



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))。

---

(57) 摘要: 本申请提出一种语音质量评价方法和装置, 该语音质量评价方法包括: 接收待评价语音数据; 提取所述待评价语音数据的评价特征; 根据所述待评价语音数据的评价特征和已构建的语音质量评价模型, 对所述待评价语音数据进行质量评价, 其中, 所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。该方法能够扩展语音质量评价的应用范围。

## 语音质量评价方法和装置

### 相关申请的交叉引用

本申请要求科大讯飞股份有限公司于 2016 年 10 月 12 日提交的、发明名称为“语音质量评价方法和装置”的、中国专利申请号“201610892176.1”的优先权。

### 技术领域

本申请涉及通信技术领域，尤其涉及一种语音质量评价方法和装置。

### 背景技术

随着技术的不断发展，通信在人们生活中的地位越来越重要，如采用通信网络进行语音数据的传输。语音质量是评价通信网络质量的一个重要因素。为了达到评价语音质量的目的，开发有效的语音质量评价算法是必须的。

相关技术中，通信网络中的语音质量评价算法包括语音质量的感知评价 ( Perceptual Evaluation of Speech Quality , PESQ ) 算法和感知客观语音质量评价 ( Perceptual Objective Listening Quality Analysis , POLQA ) 算法。这些算法在实现时需要获取输入语音数据和输出语音数据，输入语音数据一般为干净语音数据，输出语音数据一般为经过通信网络后的退化语音数据，通过对

输入语音数据和输出语音数据进行分析，对输出语音数据进行质量评价。输入语音数据一般是采用运营商的路测车采集的，但是，在小区楼层或商场等室内条件下，无法通过路测车进行采集，因此无法获取到输入语音数据，也就不能通过输入语音数据进行语音质量评价，使得上述基于输入语音数据和输出语音数据对输出语音数据进行语音质量评价的算法存在应用局限性。

## 发明内容

本申请旨在至少在一定程度上解决相关技术中的技术问题之一。

为此，本申请的一个目的在于提出一种语音质量评价方法，该方法可以在对待评价语音数据进行语音质量评价时，不需要相应的输入语音数据，实现仅依赖单端语音数据的语音质量评价，从而扩展应用范围。

本申请的另一个目的在于提出一种语音质量评价装置。

为达到上述目的，本申请第一方面实施例提出的语音质量评价方法，包括：接收待评价语音数据；提取所述待评价语音数据的评价特征；根据所述待评价语音数据的评价特征和已构建的语音质量评价模型，对所述待评价语音数据进行质量评价，其中，所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。

本申请第一方面实施例提出的语音质量评价方法，通过采用语音质量评价模型对待评价语音数据进行质量评价，可以在语音质量评价时仅需要单端语音

数据，避免依赖双端语音数据造成的应用受限问题，从而扩展应用范围。

为达到上述目的，本申请第二方面实施例提出的语音质量评价装置，包括：接收模块，用于接收待评价语音数据；提取模块，用于提取所述待评价语音数据的评价特征；评价模块，用于根据所述待评价语音数据的评价特征和已构建的语音质量评价模型，对所述待评价语音数据进行质量评价，其中，所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。

本申请第二方面实施例提出的语音质量评价装置，通过采用语音质量评价模型对待评价语音数据进行质量评价，可以在语音质量评价时仅需要单端语音数据，避免依赖双端语音数据造成的应用受限问题，从而扩展应用范围。

本申请实施例还提出了一种设备，包括：一个或多个处理器；用于存储一个或多个程序的存储器；当所述一个或多个程序被所述一个或多个处理器执行时，使得所述一个或多个处理器执行如本申请第一方面实施例任一项所述的方法。

本申请实施例还提出了一种非临时性计算机可读存储介质，当所述存储介质中的一个或多个程序由设备的一个或多个处理器执行时，使得所述一个或多个处理器执行如本申请第一方面实施例任一项所述的方法。

本申请实施例还提出了计算机程序产品，当所述计算机程序产品被设备中的一个或多个处理器执行时，使得所述一个或多个处理器执行如本申请第一方

面实施例任一项所述的方法。

本申请附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本申请的实践了解到。

## 附图说明

本申请上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

图 1 是本申请一个实施例提出的语音质量评价方法的流程示意图；

图 2 是本申请另一个实施例提出的语音质量评价方法的流程示意图；

图 3 是本申请一个实施例提出的语音质量评价装置的结构示意图；

图 4 是本申请另一个实施例提出的语音质量评价装置的结构示意图。

## 具体实施方式

下面详细描述本申请的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的模块或具有相同或类似功能的模块。下面通过参考附图描述的实施例是示例性的，仅用于解释本申请，而不能理解为对本申请的限制。相反，本申请的实施例包括落入所附加权利要求书的精神和内涵范围内的所有变化、修改和等同物。

为了解决 PESQ 算法的问题以及更好的适应 4G/LTE 时代的语音质量评价需求，ITU-T 于 2006 年开始了 POLQA 算法的开发工作，2011 年初正式发布为

ITU-T P.863 标准。主要特点可以覆盖最新的语音编码和网络传输技术，在用于 3G，4G/LTE，VoIP 网络时具有更高的准确性支持超宽带（50Hz~14KHz）语音传输，高质量语音传输。因此，POLQA 算法是目前评价通信网络的语音质量的通常被选择的算法。

深度学习的概念源于人工神经网络的研究。含多隐层的多层感知器就是一种深度学习结构。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示。目前深度学习的应用领域主要包括：计算机视觉、语音识别的声学模型训练、机器翻译和语义挖掘等自然语言处理领域。

由于深度学习是新出现的技术，依然在不断发展中，因此目前比较成功的应用领域仅如上所示的有限领域。依据发明人所知，在通信领域，特别是通信领域的语音质量评价并未应用。

本申请的发明人作为通信领域的技术人员，在需要完成语音质量评价时，之前也通常采用 POLQA 算法，但是，发明人发现 POLQA 算法需要双端的语音数据，即在评价输出语音数据的语音质量时，不仅需要输出语音数据还需要输入语音数据，由于在一些情况下难以获取输入语音数据，就会使得 POLQA 算法的应用受限。为了避免应用受限问题，需要提出新的解决方案。发明人通过进一步分析发现，深度学习构建的模型具有优良的性能，因此可以将深度学习引入到语音质量评价算法中。进一步的，为了避免双端语音数据存在的应用

局限性问题的，在通过深度学习构建模型时，可以仅采用单端语音数据作为样本进行训练，从而在采用已构建的模型进行语音质量评价时，可以仅需要待评价语音数据这一单端语音数据。

因此，本申请的主要思路是将深度学习引入到语音质量评价，特别是通信领域的语音质量评价中。将为通信领域的语音质量评价提供仅依赖单端语音数据的新的解决方案，并且在仅依赖单端语音数据时，采用深度学习方式构建模型，可以保证模型的优良性能，从而解决受限更少性能更优的语音质量评价的技术问题。进一步的，需要说明的是，虽然上述对本申请的主要思路进行了说明，但是，具体的技术方案不限于上述的主要思路，还可以与其他特征相互结合，这些不同技术特征之间的结合依然属于本申请的保护范围。

进一步的，需要说明的是，虽然上述给出了主要解决的技术问题，但是，本申请并不限于仅能解决上述技术问题，应用本申请还可以解决的其他技术问题依然属于本申请的保护范围。

进一步的，需要说明的是，虽然上述给出了本申请的主要思路，以及后续实施例会对一些特别点进行说明。但是，本申请的创新点并不限于上述的主要思路及特别点所涉及的内容，并不排除本申请中一些并未特殊说明的内容依然可能会包含本申请的创新点。

可以理解的是，虽然上述进行了一些说明，但依然不排除其他可能方案，因此，与后续本申请给出的实施例相同、相似、等同等情况的技术方案依然属

于本申请的保护范围。

下面将结合具体实施例对本申请的技术方案进行说明。

图 1 是本申请一个实施例提出的语音质量评价方法的流程示意图。

如图 1 所示，本实施例的方法包括：

S11：接收待评价语音数据。

以通信领域为例，待评价语音数据可以具体是指通信网络的输出语音数据，即输入语音数据经过通信网络后的退化语音数据。输入语音数据一般是指干净语音数据或者称为原始语音数据，而退化语音数据一般是指相对于原始语音数据存在清楚度退化、存在延迟、杂音等一项或多项内容的质量退化的语音数据。

S12：提取所述待评价语音数据的评价特征。

所述评价特征与语音质量模型构建时，对退化语音数据提取的评价特征相同，具体可以根据应用需求确定。

一般来讲，评价特征是指从人耳听觉感知的角度描述语音数据的特征，具体内容可以参见后续描述。

S13：根据所述待评价语音数据的评价特征和已构建的语音质量评价模型，对所述待评价语音数据进行质量评价，其中，所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。

其中，语音质量评价模型可以是在需要进行语音质量评价之前预先构建的，例如，通过离线方式先构建出语音质量评价模型，在需要语音质量评价时，就

可以直接采用预先构建的语音质量评价模型。当然，也并不排除语音质量评价模型是在线构建的，比如在需要进行语音质量评价时在线构建的。具体构建内容可以参见后续描述。

语音质量评价模型的输入和输出分别是单端语音数据的评价特征和质量信息，因此，在提取出待评价语音数据的评价特征后，可以将该评价特征作为语音质量评价模型的输入，从而得到的输出就是待评价语音数据的质量信息，实现语音质量评价。

进一步的，语音质量评价模型可以用回归模型或分类模型描述，在不同描述情况下，上述的质量信息的具体内容可以是不同的。例如，如果语音质量评价模型采用回归模型描述，则得到的质量信息是具体的评价得分，如 1-5 分中的一个得分；如果语音质量评价模型采用分类模型描述，则得到的质量信息是评价类别，如差、较差、一般、好、较好中的一个类别。

进一步的，一些实施例中，为了提高语音质量评价的准确度，还可以对 S13 得到的质量评价结果进行规整。以质量评价结果是评价得分为例，在规整时，可以将 S13 得到的评价得分直接作为最终的评价得分，或者，也可以结合通信网络的丢包、抖动、时延等相关参数对 S13 得到的评价得分进行规整得到最终的评价得分。具体的结合网络参数进行规整的算法可以设置，在此不再详述，比如可以在 S13 得到的评价得分的基础上乘以一个系数作为最终的评价得分，该系数与通信网络的上述参数相关。

本实施例中,通过采用语音质量评价模型对待评价语音数据进行质量评价,可以在语音质量评价时仅需要单端语音数据,避免依赖双端语音数据造成的应用受限问题,从而扩展应用范围。

图 2 是本申请另一个实施例提出的语音质量评价方法的流程示意图。

本实施例以待评价语音数据是经过通信网络后的退化语音数据为例。在构建语音质量评价模型时以深度学习方式构建为例。

参见图 2,本实施例的方法包括:

S21:获取语音数据,所述语音数据包括干净语音数据和退化语音数据。

其中,可以采用收集和/或从已有数据中直接获取的方式,获取到语音数据。为了提高构建的语音质量评价模型的准确度,此处应该获取到尽可能多的语音数据。

以收集方式为例,具体收集语音数据时,可以采用模拟通信的方式,分别收集到通话时的干净语音数据和经过通信网络后的退化语音数据,具体地先从高保真录音室采集大量干净语音数据,如 2000 小时的干净语音数据;然后利用多部手机模拟通话方式,即使用一部手机拨打电话播放所述干净语音数据,另一部手机接听这些干净语音数据,通过在通信网络上不同的接口处,还原发送的数据包,得到经过通信网络后的退化语音数据。

当然,也可直接收集真实的网络通话语音数据,分别获取相应干净语音数据和退化语音数据,具体获取方式本申请不作限定。

进一步的，在收集语音数据时，干净语音数据和退化语音数据可以分开收集，从而可以直接分别获取到干净语音数据和退化语音数据。或者，在收集语音数据时，干净语音数据和退化语音数据可以一起收集，此时可以分别对干净语音数据和退化语音数据进行标记，以区分干净语音数据和退化语音数据，如使用 1 表示干净语音数据，0 表示退化语音数据，此时，可以根据标记分别获取到干净语音数据和退化语音数据。

S22：根据所述干净语音数据获取待处理的干净语音数据，以及，根据所述退化语音数据获取待处理的退化语音数据。

可以包括：

将获取的退化语音数据直接作为待处理的退化语音数据；或者，

提取获取的退化语音数据的有效语音段，将退化语音数据的有效语音段作为待处理的退化语音数据；或者，

对获取的退化语音数据进行聚类，将聚类中心对应的退化语音数据作为待处理的退化语音数据；或者，

提取获取的退化语音数据的有效语音段，对退化语音数据的有效语音段进行聚类，将聚类中心对应的退化语音数据的有效语音段作为待处理的退化语音数据。

具体的，在获取到干净语音数据和退化语音数据后，可以将获取的干净语音数据和退化语音数据直接分别作为待处理的干净语音数据和待处理的退化语

音数据。进一步的，还可以在获取到干净语音数据和退化语音数据后，分别进行有效语音段的提取，将提取得到的干净语音数据的有效语音段作为待处理的干净语音数据，将退化语音数据的有效语音段作为待处理的退化语音数据。具体的提取有效语音段的方式不限定，例如采用语音活动检测（Voice Activity Detection，VAD）方式。通过仅处理有效语音段，可以减少运算量和复杂度。

进一步的，在获取待处理的退化语音数据时，可以将语音数据中包括的所有退化语音数据或所有退化语音数据的有效语音段作为待处理的退化语音数据，或者，也可以选择部分退化语音数据或其有效语音段作为待处理的退化语音数据。在选择时，可以采用聚类方式，对所有的退化语音数据或其有效语音段进行聚类，将聚类中心对应的退化语音数据或其有效语音段作为待处理的退化语音数据。

例如，在聚类时，提取退化语音数据的有效语音段的 ivector 特征，使用 k-means 方法对提取的 ivector 特征进行聚类，得到 k 个聚类中心，将每个聚类中心对应的退化语音数据或其有效语音段作为待处理的退化语音数据。通过聚类以及只选择聚类中心对应的退化语音数据进行处理，可以减少数据量，提高运算效率。

S23：根据待处理的干净语音数据和待处理的退化语音数据，计算待处理的退化语音数据的评价得分。

以待处理的数据是有效语音段为例，在得到干净语音数据的有效语音段和

退化语音数据的有效语音段后，可以根据干净语音数据的有效语音段，对退化语音数据每个有效语音段进行逐帧分析，计算得到退化语音数据的有效语音段的评价得分。计算方式不限定，例如，所述评价得分为语音数据的平均意见分 ( Mean Opinion Score , MOS ) 得分，具体计算方法可以与现有技术相同，如使用 POLQA 算法或 PESQ 算法计算得到，在此不再详述。

S24：提取待处理的退化语音数据的评价特征。

所述评价特征从人耳听觉感知的角度描述语音数据，具体提取时，先提取待处理的退化语音数据的时域特征，如语音数据的短时平均能量、语音的分段底噪、语音的短时波形冲击或者震荡、基频特征及基频的差分特征，如基频特征的一阶、二阶差分值等；然后再提取待处理的退化语音数据的频域特征，所述频域特征如 FilterBank 特征、线性预测编码(linear predictive coding , LPC)特征等；所述频域特征提取时，采用能够描述人听觉的耳蜗形状的滤波器，从而使得提取的频域特征能够从人耳听觉感知角度描述语音数据；为了更好的描述退化语音数据，还可以提取每种频域特征的均值、方差、最大值、最小值、及差分特征，如一阶、二阶差分值等；具体提取哪种评价特征可以根据应用需求及语音数据的退化情况确定，具体本申请不作限定。

S25：根据所述提取待处理的退化语音数据的评价特征和所述待处理的退化语音数据的评价得分进行训练，构建语音质量评价模型。

在训练时，可以具体采用深度学习方式训练得到语音质量评价模型的参数，

构建出语音质量评价模型。

深度学习方式采用的网络拓扑结构可以为深度神经网络 ( (Deep Neural Networks , DNN)、卷积神经网络(Convolutional Neural Networks , CNN)、循环神经网络 ( Recurrent Neural Networks , RNN )、长短时记忆 ( Long Short-Term Memory ,LSTM )神经网络等网络中的一种或多种组合 ,具体本申请不作限定 ; 具体网络的选择根据应用需求确定。在确定模型的输入和输出后 , 参数训练过程与现有技术相同 , 在此不再详述。

进一步的 , 语音质量评价模型可以用不同类型的模型描述 , 如可以采用回归模型或分类模型描述。在不同类型下 , 模型对应的输入和输出可以相应调整。

具体的 , 采用回归模型描述语音质量评价模型时 , 直接将上述获取的待处理的退化语音数据的评价特征和待处理的退化语音数据的评价得分分别作为模型输入和输出。

采用分类模型描述语音质量评价模型时 , 直接将上述获取的待处理的退化语音数据的评价特征作为模型输入 , 而模型输出是对待处理的退化语音数据的评价得分进行量化后 , 得到的评价类别。

具体量化时 , 可以采用固定步长或非固定步长对退化语音数据的评价得分进行量化 , 如果采用固定步长时 , 固定步长为 0.2 , 对所有退化语音数据的评价得分进行量化 , 得到量化后退化语音数据的类别 , 以 MOS 得分为例 , 以固定步长 0.2 进行量化时 , 1 分到 5 分量化后可以得到 20 个评价类别 ; 如果采用非

固定步长时，可以根据应用需求确定退化语音数据的每个范围内评价得分的量化步长，如评价得分较低范围内，可以采用大步长量化，评价得分较高范围内，可以采用小步长量化；以 MOS 得分为例，如 1 到 3 分为较低得分范围，可以采用大步长量化，如 0.5；3 到 5 分为较高得分范围，可以采用小步长量化，如 0.2，量化后共可以得到 14 个评价类别；

当然，也可以采用其它方法对所述评价得分进行量化，将评价得分划分为多个评价类别，如量化后的所述评价类别为差、较差、一般、好、较好，具体本申请不作限定。

S26：接收经过通信网络后的退化语音数据。

S27：提取所述退化语音数据的评价特征。

评价特征的提取方式与训练过程中的提取方式相同，在此不再详述。

S28：根据所述评价特征和所述已构建的语音质量评价模型，对所述退化语音数据进行质量评价。

具体的，将当前的退化语音数据的评价特征作为语音质量评价模型的输入，将模型的输出作为对当前的退化语音数据的质量评价结果。其中，如果语音质量评价模型采用回归模型描述，则质量评价结果是评价得分，如果语音质量评价模型采用分类模型描述，则质量评价结果是评价类别。

本实施例中，通过采用语音质量评价模型对待评价语音数据进行质量评价，可以在语音质量评价时仅需要单端语音数据，避免依赖双端语音数据造成的应

用受限问题,从而扩展应用范围。进一步的,通过采用深度学习方式进行训练,可以利用深度学习方式的优良性能,使得语音质量评价模型更准确,进而语音质量评价结果更准确。进一步的,通过对通信领域的语音数据进行质量评价,可以将深度学习与通信领域的语音质量评价相结合,为通信领域的语音质量评价提供新的解决思路。

图 3 是本申请一个实施例提出的语音质量评价装置的结构示意图。

如图 3 所示,本实施例的装置 30 包括:接收模块 31、提取模块 32 和评价模块 33。

接收模块 31,用于接收待评价语音数据;

提取模块 32,用于提取所述待评价语音数据的评价特征;

评价模块 33,用于根据所述待评价语音数据的评价特征和已构建的语音质量评价模型,对所述待评价语音数据进行质量评价,其中,所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。

一些实施例中,所述待评价语音数据包括:经过通信网络后的退化语音数据。

一些实施例中,参见图 4,本实施例的装置 30 还包括:用于构建语音质量评价模型的构建模块 34,所述构建模块 34 包括:

第一获取子模块 341,用于获取语音数据,所述语音数据包括干净语音数

据和退化语音数据；

第二获取子模块 342，用于根据所述干净语音数据获取待处理的干净语音数据，以及，根据所述退化语音数据获取待处理的退化语音数据；

计算子模块 343，用于根据待处理的干净语音数据和待处理的退化语音数据，计算待处理的退化语音数据的评价得分；

提取子模块 344，用于提取待处理的退化语音数据的评价特征；

训练子模块 345，用于根据所述待处理的退化语音数据的评价特征和所述待处理的退化语音数据的评价得分进行训练，构建语音质量评价模型。

一些实施例中，所述语音质量评价模型是采用深度学习方式进行训练后构建的。

一些实施例中，所述训练子模块 345 具体用于：

如果采用回归模型描述所述语音质量评价模型，则将所述待处理的退化语音数据的评价特征和所述待处理的退化语音数据的评价得分，分别作为模型输入和模型输出，训练模型参数，构建语音质量评价模型；或者，

如果采用分类型描述所述语音质量评价模型，则将所述待处理的退化语音数据的评价特征作为模型输入，对所述待处理的退化语音数据的评价得分进行量化，得到评价类别，将评价得分类别作为模型输出，训练模型参数，构建语音质量评价模型。

一些实施例中，所述第二获取子模块 342 用于根据所述干净语音数据获取

待处理的干净语音数据，包括：

将获取的干净语音数据直接作为待处理的干净语音数据；或者，

提取获取的干净语音数据的有效语音段，将干净语音数据的有效语音段作为待处理的干净语音数据。

一些实施例中，所述第二获取子模块 342 用于根据所述退化语音数据获取待处理的退化语音数据，包括：

将获取的退化语音数据直接作为待处理的退化语音数据；或者，

提取获取的退化语音数据的有效语音段，将退化语音数据的有效语音段作为待处理的退化语音数据；或者，

对获取的退化语音数据进行聚类，将聚类中心对应的退化语音数据作为待处理的退化语音数据；或者，

提取获取的退化语音数据的有效语音段，对退化语音数据的有效语音段进行聚类，将聚类中心对应的退化语音数据的有效语音段作为待处理的退化语音数据。

可以理解的是，本实施例的装置与上述方法实施例对应，具体内容可以参见方法实施例的相关描述，在此不再详细说明。

本实施例中，通过采用语音质量评价模型对待评价语音数据进行质量评价，可以在语音质量评价时仅需要单端语音数据，避免依赖双端语音数据造成的应用受限问题，从而扩展应用范围。

本申请实施例还提出了一种设备，包括：一个或多个处理器；用于存储一个或多个程序的存储器；当所述一个或多个程序被所述一个或多个处理器执行时，使得所述一个或多个处理器执行如下方法：接收待评价语音数据；提取所述待评价语音数据的评价特征；根据所述待评价语音数据的评价特征和已构建的语音质量评价模型，对所述待评价语音数据进行质量评价，其中，所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。

本申请实施例还提出了一种非临时性计算机可读存储介质，当所述存储介质中的一个或多个程序由设备的一个或多个处理器执行时，使得所述一个或多个处理器执行如下方法：接收待评价语音数据；提取所述待评价语音数据的评价特征；根据所述待评价语音数据的评价特征和已构建的语音质量评价模型，对所述待评价语音数据进行质量评价，其中，所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。

本申请实施例还提出了计算机程序产品，当所述计算机程序产品被设备中的一个或多个处理器执行时，使得所述一个或多个处理器执行如下方法：接收待评价语音数据；提取所述待评价语音数据的评价特征；根据所述待评价语音数据的评价特征和已构建的语音质量评价模型，对所述待评价语音数据进行质量评价，其中，所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。

可以理解的是，上述各实施例中相同或相似部分可以相互参考，在一些实施例中未详细说明的内容可以参见其他实施例中相同或相似的内容。

需要说明的是，在本申请的描述中，术语“第一”、“第二”等仅用于描述目的，而不能理解为指示或暗示相对重要性。此外，在本申请的描述中，除非另有说明，“多个”的含义是指至少两个。

流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为，表示包括一个或多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分，并且本申请的优选实施方式的范围包括另外的实现，其中可以不按所示出或讨论的顺序，包括根据所涉及的功能按基本同时的方式或按相反的顺序，来执行功能，这应被本申请的实施例所属技术领域的技术人员所理解。

应当理解，本申请的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如，如果用硬件来实现，和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列(PGA)，现场可编程门阵列(FPGA)等。

本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部

分步骤是可以通程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，该程序在执行时，包括方法实施例的步骤之一或其组合。

此外，在本申请各个实施例中的各功能单元可以集成在一个处理模块中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用，也可以存储在一个计算机可读存储介质中。

上述提到的存储介质可以是只读存储器，磁盘或光盘等。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

尽管上面已经示出和描述了本申请的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本申请的限制，本领域的普通技术人员在本申请的范围内可以对上述实施例进行变化、修改、替换和变型。

## 权利要求书

1、一种语音质量评价方法，其特征在于，包括：

接收待评价语音数据；

提取所述待评价语音数据的评价特征；

根据所述待评价语音数据的评价特征和已构建的语音质量评价模型，对所述待评价语音数据进行质量评价，其中，所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。

2、根据权利要求1所述的方法，其特征在于，所述待评价语音数据包括：经过通信网络后的退化语音数据。

3、根据权利要求2所述的方法，其特征在于，还包括：构建语音质量评价模型，所述构建语音质量评价模型包括：

获取语音数据，所述语音数据包括干净语音数据和退化语音数据；

根据所述干净语音数据获取待处理的干净语音数据，以及，根据所述退化语音数据获取待处理的退化语音数据；

根据待处理的干净语音数据和待处理的退化语音数据，计算待处理的退化语音数据的评价得分；

提取待处理的退化语音数据的评价特征；

根据所述待处理的退化语音数据的评价特征和所述待处理的退化语音数据的评价得分进行训练，构建语音质量评价模型。

4、根据权利要求 1-3 任一项所述的方法，其特征在于，所述语音质量评价模型是采用深度学习方式进行训练后构建的。

5、根据权利要求 4 所述的方法，其特征在于，所述根据所述待处理的退化语音数据的评价特征和所述待处理的退化语音数据的评价得分进行训练，构建语音质量评价模型，包括：

如果采用回归模型描述所述语音质量评价模型，则将所述待处理的退化语音数据的评价特征和所述待处理的退化语音数据的评价得分，分别作为模型输入和模型输出，训练模型参数，构建语音质量评价模型；或者，

如果采用分类型描述所述语音质量评价模型，则将所述待处理的退化语音数据的评价特征作为模型输入，对所述待处理的退化语音数据的评价得分进行量化，得到评价类别，将评价得分类别作为模型输出，训练模型参数，构建语音质量评价模型。

6、根据权利要求 3 所述的方法，其特征在于，所述根据所述干净语音数据获取待处理的干净语音数据，包括：

将获取的干净语音数据直接作为待处理的干净语音数据；或者，

提取获取的干净语音数据的有效语音段，将干净语音数据的有效语音段作为待处理的干净语音数据。

7、根据权利要求 3 所述的方法，其特征在于，所述根据所述退化语音数据获取待处理的退化语音数据，包括：

将获取的退化语音数据直接作为待处理的退化语音数据；或者，

提取获取的退化语音数据的有效语音段，将退化语音数据的有效语音段作为待处理的退化语音数据；或者，

对获取的退化语音数据进行聚类，将聚类中心对应的退化语音数据作为待处理的退化语音数据；或者，

提取获取的退化语音数据的有效语音段，对退化语音数据的有效语音段进行聚类，将聚类中心对应的退化语音数据的有效语音段作为待处理的退化语音数据。

8、一种语音质量评价装置，其特征在于，包括：

接收模块，用于接收待评价语音数据；

提取模块，用于提取所述待评价语音数据的评价特征；

评价模块，用于根据所述待评价语音数据的评价特征和已构建的语音质量评价模型，对所述待评价语音数据进行质量评价，其中，所述语音质量评价模型用于表明单端语音数据的评价特征与所述单端语音数据的质量信息之间的关系。

9、根据权利要求 8 所述的装置，其特征在于，所述待评价语音数据包括：经过通信网络后的退化语音数据。

10、根据权利要求 9 所述的装置，其特征在于，还包括：用于构建语音质量评价模型的构建模块，所述构建模块包括：

第一获取子模块，用于获取语音数据，所述语音数据包括干净语音数据和退化语音数据；

第二获取子模块，用于根据所述干净语音数据获取待处理的干净语音数据，以及，根据所述退化语音数据获取待处理的退化语音数据；

计算子模块，用于根据待处理的干净语音数据和待处理的退化语音数据，计算待处理的退化语音数据的评价得分；

提取子模块，用于提取待处理的退化语音数据的评价特征；

训练子模块，用于根据所述待处理的退化语音数据的评价特征和所述待处理的退化语音数据的评价得分进行训练，构建语音质量评价模型。

11、根据权利要求 8-10 任一项所述的装置，其特征在于，所述语音质量评价模型是采用深度学习方式进行训练后构建的。

12、根据权利要求 11 所述的装置，其特征在于，所述训练子模块具体用于：

如果采用回归模型描述所述语音质量评价模型，则将所述待处理的退化语音数据的评价特征和所述待处理的退化语音数据的评价得分，分别作为模型输入和模型输出，训练模型参数，构建语音质量评价模型；或者，

如果采用分类型描述所述语音质量评价模型，则将所述待处理的退化语音数据的评价特征作为模型输入，对所述待处理的退化语音数据的评价得分进行量化，得到评价类别，将评价得分类别作为模型输出，训练模型参数，构建语音质量评价模型。

13、根据权利要求 10 所述的装置，其特征在于，所述第二获取子模块用于根据所述干净语音数据获取待处理的干净语音数据，包括：

将获取的干净语音数据直接作为待处理的干净语音数据；或者，

提取获取的干净语音数据的有效语音段，将干净语音数据的有效语音段作为待处理的干净语音数据。

14、根据权利要求 10 所述的装置，其特征在于，所述第二获取子模块用于根据所述退化语音数据获取待处理的退化语音数据，包括：

将获取的退化语音数据直接作为待处理的退化语音数据；或者，

提取获取的退化语音数据的有效语音段，将退化语音数据的有效语音段作为待处理的退化语音数据；或者，

对获取的退化语音数据进行聚类，将聚类中心对应的退化语音数据作为待处理的退化语音数据；或者，

提取获取的退化语音数据的有效语音段，对退化语音数据的有效语音段进行聚类，将聚类中心对应的退化语音数据的有效语音段作为待处理的退化语音数据。

15、一种设备，其特征在于，包括：

一个或多个处理器；

用于存储一个或多个程序的存储器；

当所述一个或多个程序被所述一个或多个处理器执行时，使得所述一个或

多个处理器执行如权利要求 1-7 任一项所述的方法。

16、一种非临时性计算机可读存储介质，其特征在于，当所述存储介质中的一个或多个程序由设备的一个或多个处理器执行时，使得所述一个或多个处理器执行如权利要求 1-7 任一项所述的方法。

17、一种计算机程序产品，其特征在于，当所述计算机程序产品被设备中的一个或多个处理器执行时，使得所述一个或多个处理器执行如权利要求 1-7 任一项所述的方法。

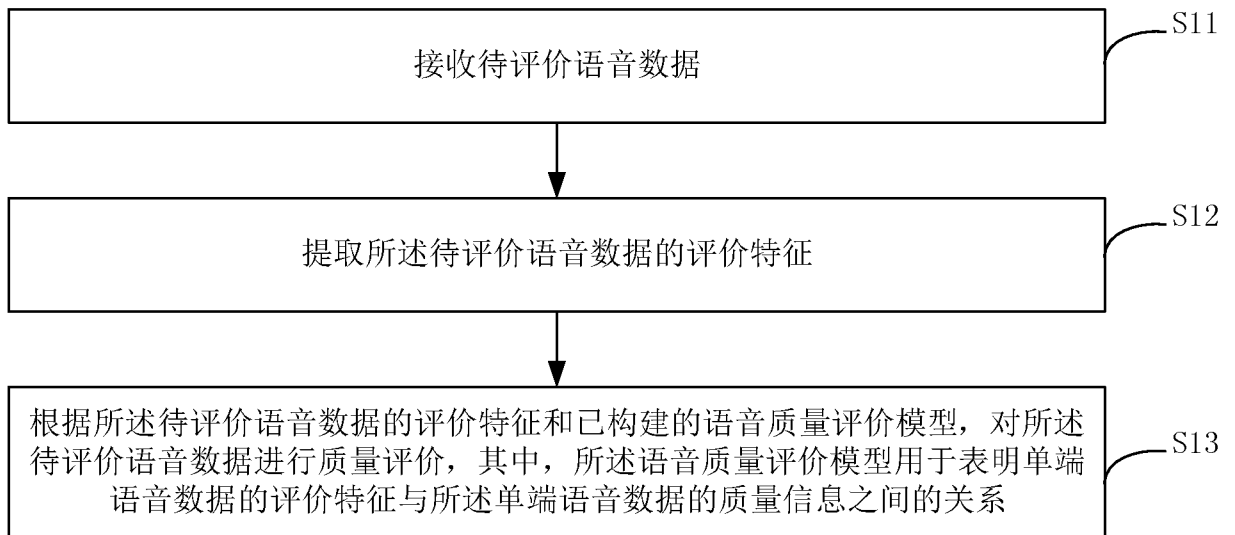


图 1

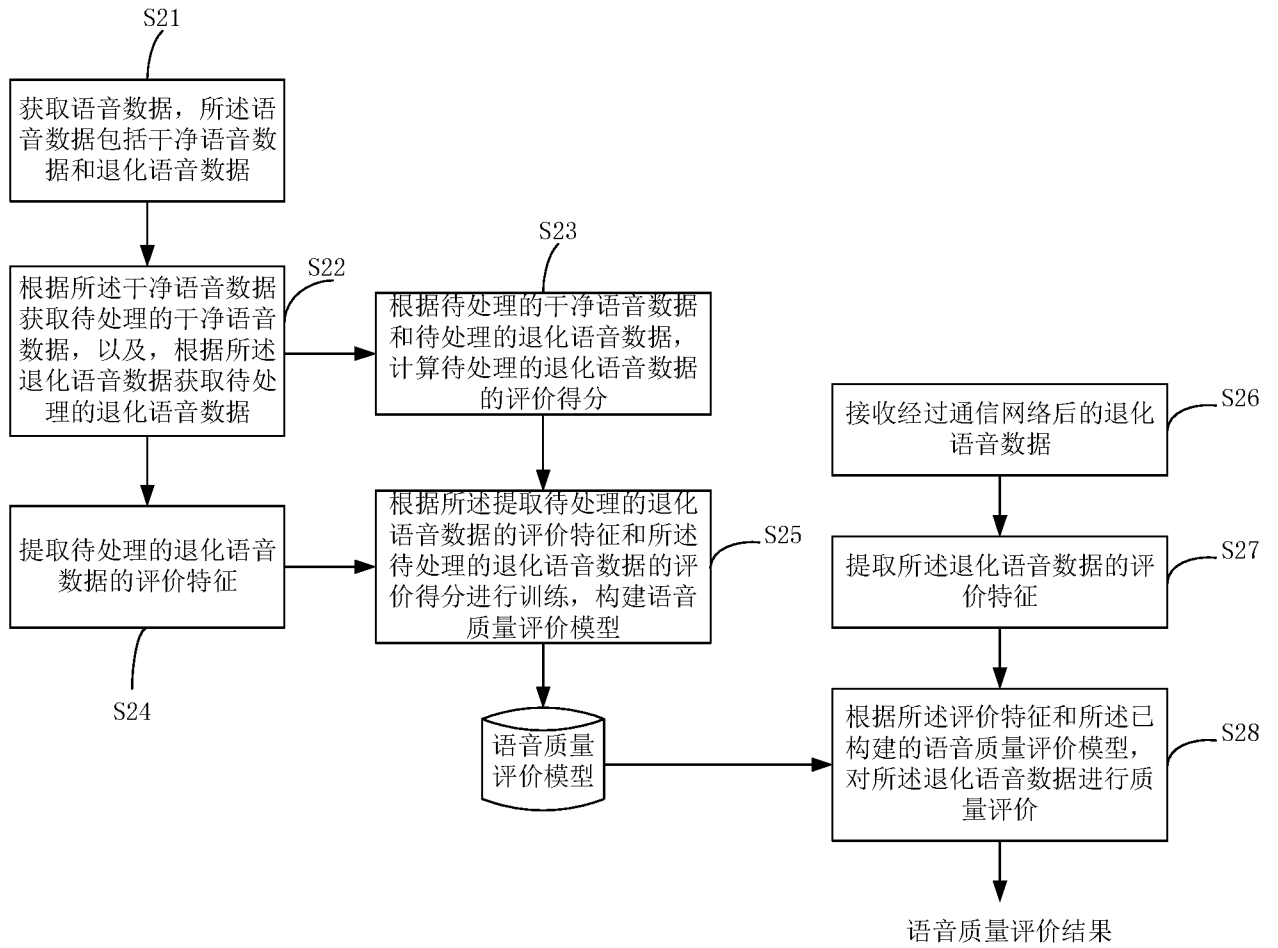


图 2

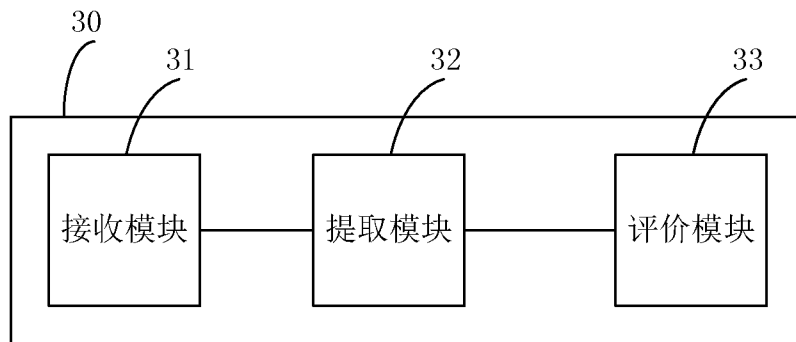


图 3

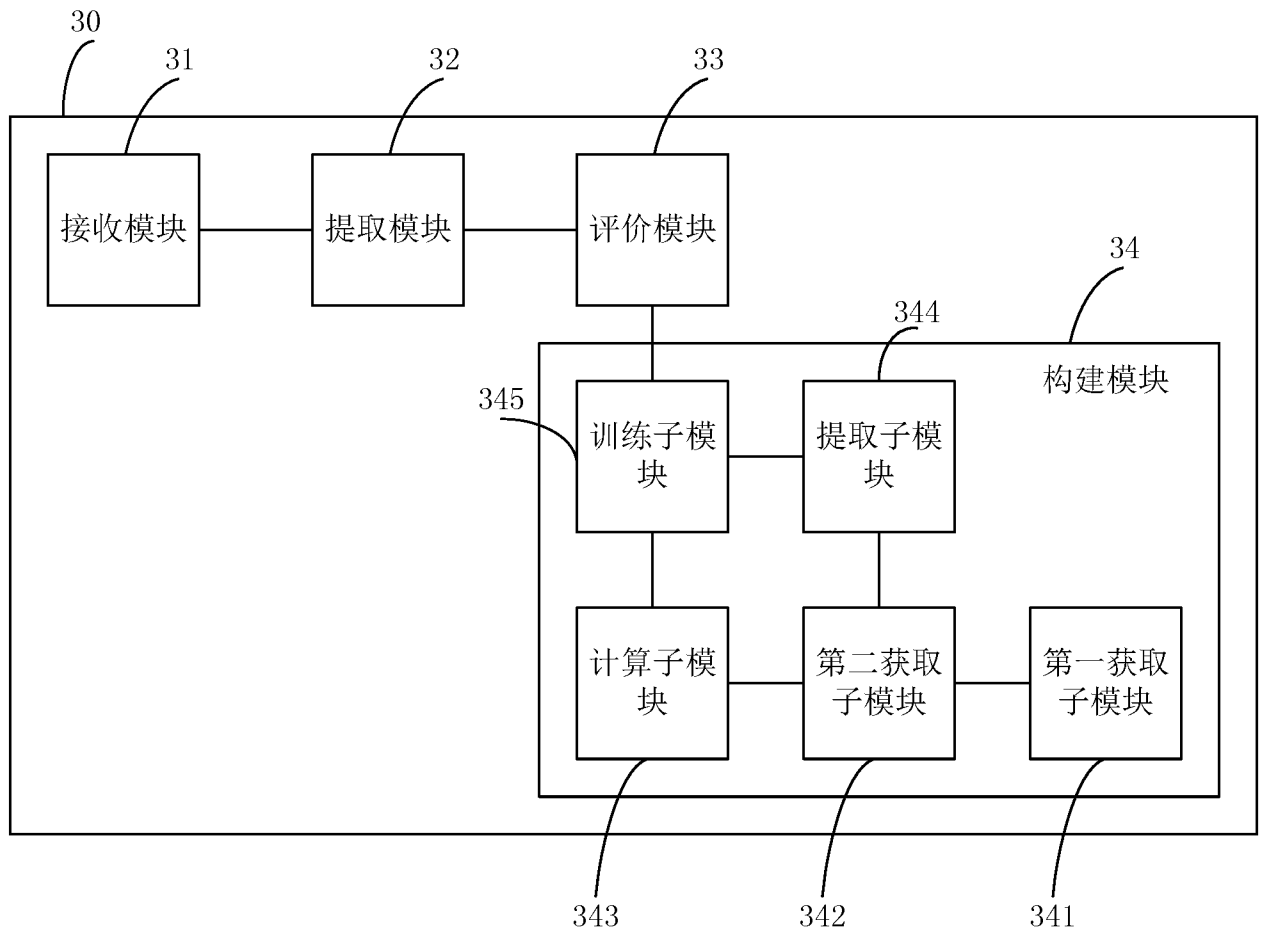


图 4

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/CN2016/111050

## A. CLASSIFICATION OF SUBJECT MATTER

G10L 25/60 (2013.01) i; G10L 15/06 (2013.01) 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, H04M

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)

CNXTX; CNABS, CNKI: 单端, 非侵入, 语音, 质量, 评价, 评估, 建模, 模型, 得分, 分数, 分值, 特征, 音频, paragraphs, 回归模型;

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104361894 A (HUNAN INSTITUTE OF METROLOGY AND TEST et al.), 18 February 2015 (18.02.2015), description, paragraphs 106-189 and 210-214, and figure 1	1, 2, 4, 8, 9, 11, 15-17
X	尹伟, “基于模型的语音增强方法及质量评估研究”, 武汉大学博士学位论文, 30 September 2009 (30.09.2009), pages 95-104, (YIN, Wei, Research on Speech Enhancement Based on Speech Modeling and Speech Quality Assessment, Dissertation Submitted to Wuhan University for the Doctoral Degree)	1, 2, 4, 8, 9, 11, 15-17
A	CN 105282347 A (CHINA MOBILE COMMUNICATIONS CORPORATION), 27 January 2016 (27.01.2016), entire document	1-17
A	US 2015348571 A1 (NEC CORP.), 03 December 2015 (03.12.2015), entire document	1-17

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

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” document member of the same patent family</p>
---	---

<p>Date of the actual completion of the international search</p> <p style="text-align: center;">15 June 2017</p>	<p>Date of mailing of the international search report</p> <p style="text-align: center;">30 June 2017</p>
<p>Name and mailing address of the ISA</p> <p>State Intellectual Property Office of the P. R. China</p> <p>No. 6, Xitucheng Road, Jimenqiao</p> <p>Haidian District, Beijing 100088, China</p> <p>Facsimile No. (86-10) 62019451</p>	<p>Authorized officer</p> <p style="text-align: center;">LIU, Jia</p> <p>Telephone No. (86-10) 62412046</p>

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
PCT/CN2016/111050

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104361894 A	18 February 2015	None	
CN 105282347 A	27 January 2016	None	
US 2015348571 A1	03 December 2015	JP 2016006504 A	14 January 2016

<p><b>A. 主题的分类</b></p> <p>G10L 25/60(2013.01)i; G10L 15/06(2013.01)i</p> <p>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</p>																	
<p><b>B. 检索领域</b></p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G10L, H04M</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNXTX;CNABS, CNKI:单端, 非侵入, 语音, 质量, 评价, 评估, 建模, 模型, 得分, 分数, 分值, 特征, 音频, 段, 回归模型; VEN:single end, non-instructive, speech, quality, estimate, model+,score, characteristic, section, regression model</p>																	
<p><b>C. 相关文件</b></p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 104361894 A (湖南省计量检测研究院等) 2015年 2月 18日 (2015 - 02 - 18) 说明书106-189段, 第210-214段, 附图1</td> <td>1、2、4、8、 9、11、15-17</td> </tr> <tr> <td>X</td> <td>尹伟. “基于模型的语音增强方法及质量评估研究” 武汉大学博士学位论文, 2009年 9月 30日 (2009 - 09 - 30), 第95-104页</td> <td>1、2、4、8、 9、11、15-17</td> </tr> <tr> <td>A</td> <td>CN 105282347 A (中国移动通信集团公司) 2016年 1月 27日 (2016 - 01 - 27) 全文</td> <td>1-17</td> </tr> <tr> <td>A</td> <td>US 2015348571 A1 (NEC CORP.) 2015年 12月 3日 (2015 - 12 - 03) 全文</td> <td>1-17</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 104361894 A (湖南省计量检测研究院等) 2015年 2月 18日 (2015 - 02 - 18) 说明书106-189段, 第210-214段, 附图1	1、2、4、8、 9、11、15-17	X	尹伟. “基于模型的语音增强方法及质量评估研究” 武汉大学博士学位论文, 2009年 9月 30日 (2009 - 09 - 30), 第95-104页	1、2、4、8、 9、11、15-17	A	CN 105282347 A (中国移动通信集团公司) 2016年 1月 27日 (2016 - 01 - 27) 全文	1-17	A	US 2015348571 A1 (NEC CORP.) 2015年 12月 3日 (2015 - 12 - 03) 全文	1-17
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 104361894 A (湖南省计量检测研究院等) 2015年 2月 18日 (2015 - 02 - 18) 说明书106-189段, 第210-214段, 附图1	1、2、4、8、 9、11、15-17															
X	尹伟. “基于模型的语音增强方法及质量评估研究” 武汉大学博士学位论文, 2009年 9月 30日 (2009 - 09 - 30), 第95-104页	1、2、4、8、 9、11、15-17															
A	CN 105282347 A (中国移动通信集团公司) 2016年 1月 27日 (2016 - 01 - 27) 全文	1-17															
A	US 2015348571 A1 (NEC CORP.) 2015年 12月 3日 (2015 - 12 - 03) 全文	1-17															
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																	
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” 同族专利的文件</p>																	
<p>国际检索实际完成的日期</p> <p>2017年 6月 15日</p>		<p>国际检索报告邮寄日期</p> <p>2017年 6月 30日</p>															
<p>ISA/CN的名称和邮寄地址</p> <p>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</p> <p>传真号 (86-10)62019451</p>		<p>受权官员</p> <p>刘佳</p> <p>电话号码 (86-10)62412046</p>															

国际检索报告  
关于同族专利的信息

国际申请号  
PCT/CN2016/111050

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	104361894	A	2015年 2月 18日	无			
CN	105282347	A	2016年 1月 27日	无			
US	2015348571	A1	2015年 12月 3日	JP	2016006504	A	2016年 1月 14日