

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局(43) 国际公布日
2013 年 7 月 11 日 (11.07.2013)

W I P O | P C T

(10) 国际公布号
W O 2013/102403 A 1

- (51) 国际分类号 : G10L 19/20 (2013.01) G10L 19/22 (2013.01)
- (21) 国际申请号 : PCT/CN2012/086953
- (22) 国际申请日 : 2012 年 12 月 19 日 (19.12.2012)
- (25) 中 介 言 : 中文
- (26) 公布语言 : 中文
- (30) 优先权 : 2012 10001235.3 2012 年 1 月 4 日 (04.01.2012) CN
- (71) 申请人 : 中国移动通信集团公司 (CHINA MOBILE COMMUNICATIONS CORPORATION) [CN/CN]; 中国北京市西城区金融大街 29 号 , Beijing 100032 (CN)。
- (72) 发明人 刘霖 (LIU, Lin); 中国北京市西城区金融大街 29 号 , Beijing 100032 (CN)。
- (74) 代理人 : 北京集佳知识产权代理有限公司 (UNITALEN ATTORNEYS AT LAW); 中国北京市朝阳区建国门外大街 22 号赛特广场 7 层 , Beijing 100004 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, ML, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH,

[见续页]

(54) Title: AUDIO SIGNAL PROCESSING METHOD AND DEVICE, AND TERMINAL

(54) 发明名称 : 一种音频信号处理方法、装置及终端

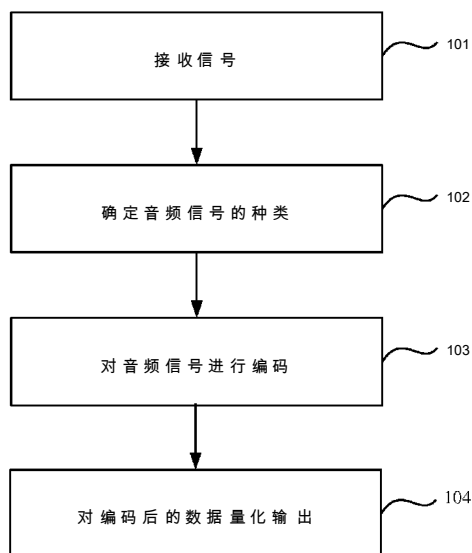


图 1/ Fig. 1

101 RECEIVE A SIGNAL
102 DETERMINE THE TYPE OF THE AUDIO SIGNAL
103 ENCODE THE AUDIO SIGNAL
104 QUANTIFY AND OUTPUT THE ENCODED DATA

(57) Abstract: An audio signal processing method and device, and a terminal. The method comprises: when an audio signal is being encoded, determining the type of the received audio signal according to a received video signal; when it is determined that the received audio signal is a voice signal, encoding the audio signal through time domain encoding; when it is determined that the received audio signal is a non-voice signal, encoding the audio signal through frequency domain encoding. Therefore, the identified voice signal and the identified non-voice signal are encoded separately, and transmission of sound is realized. The solution also provides an audio signal processing device and a terminal.

(57) 摘要: 一种音频信号处理方法、装置及终端, 包括: 在对音频信号进行编码时, 通过接收到的视频信号确定接收到的音频信号的种类, 在确定接收到的音频信号为语音信号时, 利用时域编码的方式对该音频信号进行编码, 在确定接收到的音频信号为非语音信号时, 利用频域编码的方式对该音频信号进行编码, 从而对识别出的语音信号和非语音信号分别进行编码, 并实现声音的传输。本方案还提供一种音频信号处理装置以及终端。

2013/102403 A1



CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, 本国际公布 :
IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,
RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, - 包括国际检索报告(条约第21条(3))。
CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

一种音频信号处理方法、装置及终端

本申请要求于 2012 年 1 月 4 日提交中国专利局、申请号为 201210001235.3、发明名称为“一种音频信号处理方法、装置及终端”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本发明涉及终端领域，尤其涉及一种音频信号处理方法、装置及终端。

背景技术

随着第三代移动通信技术 (3G, 3rd-generation) 的快速发展，可视电话逐步在 3G 网络中得到了较多的应用。在目前的可视电话技术中，低码率音频编
10 码技术是可视电话应用中的一个关键技术。

在低码率音频编码领域，存在 2 个主要的技术路线，一种是时域编码，一种是频域编码。

时域编码是针对音频信号的波形，进行编码。针对时域编码比较典型的有国际电信联盟 (ITU, International Telecommunication Union) G.729、G.723.1
15 和 G.728 等编码标准，这些编码标准广泛采用了码激励线性预测 (CELP, Code Excited Linear Prediction) 技术，从原理上根据人类的发生机理建模，利用人类声门、声道固有的特性，去除音频信号里面的冗余信息，从而在保持较高的音频质量的同时，大幅度的降低了音频编码所需的比特率。

在这类音频编码方法中，最致命的缺陷在于该方法主要适用于人类发声
20 (语音信号)，对于杂乱无章 (包括音乐、噪声以及其他声音) 的音频信号 (非语音信号)，编码效果较差。

频域编码的原理在于，利用人耳对于声音的接受原理，在频域对于音频信号进行编码。重点编码人类关注的频段，而对于被其他频段掩蔽或是人类不易感知的频段，采用粗略量化或是不量化的策略。

25 频域编码的优势在于根据人耳的特性，去除了一定的冗余，因此对各种音频信号的编码效果几乎相当，尤其对于音乐等信号的编码质量要高于时域编码。但是在语音信号上，其编码时并未考虑人类发声机理，无法去除发声冗余，因此编码效果要远低于基于 CELP 技术的时域编码。

现有的可视电话技术中，由于语音信息相对重要，因此通常采用基于 CELP

技术的时域编码。基于时域编码的低码率音频编码可以在很低的码率上为可视电话应用提供较高质量的语音编码质量，确保可视电话中较为清晰、易懂的语音通信能力。但是，可视电话在进行语音通信的同时，经常会伴随其他的声音（非语音），比如通话方要让对方听音乐或是其他声音的情况，此时，采用基于时域编码的低码率音频编码导致编码质量很差，声音失真严重。

发明内容

本发明实施例提供一种音频信号处理方法、装置及终端，用于解决声音传输过程中采用单一编码导致声音传输质量较差的问题。

一种低码率音频编码方法，所述方法包括：

在接收音频信号的同时，接收视频信号；

根据接收到的视频信号，确定所述音频信号为语音信号或非语音信号；

若确定所述音频信号为语音信号，则利用基于时域编码的低码率音频编码对所述音频信号进行编码；

若确定所述音频信号为非语音信号，则利用基于频域编码的低码率音频编码对所述音频信号进行编码。

一种低码率音频编码装置，所述装置包括：

第一接收模块，用于接收音频信号；

第二接收模块，用于接收视频信号；

确定模块，用于根据接收到的视频信号，确定所述音频信号为语音信号或非语音信号；

第一编码模块，用于在确定模块确定所述音频信号为语音信号时，利用基于时域编码的低码率音频编码对所述音频信号进行编码；

第二编码模块，用于在确定模块确定所述音频信号为非语音信号时，利用基于频域编码的低码率音频编码对所述音频信号进行编码。

一种终端，所述终端包括上述低码率音频编码装置。

根据本发明实施例提供的方案，在对音频信号进行编码时，通过接收到的视频信号确定接收到的音频信号的种类，在确定接收到的音频信号为语音信号时，利用时域编码的方式对该音频信号进行编码，在确定接收到的音频信号为非语音信号时，利用频域编码的方式对该音频信号进行编码，从而对识别出的

语音信号和非语音信号分别进行编码，并实现声音的传输。

附图说明

图 1 为本发明实施例一提供的音频信号处理方法的步骤流程图；

图 2 为本发明实施例一提供的码流示意图；

5 图 3 为本发明实施例二提供的音频信号处理装置的结构示意图；

图 4 为本发明实施例三提供的终端的结构示意图。

具体实施方式

本发明实施例中，在可视电话环境下，利用可视电话中的图像捕捉，根据图像的信息，判别音频是无规律音频还是语音，从而指导音频编码。实现在编
10 码码率不变的情况下，提高音频编码质量。

下面结合说明书附图和各实施例对本发明方案进行说明。

实施例一：

本发明实施例一提供一种音频信号处理方法，该方法可以但不限于应用于可视电话音频编码领域，该方法的步骤如图 1 所示，包括：

15 步骤 101、接收信号。

在本步骤中，不仅需要接收音频信号，还需要接收视频信号。因此，本步骤包括：在接收音频信号的同时，接收视频信号。所述视频信号可以是可视电话中配置的摄像头针对设定区域进行拍摄获得的。

步骤 102、确定音频信号的种类。

20 在本步骤中，可以根据接收到的视频信号，确定所述音频信号为语音信号或非语音信号。

在本步骤中，可以确定当前接收到的视频信号（当前视频帧）中是否存在指定的图像，即确定摄像头当前拍摄的设定区域中是否包含指定的图像，具体的，可以根据像素信息，确定当前接收到的视频信号（当前视频帧）中是否存在指定的图像，若该视频信号中存在指定的图像，确定距离该视频信号时间最短的一个已接收的视频信号（上一视频帧）：
25

若该已接收的视频信号中存在指定的图像，在该已接收的视频信号中指定的图像与当前接收到的视频信号中指定的图像的相对位移的绝对值满足设定的阈值时，确定当前接收到的音频信号为语音信号，否则，确定当前接收到的

音频信号为非语音信号。

所述当前接收到的音频信号可以是指在本次确定出音频信号种类的时刻到下次确定出音频信号种类的时刻之间接收到的音频信号。此时，由于在目前技术和设备硬件能力下，采集一帧视频帧的时间非常短，如 20 毫秒，对视频信号的处理速度非常快，且在利用可视电话进行通话过程中，一段音频信号的时间一般较长，因此可以对音频信号开始的一段延迟忽略不计。当然，也可以在利用可视电话进行的一次通话过程中，在利用视频信号初次确定音频信号种类的短时间内，设定该时间内接收到的音频信号的种类为语音信号或非语音信号。

为了利用视频信号确定音频信号的种类，所述指定的图像可以但不限于是嘴唇、喉咙等发声器官。并可以在当前视频帧与上一视频帧中指定的图像的相对位移的绝对值满足设定的阈值时，具体的，所述指定的图像为嘴唇时，可以根据人类发声时，嘴唇面积（上嘴唇和下嘴唇围成的区域所占的面积）会发生变化的特点，判断嘴唇面积变化的绝对值是否满足设定的阈值，如大于第一阈值，确定当前音频信号是人类发出的语音信号，否则，确定当前音频信号不是人类发出的语音信号，属于非语音信号。当然，也可以根据人类发声时，上（下）嘴唇会发生上下移动的特点，判断上（或下）嘴唇移动的位移的绝对值是否满足设定的阈值，如是否大于第二阈值，并在判断上（或下）嘴唇移动的位移的绝对值满足设定的阈值时，确定当前音频信号是人类发出的语音信号，否则，确定当前音频信号不是人类发出的语音信号，属于非语音信号。

进一步的，若确定当前接收到的视频信号中不存在指定的图像，可以确定当前接收到的音频信号为非语音信号。若确定当前接收到的视频信号中存在指定的图像，且所述已接收的视频信号中不存在指定的图像，确定当前接收到的音频信号为语音信号。

当然，除了可以结合上一视频帧和当前视频帧来确定当前接收到的音频信号的种类，也可以仅根据当前接收到的视频信号来确定当前接收到的音频信号的种类，具体的，可以确定当前接收到的视频信号中是否存在指定的图像，若不存在，确定当前接收到的音频信号为非语音信号，否则，确定当前接收到的音频信号为语音信号。

可以采用现有的图像识别方法从视频帧中识别指定的图像。例如，在识别嘴唇时，可以根据嘴唇在色彩上与通话者皮肤及其他器官存在较大差异，在采集到的视频帧中，嘴唇图像像素中的红色分量（R 分量）与绿色分量（G 分量）的差异与其他区块有明显的不同的特点，利用 R 分量与 G 分量的差异作为从

5 视频帧中识别嘴唇图像的方法。

具体的，可以通过如下公式（1）实现嘴唇图像的识别：

$$h(x, y) = \frac{R(x, y)}{G(x, y) + R(x, y)} \quad (1)$$

其中， $R(x, y)$ 表示在像素点 (x, y) 上的 R 分量值， $G(x, y)$ 表示在像素点 (x, y) 上的 G 分量值。 $h(x, y)$ 表示像素点 (x, y) 上的红、绿分量的差异。

10 可以利用 $h(x, y)$ 分量对图像进行二值化，二值化的门限值可以根据多人训练得到（可以以不同肤色，不同性别，不同年龄的人）二值化的最佳门限值。对二值化后的像素信息进行整理，去除零散的噪声点即可以得到嘴唇的估计区域（上嘴唇和下嘴唇围成的区域），实现对嘴唇图像的识别。

且进一步的，可以通过以下方法确定当前视频帧与上一视频帧中指定的图

15 像的相对位移：

若在当前视频帧搜索到嘴唇区域（嘴唇图像）后，根据该区域的坐标点，裁切出该区域对应的二值化点阵，设嘴唇区域对应的二值化点阵用 P 表示，该点阵的面积可以用 $A(P)$ 表示。对于点阵 P 中任意一个像素点 (x, y) ，在上一视频帧二值化像素值为 $h'(x, y)$ ，在当前视频帧的二值化像素值为 $h(x, y)$ ，可以通过如下公式（2）计算上一视频帧和当前视频帧中嘴唇区域的差别，用 D 表示：

20

$$D = \frac{\sum_p (h(x, y) - h'(x, y))^2}{A(P)} \quad (2)$$

并可以在确定 D 满足设定的阈值时，确定当前音频信号是人类发出的语音信号，否则，确定当前音频信号不是人类发出的语音信号，属于非语音信号。

步骤 103、对音频信号进行编码。

25 在确定所述音频信号为语音信号时，利用基于时域编码的低码率音频编码对所述音频信号进行编码，具体的，可以采用现有的编码方式，如根据 ITU G.729/728/723.1，3GPP AMR-NB/WB 或是其他基于 CELP 技术的编码方式进

行编码，否则，在确定所述音频信号为非语音信号时，利用基于频域编码的低码率音频编码对所述音频信号进行编码，具体的，可以采用现有的编码方式，如使用感知权重，在快速傅里叶变换（FFT，Fast Fourier Transform）域进行格型矢量量化的编码方式。

5 步骤 104、对编码后的数据量化输出。

在对音频信号进行编码后，可以对编码后获得的数据进行量化，组织码流并输出。且可以在码流头设置标识位，对采用时域编码获得的码流和对采用频域编码获得的码流进行区分，用于后续的解码操作。具体的，如图 2 所示为带有标识位的码流，在对语音信号采用 CELP 编码（基于 CELP 技术的编码方式），
10 对非语音信号采用变换域编码（基于频域编码的编码方式）时，在编码完成后，可以在码流头设置一个标识位，该标识位为 0，标识该码流是 CELP 码流（语音码流），该标识位为 1，标识该码流是变换域编码码流（非语音码流）。

在解码端，可以根据标识位，选择使用变换域解码器还是 CELP 解码器，从而得到正确的解码码流。

15 与本发明实施例一基于同一发明构思，提供以下的装置和终端。

实施例二：

本发明实施例二提供一种音频信号处理装置，该装置可以但不限于应用于可视电话音频编码领域，该装置的结构如图 3 所示，包括：

第一接收模块 11 用于接收音频信号；第二接收模块 12 用于接收视频信号；
20 确定模块 13 用于根据接收到的视频信号，确定所述音频信号为语音信号或非语音信号；第一编码模块 14 用于在确定模块确定所述音频信号为语音信号时，利用基于时域编码的低码率音频编码对所述音频信号进行编码；第二编码模块 15 用于在确定模块确定所述音频信号为非语音信号时，利用基于频域编码的低码率音频编码对所述音频信号进行编码。

25 所述确定模块 13 具体用于确定当前接收到的视频信号中是否存在指定的图像，若该视频信号中存在指定的图像，确定距离该视频信号时间最短的一个已接收的视频信号；若该已接收的视频信号中存在指定的图像，在该已接收的视频信号中指定的图像与当前接收到的视频信号中指定的图像的相对位移的绝对值满足设定的阈值时，确定当前接收到的音频信号为语音信号，否则，确

定当前接收到的音频信号为非语音信号。

所述确定模块 13 还用于在确定当前接收到的视频信号中不存在指定的图像时，确定当前接收到的音频信号为非语音信号；以及，在确定当前接收到的视频信号中存在指定的图像，且所述已接收的视频信号中不存在指定的图像时，确定当前接收到的音频信号为语音信号。

所述确定模块 13 具体用于确定当前接收到的视频信号中是否存在指定的图像，若不存在，确定当前接收到的音频信号为非语音信号，否则，确定当前接收到的音频信号为语音信号。

所述装置还包括：

码流输出模块 16 用于对编码后获得的数据进行量化，并组织码流输出，所述码流中包括标识位，用于标识该码流对应的数据的编码方式。如，可以将标识位设置为 0，标识该码流为采用时域编码获得的码流，将标识位设置为 1，标识该码流为采用频域编码获得的码流。

实施例三、

本发明实施例三提供一种终端，该终端的结构可以如图 4 所示，该终端中可以集成有本发明实施例二提供的装置，且所述终端中还可以包括进一步包括视频信号采集模块 21 和音频信号采集模块 22：

视频信号采集模块 21 用于向所述第二接收模块提供视频信号；

音频信号采集模块 22 用于向所述第一接收模块提供音频信号。

所述终端还可以包括音频信号输出模块 23 用于输出编码后的音频信号。当然，所述终端还可以进一步包括视频信号输出模块 24 用于输出视频信号。即所述终端可以仅传输编码后的音频信号，也可以在传输编码后的音频信号的同时，传输视频信号。

具体的，本发明实施例二提供的装置可以集成在可视电话中，该装置可以独立于可视电话的摄像头，且该装置的第二接收模块可以利用摄像头（可以作为视频信号采集模块）采集的视频信号来确定音频信号的种类。当然，可视电话的摄像头也可以作为第二接收模块集成在该装置中，用于采集视频信号来确定音频信号的种类。

根据本发明实施例一至实施例三提供的方案，可以通过视频信号来确定音

频信号的种类，从而确定对音频信号的编码方法，提高音频编码质量，减少声音失真。

显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

5

权 利 要 求

1、一种音频信号处理方法，其特征在于，所述方法包括：

在接收音频信号的同时，接收视频信号；

根据接收到的视频信号，确定所述音频信号为语音信号或非语音信号；

5 若确定所述音频信号为语音信号，则利用基于时域编码的低码率音频编码对所述音频信号进行编码；

若确定所述音频信号为非语音信号，则利用基于频域编码的低码率音频编码对所述音频信号进行编码。

2、如权利要求 1 所述的方法，其特征在于，根据接收到的视频信号，确定所述音频信号为语音信号或非语音信号，具体包括：

确定当前接收到的视频信号中是否存在指定的图像，若该视频信号中存在指定的图像，则确定距离该视频信号时间最短的一个已接收的视频信号；

15 若该已接收的视频信号中存在指定的图像，且在该已接收的视频信号中指定的图像与当前接收到的视频信号中指定的图像的相对位移的绝对值满足设定的阈值，则确定当前接收到的音频信号为语音信号，否则，确定当前接收到的音频信号为非语音信号。

3、如权利要求 2 所述的方法，其特征在于，所述方法还包括：

若确定当前接收到的视频信号中不存在指定的图像，则确定当前接收到的音频信号为非语音信号；

20 若确定当前接收到的视频信号中存在指定的图像，且所述已接收的视频信号中不存在指定的图像，则确定当前接收到的音频信号为语音信号。

4、如权利要求 1 所述的方法，其特征在于，根据接收到的视频信号，确定所述音频信号为语音信号或非语音信号，具体包括：

25 确定当前接收到的视频信号中是否存在指定的图像，若不存在，则确定当前接收到的音频信号为非语音信号，若存在，则确定当前接收到的音频信号为语音信号。

5、如权利要求 1 至 4 中任一项所述的方法，其特征在于，利用基于时域编码的低码率音频编码对所述音频信号进行编码之后，或利用基于频域编码的低码率音频编码对音频信号进行编码之后，所述方法还包括：

对编码后获得的数据进行量化,并组织码流输出,所述码流中包括标识位,用于标识该码流对应的数据的编码方式。

6、一种音频信号处理装置,其特征在于,所述装置包括:

第一接收模块,用于接收音频信号;

5 第二接收模块,用于接收视频信号;

确定模块,用于根据接收到的视频信号,确定所述音频信号为语音信号或非语音信号;

第一编码模块,用于在确定模块确定所述音频信号为语音信号时,利用基于时域编码的低码率音频编码对所述音频信号进行编码;

10 第二编码模块,用于在确定模块确定所述音频信号为非语音信号时,利用基于频域编码的低码率音频编码对所述音频信号进行编码。

7、如权利要求6所述的装置,其特征在于,

15 所述确定模块,具体用于确定当前接收到的视频信号中是否存在指定的图像,若该视频信号中存在指定的图像,确定距离该视频信号时间最短的一个已接收的视频信号:

若该已接收的视频信号中存在指定的图像,且在该已接收的视频信号中指定的图像与当前接收到的视频信号中指定的图像的相对位移的绝对值满足设定的阈值,则确定当前接收到的音频信号为语音信号,否则,确定当前接收到的音频信号为非语音信号。

20 8、如权利要求7所述的装置,其特征在于,

所述确定模块,还用于在确定当前接收到的视频信号中不存在指定的图像时,确定当前接收到的音频信号为非语音信号;以及,在确定当前接收到的视频信号中存在指定的图像,且所述已接收的视频信号中不存在指定的图像时,确定当前接收到的音频信号为语音信号。

25 9、如权利要求6所述的装置,其特征在于,

所述确定模块,具体用于确定当前接收到的视频信号中是否存在指定的图像,若不存在,则确定当前接收到的音频信号为非语音信号,若存在,则确定当前接收到的音频信号为语音信号。

10、如权利要求6所述的装置,其特征在于,所述装置还包括:

码流输出模块，用于对编码后获得的数据进行量化，并组织码流输出，所述码流中包括标识位，用于标识该码流对应的数据的编码方式。

11、一种终端，其特征在于，所述终端包括如权利要求 6 至 10 中任一项所述的音频信号处理装置。

5 12、如权利要求 11 所述的终端，其特征在于，所述终端还包括视频信号采集模块和音频信号采集模块：

 视频信号采集模块，用于向所述第二接收模块提供视频信号；

 音频信号采集模块，用于向所述第一接收模块提供音频信号。

10 13、如权利要求 11 所述的终端，其特征在于，所述终端还包括音频信号输出模块，用于输出编码后的音频信号。

 14、如权利要求 13 所述的终端，其特征在于，所述终端还包括视频信号输出模块，用于输出视频信号。

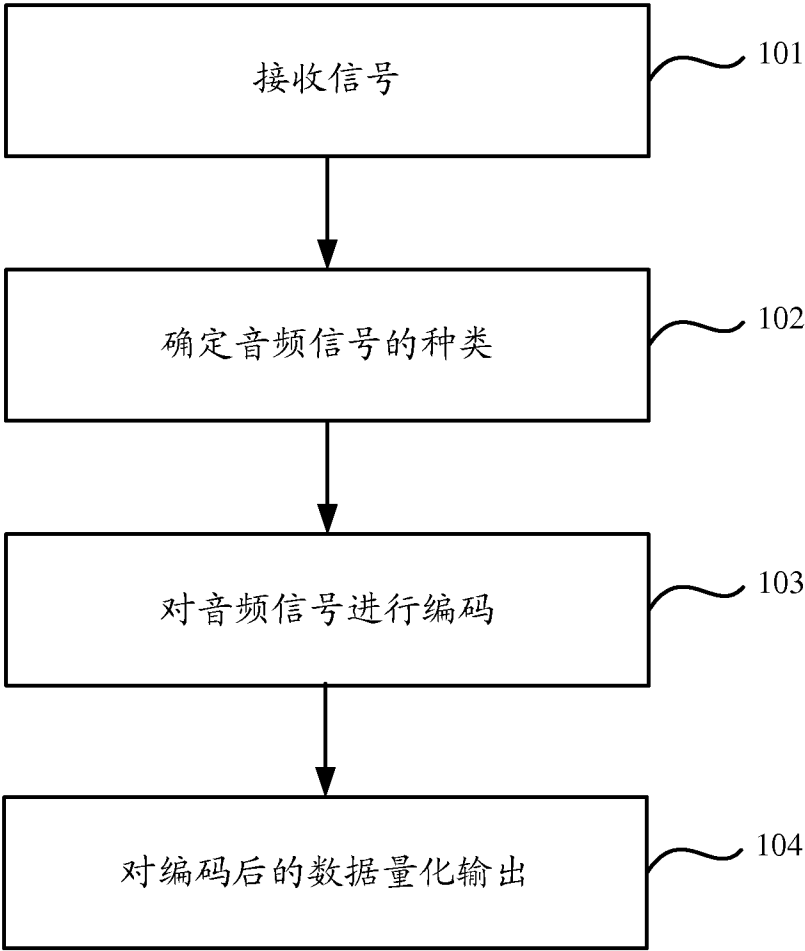


图 1

0	CELP编码码流
1	变换域编码码流

图 2

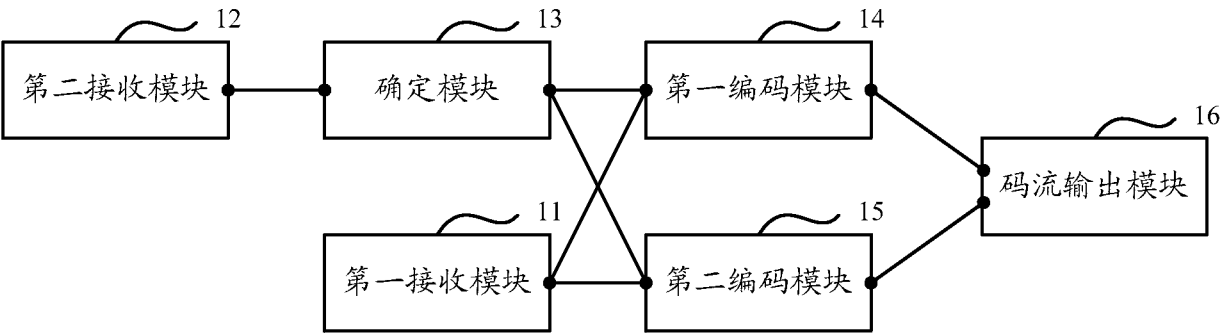


图 3

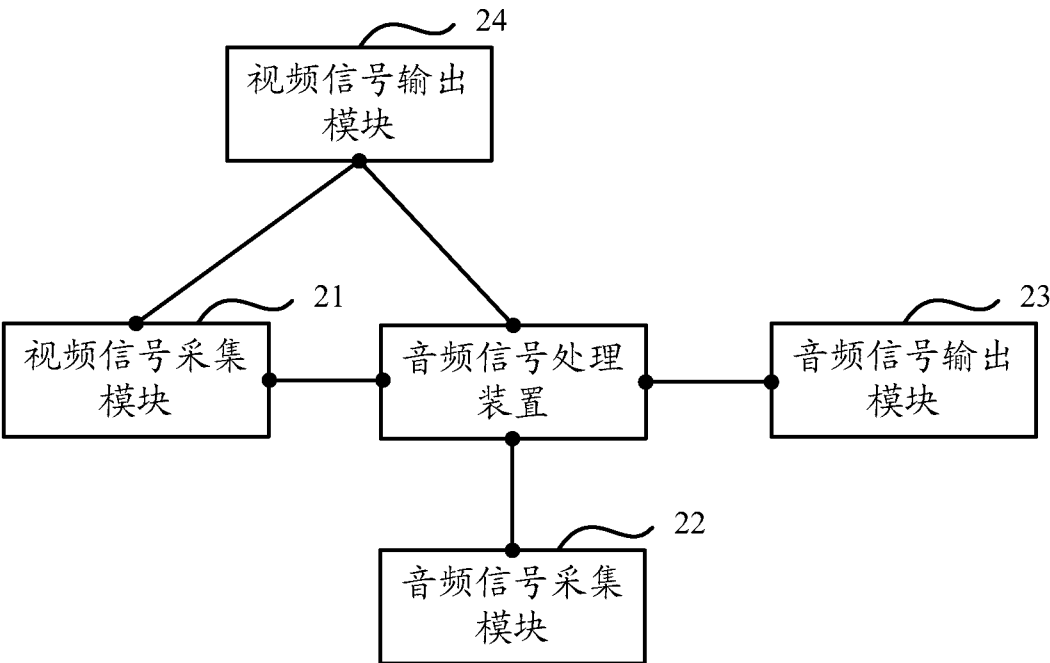


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2012/086953

A. CLASSIFICATION OF SUBJECT MATTER

See the extra sheet

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC; EC: G10L

UC: 704/205, 704/208, 704/211, 704/21,

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS, CNTXT, CNABS, CNKI, VEN, USTXT, WOTXT, EPTXT: speech or voice?, non?speech or un?speech or non?voice? or un?voice?, (time w domain) or celp, (frequency w domain) or (transform- t w domain) or ate, (hybrid or mixed or multi? mode?) 3w (coding or coder? or vocoder? or vocoding), video? or image? or picture?, mouth or lip or lips or throat

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2003009325 A I (DEUT TELEKOM AG. et al.) 09 January 2003 (09.01 .2003) see description, paragraphs [0004] to [0049] and figures 1 and 2	1-14
Y	CN 101656070 A (SPREADTRUM COMMUNICATIONS SHANGHAI CO., LTD.) 24 February 2010 (24.02.2010) see the abstract	1-14

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 12 March 2013 (12.03.2013)	Date of mailing of the international search report 28 March 2013 (28.03.2013)
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer YU, Xin Telephone No. (86-10) 62085168

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2012/086953

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6754373 B1 (INT BUSINESS MACHINES CORP.) 22 June 2004 (22.06.2004) see description, column 3, line 23 to column 4, line 49 and figure 1	3-4, 7-8
A	US 2007174051 A1 (SAMSUNG ELECTRONICS CO., LTD.) 26 July 2007 (26.07.2007) see the whole document	1-14
A	CN 101615393 A (DEUT THOMSON OHG.) 30 December 2009 (30.12.2009) see the whole document	1-14
A	US 2004267521 A1 (MICROSOFT CORP. et al.) 30 December 2004 (30.12.2004) see the whole document	1-14

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2012/086953

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
U S 2003009325 A I	09.01.2003	ES 2247741 T 3	01.03.2006
		DE 69926821 D I	29.09.2005
		DE 69926821 T 2	06.12.2007
		EP 0932141 A 2	28.07.1999
		EP 0932141 B I	24.08.2005
CN 101656070 A	24.02.2010	CN 101656070 B	04.01.2012
U S 6754373 B I	22.06.2004	None	
U S 2007174051 A I	26.07.2007	EP 1982329 A I	22.10.2008
		JP 2009524846 T	02.07.2009
		W O 2007086646 A I	02.08.2007
		K R 20070077652 A	27.07.2007
CN 101615393 A	30.12.2009	EP 2139000 B I	25.05.2011
		EP 2139000 A I	30.12.2009
U S 2004267521 A I	30.12.2004	U S 7343289 B 2	11.03.2008

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2012/086953

Continued: CLASSIFICATION OF SUBJECT MATTER

GIOL 19/20 (2013.01) i

G10L 19/22 (2013.01) i

A. 主题的分类

参见附加页

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC, EC: G10L

UC: 704/205, 704/208, 704/21 1, 704/21,

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称,和使用的检索词(如使用))

CPRSABS, CNTXT, CNABS, CNKI: 语音, 话音, 非语音, 非话音, 音乐, 编码, 压缩, 时域, 频域, 混合, 单一, 视频, 图像, 影像, 图象, 影象, 嘴唇, 嘴巴, 嘴, 唇, 喉咙, 喉

VEN, USTXT, WOTXT, EPTXT: speech or voice?, non?speech or un?speech or non?voice? or un?voice?, (time w domain) or celp, (frequency W domain) or (transform+ w domain) or ate, (hybrid or mixed or multi? mode?) 3W (coding or coder? or vocoder? or vocoding), video? or image? or picture?, mouth or lip or lips or throat

C. 相关文件

类 型 *	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	US2003009325 A I (DEUT TELEKOM AG. 等) 09. 1 月 2003 (09.01.2003) 说明书第 0004-0049 段及图 1-2	1-14
Y	CN101656070 A (展讯通信(上海)有限公司) 24.2 月 2010 (24.02.2010) 摘要	1-14
Y	US6754373 B I (INT BUSINESS MACHINES CORP.) 22.6 月 2004 (22.06.2004) 说明书第 3 栏第 23 行-第 4 栏第 49 行及附图 1	3-4, 7-8
A	US2007 174051 A I (SAMSUNG ELECTRONICS CO., LTD.) 26.7 月 2007 (26.07.2007) 全文	1-14

☒ 其余文件在 C 栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期
12.3 月 2013 (12.03.2013)国际检索报告邮寄日期
28.3 月 2013 (28.03.2013)ISA/CN 的名称和邮寄地址:
中华人民共和国国家知识产权局
中国北京市海淀区蓟门桥西土城路 6 号 100088
传真号: (86-10)62019451受权官员
喻新
电话号码: (86-10) 62085 168

C (续). 相关文件		
类 型	引用文件，必要时，指明相关段落	相关的权利要求
A	CN101615393 A (汤姆森许可贸易公司) 30.12 月 2009 (30.12.2009) 全文	1-14
A	US2004267521 A I (MICROSOFT CORP. 等) 30.12 月 2004 (30.12.2004) 全文	1-14

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2012/086953

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
US2003009325 A 1	09.01.2003	ES2247741 T 3	01.03.2006
		DE69926821 D 1	29.09.2005
		DE69926821 T 2	06. 12.2007
		EP0932141 A 2	28.07. 1999
		EP0932141 B 1	24.08.2005
CN101656070 A	24.02.2010	CN101656070 B	04.01.2012
US6754373 B 1	22.06.2004	无	
US2007 17405 1 A I	26.07.2007	EP1982329 A 1	22. 10.2008
		JP2009524846 T	02.07.2009
		WO2007086646 A 1	02.08.2007
		KR20070077652 A	27.07.2007
CN101615393 A	30. 12.2009	EP2139000 B 1	25.05.2011
		EP2139000 A I	30. 12.2009
US2004267521 A I	30. 12.2004	US7343289 B 2	11.03.2008

续：主题的分类

G10L 19/20(2013.01) i

G10L 19/22(2013.01) i