

(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
2018 年 10 月 25 日 (25.10.2018)



(10) 国际公布号
WO 2018/192186 A1

(51) 国际专利分类号:
G10L 15/02 (2006.01)

(21) 国际申请号: PCT/CN2017/104382

(22) 国际申请日: 2017 年 9 月 29 日 (29.09.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710254628.8 2017 年 4 月 18 日 (18.04.2017) CN

(71) 申请人: 广州视源电子科技股份有限公司 (GUANGZHOU SHIYUAN ELECTRONICS CO., LTD.) [CN/CN]; 中国广东省广州市黄埔区云埔四路 6 号, Guangdong 510530 (CN)。

(72) 发明人: 李忠杰(LI, Zhongjie); 中国广东省广州市黄埔区云埔四路 6 号, Guangdong 510530 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409 室, Beijing 100085 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,

(54) Title: SPEECH RECOGNITION METHOD AND APPARATUS

(54) 发明名称: 语音识别方法及装置

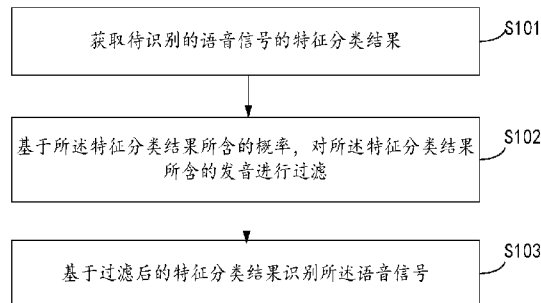


图 1

- S101 Acquire a feature classification result of a speech signal to be recognized
- S102 Filter, based on probabilities contained in the feature classification result, pronunciations contained in the feature classification result
- S103 Recognize the speech signal based on a filtered feature classification result

(57) Abstract: Disclosed are a speech recognition method and apparatus. The method comprises: acquiring a feature classification result of a speech signal to be recognized, wherein the feature classification result contains a pronunciation for describing a pronunciation feature of each speech signal frame and the probability of mapping each speech signal frame to a corresponding pronunciation (S201); if the probability of mapping any speech signal frame to a corresponding pronunciation satisfies a predetermined filtering rule, determining the pronunciation as a candidate pronunciation (S202); if the probability of mapping any of a predetermined number of

PT, QA, RO, RS, RU, RW, SA, 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, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, 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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, 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))。

speech signal frames adjacent to the speech signal frame to the candidate pronunciation satisfies the predetermined filtering rule, deleting the candidate pronunciation from pronunciations contained in the feature classification result (S203); if none of the probabilities of mapping the predetermined number of speech signal frames adjacent to the speech signal frame to the candidate pronunciation satisfies the predetermined filtering rule, retaining the candidate pronunciation in the pronunciations contained in the feature classification result (S204); and recognizing the speech signal based on a filtered feature classification result (S205). During the process of speech signal recognition, a recognition operation related to a pronunciation filtered out is not required to be executed, thereby improving the efficiency of speech recognition.

(57) 摘要: 一种语音识别方法及装置, 所述方法包括: 获取待识别的语音信号的特征分类结果: 特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率 (S201); 如果任一语音信号帧映射到对应的发音的概率, 满足预定过滤规则, 将该发音确定为候选发音 (S202); 如果该语音信号帧的预定帧数的相邻语音信号帧中的任一帧, 映射到该候选发音的概率满足预定过滤规则, 则将该候选发音从所述特征分类结果所含的发音中删除 (S203); 如果该语音信号帧的预定帧数的相邻语音信号帧, 映射到该候选发音的概率均不满足预定过滤规则, 则将该候选发音保留在所述特征分类结果所含的发音中 (S204); 基于过滤后的特征分类结果识别所述语音信号 (S205)。在识别语音信号的过程中, 无需再执行与被过滤掉的发音相关的识别操作, 进而能提高语音识别效率。

语音识别方法及装置

技术领域

[01] 本发明涉及计算机技术领域，尤其涉及语音识别方法及装置。

5 背景技术

[02] 随着计算机技术的发展，语音识别（Automatic Speech Recognition, ASR）技术在人机交互等领域的应用越来越多。目前，语音识别技术主要通过信号处理模块、特征提取模块、声学模型、语言模型（Language Model, LM）、字典和解码器(Decoder)，将待识别的语音信号转换为文本信息，完成语音识别。

10 [03] 在语音识别过程中，信号处理模块和特征提取模块，先将待识别的语音信号划分成多个语音信号帧，然后通过消除噪音、信道失真等处理对各语音信号帧进行增强，再将各语音信号帧从时域转化到频域，并从转换后的语音信号帧内提取合适的声学特征。而根据训练语音库的特征参数训练出的声学模型，以特征提取模块所提取的声学特征为输入，映射到能够描述语音信号帧的发音特征的发音、并计算出语音信号帧映射到各发音的概率，得到特征分类
15 结果。

[04] 语言模型含有不同的字词（如：字、词、短语）之间关联关系、及其概率（可能性），用于估计由不同字词组成的各种文本信息的可能性。解码器可以基于已经训练好的声学模型、语言模型及字典建立一个识别网络，识别网络中的各路径分别与各种文本信息、以及各文本信息的发音对应，然后针对声学模型输出的发音，在该识别网络中寻找最佳的一条路径，基
20 于该路径能够以最大概率输出该语音信号对应的文本信息，完成语音识别。

[05] 但是，语言模型一般是基于大量语料训练出来的模型，包含大量字词之间的关联关系和可能性，所以，基于语音模型建立的识别网络包含的节点较多，每个节点的分支数量也非常多。在识别网络中进行路径搜索时，各语音信号帧的发音涉及的节点数会以指数形式暴增，导致路径搜索量极大，搜索过程耗费的时间较多，进而会降低语音识别效率。

25

发明内容

[06] 有鉴于此，本发明提供一种语音识别方法及装置，以解决语音识别过程耗时多、识别效

率低的问题。

[07]根据本发明的第一方面，提供一种语音识别方法，包括步骤：

[08]获取待识别的语音信号的特征分类结果；所述特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率；

5 [09]基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤；

[10]基于过滤后的特征分类结果识别所述语音信号。

[11]在一个实施例中，所述基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤，包括：

[12]判断任一语音信号帧映射到对应的发音的概率是否满足预定过滤规则；

10 [13]如果所述对应的发音满足预定过滤规则，对所述对应的发音进行滤掉。

[14]在一个实施例中，如果任一语音信号帧映射到对应的发音的概率，与该语音信号帧的最大映射概率之间的概率差，在预定的差值范围内，则确定所述对应的发音满足预定过滤规则；

[15]如果任一语音信号帧映射到对应的发音的概率，小于该语音信号帧映射到预定数目的发音中各发音的概率，则确定所述对应的发音满足预定过滤规则。

15 [16]在一个实施例中，所述预定数目为以下任一：

[17]该帧语音信号帧对应的发音中被保留在特征分类结果内的发音的数量；

[18]预定的比例阈值与该帧语音信号帧对应的发音的总数目的乘积。

[19]在一个实施例中，所述基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤，包括：

20 [20]获取任一语音信号帧映射到各发音的概率的直方图分布；

[21]获取与所述直方图分布对应的束宽；

[22]将概率分布在所述束宽之外的发音，确定为满足所述预定过滤规则的发音；

[23]将满足所述预定过滤规则的发音，从所述特征分类结果所含的发音中删除。

[24]在一个实施例中，所述将满足所述预定过滤规则的发音从所述特征分类结果所含的发音
25 中删除，包括：

[25]如果任一语音信号帧映射到对应的发音的概率满足预定过滤规则，将该发音确定为候选

发音;

[26]如果该语音信号帧的预定帧数的相邻语音信号帧中的任一帧,映射到该候选发音的概率满足预定过滤规则,则将该候选发音从所述特征分类结果所含的发音中删除;

5 [27]如果该语音信号帧的预定帧数的相邻语音信号帧,映射到该候选发音的概率均不满足预定过滤规则,则将该候选发音保留在所述特征分类结果所含的发音中。

[28]根据本发明的第二方面,提供一种语音识别装置,包括:

[29]分类结果获取模块,用于获取待识别的语音信号的特征分类结果;所述特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率;

10 [30]发音过滤模块,用于基于所述特征分类结果所含的概率,对所述特征分类结果所含的发音进行过滤;

[31]语音识别模块,用于基于过滤后的特征分类结果识别所述语音信号。

[32]在一个实施例中,所述发音过滤模块还包括:

[33]第一过滤模块,用于在任一语音信号帧映射到对应的发音的概率,与该语音信号帧的最大映射概率之间的概率差,在预定的差值范围内时,对所述对应的发音进行过滤;

15 [34]第二过滤模块,用于在任一语音信号帧映射到对应的发音的概率,小于该语音信号帧映射到预定数目的发音中各发音的概率时,对所述对应的发音进行过滤。

[35]在一个实施例中,所述发音过滤模块包括:

[36]概率分布模块,用于获取任一语音信号帧映射到各发音的概率的直方图分布;

[37]束宽确定模块,用于获取与所述直方图分布对应的束宽;

20 [38]发音确定模块,用于将概率分布在所述束宽之外的发音,确定为满足所述预定过滤规则的发音;

[39]发音删除模块,用于将满足所述预定过滤规则的发音从所述特征分类结果所含的发音中删除。

[40]在一个实施例中,所述发音过滤模块包括:

25 [41]候选发音模块,用于在任一语音信号帧映射到对应的发音的概率满足预定过滤规则时,将该发音确定为候选发音;

[42]候选发音删除模块,用于在该语音信号帧的预定帧数的相邻语音信号帧中的任一帧,映

射到该候选发音的概率满足预定过滤规则时，将该候选发音从所述特征分类结果所含的发音中删除；

[43] 候选发音保留模块，用于在该语音信号帧的预定帧数的相邻语音信号帧，映射到该候选发音的概率均不满足预定过滤规则时，将该候选发音保留在所述特征分类结果所含的发音中。

- 5 [44] 实施本发明提供的实施例，在识别语音信号时，先获取该语音信号的特征分类结果，然后基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤，那么在识别语音信号的过程中，无需再执行与被过滤掉的发音相关的识别操作，如无需再在识别网络中搜索与被过滤掉的发音相关的路径，因此能有效降低语音识别过程耗费的时间，进而能提高语音识别效率。

10

附图说明

[45] 图 1 是本发明一示例性实施例示出的语音识别方法的流程图；

[46] 图 2 是本发明另一示例性实施例示出的语音识别方法的流程图；

[47] 图 3 是本发明一示例性实施例示出的语音识别装置的逻辑框图；

- 15 [48] 图 4 是本发明另一示例性实施例示出的语音识别装置的逻辑框图；

[49] 图 5 是本发明一示例性实施例示出的语音识别装置的硬件结构图。

具体实施方式

- [50] 这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，
20 除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的
实施方式并不代表与本发明相一致的所有实施方式。相反，它们仅是与如所附权利要求书
中所详述的、本发明的一些方面相一致的装置和方法的例子。

- [51] 在本发明使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本发明。在本发
明和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形
25 式，除非上下文清楚地表示其他含义。还应当理解，本文中使用的术语“和/或”是指并包含
一个或多个相关联的列出项目的任何或所有可能组合。

[52]应当理解，尽管在本发明可能采用术语第一、第二、第三等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本发明范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以被称为第一信息。取决于语境，如在此所使用的词语“如果”可以被解释成为“在...时”或“当...时”或“响应于确定”。

[53]本发明实施例的语音识别，在识别过程中会涉及到声学模型和语言模型，其中，声学模型是对声学、语音学、环境的变量、以及发出语音的人员的性别、口音等的差异进行的知识表示，可以通过 LSTM (Long Short-Term Memory, 时间递归神经网络)、CTC(Connectionist temporal classification)模型、或者隐马尔可夫模型 HMM，对训练语音库所含的语音进行训练，获得语音的声学特征到发音的映射，构成声学模型，该发音与建模单元相关。如果建模单元为音节，该发音为音节；如果建模单元为音素，该发音为音素；如果建模单元为构成音素的状态，该发音为状态。

[54]而训练声学模型时，考虑到发音会随着字词、语速、语调、轻重音、以及方言等影响发音的因素不同而不同，训练语音库需要涵盖不同的字词、语速、语调、轻重音、以及方言等影响发音的因素的大量语音。此外，考虑到语音识别的精确性，可以选择音节、音素、状态等较小的发音单位为建模单元。因此，基于训练语音库所含的大量语音以及预定的建模单元，进行模型训练，会构建出大量的声学模型。语音识别过程中，通过大量的声学模型对待识别的语音信号进行特征分类，所获得特征分类结果会包含大量的发音(类别)，如：3000 到 10000 个发音。

[55]此外，目前的语音识别技术要识别出语音信号对应的文本信息，需针对每一个发音，在识别网络中搜索所有可能的路径，在这个搜索过程中会产生指数形式的路径增量。如果在识别网络中搜索 3000 到 10000 个发音涉及的所有可能的路径，搜索所需的存储资源和计算量可能超出语音识别系统所能承受的极限，因此，目前的语音识别技术会耗费大量的时间和资源，存在语音识别效率低的问题，本发明针对如何提高语音识别效率，提出解决方案。

[56]本发明的方案，为了解决语音识别效率低这个问题，针对语音识别过程所得的特征分类结果进行改进，预先根据语音识别涉及的设备资源、识别效率需求设定过滤规则，然后在识别语音信号时，先获取该语音信号的特征分类结果，然后基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤，那么在识别语音信号的过程中，无需再在识别网络中搜索与被过滤掉的发音相关的路径，因此能有效降低搜索过程耗费的时间，进而能提高语音识别效率。以下结合附图详细说明本发明的语音识别过程。

[57]请参阅图 1，图 1 是本发明一示例性实施例示出的语音识别方法的流程图，该实施例能应用于具备语音处理能力的各种电子设备上，可以包括以下步骤 S101-S103：

[58]步骤 S101、获取待识别的语音信号的特征分类结果；所述特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率。

5 [59]步骤 S102、基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤。

[60]步骤 S103、基于过滤后的特征分类结果识别所述语音信号。

[61]本发明实施例中，所述语音信号可以是本地语音采集设备所实时采集的用户发出的语音，也可以是其语音采集设备远程传送过来的语音。在获取语音信号的特征分类结果时，可以实时通过本领域的语音预处理模块对语音信号进行预处理，通过特征提取模块对预处理后的语音信号进行特征提取，所提取的特征可以包括 PLP（Perceptual Linear Predictive，感知线性预测）、LPCC（Linear Predictive Cepstral Coefficient，线性预测倒谱系数）、FBANK（Mel-Scale Filter Bank，梅尔标度滤波器组）、MFCC（Mel-Frequency Cepstral Coefficients，梅尔倒谱系数）等，然后通过声学模型对提取的特征进行相应处理，获得特征分类结果，特征分类结果所含的概率，用于表示语音信号帧映射到对应的发音的可能性。在其他例子中，也可以直接接收其他终端设备传送过来的特征分类结果。

10

15

[62]在得到特征分类结果后，本发明的方案，考虑到特征分类结果所含的部分发音，与待识别的语音信号的语音信号帧相关性较低，对语音识别准确率的影响较小，在降低特征分类结果所含的大量发音对语音识别效率的影响时，可以在基于特征分类结果进行语音识别前，将这些对语音识别准确率影响较小的发音从所述特征分类结果中过滤掉，来减少特征分类结果所含的发音的数量，进而提高语音识别效率。

20

[63]一般情况下，发音与待识别的语音信号帧的相关性越低，在通过声学模型对语音信号的声学特征进行分类时，语音信号帧映射到该发音的概率越低。因此，可以基于语音信号帧映射到各发音的概率，来过滤特征分类结果所含的发音，过滤后，任一语音信号帧映射到被过滤掉的发音的概率，小于该语音信号帧映射到其他发音的概率。

25

[64]此外，在过滤相关性较低的发音时，考虑到不同应用场景对语音识别准确率的需求，还需要衡量所过滤掉的发音对语音识别准确率的影响，因此，可以根据语音识别准确率的需求，预先设定能限制过滤掉的发音对识别准确率的影响程度的各种过滤规则。针对各种预定过滤规则，在过滤特征分类结果所含的发音时，判断任一语音信号帧映射到对应的发音的概率是

否满足预定过滤规则，如果所述对应的发音满足预定过滤规则，对所述对应的发音进行滤掉。
过滤掉的发音一般指从特征分类结果中删除掉的发音。

[65]以下列举几种对所述特征分类结果所含的发音进行过滤的方式：

[66]过滤方式一：按预定数目过滤掉低概率的发音，该预定数目可以指语音信号帧对应的发
5 音中被保留在特征分类结果内的发音的数量；也可以指预定的比例阈值与语音信号帧对应的
发音的总数目的乘积。在过滤时，如果任一语音信号帧映射到对应的发音的概率，小于该语
音信号帧映射到预定数目的发音中各发音的概率，则确定所述对应的发音满足预定过滤规则。

[67]其中，预定的比例阈值，可以由本发明的设计人员根据需要达到的语音识别准确率来设
定，例如，设定为 $1/4$ ，指被保留的发音与所有发音的数量比例。

10 [68]在一例子中，实际过滤时，可以按概率从小到大的顺序，从特征分类结果中删除发音，
当保留的发音的数量与原来所有发音的数量的比例，满足预定的比例阈值，完成对特征分类
结果的过滤。

[69]在其他例子中，预定的比例阈值可以指未被过滤掉的发音与被过滤掉的发音的数量比例。
实际过滤时，可以按概率从大到小的顺序，在特征分类结果中挑选发音，当挑选出的发音的
15 数量与剩余的发音的数量的比例，满足预定的比例阈值时，完成对特征分类结果的过滤。

[70]实际应用中，预定数目指该帧语音信号帧对应的发音中被保留在特征分类结果内的发音
的数量时，可以由本发明的设计人员根据需要达到的语音识别准确率来设定预定数目，例如，
设定为 2000 至 9000 中的任一数值。过滤时，可以按概率从小到大的顺序，将每一语音信号
帧所映射到的发音进行排列，然后将排列在前预定位数的发音从特征分类结果中给删除，完
20 成对特征分类结果的过滤，所述预定位数与所述预定数目的数值相等。

[71]在其他例子中，预定数目可以指未被过滤掉的发音的数量，例如，设定为 1000。实际过
滤时，可以按概率从大到小的顺序，将每一语音信号帧所映射到的发音进行排列，然后将排
列在前预定位数的发音保留在特征分类结果中，将其他发音从特征分类结果中删除，完成对
特征分类结果的过滤，所述预定位数与所述数量阈值的数值相等。在其他实施例，还可以
25 采取其他技术手段按过滤方式一对特征分类结果进行过滤，本发明对此不做限制。

[72]过滤方式二：按预定的差值阈值过滤掉低概率的发音，该差值阈值可以由本发明的设计
人员根据需要达到的语音识别准确率来设定，例如，设定为 -0.5，指被过滤掉的发音的概率与
同一语音信号帧映射到的概率最大的发音之间的概率差。过滤时，如果任一语音信号帧映射
到对应的发音的概率，与该语音信号帧的最大映射概率之间的概率差，在预定的差值范围内，

则确定所述对应的发音满足预定过滤规则，可以对所述对应的发音进行过滤。

[73]在一例子中，实际过滤时，可以按概率从大到小的顺序，将每一语音信号帧所映射到的发音进行排列，将该语音信号帧映射到排列在第一位的发音的概率，确定为最大概率，然后从排列在最后一位的发音开始，依次获得该语音信号帧映射到每个发音的概率与最大概率的差值，如果差值小于-0.5，则将该发音从特征分类结果中删除。在其他实施例

5 中，还可以采取其他技术手段按过滤方式二对特征分类结果进行过滤，本发明对此不做限制。

[74]过滤方式三：按概率的直方图分布过滤分布在所述束宽之外的发音，实际过滤时，可以先获取任一语音信号帧映射到各发音的概率的直方图分布；获取与所述直方图分布对应的束宽；然后将概率分布在所述束宽之外的发音，确定为满足所述预定过滤规则的发音；最终将

10 满足所述预定过滤规则的发音，从所述特征分类结果所含的发音中删除。实际应用中，束宽可以由本发明的设计人员根据需要达到的语音识别准确率、以及直方图的分布状况来确定，如：预先设定需要过滤掉 8000 个低概率的发音，可以从直方图中低概率一侧开始查找 8000 个发音，将第 8000 个发音所在位置确定为束宽边界。在其他实施例

中，还可以采取其他技术手段按过滤方式三对特征分类结果进行过滤，本发明对此不做限制。

15 [75]在按以上任一过滤方式，对所述特征分类结果所含的发音进行过滤后，可以直接调取预定的识别网络，搜索与过滤后的特征分类结果所含的发音相关的路径，寻找最佳的一条路径，基于该路径以最大概率输出待识别的语音信号对应的文本信息，完成语音识别，这里提到的识别网络，可以指解码器针对待识别的语音信号，根据已经训练好的声学模型、语言模型及字典建立的识别网络。

[76]在寻找最佳的一条路径时，可以将特征分类结果所含的概率（声学得分）转换到和语音模型所含的字词（如：字、词、短语）之间关联概率（语言得分）相近的数值空间，并加权相加，作为路径搜索过程的综合分值，每一语音信号帧都会用一个预设的门限值来限制，与最佳路径的差值大于这个门限值，则该路径丢弃，否则保留；每一语音信号帧完成搜索后，会根据预设的最大路径数量，对所有路径进行排序，只保留此数量的最优路径，直至最后一

20 帧完成，由此得出最后的路径图。

[77]在某些例子中，输出特征分类结果的声学模型的建模单元较小，如以状态为建模单元，由于单个音素可以由三到五个状态组成，一个音素的发音所成的语音信号可以分割为多个语音信号帧，因此，易出现多个连续的语音信号帧的声学特征较类似的状况，那么特征分类结果中描述这些连续的语音信号帧中的各帧的发音，易出现类似状况。针对这种状况，如果本

25 发明基于特征分类结果所含的概率和预定过滤规则，分别对每帧语音信号帧映射到的发音进

行过滤，易将对识别准确率影响较大的发音过滤掉，为了避免误过滤这类发音，在过滤特征分类结果时，可以综合考虑连续的语音信号帧的过滤状况，具体实现过程可以参阅图 2 所示的方法，包括以下步骤 S201-S205：

5 [78] 步骤 S201、获取待识别的语音信号的特征分类结果；所述特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率。

[79] 步骤 S202、如果任一语音信号帧映射到对应的发音的概率，满足预定过滤规则，将该发音确定为候选发音。

[80] 步骤 S203、如果该语音信号帧的预定帧数的相邻语音信号帧中的任一帧，映射到该候选发音的概率满足预定过滤规则，则将该候选发音从所述特征分类结果所含的发音中删除。

10 [81] 步骤 S204、如果该语音信号帧的预定帧数的相邻语音信号帧，映射到该候选发音的概率均不满足预定过滤规则，则将该候选发音保留在所述特征分类结果所含的发音中。

[82] 步骤 S205、基于过滤后的特征分类结果识别所述语音信号。

[83] 本发明实施例中，预定过滤规则可以是以上所述的过滤方式一至过滤方式四涉及的任何一种规则，还可以是能限制过滤掉的发音对识别准确率的影响程度的其他过滤规则。

15 [84] 连续的语音信号帧的预定帧数可以由本发明的设计人员根据需要达到的语音识别准确率来设定，例如，设定为 6，相邻的前三帧以及相邻的后三帧。

[85] 由上述实施例可知：本发明的语音识别方法在识别语音信号时，先获取该语音信号的特征分类结果，然后基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤，那么在识别语音信号的过程中，无需再执行与被过滤掉的发音执行相关的识别操作，
20 如无需再在识别网络中搜索与被过滤掉的发音相关的路径，因此能有效降低语音识别过程耗费的时间，进而能提高语音识别效率。

[86] 进而，本发明实施例的语音识别方法可以应用于各种电子设备的人机交互软件内，例如：智能手机内的语音拨号、语音操控、语音查找，应用于智能手机内的语音查找时，如果用户在距离智能手机的预定范围内发出一段语音，那么应用于语音查找上的语音识别方法，可以在接收到语音采集设备采集的用户语音后，先获取该语音的特征分类结果，然后基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤，然后在识别网络中只搜索未被过滤掉的发音相关的路径，通过路径搜索快速识别出用户语音对应的文本信息，进而使语音助手基于该识别结果快速响应用户。

[87]与前述方法的实施例相对应，本发明还提供了装置的实施例。

[88]参见图 3，图 3 是本发明一示例性实施例示出的语音识别装置的逻辑框图，该装置可以包括：分类结果获取模块 310、发音过滤模块 320 和语音识别模块 330。

5 [89]其中，分类结果获取模块 310，用于获取待识别的语音信号的特征分类结果；所述特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率。

[90]发音过滤模块 320，用于基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤。

[91]语音识别模块 330，用于基于过滤后的特征分类结果识别所述语音信号。

10 [92]一些例子中，发音过滤模块 320 可以包括：

[93]第一过滤模块，用于在任一语音信号帧映射到对应的发音的概率，与该语音信号帧的最大映射概率之间的概率差，在预定的差值范围内时，对所述对应的发音进行过滤。

[94]第二过滤模块，用于在任一语音信号帧映射到对应的发音的概率，小于该语音信号帧映射到预定数目的发音中各发音的概率时，对所述对应的发音进行过滤。

15 [95]另一些例子中，发音过滤模块 320 还可以包括：

[96]概率分布模块，用于获取任一语音信号帧映射到各发音的概率的直方图分布。

[97]束宽确定模块，用于获取与所述直方图分布对应的束宽。

[98]发音确定模块，用于将概率分布在所述束宽之外的发音，确定为满足所述预定过滤规则的发音。

20 [99]发音删除模块，用于将满足所述预定过滤规则的发音从所述特征分类结果所含的发音中删除。

[100] 参见图 4，图 4 是本发明另一示例性实施例示出的语音识别装置的逻辑框图，该装置可以包括：分类结果获取模块 410、发音过滤模块 420 和语音识别模块 430。发音过滤模块 420 可以包括候选发音确定模块 421、候选发音删除模块 422 和候选发音保留模块 423。

25 [101] 其中，分类结果获取模块 410，用于获取待识别的语音信号的特征分类结果；所述特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率。

[102] 候选发音确定模块 421, 用于在任一语音信号帧映射到对应的发音的概率满足预定过滤规则时, 将该发音确定为候选发音。

[103] 候选发音删除模块 422, 用于在该语音信号帧的预定帧数的相邻语音信号帧中的任一帧, 映射到该候选发音的概率满足预定过滤规则时, 将该候选发音从所述特征分类结果所含的发音中删除。

[104] 候选发音保留模块 423, 用于在该语音信号帧的预定帧数的相邻语音信号帧, 映射到该候选发音的概率均不满足预定过滤规则时, 将该候选发音保留在所述特征分类结果所含的发音中。

[105] 语音识别模块 430, 用于基于过滤后的特征分类结果识别所述语音信号。

10 [106] 上述装置中各个单元(或模块)的功能和作用的实现过程具体详见上述方法中对应步骤的实现过程, 在此不再赘述。

[107] 对于装置实施例而言, 由于其基本对应于方法实施例, 所以相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的, 其中所述作为分离部件说明的单元或模块可以是或者也可以不是物理上分开的, 作为单元或模块显示的部件可以是或者也可以不是物理单元或模块, 即可以位于一个地方, 或者也可以分布到多个网络单元或模块上。可以根据实际的需要选择其中的部分或者全部模块来实现本发明方案的目的。本领域普通技术人员在不付出创造性劳动的情况下, 即可以理解并实施。

15 [108] 本发明语音识别装置的实施例可以应用在电子设备上。具体可以由计算机芯片或实体实现, 或者由具有某种功能的产品来实现。一种典型的实现中, 电子设备为计算机, 计算机的具体形式可以是个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件收发设备、游戏控制台、平板计算机、可穿戴设备、互联网电视、智能机车、无人驾驶汽车、智能冰箱、其他智能家居设备或者这些设备中的任意几种设备的组合。

20 [109] 装置实施例可以通过软件实现, 也可以通过硬件或者软硬件结合的方式实现。以软件实现为例, 作为一个逻辑意义上的装置, 是通过其所在电子设备的处理器将非易失性存储器等可读介质中对应的计算机程序指令读取到内存中运行形成的。从硬件层面而言, 如图 5 所示, 为本发明语音识别装置所在电子设备的一种硬件结构图, 除了图 5 所示的处理器、内存、网络接口、以及非易失性存储器之外, 实施例中装置所在的电子设备通常根据该电子设备的实际功能, 还可以包括其他硬件, 对此不再赘述。电子设备的存储器可以存储处理器可执行

的程序指令；处理器可以耦合存储器，用于读取所述存储器存储的程序指令，并作为响应，执行如下操作：获取待识别的语音信号的特征分类结果；所述特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率；基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤；基于过滤后的特征分类结果识别所述语音信号。

[110] 在其他实施例中，处理器所执行的操作可以参考上文方法实施例中相关的描述，在此不予赘述。

[111] 以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明保护的范围之内。

权利要求书

1.一种语音识别方法，其特征在于，包括步骤：

获取待识别的语音信号的特征分类结果；所述特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率；

5 基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤；

基于过滤后的特征分类结果识别所述语音信号。

2.根据权利要求 1 所述的方法，其特征在于，所述基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤，包括：

判断任一语音信号帧映射到对应的发音的概率是否满足预定过滤规则；

10 如果所述对应的发音满足预定过滤规则，对所述对应的发音进行滤掉。

3.根据权利要求 2 所述的方法，其特征在于：

如果任一语音信号帧映射到对应的发音的概率，与该语音信号帧的最大映射概率之间的概率差，在预定的差值范围内，则确定所述对应的发音满足预定过滤规则；

15 如果任一语音信号帧映射到对应的发音的概率，小于该语音信号帧映射到预定数目的发音中各发音的概率，则确定所述对应的发音满足预定过滤规则。

4.根据权利要求 3 所述的方法，其特征在于，所述预定数目为以下任一：

该帧语音信号帧对应的发音中被保留在特征分类结果内的发音的数量；

预定的比例阈值与该帧语音信号帧对应的发音的总数目的乘积。

20 5.根据权利要求 1 所述的方法，其特征在于，所述基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤，包括：

获取任一语音信号帧映射到各发音的概率的直方图分布；

获取与所述直方图分布对应的束宽；

将概率分布在所述束宽之外的发音，确定为满足预定过滤规则的发音；

将满足预定过滤规则的发音，从所述特征分类结果所含的发音中删除。

25 6.根据权利要求 1 至 5 中任一项所述的方法，其特征在于，所述基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤，包括：

如果任一语音信号帧映射到对应的发音的概率满足预定过滤规则，将该发音确定为候选发音；

30 如果该语音信号帧的预定帧数的相邻语音信号帧中的任一帧，映射到该候选发音的概率满足预定过滤规则，则将该候选发音从所述特征分类结果所含的发音中删除；

如果该语音信号帧的预定帧数的相邻语音信号帧，映射到该候选发音的概率均不满足预

定过滤规则，则将该候选发音保留在所述特征分类结果所含的发音中。

7.一种语音识别装置，其特征在于，包括：

分类结果获取模块，用于获取待识别的语音信号的特征分类结果；所述特征分类结果包含用于描述各语音信号帧的发音特征的发音以及各语音信号帧映射到对应的发音的概率；

5 发音过滤模块，用于基于所述特征分类结果所含的概率，对所述特征分类结果所含的发音进行过滤；

语音识别模块，用于基于过滤后的特征分类结果识别所述语音信号。

8.根据权利要求7所述的装置，其特征在于，所述发音过滤模块还包括：

10 第一过滤模块，用于在任一语音信号帧映射到对应的发音的概率，与该语音信号帧的最大映射概率之间的概率差，在预定的差值范围内时，对所述对应的发音进行过滤；

第二过滤模块，用于在任一语音信号帧映射到对应的发音的概率，小于该语音信号帧映射到预定数目的发音中各发音的概率时，对所述对应的发音进行过滤。

9.根据权利要求7所述的装置，其特征在于，所述发音过滤模块包括：

概率分布模块，用于获取任一语音信号帧映射到各发音的概率的直方图分布；

15 束宽确定模块，用于获取与所述直方图分布对应的束宽；

发音确定模块，用于将概率分布在所述束宽之外的发音，确定为满足所述预定过滤规则的发音；

发音删除模块，用于将满足所述预定过滤规则的发音从所述特征分类结果所含的发音中删除。

20 10.根据权利要求7至9中任一项所述的装置，其特征在于，所述发音过滤模块包括：

候选发音确定模块，用于在任一语音信号帧映射到对应的发音的概率满足预定过滤规则时，将该发音确定为候选发音；

25 候选发音删除模块，用于在该语音信号帧的预定帧数的相邻语音信号帧中的任一帧，映射到该候选发音的概率满足预定过滤规则时，将该候选发音从所述特征分类结果所含的发音中删除；

候选发音保留模块，用于在该语音信号帧的预定帧数的相邻语音信号帧，映射到该候选发音的概率均不满足预定过滤规则时，将该候选发音保留在所述特征分类结果所含的发音中。

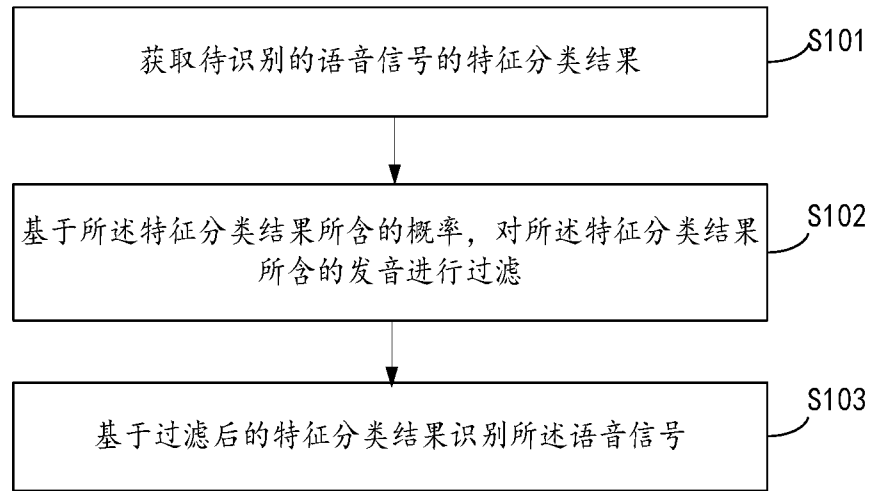


图 1

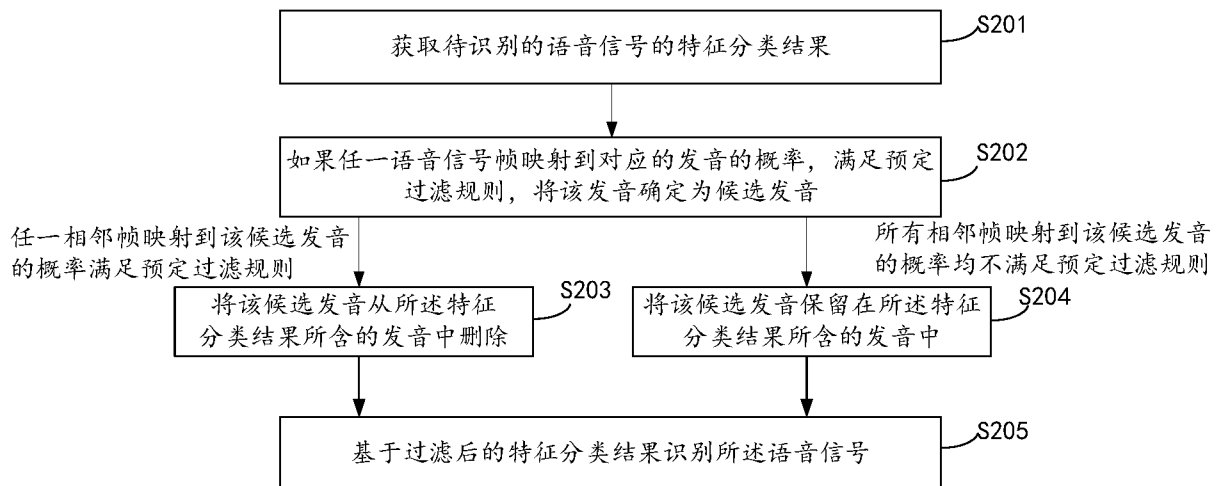


图 2

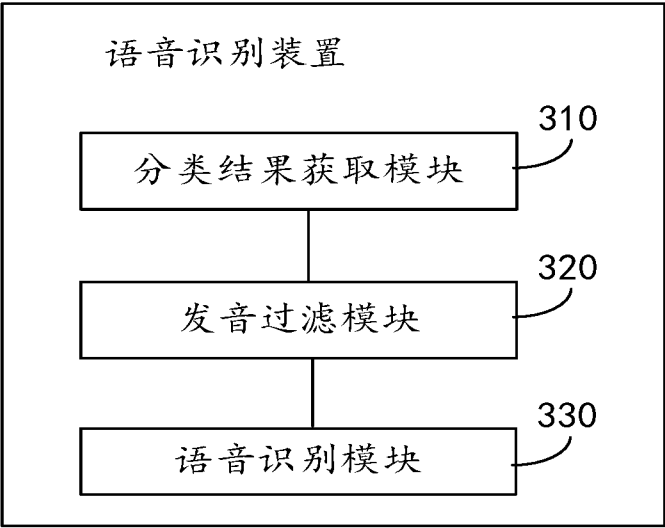


图 3

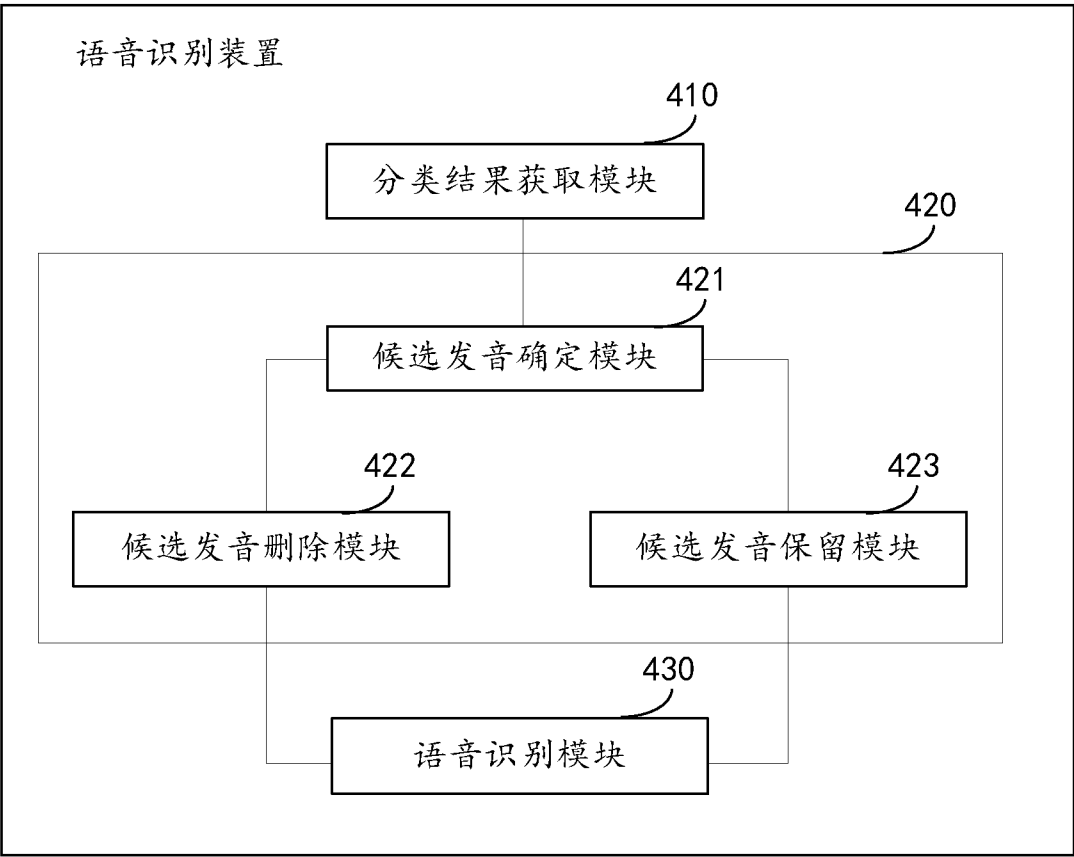


图 4

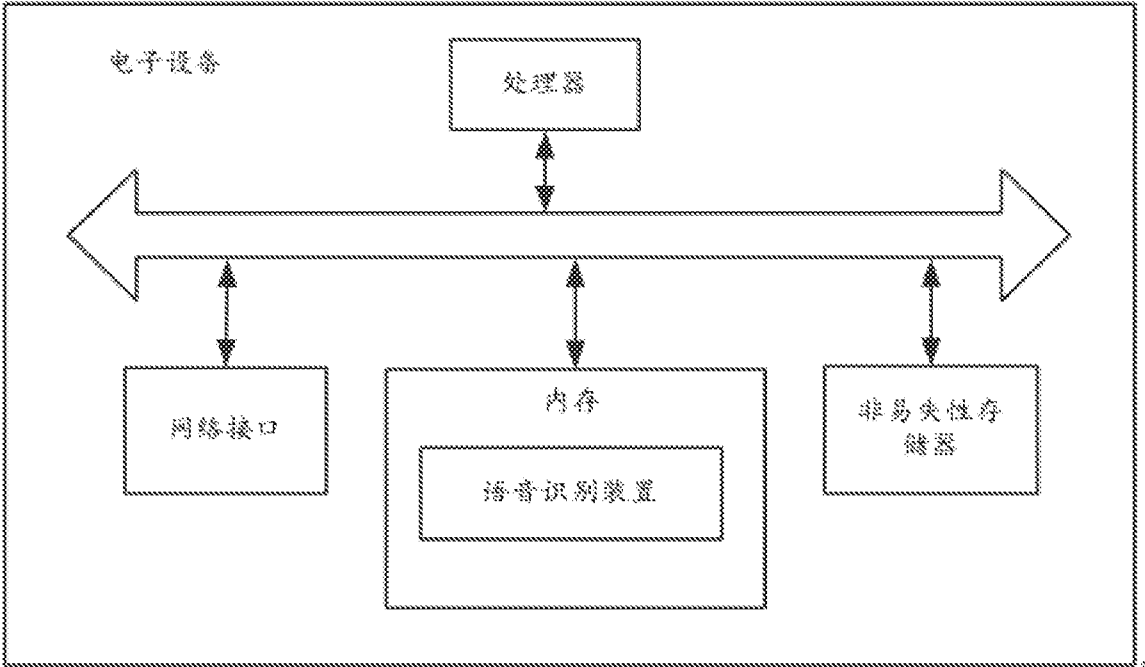


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/104382

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/02 (2006.01) i

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)

G10L 15/-

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)

CNPAT; CNKI; EPODOC; WPI: 语音, 音频, 话音, 特征, 过滤, 识别, 滤除, 删除, sound, voice, speech, character, earmark, filtrate, delete

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106875936 A (GUANGZHOU CVTE ELECTRONIC TECHNOLOGY CO., LTD.) 20 June 2017 (20.06.2017), claims 1-10	1-10
X	CN 102779510 A (SOUTHEAST UNIVERSITY) 14 November 2012 (14.11.2012), claim 1, and figure 1	1-10
A	US 9396722 B2 (ELECTRONIC AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 19 July 2016 (19.07.2016), entire document	1-10
A	CN 101030369 A (TSINGHUA UNIVERSITY) 05 September 2007 (05.09.2007), entire document	1-10
A	CN 101101751 A (TOSHIBA CORP.) 09 January 2008 (09.01.2008), entire document	1-10

☐ 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
27 December 2017

Date of mailing of the international search report
16 January 2018

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
LIU, Zihan
Telephone No. (86-10) 62414489

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/104382

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106875936 A	20 June 2017	None	
CN 102779510 A	14 November 2012	CN 102779510 B	18 December 2013
US 9396722 B2	19 July 2016	US 2014379345 A1	25 December 2014
		KR 20140147587 A	30 December 2014
CN 101030369 A	05 September 2007	CN 101030369 B	29 June 2011
CN 101101751 A	09 January 2008	US 2008281595 A1	13 November 2008
		JP 4241771 B2	18 March 2009
		US 7895040 B2	22 February 2011
		JP 2008015120 A	24 January 2008

国际检索报告

国际申请号

PCT/CN2017/104382

A. 主题的分类 G10L 15/02 (2006.01) i 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) G10L15/- 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)) CNPAT;CNKI;EPODOC;WPI: 语音, 音频, 话音, 特征, 过滤, 识别, 滤除, 删除, sound, voice, speech, character, earmark, filtrate, delete		
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
PX	CN 106875936 A (广州视源电子科技有限公司) 2017年 6月 20日 (2017 - 06 - 20) 权利要求1-10	1-10
X	CN 102779510 A (东南大学) 2012年 11月 14日 (2012 - 11 - 14) 权利要求1、附图1	1-10
A	US 9396722 B2 (ELECTRONIC AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 2016年 7月 19日 (2016 - 07 - 19) 全文	1-10
A	CN 101030369 A (清华大学) 2007年 9月 5日 (2007 - 09 - 05) 全文	1-10
A	CN 101101751 A (株式会社东芝) 2008年 1月 9日 (2008 - 01 - 09) 全文	1-10
☐ 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 2017年 12月 27日		国际检索报告邮寄日期 2018年 1月 16日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (ISA/CN) 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 刘子茜 电话号码 (86-10)62414489

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/104382

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	106875936	A	2017年 6月 20日	无	
CN	102779510	A	2012年 11月 14日	CN	102779510 B 2013年 12月 18日
US	9396722	B2	2016年 7月 19日	US	2014379345 A1 2014年 12月 25日
				KR	20140147587 A 2014年 12月 30日
CN	101030369	A	2007年 9月 5日	CN	101030369 B 2011年 6月 29日
CN	101101751	A	2008年 1月 9日	US	2008281595 A1 2008年 11月 13日
				JP	4241771 B2 2009年 3月 18日
				US	7895040 B2 2011年 2月 22日
				JP	2008015120 A 2008年 1月 24日