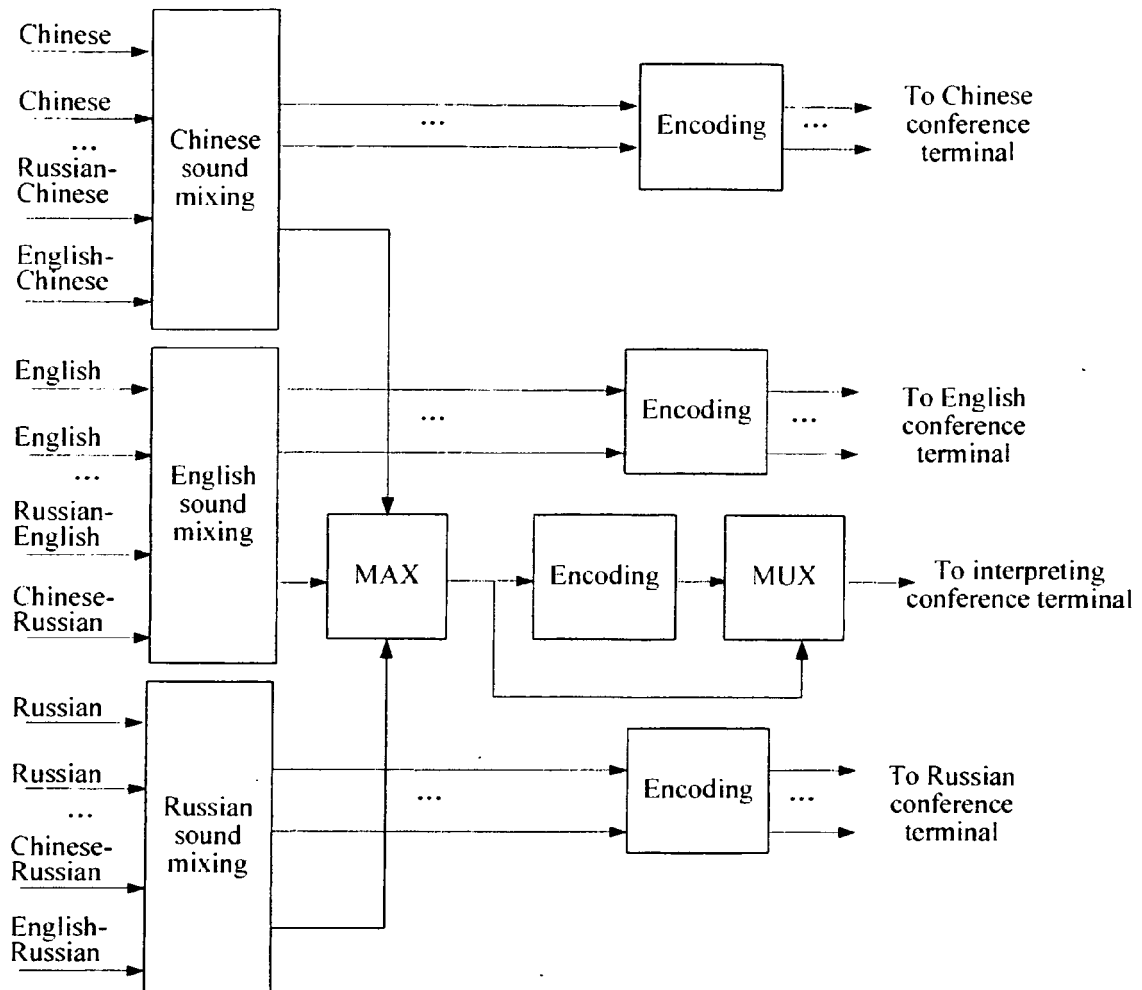


FIG. 15

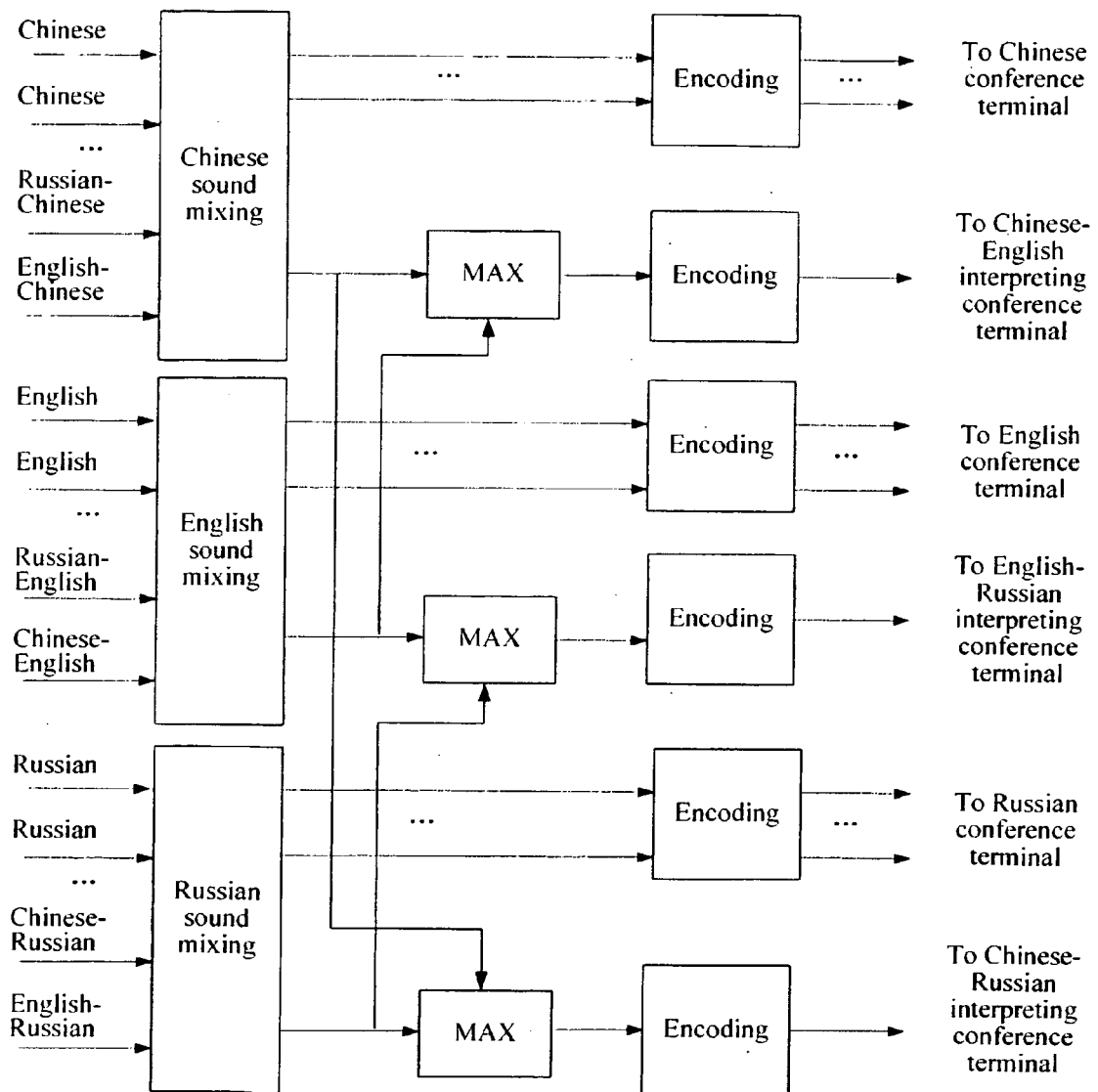


FIG. 16

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 20060120307 A1 [0013]