

(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
2010 年 12 月 23 日 (23.12.2010)

PCT

(10) 国际公布号
WO 2010/145335 A1

- (51) 国际专利分类号:
H04N 7/14 (2006.01) *H04L 29/06* (2006.01)
- (21) 国际申请号: PCT/CN2010/072248
- (22) 国际申请日: 2010 年 4 月 27 日 (27.04.2010)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
200910208719.3 2009 年 10 月 27 日 (27.10.2009) CN
- (71) 申请人 (对除美国外的所有指定国): **中兴通讯股份有限公司 (ZTE CORPORATION)** [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。
- (72) 发明人: 及
- (75) 发明人/申请人 (仅对美国): **廖芯 (LIAO, Xin)** [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。
李满海 (LI, Manhai) [CN/CN]; 中国广东省深圳市

南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。
周龙江 (ZHOU, Longjiang) [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。
沈玉娜 (Shen, Yuna) [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。

(74) 代理人: **北京康信知识产权代理有限公司 (KANGXIN PARTNERS, P.C.)**; 中国北京市海淀区知春路甲 48 号盈都大厦 A 座 16 层, Beijing 100098 (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, 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, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV,

[见续页]

(54) Title: SYSTEM AND METHOD FOR GENERATING INTERACTIVE VOICE AND VIDEO RESPONSE MENU

(54) 发明名称: IVVR 菜单生成系统和方法

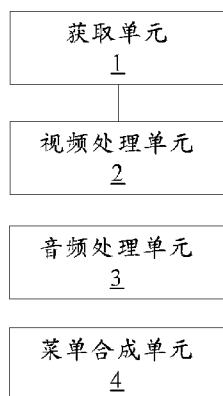


图 2 / Fig. 2

1 OBTAINING UNIT
2 VIDEO PROCESSING UNIT
3 AUDIO PROCESSING UNIT
4 MENU COMBINATION UNIT

(57) Abstract: A system and method for generating interactive voice and video response menu are provided by present invention. The system comprises: an obtaining unit, used for obtaining the files uploaded by the user side, wherein the files comprise: the first files and the second files, the first files comprise video files and/ or image files, and the second files comprise: audio files and/ or text files; a video processing unit, used for converting the first files uploaded by the user side into the first preset format video stream, an audio processing unit, used for converting the second files uploaded by the user side into the second preset format audio stream; a menu combination unit, used for combining the video stream and the audio stream into IVVR (Interactive Voice and Video Response) menu. According to the technical solution provided by the invention, the problems in related technology, such as making the IVVR menu offline, low flexibility, large limitation, possibility of not meeting the requirement of IVVR system are solved, therefore the flexibility is improved, the limitation is reduced and user making is convenient.

(57) 摘要: 本发明公开了一种交互式语音及视频应答菜单生成系统和方法, 上述系统包括: 获取单元, 用于获取用户上传的文件, 其中, 该文件包括: 第一文件和第二文件, 第一文件包括: 视频文件和/或图片文件, 第二文件包括: 音频文件和/或文字文件; 视频处理单元, 用于将用户上传的第一文件转换为第一预定格式的视频流; 音频处理单元, 用于将用户上传的第二文件转换为第二预定格式的音频流; 菜单合成单元, 用于将视频流和音频流生成 IVVR 菜单。根据本发明提供的技术方案, 解决了相关技术中 IVVR 菜单制作在线下制作, 灵活性低、局限性大、并可方便用户制作。



WO 2010/145335 A1



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, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ,
BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。
- 在修改权利要求的期限届满之前进行, 在收到该
修改后将重新公布(细则 48.2(h))。
- 根据申请人的请求, 在条约第 21 条(2)(a)所规定的
期限届满之前进行。

IVVR 菜单生成系统及方法

技术领域

本发明涉及通信领域，具体而言，涉及一种交互式语音及视频应答（Interactive Voice and Video Response，简称为 IVVR）菜单生成系统及方法。

5 背景技术

交互式语音及视频应答（IVVR），作为一种全新的无线语音及视频应答增值服务，正在跟随 3G 的快速发展被广泛推广。视频电话用户通过拨打指定号码，获得所需信息或者参与互动式服务，并借助视频和语音的特色，突出其“互动”型功能。

10 但是，现有 IVVR 制作方式较为复杂，具体可以参见图 1。

图 1 为相关技术中 IVVR 菜单生成方法的流程图。如图 1 所示，该流程主要包括以下步骤（步骤 S101-步骤 S107）：

步骤 S101：用户首先使用 DIY 软件制作好资源，需要通过互联网上传到服务器上；

15 步骤 S103：IVVR 系统判断该资源单路的视频总带宽是否符合要求（即是否小于 64K），如果是，执行步骤 S105，否则，执行步骤 S107；

步骤 S105：IVVR 系统判断用户上传的资源是否为 QCIF 视频流，如果是，则执行步骤 S107，否则，返回执行步骤 S101；

20 步骤 S107：执行除上述判断以外的其他判断。如果不符合，则需要用户重新上传资源或通过系统转换资源，返回执行步骤 S101，如果是，则上传成功。

因而，从上述现有技术的实现原理来看，现有 IVVR 生成方法存在以下不足：

（1）IVVR 菜单制作在线下制作，因而灵活性较低，局限性较大。

25 （2）制作者手动处理 IVVR 菜单中的视频和音频合成，花费大量的制

作时间和精力。

(3) 线下制作的资源极有可能不符合 IVVR 系统要求, 需要用户重新上传或系统重新转换。

发明内容

5 本发明的主要目的在于提供一种改进的 IVVR 菜单生成系统及方法, 以至少解决上述问题。

10 根据本发明的一个方面, 提供了一种交互式语音及视频应答菜单生成系统, 包括: 获取单元, 用于获取用户上传的文件, 其中, 文件包括: 第一文件和第二文件, 第一文件包括: 视频文件和/或图片文件, 第二文件包括: 音频文件和/或文字文件; 视频处理单元, 用于将用户上传的第一文件转换为第一预定格式的视频流; 音频处理单元, 用于将用户上传的第二文件转换为第二预定格式的音频流; 菜单合成单元, 用于将视频流和音频流生成 IVVR 菜单。

15 根据本发明的另一方面, 提供了一种交互式语音及视频应答菜单生成方法, 包括: IVVR 菜单生成系统获取用户上传的文件, 其中, 文件包括: 第一文件和第二文件, 第一文件包括: 视频文件和/或图片文件, 第二文件包括: 音频文件和/或文字文件; IVVR 菜单生成系统将用户上传的第一文件转换为第一预定格式的视频流, 并将用户上传的第二文件转换为第二预定格式的音频流; IVVR 菜单生成系统将视频流和音频流生成 IVVR 菜单。

20 通过本发明, 用户只需呼入 IVVR 菜单生成系统, 通过上传视频、音频、图片、或文字文件, 可实现自动生成 IVVR 菜单, 解决了相关技术中 IVVR 菜单制作在线下制作, 灵活性低、局限性大、并可能不符合 IVVR 系统要求的问题, 进而可以提高灵活性、降低局限性、并且方便用户制作。

25 本发明的其它特征和优点将在随后的说明书中阐述, 并且, 部分地从说明书中变得显而易见, 或者通过实施本发明而了解。本发明的目的和其他优点可通过在所写的说明书、权利要求书、以及附图中所特别指出的结构来实现和获得。

附图说明

此处所说明的附图用来提供对本发明的进一步理解,构成本申请的一部分,本发明的示意性实施例及其说明用于解释本发明,并不构成对本发明的不当限定。在附图中:

- 5 图 1 为相关技术中 IVVR 菜单生成系统的流程图;
- 图 2 为根据本发明实施例的 IVVR 菜单生成系统的结构框图;
- 图 3 为根据本发明优选实施例的 IVVR 菜单生成系统的结构框图;
- 图 4 为根据本发明实施例的 IVVR 菜单生成方法的流程图;
- 图 5 为根据本发明优选实施例的 IVVR 菜单生成方法的流程图。

10 具体实施方式

需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本发明。

- 图 2 为根据本发明实施例的 IVVR 菜单生成系统的结构框图。图 3 为根据本发明优选实施例的 IVVR 菜单生成系统的结构框图。如图 2 所示,根据
15 本发明实施例的 IVVR 菜单生成系统包括:获取单元 1、视频处理单元 2、音频处理单元 3、菜单合成单元 4,以下结合图 3 进行描述。

获取单元 1,用于获取用户上传的文件,其中,文件包括:第一文件和第二文件,第一文件包括:视频文件和/或图片文件,第二文件包括:音频文件和/或文字文件;

- 20 在具体实施过程中,上述获取单元 1 可以通过业务引导的方式,向用户提供语音交互界面,引导用户按一定逻辑流程上传 IVVR 菜单生成所需要的文件,并将该文件返回给视频处理单元和音频处理单元。

其中,上述获取单元 1 还需要在视频文件和音频文件中分别加入视频编号和音频编号,并把这些文件分别发送给视频处理单元 2 和音频处理单元 3。

- 25 视频处理单元 2,与获取单元 1 相连接,用于将用户上传的第一文件转换为第一预定格式的视频流;

优选地，视频处理单元 2，用于获取第一文件中的视频流，加入时间同步标记，转换成第一预定格式的视频流，并根据视频编号存储该视频流。

优选地，如图 3 所示，视频处理单元 2 包括：视频判断模块 20、视频过滤模块 22、视频转换模块 26、和视频预存模块 28。

- 5 视频判断模块 20，用于判断第一文件是否包括视频文件，如果是，则将该视频文件发送至视频过滤模块 22；

优选地，若第一文件中包括不可识别的文件，则视频判断模块 20 将返回该判断结果给业务引导模块，或者不对该文件作任何处理。

- 10 视频过滤模块 22，用于对输入的视频文件进行过滤，获取纯视频流，并发送至视频转换模块 26；

视频转换模块 26，与视频过滤模块 22 相连接，用于在输入的纯视频流中加入时间同步标记，转换成第一预定格式的视频流，并输出至视频预存模块 28；

- 15 视频预存模块 28，连接至视频转换模块 26，用于获取输入的第一预定格式的视频流的视频编号，并根据视频编号存储该视频流。

优选地，该视频处理单元 2 还可以包括图片转换模块 24；其中：

视频判断模块 20，还用于判断第一文件是否包括图片文件，如果是，则将图片文件发送至图片转换模块 24；

- 20 图片转换模块 24，用于将输入的图片文件按照预定帧间隔排序组成纯视频流，加入时间同步标记，转换成第一预定格式的视频流，并输出至视频预存模块 28；

视频预存模块 28，还用于获取图片转换模块 24 输入的第一预定格式的视频流的视频编号，并根据该视频编号存储该视频流。

- 25 音频处理单元 3，与获取单元 1 相连接，用于将用户上传的第二文件转换为第二预定格式的音频流；

优选地，音频处理单元 3，用于获取第二文件中的音频流，加入时间同步标记，转换成第二预定格式的音频流，并根据音频编号存储该音频流。

优选地，如图 3 所示，音频处理单元 3 可以进一步包括：音频判断模块 30、音频转换模块 32、音频预存模块 36。

音频判断模块 30，用于对第二文件进行识别，对识别出的音频文件的音频格式的有效性进行判断，将有效音频格式的音频文件发送至音频转换模块 32，将识别出的图片文件发送至音频转换模块 32；

音频转换模块 32，与音频判断模块 30 相连接，用于将接收到的音频文件加入时间同步标记，转换成第二预定格式的音频流，并发送至音频预存模块 36；

音频预存模块 36，用于获取音频转换模块 32 输入的第二预定格式的音频流的音频编号，并根据音频编号存储该音频流。

优选地，音频处理单元 3 还可以包括：文语转换（Text To Speech，简称为 TTS）转换模块 34；其中：

音频判断模块 30，还用于判断第二文件是否包括文本文件，如果是，则将文本文件发送至 TTS 转换模块 34；

TTS 转换模块 34，与音频判断模块 30 相连接，用于将接收到的文本文件通过文语转换 TTS 引擎转换为音频文件，并将该音频文件加入时间同步标记，转换成第二预定格式的音频流，并发送至音频预存模块 36；

音频预存模块 36，与 TTS 转换模块 34 相连接，还用于获取 TTS 转换模块 34 输入的第二预定格式的音频流的视频编号，并根据该音频编号存储该音频流。

菜单合成单元 4，分别与视频处理单元 2 和音频处理单元 3 相连接，用于将视频流和音频流生成 IVVR 菜单。

考虑到相关技术中 IVVR 菜单制作在线下制作，灵活性较低、局限性较大、并可能不符合 IVVR 系统要求的问题，本发明实施例提供了一种交互式语音及视频应答（IVVR）菜单生成系统，用户只需呼入 IVVR 菜单生成系统，通过上传视频、音频、图片、文字文件，即可实现自动生成 IVVR 菜单。

优选地，菜单合成单元 4，用于将对应的音频流和视频流组成至少一个媒体流组，将至少一个媒体流组组合成媒体流组集合，将媒体流组集合生成

IVVR 菜单，其中，音频流、视频流、至少一个媒体流组、媒体流组集合均小于预定阈值。

上述预定阈值可以是相同或不同的阈值，例如下文中描述的第一阈值、第二阈值、第三阈值，且可以满足以下关系：第三阈值 $\geq$ 第二阈值 $\geq$ 第一阈值。具体根据实际情况预先设定。

优选地，如图 3 所示，菜单合成单元 4 可以进一步包括：

IVVR 处理模块 40，用于判断接收到的音频流或视频流的大小是否小于第一阈值，如果是，则将音频流或视频流发送至 IVVR 合成模块 42，否则，将音频流或视频流发送至 IVVR 压缩模块 44；

IVVR 合成模块 42，用于将接收到的对应的音频流和视频流组成至少一个媒体流组，分别判断各个媒体流组的大小是否大于第二阈值，在各个媒体流组的大小都小于第二阈值的情况下，将至少一个媒体流组合成媒体流组集合，判断媒体流组集合是否小于第三阈值，如果是，则生成 IVVR 菜单，否则，发送至 IVVR 压缩模块，其中，第三阈值 $\geq$ 第二阈值 $\geq$ 第一阈值；

IVVR 压缩模块 44，用于根据预定的压缩指令对接收到的媒体流进行压缩，并将压缩完成的媒体流发送至 IVVR 处理模块 40。

在具体实施过程中，IVVR 处理模块主要负责接收获取单元发送的生成指令并返回处理结果；通过调用视频预存模块的视频流和音频预存模块的音频流的编号，打包成一组媒体流；判断该组媒体流大小，例如，以 64k 为预定阈值，小于 64k 的媒体流发送到 IVVR 合成模块，大于 64k 的媒体流发送到 IVVR 压缩模块。IVVR 合成模块主要负责接收 IVVR 处理模块发送来的媒体流，读取媒体流和音频流中的第一帧的时间同步标记，保证该组视频音频的同步开始；IVVR 合成模块通过读取时间同步标记，通过计算视频音频流中的时长，统一调整该媒体流集合中每组帧的时间间隔，把多组媒体流合成媒体流集合，在把该集合的视频流格式转换为 QCIF 后，返回给 IVVR 处理模块；判断合成的媒体流集合的总大小，超过规定大小的视频，发送给 IVVR 压缩模块，否则生成 IVVR 菜单。IVVR 压缩模块主要收到 IVVR 处理模块的压缩指令后，对接收的媒体流进行压缩，压缩完成的媒体流返回给 IVVR 处理模块。

通过上述实施例，提供了一种 IVVR 菜单生成系统，该系统使用非常方

便，用户可以通过任何视频电话在在任意时刻制作 IVVR 菜单。且 IVVR 菜单制作简单，只需要上传制作需要的视频音频或文本文件即可完成制作。

方法实施例

根据本发明实施例，还提供了一种 IVVR 菜单生成方法。

- 5 图 4 为根据本发明实施例的 IVVR 菜单生成方法的流程图。其中，该方法应用于上述 IVVR 菜单生成系统，如图 4 所示，根据本发明实施例的 IVVR 菜单生成方法包括以下处理（步骤 S401-步骤 S405）：

步骤 S401：IVVR 菜单生成系统获取来自于用户端上传的文件，其中，文件包括：第一文件和第二文件，第一文件包括：视频文件和/或图片文件，
10 第二文件包括：音频文件和/或文字文件；

其中，用户可以根据引导菜单提示顺序上传文件，上传第一文件和第二文件的顺序根据引导界面预定方式而定，先后顺序在系统中可以调整。

步骤 S403：IVVR 菜单生成系统将用户端上传的第一文件转换为第一预定格式的视频流，并将用户端上传的第二文件转换为第二预定格式的音频流；

- 15 优选地，将用户端上传的第一文件转换为第一预定格式的视频流包括以下处理：

（1）判断第一文件是否包括视频文件和/或图片文件；

- （2）在第一文件包括视频文件的情况下，对视频文件进行过滤，获取纯视频流，加入时间同步标记，转换成第一预定格式的视频流，和/或，在第一文件包括图片文件的情况下，将图片文件按照预定帧间隔排序组成视频流，
20 加入时间同步标记，转换成第一预定格式的视频流；

（3）获取转换成第一预定格式的视频流的视频编号，并根据视频编号存储该视频流。

- 25 优选地，将用户端上传的第二文件转换为第二预定格式的音频流包括以下处理：

（1）判断第二文件是否包括音频文件和/或文本文件；

(2) 在第二文件包括音频文件的情况下, 对该音频文件的音频格式的有效性进行判断, 获取有效音频格式的音频文件, 加入时间同步标记, 转换成第二预定格式的音频流, 和/或, 在第二文件包括文本文件的情况下, 将该文本文件通过文语转换 TTS 引擎转换为音频文件, 并将该音频文件加入时间同步标记, 转换成第二预定格式的音频流;

(3) 获取转换成第二预定格式的音频流的音频编号, 并根据音频编号存储该音频流。

步骤 S405: IVVR 菜单生成系统将视频流和音频流生成 IVVR 菜单。

优选地, 步骤 S405 具体可以包括以下处理:

(1) 判断音频流或视频流的大小是否大于第一阈值;

(2) 如果是, 根据预定的压缩指令对该音频流或视频流进行压缩, 否则, 将接收到的对应的音频流和视频流组成至少一个媒体流组, 分别判断各个媒体流组的大小是否大于第二阈值;

(3) 在各个媒体流组的大小都小于第二阈值的情况下, 将至少一个媒体流组合成媒体流组集合, 判断媒体流组集合是否小于第三阈值, 其中, 第三阈值 $\geq$ 第二阈值 $\geq$ 第一阈值;

(4) 在媒体流组集合小于第三阈值的情况下, 生成 IVVR 菜单, 否则, 根据预定的压缩指令对媒体流组集合进行压缩。

图 5 为根据本发明优选实施例的 IVVR 菜单生成方法的流程图。如图 5 所示, 根据本发明优选实施例的 IVVR 菜单生成方法包括以下处理 (步骤 S501-步骤 S527):

假设用户要制作一个 IVVR 菜单。其中, 该菜单由一个视频文件、一个图片、一个音频、和一段文字转语音的音频合成。一个视频文件名是 “a” (格式为 3GP, 帧速率 10fps, I 帧间隔 5 秒, 时长 20 秒, 大小 70k), 一个图片文件名是 “b” (格式为 gif, 帧速率 1fps, 时长 5 秒, 大小 12k), 一个音频文件名是 “c” (格式为 AMR, 时长 10 秒, 大小 12k), 一段文本内容是 “您好, 欢迎使用 IVVR 菜单制作系统”。该用户用 3G 手机拨打 IVVR 菜单生成系统的接入码, 进入 IVVR 系统。以下结合图 2 和图 3 描述的系统描述该方案。

步骤 S501: 用户的手机界面接收到获取单元 1 的引导提示“请上传视频文件或图片文件”，用户按具体引导步骤将“a”文件上传，获取单元接收到该文件后，将该文件设置编号为 1-1，发送给视频处理单元 2；

5 步骤 S503: 视频处理单元 2 中的视频判断模块 20 在接收“a”后，判断出该文件是视频文件，返回指令给获取单元 1。若判断出该文件是图像，则跳到步骤 S525；

步骤 S505: 视频判断模块 20 判断出该视频文件格式为 3GP，符合系统指定视频格式，发送该视频文件到视频过滤模块 22；

10 步骤 S507: 视频过滤模块 22 将该视频文件中的音频内容过滤掉，生成纯视频流，并发送给视频转换模块 26；

步骤 S509: 视频转换模块 26 接收到纯视频流，并将该视频流媒体格式转换成 H263 格式，调整帧速率为 8fps，设置第一个帧为 2 秒，在加入时间同步标记后，发送到视频预存模块 28。视频预存模块 28 将接收到的编号为 1-1 的视频流存储到视频处理单元 2 内第一个位置；

15 步骤 S511: 获取单元 1 接收到视频上传成功的指令，提示用户上传音频或输入文字。用户按引导步骤将“c”文件上传，获取单元 1 接收到后，设置该文件编号 1-2 发送到音频处理单元 3；

20 步骤 S513: 音频处理单元 3 中的音频判断模块 30 接收到用户上传文件“c”，判断出该文件为音频文件，发送该音频文件给音频处理单元 3 中的音频转换模块 32。若音频判断模块 30 判断出用户输入文本，发送给 TTS 转换模块 34，则跳回步骤 S527；

步骤 S515: 音频转换模块 32 在接收到音频文件“c”后，判断出该音频媒体格式为 AMR，返回接收成功指令给获取单元 1；若判断出该音频不是 AMR 格式，则跳回步骤 S511；

25 步骤 S517: 音频转换模块 32 把音频流转换音频码率为系统统一指定码率，并加入时间同步标记后，发送到音频预存模块 36。音频预存模块 36 将接收到的编号为 1 的音频流存储到音频处理单元 2 内第一个位置；

步骤 S519: 获取单元 1 接收到音频上传成功的指令，提示用户“音频上传成功，按 1 上传下一个视频或图像，按#结束”；

步骤 S521: 获取单元 1 接收到用户选择“1”的指令, 重复步骤 S501 到步骤 S513, 完成一个图像“b”的上传和一段文本“您好, 欢迎使用 IVVR 菜单制作系统”的输入, 并设置编号为 2-1 和 2-2;

5 步骤 S523: 获取单元 1 引导用户按“#”, 确认文件全部上传完毕, 同时发送生成 IVVR 菜单命令给菜单合成单元。菜单合成模块 4 的 IVVR 处理模块 40 接收到从视频预存模块 28 和音频预存模块 36 提取编号为 1 的视频流和音频流, 读取视频流和音频流第一帧的时间同步信息, 以时间长的视频或音频为改组媒体流的标准时长, 打包成第一组媒体流组, 并把该媒体流组的时间同步标记发送给 IVVR 合成模块 42; 对该组媒体流大小进行计算, 若
10 计算出该组媒体流大于第一阈值(例如, 64k)后, 则发送到 IVVR 压缩模块 44 进行压缩。IVVR 压缩模块 44 接收到这组媒体流, 压缩成小于第二阈值(例如, 64k)的媒体流后, 返回给 IVVR 处理模块 40。IVVR 合成模块 42 在读取 IVVR 处理模块 40 的多组媒体流的组同步标记后, 调整 I-帧的时间间隔, 合成多组媒体流集合, 并返回该媒体流集合到 IVVR 处理模块 40。IVVR 处
15 理模块 40 计算该媒体流集合的大小, 判断是否大于第三阈值(例如, 64k), 大于 64k 的媒体流集合发送到 IVVR 压缩模块 44 进行压缩后, 否则, 将该媒体流集合发送给 IVVR 合成模块 42, IVVR 合成模块 42 接收媒体流集合, 将其转换成四分之一通用中间格式(Quarter Common Intermediate Format, 简称为 QCIF)格式后, 合并成 IVVR 菜单;

20 步骤 S525: 视频转换模块 26 接收的为图片文件, 首先把图片文件按一定帧间隔组成存视频流后, 跳回步骤 S509;

步骤 S527: TTS 转换模块 34 接收到用户输入文字, 通过 TTS 引擎, 将文字转换成语音, 并将语音转换成格式 AMR 音频流, 同时, 在音频流帧中加入时间同步标记后, 跳回步骤 S517。发送到音频预存模块 36。

25 通过上述实施例, 提供了一种 IVVR 菜单生成方法, 实现了用户制作方便的效果, 生成的 IVVR 菜单符合 IVVR 系统要求, 其视频总带宽小于预定阈值(例如, 64K), 视频流为 QCIF, 格式为 3GP, 系统可以直接读取使用。

综上所述, 通过本发明的上述实施例, 提供的 IVVR 菜单生成方案, 用户可以通过任何视频电话在任何时刻制作 IVVR 菜单, 用户只需要上传制作
30 需要的视频音频或文本文件即可完成制作。达到了用户制作方便的效果, 提高了语音拨号在家庭和办公中的普及应用。

显然，本领域的技术人员应该明白，上述的本发明的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的程序代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制成单个集成电路模块来实现。这样，本发明不限制于任何特定的硬件和软件结合。

以上所述仅为本发明的优选实施例而已，并不用于限制本发明，对于本领域的技术人员来说，本发明可以有各种更改和变化。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权利要求书

1. 一种交互式语音及视频应答 IVVR 菜单生成系统，其特征在于，所述系统包括：

获取单元，用于获取用户上传的文件，其中，所述文件包括：第一文件和第二文件，所述第一文件包括：视频文件和/或图片文件，所述第二文件包括：音频文件和/或文字文件；

视频处理单元，用于将用户上传的所述第一文件转换为第一预定格式的视频流；

音频处理单元，用于将用户上传的所述第二文件转换为第二预定格式的音频流；

菜单合成单元，用于将所述视频流和所述音频流生成 IVVR 菜单。

2. 根据权利要求 1 所述的系统，其特征在于，

所述视频处理单元，用于获取所述第一文件中的视频流，加入时间同步标记，转换成所述第一预定格式的视频流，并根据视频编号存储该视频流。

3. 根据权利要求 1 所述的系统，其特征在于，所述视频处理单元包括：

视频判断模块，用于判断所述第一文件是否包括视频文件，如果是，则将所述视频文件发送至所述视频过滤模块；

所述视频过滤模块，用于对所述视频文件进行过滤，获取纯视频流，并发送至视频转换模块；

所述视频转换模块，用于在输入的所述纯视频流中加入时间同步标记，转换成所述第一预定格式的视频流，并输出至视频预存模块；

所述视频预存模块，用于获取所述视频转换模块输入的所述第一预定格式的视频流的视频编号，并根据该视频编号存储该视频流。

4. 根据权利要求 3 所述的系统，其特征在于，所述视频处理单元还包括：图片转换模块；其中：

所述视频判断模块，还用于判断所述第一文件是否包括图片文件，如果是，则将所述图片文件发送至图片转换模块；

所述图片转换模块，用于将输入的所述图片文件按照预定帧间隔排序组成纯视频流，加入时间同步标记，转换成所述第一预定格式的视频流，并输出至所述视频预存模块；

所述视频预存模块，还用于获取所述图片转换模块输入的所述第一预定格式的视频流的视频编号，并根据该视频编号存储该视频流。

5. 根据权利要求 1 所述的系统，其特征在于，

所述音频处理单元，用于获取所述第二文件中的音频流，加入时间同步标记，转换成所述第二预定格式的音频流，并根据音频编号存储该音频流。

6. 根据权利要求 1 所述的系统，其特征在于，所述音频处理单元包括：

音频判断模块，用于判断所述第二文件是否包括音频文件，如果是，则将所述音频文件发送至音频转换模块；

所述音频转换模块，用于将所述音频判断模块输入的音频文件加入时间同步标记，转换成所述第二预定格式的音频流，并发送至音频预存模块；

所述音频预存模块，用于获取所述音频转换模块输入的第二预定格式的音频流的音频编号，并根据所述音频编号存储该音频流。

7. 根据权利要求 6 所述的系统，其特征在于，所述音频处理单元还包括：文语转换 TTS 模块；其中：

所述音频判断模块，还用于判断所述第二文件是否包括文本文件，如果是，则将所述文本文件发送至所述 TTS 转换模块；

所述 TTS 转换模块，用于将接收到的文本文件通过文语转换 TTS 引擎转换为音频文件，并将该音频文件加入时间同步标记，转换成所述第二预定格式的音频流，并发送至所述音频预存模块；

所述音频预存模块，还用于获取所述 TTS 转换模块输入的第二预定格式的音频流的视频编号，并根据该音频编号存储该音频流。

8. 根据权利要求 1 所述的系统，其特征在于，

所述菜单生成单元，用于将对应的音频流和视频流组成至少一个媒体流组，将所述至少一个媒体流组组合成媒体流组集合，将所述媒体流组集合生成所述 IVVR 菜单，其中，所述音频流、所述视频流、所述至少一个媒体流组、所述媒体流组集合均小于预定阈值。

9. 根据权利要求 1 所述的系统，其特征在于，所述菜单合成单元包括：

IVVR 处理模块，用于判断接收到的音频流或视频流的大小是否小于第一阈值，如果是，则将所述音频流或视频流发送至 IVVR 合成模块，否则，将所述音频流或视频流发送至 IVVR 压缩模块；

所述 IVVR 合成模块，用于将接收到的对应的音频流和视频流组成至少一个媒体流组，分别判断各个所述媒体流组的大小是否大于第二阈值，在各个所述媒体流组的大小都小于所述第二阈值的情况下，将所述至少一个媒体流组组合成媒体流组集合，判断所述媒体流组集合是否小于第三阈值，如果是，则生成所述 IVVR 菜单，否则，发送至 IVVR 压缩模块，其中，所述第三阈值 $\geq$ 所述第二阈值 $\geq$ 所述第一阈值；

所述 IVVR 压缩模块，用于根据预定的压缩指令对接收到的媒体流进行压缩，并将压缩完成的所述媒体流发送至所述 IVVR 处理模块。

10. 一种交互式语音及视频应答 IVVR 菜单生成方法，其特征在于，包括：

所述 IVVR 菜单生成系统获取用户上传的文件，其中，所述文件包括：第一文件和第二文件，所述第一文件包括：视频文件和/或图片文件，所述第二文件包括：音频文件和/或文字文件；

将用户上传的所述第一文件转换为第一预定格式的视频流，并将用户上传的所述第二文件转换为第二预定格式的音频流；

将所述视频流和所述音频流生成 IVVR 菜单。

11. 根据权利要求 10 所述的方法，其特征在于，所述将用户上传的所述第一文件转换为第一预定格式的视频流包括：

判断所述第一文件是否包括视频文件；

在所述第一文件包括视频文件的情况下，对所述视频文件进行过滤，获取纯视频流；

在所述纯视频流中加入时间同步标记，并转换成所述第一预定格式的视频流；

获取所述转换成所述第一预定格式的视频流的视频编号，并根据所述视频编号存储该视频流。

12. 根据权利要求 10 所述的方法，其特征在于，所述将用户上传的所述第一文件转换为第一预定格式的视频流还包括：

判断所述第一文件是否包括图片文件；

在所述第一文件包括图片文件的情况下，将所述图片文件按照预定帧间隔排序组成纯视频流；

在所述纯视频流中加入时间同步标记，并转换成第一预定格式的视频流；

获取所述转换成第一预定格式的视频流的视频编号，并根据所述视频编号存储该视频流。

13. 根据权利要求 10 所述的方法，其特征在于，将用户上传的所述第二文件转换为第二预定格式的音频流包括：

判断所述第二文件是否包括音频文件；

在所述第二文件包括音频文件的情况下，对该音频文件的音频格式的有效性进行判断，获取有效音频格式的音频文件；

在所述有效音频格式的音频文件中加入时间同步标记，并转换成第二预定格式的音频流；

获取所述转换成第二预定格式的音频流的音频编号，并根据所述音频编号存储该音频流。

14. 根据权利要求 10 所述的方法，其特征在于，将用户上传的所述第二文件转换为第二预定格式的音频流还包括：

判断所述第二文件是否包括文本文件；

在所述第二文件包括文本文件的情况下，将该文本文件通过文语转换 TTS 引擎转换为音频文件；

在所述音频文件中加入时间同步标记，转换成第二预定格式的音频流；

获取所述转换成第二预定格式的音频流的音频编号，并根据所述音频编号存储该音频流。

15. 根据权利要求 10 至 14 中任一项所述的方法，其特征在于，将所述视频流和所述音频流生成 IVVR 菜单包括：

在所述音频流或所述视频流的大小分别小于第一阈值的情况下，将接收到的对应的音频流和视频流组成至少一个媒体流组，分别判断各个所述媒体流组的大小是否大于第二阈值；

在各个所述媒体流组的大小都小于所述第二阈值的情况下，将所述至少一个媒体流组组合成媒体流组集合，并判断所述媒体流组集合是否小于第三阈值，其中，所述第三阈值 $\geq$ 所述第二阈值 $\geq$ 所述第一阈值；

在所述媒体流组集合小于所述第三阈值的情况下，生成所述 IVVR 菜单，否则，根据预定的压缩指令对所述媒体流组集合进行压缩。

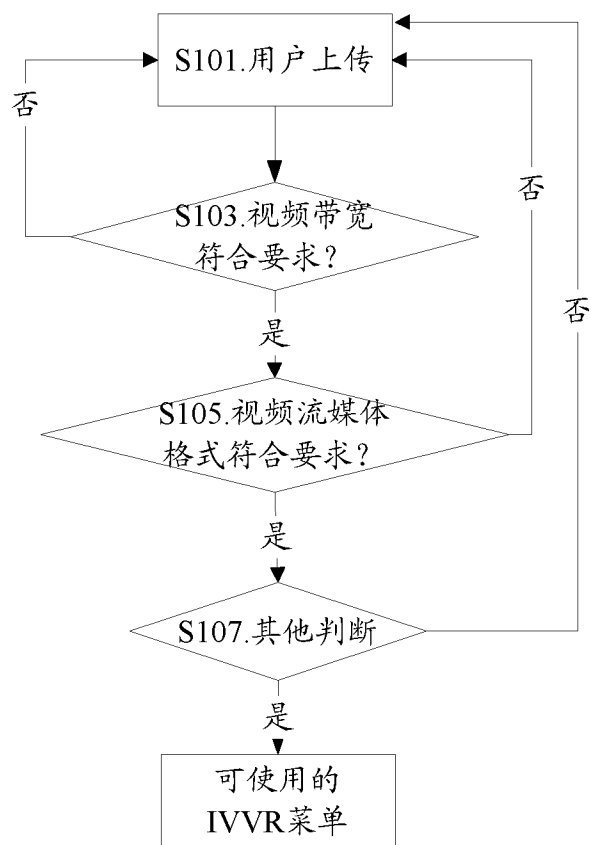


图 1

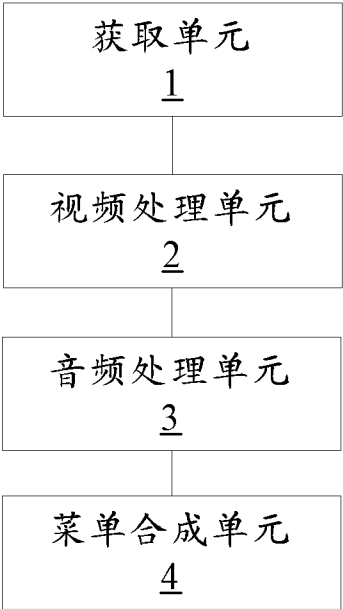


图 2

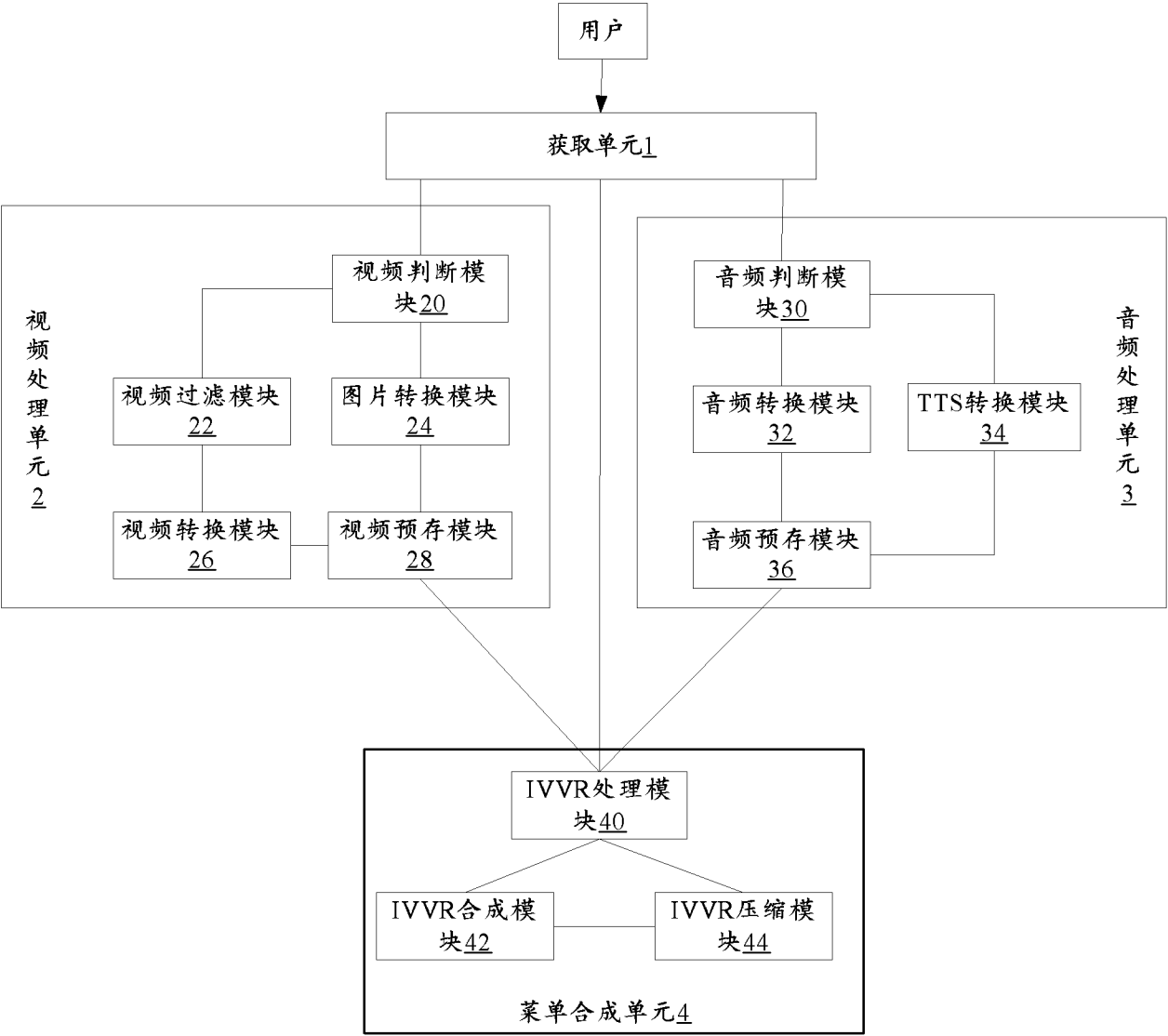


图 3

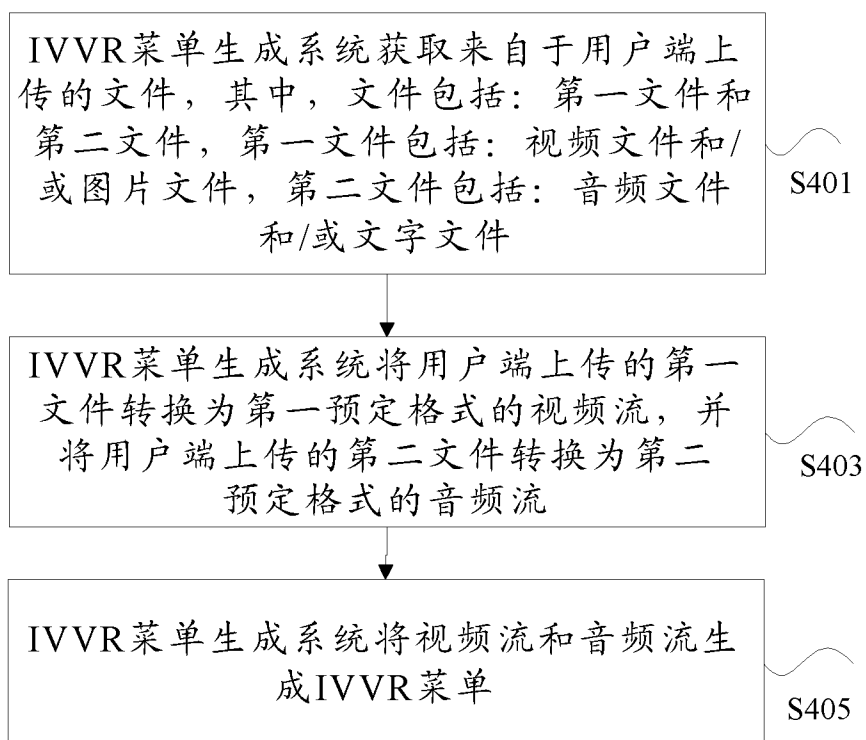


图 4

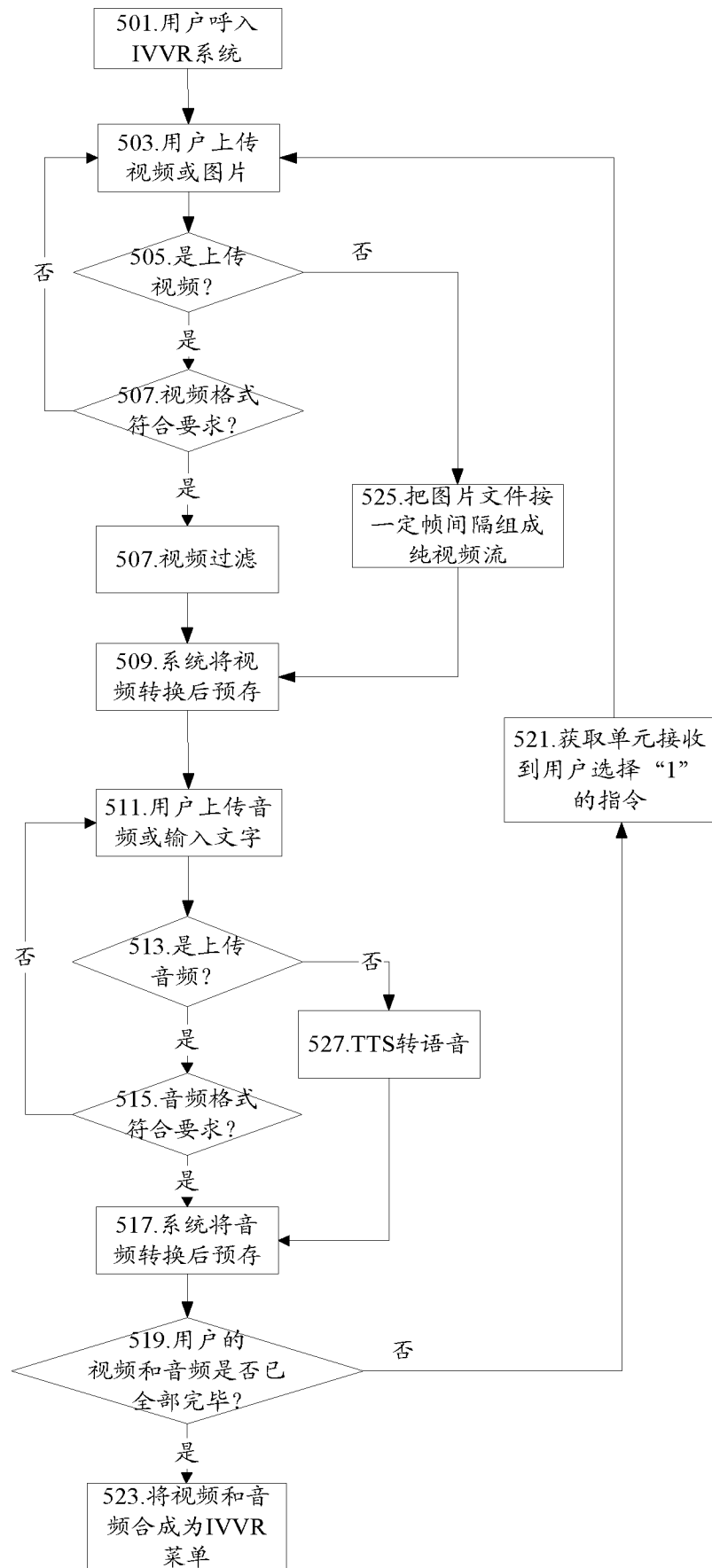


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2010/072248

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: H04N, H04L

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)

WPI, EPODOC, CNPAT, CNKI: video, audio, sound, imag+, interact+, response, menu, upload

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN1992884A (ALCATEL LUCENT) 04 July 2007 (04.07.2007) the whole document	1-15
A	CN1960408A (CHINA TELECOM CO LTD) 09 May 2007 (09.05.2007) the whole document	1-15
A	CN1770857A (LIANG SONG) 10 May 2006 (10.05.2006) the whole document	1-15

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search
07 July 2010 (07.07.2010)Date of mailing of the international search report
12 Aug. 2010 (12.08.2010)Name and mailing address of the ISA/CN
The State Intellectual Property Office, the P.R.China
6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China
100088
Facsimile No. 86-10-62019451Authorized officer
WANG Qian
Telephone No. (86-10)62411517

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2010/072248

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN1992884A	04.07.2007	EP1802118A1	27.06.2007
		US2007165793A1	19.07.2007
		WO2007071433A1	28.06.2007
		INDELNP200805246E	08.08.2008
		KR20080083009A	12.09.2008
		JP2009521024T	28.05.2009
CN1960408A	09.05.2007	none	
CN1770857A	10.05.2006	none	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2010/072248

A. CLASSIFICATION OF SUBJECT MATTER

H04N7/14(2006.01)i

H04L29/06(2006.01)i

国际检索报告

国际申请号
PCT/CN2010/072248

A. 主题的分类

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: H04N, H04L

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))

WPI, EPODOC: video, audio, sound, imag+, interact+, response, menu, upload

CNPAT, CNKI: 视频, 音频, 声, 音, 图像, 图象, 交互, 应答, 响应, 菜单, 上传

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
A	CN1992884A (阿尔卡特朗讯公司) 04.7 月 2007 (04.07.2007) 全文	1-15
A	CN1960408A (中国电信股份有限公司) 09.5 月 2007 (09.05.2007) 全文	1-15
A	CN1770857A (梁松) 10.5 月 2006 (10.05.2006) 全文	1-15

☐ 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
07.7 月 2010 (07.07.2010)

国际检索报告邮寄日期
12.8 月 2010 (12.08.2010)

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

王倩
电话号码: (86-10) 62411517

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2010/072248

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN1992884A	04.07.2007	EP1802118A1	27.06.2007
		US2007165793A1	19.07.2007
		WO2007071433A1	28.06.2007
		INDELNP200805246E	08.08.2008
		KR20080083009A	12.09.2008
		JP2009521024T	28.05.2009
CN1960408A	09.05.2007	无	
CN1770857A	10.05.2006	无	

A. 主题的分类

H04N7/14(2006.01)i

H04L29/06(2006.01)i