

(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
2016 年 8 月 4 日 (04.08.2016)



(10) 国际公布号
WO 2016/119604 A1

- (51) 国际专利分类号:
G10L 15/08 (2006.01)
- (21) 国际申请号: PCT/CN2016/071164
- (22) 国际申请日: 2016 年 1 月 18 日 (18.01.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510037176.9 2015 年 1 月 26 日 (26.01.2015) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [GB/CN]; 英国开曼群岛大开曼乔治城资本大厦一座四层 847 号邮箱, Cayman Islands (GB)。
- (72) 发明人: 及
- (71) 申请人 (仅对美国): 闻乃松 (WEN, Naisong) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区西坝河西里 28 号英特公寓 C 座 104, Beijing 100028 (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, IR, IS, JP, KE, KG, KN, KP, KR, 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, 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。

[见续页]

(54) Title: VOICE INFORMATION SEARCH METHOD AND APPARATUS, AND SERVER

(54) 发明名称: 一种语音信息搜索方法、装置及服务器

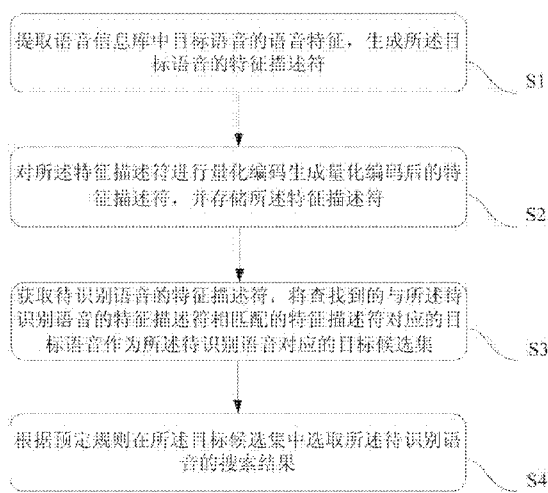


图 1

- S1 Extracting voice features of target voices in a voice information base to generate feature descriptors of the target voices
- S2 Carrying out quantisation coding on the feature descriptors to generate quantisation-coded feature descriptors, and storing the feature descriptors
- S3 Acquiring a feature descriptor of a voice to be identified, and taking a found target voice corresponding to a feature descriptor matching the feature descriptor of the voice to be identified as a target candidate set corresponding to the voice to be identified
- S4 Selecting a search result of the voice to be identified in the target candidate set according to a pre-set rule

(57) Abstract: A voice information search method and apparatus, and a server. The method comprises: extracting voice features of target voices in a voice information base to generate feature descriptors of the target voices (S1); acquiring a feature descriptor of a voice to be identified, and taking a found target voice corresponding to a feature descriptor matching the feature descriptor of the voice to be identified as a target candidate set corresponding to the voice to be identified (S3); and selecting a search result of the voice to be identified in the target candidate set according to a pre-set rule (S4). By utilizing the method, query efficiency can be optimised, and the voice search speed can be increased.

(57) 摘要: 一种语音信息搜索方法、装置及服务器, 方法包括: 提取语音信息库中目标语音的语音特征, 生成所述目标语音的特征描述符 (S1); 获取待识别语音的特征描述符, 将查找到的与所述待识别语音的特征描述符相匹配的特征描述符对应的目标语音作为所述待识别语音对应的目标候选集 (S3); 根据预定规则在所述目标候选集中选取所述待识别语音的搜索结果 (S4)。利用该方法, 可以优化查询效率, 提高语音搜索速度。



WO 2016/119604 A1



(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))。

一种语音信息搜索方法、装置及服务器

技术领域

本申请属于电子信息信号处理技术领域，尤其涉及一种语音信息搜索方法、装置及服务
5 器。

背景技术

在未来，语音识别将逐步成为电子信息技术中人机交互的关键技术。目前在银行自助服务、公共自助服务、微信的终端应用、即时语音通信等领域对语音识别技术的需求越来越强
10 烈，尤其是在移动互联网领时代基于安全考虑的语音内容审核。

例如在目前的众多社交应用 App 中，用户可以发布包括各种内容的语音信息，其中有些可能涉及恐怖、色情、广告推销、诈骗等违法信息。目前，常用的方法是基于特定语种和关键词的语音识别技术进行语音信息内容自动审核。在该技术中通常的做法可以包括获取审核语音的语音特征，通过训练产生声学模型，目的是为每个发音建立发音模板。在识别时将待
15 识别的语音特征与审核语音的声学模型逐个进行匹配，选取与待识别语音最接近的发音模板作为待识别语音的所表达的含义。

在实际的语音识别过程中，通常将语音信息分成多个音频特征，例如以 20 毫秒为一个帧长，一段 10 秒是语音将产生 500 个音频特征。而存储的审核语音常常多达成千上万，同一个含义的审核语音又可以包括多个不同方言、不同语气的表述方式，同时每个审核语音的
20 发音模块中又存在大量音频特征，在大规模数据集的情况下，现有的基于发音模块进行音频识别的方法面临着高维特征索引和查询过程复杂，查询时间长的问题，降低了查询效率。

发明内容

本申请目的在于提供一种语音信息搜索方法、装置及服务器，可以提取与特定人无关的
25 语音底层特征，然后进行量化编码，建立索引，通过 K-d 数搜索数据库中已有的目标语音，达到语音内容快速搜索的目的，提高查询效率。

本申请提供的一种语音信息搜索方法、装置及服务器是这样实现的：

一种语音信息搜索方法，所述方法包括：

提取语音信息库中目标语音的语音特征，生成所述目标语音的特征描述符；

30 对所述特征描述符进行量化编码生成量化编码后的特征描述符，并存储所述特征描述

符；

获取待识别语音的特征描述符，在所述存储的所述特征描述符中，查找与所述待识别语音的特征描述符相匹配的特征描述符对应的目标语音，将查找到的目标语音作为所述待识别语音对应的目标候选集；

5 根据预定规则在所述目标候选集中选取所述待识别语音的搜索结果。

一种语音信息搜索装置，所述装置包括：

信息获取模块，用于获取目标语音，并提取所述目标语音的语音特征；

描述符模块，用于基于所述目标语音的语音特征生成所述目标语音的特征描述符；

量化编码模块，用于对所述特征描述符进行量化编码，生成量化编码后的特征描述符，

10 并存储所述特征描述符；

识别信息模块，用于获取待识别语音的特征描述符；

第一搜索模块，用于在所述存储的所述特征描述符中，查找与所述待识别语音的特征描述符相匹配的特征描述符对应的目标语音，将查找到的目标语音作为所述待识别语音对应的目标候选集；

15 第二搜索模块，用于根据预定规则在所述目标候选集中选取所述待识别语音的搜索结果。

一种语音信息搜索服务器，所述服务器被设置成包括：

第一处理单元，用于获取目标语音，生成所述目标语音的特征描述符；还用于对所述特征描述符进行量化编码；

20 存储单元，用于分别存储所述量化编码后的特征描述符中路径相同的特征描述符；

第二处理单元，用于获取待识别语音的特征描述符；还用于在所述存储的特征描述符中查找与所述待识别语音相匹配的特征描述符的目标语音，获取候选集；还用于根据预定规则在所述候选集中选取所述待识别语音的搜索结果。

25 本申请提供一种语音信息搜索方法、装置及服务器，可以将语音信息库中存储的需要审核的目标关键词或者短语的目标语音信息行音素级别的模型学习和表述，并生成特征描述符，建立索引。本申请中可以对生成特征描述符进行量化编码，降低特征描述符索引维度和信息长度，可以提高信息索引时的处理速度。在查询时，本申请利用 K-d 树获取搜索范围更小的待识别语音的目标候选集，然后进一步筛选出搜索结果。本申请提供的语音信息搜索方法，将传统高纬度、复杂的发音模块语音识别转化成相似音频特征的搜索，而且通过特征描述符降低索引维度和 K-d 树优化查询效率，可以大大提高语音信息搜索速度。

30

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请中记载的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图1是本申请一种语音信息搜索方法一种实施例的流程示意图；

图2是本申请一种对特征描述符进行量化编码的示意图；

图3是本申请建立特征描述符的索引示意图；

图4是本申请一种语音信息搜索装置的一种实施例的模块结构示意图；

图5是本申请提供的一种量化编码模块的模块结构示意图；

图6是本申请提供的一种第二搜索模块的模块结构示意图。

具体实施方式

为了使本技术领域的人员更好地理解本申请中的技术方案，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都应当属于本申请保护的范围。

本申请中提供的一种语音信息搜索方法，可以将需要审核的目标关键词或者短语对应的不同语种的语音底层特征添加到数据库，进行音素级别的模型学习和表述，并生成特征描述符，建立索引。在查询时，利用 K-d 树获取待识别语音的目标候选集，然后进一步筛选出搜索结果。基于上述的检索方案，可以通过特征描述符量化编码降低索引维度和 K-d 树优化查询效率，可以大大提高语音信息搜索速度。

图 1 是本申请所述一种语音信息搜索方法的一个实施例的方法流程图，如图 1 所述，所述方法可以包括：

S1：提取语音信息库中目标语音的语音特征，生成所述目标语音的特征描述符。

所述的语音信息库可以包括存储的目标语音。所述目标语音可以预先采集或设置。所述目标语音具体的在不同的应用场景中可以包括不同的内容。例如在基于安全考虑的语音内容审核中，所述语音信息库中的目标语音可以为包括多个方言或多种语气的涉及恐怖、色情、广告、诈骗等内容的关键词或者短语。在家庭音像或者汽车智能控制终端中，所述语音信息库中的目标语音可以包括对家庭智能设备如智能电视、音响等或者汽车驾驶控制设备进行功能控制的语音关键词或者短语。又或者收集存储的在智能终端社交、购物、聊天等应用中常

用的例如“天气”、“苹果”“双十一”等关键词或者短语的目标语音信息。所述的语音信息库中存储的目标语音可以根据不同的应用场景进行设置，本申请中所述的语音信息搜索方法可以适用但并不限于基于安全考虑的语音内容审核的应用场景。

5 基于语音内容自动审核的语音识别通常需要提取与特定人无关的语音底层特征，这样可以更加准确识别不同的人说的相同的话，或者相同的人因为自身状态和场合不同但说的内容相同的话。本申请中提取跟特定人无关的语音底层特征的方法可以包括 MFCC (Mel-Frequency Cepstrum Coefficients, Mel 频率倒谱系数)和 PLP (Perceptual Linear Predictive, 感觉加权线性预测)方法。其中 MFCC 是建立在傅立叶和倒谱分析基础上，对短时音频帧中的采样点进行傅立叶变换，得到这个短时音频帧在每个频率上的能量大小，可以
10 较好的反应音频信号的频域特征。因此，在本实施例中可以采用 MFCC 的方法获取语音信息库中的目标语音的语音特征。

具体的，所述提取语音信息库中目标语音的语音特征可以包括如下的处理步骤：

S101：对所述目标语音进行预处理。

所述的预处理可以包括对所述目标语音进行的语音格式转换、预加重、分帧、加窗处理
15 等操作。在本实施例中，所述的对目标语音进行的预处理具体的实施过程可以包括：

S1011：语音格式转换。

所述语音信息库中存储的目标语音可以包括采集获取的多种格式的语音信息，如 amr 格式、wav 格式等。本实施例中可以不同的目标语音格式统一转换为 wav 格式，方便后续数据的统一、快速处理。

20 S1012：预加重。

所述的预加重通常是指将语音信号通过高通滤波器使信号的频谱变的平坦，保证从低频到高频的整个频带中能用相同或相近的信噪比求取频谱。同时，预加重还可以消除发声过程中的声带和嘴唇效应，用来补偿语音信号受到发音系统所抑制的高频部分和突出高频的共振峰。滤波器公式如下：

25
$$H(z) = 1 - \mu z^{-1}$$

上式中， μ 的值介于 0.9-1.0 之间，本实施例中可以取值为 0.97。

S1013：分帧。

所述分帧可以包括将目标语音的所有采集点每 N 个采样点作为一个帧。本实施例中为了避免相邻两帧的变化过大，可以使相邻帧之间有一段重叠区域，所述重叠区域可以包括 M 个
30 采样点。具体的，例如本实施例中的语音信号采样频率可以为 8000Hz，每帧包括的采样点 N 为 512，重叠区域 M 为 256。

S1014: 加窗。

通常为了高效处理语音信号,可以对语音信号进行加窗,也就一次仅处理窗中的数据。因为实际的语音信号是很长的,在实际信号处理时通常可以不需要对非常长的数据进行一次
5 性处理。可以每次取一段数据,进行分析,然后再取下一段数据,再进行分析。在此过程中可以构造一个函数,这个函数在某一区间有非零值,而在其余区间皆为 0。汉明窗就是这样的一种函数,本实施例中可以将当前分帧的信号乘以汉明窗,在处理下一帧时通常可以每次移动窗口的三分之一或者二分之一,以增加帧左端和右端的连续性。本实施例中提供的加窗公式可以为:

$$S'(n) = S(n) \times W(n)$$

10 上式中,

$$W(n, a) = (1 - a) - a \times \cos \left[\frac{2\pi n}{N - 1} \right], 0 \leq n \leq N - 1$$

上式中, N 为帧的采样点个数, S(n)表示语音信号, S'(n)为加窗处理后的语音信号, a 可以取值为 0.46。

通过上述预处理后,可以得到取预处理后的目标语音的能量。

15 S102: 计算所述预处理后的目标语音的能量谱。

本实施例中可以对预处理后的各帧语音信号进行快速傅里叶变换得到各帧的频谱,然后可以对语音信号的频谱取模平方得到语音信号的功率谱。

S103: 对所述能量谱进行 Mel 滤波, 计算所述 Mel 滤波后的能量谱的对数。

20 可以将所述能量谱通过预先设置的一组 Mel 尺度的三角形滤波器, 对频谱进行平滑化, 消除谐波的作用, 突显原先语音的共振峰。然后可以采用下式计算每个滤波器输出的对数:

$$S(m) = \ln \left(\sum_{k=0}^{N-1} |X_a(k)|^2 H_m(k) \right), 0 \leq m \leq M$$

上式中, N 为帧的采样点个数, M 为滤波器个数, Xa 为经过傅里叶变换后的频谱, Hm 为第 m 个滤波器。

25 S104: 对所述能量谱的对数进行 DCT 变换得到 MFCC 系数, 获取所述目标语音的语音特征。

本实施例中可以采用下式进行 DCT (Discrete Cosine Transform, 离散余弦变换) 变换得到 MFCC 系数:

$$C(n) = \sum_{m=0}^{N-1} s(m) \cos \left(\frac{\pi n(m - 0.5)}{M} \right), n = 1, 2, \dots, L$$

上式中, N 为帧的采样点的个数, M 为滤波器个数, L 阶指 MFCC 系数阶数, 通常取

30 12-16。

可以将上式变换得到的 MFCC 系数作为目标语音的语音特征。

另一种优选的实施方式中，还可以在所述语音特征中添加动态参数，提高语音识别性。本实施例所述的方法中可以在所述语音特征中加入表征语音动态特性的差分参数，屏蔽可以区分不同的人说的同样的话的特征，提高系统的语音识别性能。因此，本实施例中所述语音信息搜索方法还可以包括：

S105：计算所述 MFCC 系数的一阶和二阶差分系数，将所述一阶和二阶差分系数添加至所述语音特征中。即所述语音特征还可以包括所述 MFCC 系数的一阶和二阶差分系数。所述 MFCC 系数的一阶和二阶差分系数具体的可以采用下述公式计算：

$$d_t = \frac{\sum_{n=1}^N n(C_{t+n} - C_{t-n})}{2 \sum_{n=1}^N n^2}$$

10 上式中 d_t 表示第 t 个一阶差分， C_t 表示第 t 个倒谱系数， N 表示一阶导数的时间差，可取 1 或 2。将上式的结果再代入就可以得到二阶差分的参数。

提取所述语音信息库中目标语音的语音特征后，可以生成所述目标语音的特征描述符。本申请中所述的特征描述符可以包括语音特征的 VLAD(vector of locally aggregated descriptors, VLAD, 局部特征聚合描述符)特征描述符。下述是本实施例提供的一种生成特征描述符的实施方法，可以将语音特征聚合到单独特征向量中。具体的实施过程可以包括：

S101'：对所述提取的语音特征通过 k-means 聚类方法获取所述目标语音的码本；

生成所述 VLAD 描述符通常需要先训练码本，可以从所述提取的语音特征中随机选取 N 个语音特征，通过 k-means 聚类方法得到码字数量为 k 的码本 $\{\mu_1, \dots, \mu_k\}$ 。所述码本中的每一项为码字，可以表示为一个或者多个相同或者相近语音样本的聚合，例如 μ_1 可以表示为语音信息库中多个不同语气但表示相同含义的目标语音的聚合。这样可以将所述语音信息库中的大量目标语音聚合形成码字数量为 k 的码本。所述 k 可以远远小于所述语音信息库中目标语音的数量。

S102'：获取所述目标语音的语音特征集合，计算所述语音特征集合与所述码本中距离最近的码字的所有残差向量之和。

25 所述语音信息库中一个目标语音的一个或多个语音特征可以形成一个特征集合 $\{x_1, \dots, x_p\}$ 。对于某个目标语音的语音特征形成的特征集合，依次查找所述特征集合中的每个语音特征在所述码本中距离最近的码字，并计算所述特征集合中的语音特征与当前查找到的距离最近的码字的残差向量，然后累加属于同一个码字的所有残差向量。如下式所示：

$$v_i = \sum_{x_t \in NN(x_t)=i} x_t - \mu_i$$

上式中, x_t 为所述特征集合 $\{x_1, \dots, x_p\}$ 中的第 t 个特征, $1 \leq t \leq P$, μ_i 为所述码本 $\{\mu_1, \dots, \mu_k\}$ 中的第 i 个码字, $1 \leq i \leq k$. $x_t: NN(x_t) = i$ 表示映射为同一个码字的特征子集。

S103': 对所述码字的残差向量之和进行归一化, 生成所述目标语音的特征描述符。

对所述累加后的码字的残差向量进行归一化。可以连接所述归一化后的码字残差向量之和组成目标语音的 VLAD 总特征描述符 V 。所述 VLAD 总特征描述符 V 可以表示为:

$$V = \{v'_1, \dots, v'_k\}$$

其中 v'_i 为所述归一化处理后的特征描述符的向量数据, 所述的归一化处理可以采用下式进行处理:

$$v'_i = \frac{v_i}{\|v_i\|_2}$$

本实施例中可以提取语音信息库中采集存储的目标语音的语音特征, 然后可以生成所述目标语音的特征描述符。所述的总特征描述符 V 中的每一个 VLAD 特征描述符可以为一个多维的特性向量, 例如可以为一个 128 维的特征向量 $\{l1, l2, l3, \dots, l128\}$, 其中每一个特征向量的每个维度都是目标语音的聚类索引, 可以根据所述多维的特征向量 v'_i 定位找到相应的目标语音。

S2: 对所述特征描述符进行量化编码, 生成量化编码后的特征描述符, 并存储所述特征描述符。

实际的实施过程中, 所述得到的一个 VLAD 特征描述符通常达到上千比特, 为了加快搜索速度, 尤其是在大数据集的应用场景下降低语音信息检索的困难和复杂性, 本申请可以对所述获取的 VLAD 特征描述符进行量化编码, 降低所述特征描述符的维度和信息长度, 优化搜索查询效率。

具体的所述对所述特征描述符进行量化编码生成量化编码后的土地证描述符可以包括:

S201: 将每个所述特征描述符等分成 L 份子向量, 对所述 L 个子向量分别进行聚类, 并设置所述子向量聚类后的索引编号, $L \geq 2$;

S202: 将每个所述特征描述符的 L 个子向量分别用与所述子向量距离最近的所述聚类的索引编号表示, 生成量化编码后的特征描述符。

图 2 是本实施例中所述一种对特征描述符进行量化编码的示意图。图 2 中以 VLAD 特征描述符为 128 维特征向量为例, 可以将其分成 8 等份 ($y1 \sim y8$), 每一等份中包括 128 维特征向量的中的 16 维分量 (16 components)。可以分别对所述每个分量进行单独聚类并映射到 256 个聚类中心上 (256 centroids)。聚类后的每个子向量可以用于索引编号表示, 例如图 2 中的 $q1$ ($y1$) 可以表示包括所述 128 维特征向量的前 16 维分量。分成的 8 等份的子向量

中每个子向量可以用 8 位二进制表示,这样 128 维的特征向量就可以用于 8 位的 $8\text{bit} \times 8 = 64\text{bit}$ 的信息表示,降低了信息处理维度和信息数据长度,提高处理效率。

通过上述方法可以将高维度的 VLAD 特征描述符表示为长度为 L,分量为二级聚类索引编号的低维特征向量,可以大大提高后续语音信息的搜索速度。

5 对所述特征描述符进行量化编码形成所述量化编码后特征描述符后,可以存储所述量化编码后的特征描述符。具体的存储时可以包括以所述量化编码后的特征描述符的索引的方式实现。本实施例中优选的可以采用建立 K-d 树索引,具体的所述存储所述特征描述符实施方式可以包括:

S201': 建立高度为 (L+1) 的 K-d 树;

10 S202': 为所述 K-d 树的非叶子节点划分索引维度和与所述索引维度相对应的划分值;并建立与所述划分值进行比较的结果路径指向;

S203': 从所述 K-d 树的根节点开始,将与非叶子节点的索引维度相对应的特征描述符的值与所述非叶子节点的划分值进行比较,并基于比较的结果与所述结果路径指向将所述特征描述符存储至所述 K-d 树的叶子节点中。

15 图 3 是本申请所述建立特征描述符的索引示意图。在图 3 的示例中,可以建立一个高度为 (3+1) 的 K-d 树,所述 K-d 树最后一层为叶子节点。需要进行存储的特征描述符为 3 维特征向量,可以预先划分每个非叶子节点的索引维度和划分值,并可以预先设置比较的结果路径指向,例如如果所述特征描述符在该非叶子节点上相应索引维度的值小于该非叶子节点的划分值,则进入该非叶子节点的左子树继续进行比较,否则进行右子树继续进行比较,直至进入叶子节点。需要说明的是,所述的索引维度可以表示为特征描述符中索引维度的值所指示的那一个维度的值,如果某一非叶子节点的上层节点已经比较过的维度则不再进行比较。因此,在非叶子节点索引维度划分的时候,从所述 K-d 树根节点到当前非叶子节点的路径上已经划分过的索引维度值不再划分给当前的非叶子节点。

25 上述为非叶子节点划分的划分值,在具体的实施中可以取值上述量化编码时设置的所述聚类中心个数的一半,如上述量化编码时设置的聚类中心个数为 256,在设置所述划分值时可以取值为 128。

30 在上述例子中,可以分别将 3 维特征向量的特征描述符存储到叶子节点中。如一个特征描述符为 (45,210,60) 的 3 维特征向量,按照图 3 建立的所述 K-d 树可以先从根节点开始,将特征描述符 (45,210,60) 与根节点中所示的索引维度为 2、划分值为 128 进行相应索引维度的值的比较。当前节点索引维度为 2 所对应的所述特征描述符 (45,210,60) 的值为 210,大于当前节点划分值 128,则进入根节点的右子树。然后所述特征描述符 (45,210,60) 继续

与当前非叶子节点设置的索引维度为 1、划分值为 128 进行相应索引维度值的比较，所述特征描述符 (45,210,60) 索引维度为 1 所对应的值为 45，小于当前非叶子节点的划分值 128，则进入当前非叶子节点的左子树。同样的方法，可以按照当前非叶子节点设置索引维度为 3、划分值为 128 将所述特征描述符 (45,210,60) 的值 60 与 128 进行比较，然后进入当前非叶子节点的左侧，到达所述 K-d 树的叶子节点。此时，可以将所述特征描述符 (45,210,60) 存入至该叶子节点中。对应更高维的特征描述符，可以建立相应的 K-d 树，仍然按照上述方法进行存储到叶子节点中。

上述 K-d 中的最后一层为叶子节点，每个叶子节点存储的为所有经过从根节点到该叶子节点这条搜索路径的特征描述符的特征向量。其中每一个特征向量的每个维度都是目标语音的聚类索引。这样，每个叶子节点存储的是上述建立的 K-d 索引中相同路径的特征描述符的集合，在此可以将一个叶子节点存储的特征描述符作为一个目标候选集。

本申请所述方法优选的实施例中，为保证每个所述叶子节点存储的目标候选集容量相当，均衡各个叶子节点存储的特征描述符的数量，所述为非叶子节点划分索引维度可以包括：为非叶子节点划分的索引维度值 S 为随机生成的取值范围为 $1 \leq S \leq L$ 的整数，并且当前非叶子节点的索引维度值 S 为从所述 K-d 树根节点到所述当前非叶子节点的路径上未划分过的索引维度值。上述中的 L 为所述特征描述符的长度，即所述特征描述符中特征向量的维数。这样，非叶子节点所述索引维度的随机划分，提高 K-d 树左右两侧节点的均匀分布性，可以避免部分叶子节点存储的目标候选集内的特征描述符数量过多，可以提高后续在目标候选集内的搜索速度，均衡各个目标候选集的负载。

S3: 获取待识别语音的特征描述符，在所述存储的所述特征描述符中，查找与所述待识别语音的特征描述符相匹配的特征描述符对应的目标语音，将查找到的目标语音作为所述待识别语音对应的目标候选集；

对于给定的待识别语音，可以按照上述所述的方法提取待识别语音的语音特征，并获取所述待识别语音的量化编码后的特征描述符。可以将所述待识别语音的特征描述符与所述目标语音的特征描述符匹配，查找到与所述待识别语音的特征描述符相匹配的特征描述符对应的目标语音，将查找到的目标语音作为所述待识别语音对应的目标候选集。所述与所述待识别语音的特征描述符相匹配的特征描述符对应的目标语音在本实施例中可以包括在建立的 K-d 树中与所述待识别语音的特征描述符路径相同的叶子节点中存储的特征描述符所对应的目标语音。

例如通过类似上述图 3 所示的实施方式，可以将待识别语音生成与叶子节点存储的特征描述符的维度相同的特征描述符，例如 (90,135,78)。然后可以按照图 3 所示的 K-d 树及设

置的索引维度和分值进行特征描述符的查找，最终获取到一个叶子节点的与待识别语音的特征描述符相匹配的目标候选集。所述待识别语音的特征描述符在所述 K-d 树中的查找过程可以参照前述目标语音的特征描述符 (45,210,60) 的存储过程，可以查找到待识别语音的特征描述符 (90,135,78) 对应的目标语音，确定目标候选集，如图 3 所示的 (90,135,75) 所在的叶子节点。这样实现了将待识别语音定位到一个范围较小的目标候选集所对应的目标语音集中。

S4: 根据预定规则在所述目标候选集中选取所述待识别语音的搜索结果。

获取目标候选集后，已经大大缩小了搜索范围。然后可以根据预先设置的预定做进一步精选，获取所述待识别语音的搜索结果。本实施例中提供一种在所述目标候选集中进一步精选搜索结果的方法，因此，本实施例中具体所述根据预定规则在所述目标候选集中选取所述待识别语音的搜索结果可以包括：

在所述目标候选集中选取与所述待识别语音的特征描述符欧氏距离最小的前 R 个特征描述符作为搜索结果集，以所述搜索结果集所对应的目标语音作为所述待识别语音的搜索结果， $R \geq 1$ 。

本实施例中可以计算所述待识别语音的特征描述符与所述目标候选集中的特征向量的欧式距离，可以将计算得出的欧式距离按照递增的顺序进行排序，选取所述欧式距离最小的前 R 个特征描述符作为搜索结果集。当然，所述搜索结果集内的特征描述符所对应的语音信息库中的目标语音即为所述待识别语音的搜索结果。

这里所述的选取的欧式距离最小的前 R 个结果中 R 的取值范围可以根据需求进行自行设置。例如所述 R 可以取值为 1，可以表示为选取所述欧式距离最小的特征描述符，如待识别语音的特征描述符为 (90,135,78)，在获取的目标候选集中可以选出欧式距离最小的特征描述符 (90,135,75) 作为所述搜索结果集。当然，在实际的应用中，所述 R 也可以取值大于 1，小于所述目标候选集内特征描述符的个数，例如可以取值为 3，这样可以从所述目标候选集中选出欧式距离最小的前 3 个特征描述符 (90,135,78)、(87,135,80)、(101,137,81) 作为搜索结果集，可以将欧式距离最小的特征描述符 (90,135,78) 所对应的目标语音作为待识别语音的优选搜索结果，将其余两个搜索结果选集的特征描述符所对应的目标语音作为参考或者备选搜索结果。

本申请提供的一种语音信息搜索方法，可以提取目标语音的语音底层特征，生成相应的 VLAD 特征描述符，用于表示目标语音的特征集合，可以根据所述特征描述符定位到相应的目标语音。本申请引入 VLAD 特征描述符将语音特征集合转换成长度固定的整体特征，然后将信息长度较长、维度较高的 VLAD 特征描述符经过量化编码转换为向量长度较小、维度更

低的特征向量，大大提高了信息读取、解析、索引速度。最后通过 K-d 树索引，建立目标候选集，将待识别语音通过 K-d 树索引进行搜索获取搜索范围更小的目标候选集，在所述目标候选集中进一步精选得到搜索结果，进一步加快了搜索速度。

基于本申请所述语音信息搜索方法，本申请还提供一种语音信息搜索装置，该装置可以将语音信息库中的目标语音按照规则分别存储到相应的候选集模块中，并建立相应的索引机制，在获取待识别语音后可以快速的通过索引获取目标候选集，进一步的获取搜索结果，从而将传统的发音模板的语音识别转换成语音特征的搜索，并且通过量化编码和优化索引，提高了搜索速度，优化查询效率。图 4 是本申请所述一种语音信息搜索装置的模块结构示意图，如图 4 所示，所述装置可以包括：

信息获取模块 101，可以用于获取目标语音，并提取所述目标语音的语音特征；

描述符模块 102，可以用于基于所述目标语音的语音特征生成所述目标语音的特征描述符；

量化编码模块 103，可以用于对所述特征描述符进行量化编码，生成量化编码后的特征描述符，并存储所述特征描述符；

识别信息模块 104，可以用于获取待识别语音的特征描述符；

第一搜索模块 105，可以用于在所述存储的所述特征描述符中，查找与所述待识别语音的特征描述符相匹配的特征描述符对应的目标语音，将查找到的目标语音作为所述待识别语音对应的目标候选集；

第二搜索模块 106，可以用于根据预定规则在所述目标候选集中选取所述待识别语音的搜索结果。

所在语音搜索装置中，所述信息获取模块 101 提取目标语音的语音特征的方式可以包括基于 MFCC 和 PLP 方法的提取方式。例如利用 MFCC 提取语音特征的过程可以包括：

对所述目标语音进行预处理；

计算所述预处理后的目标语音的能量谱；

对所述能量谱进行 Mel 滤波，计算所述 Mel 滤波后的能量谱的对数；

对所述能量谱的对数进行 DCT 变换得到 MFCC 系数，获取所述目标语音的语音特征。

其中，所述对所述目标语音进行预处理具体的可以包括：对所述目标语音进行语音格式转换、预加重、分帧、加窗处理。

当然，如上述，为进一步提高装置语音识别率，所述利用 MFCC 提取语音特征的过程还可以包括：

计算所述 MFCC 系数的一阶和二阶差分系数，将所述一阶和二阶差分系数添加至所述语

音特征中。

本实施例具体的提取目标语音的语音特征的实施过程可以参照本申请其他实施例中的叙述，在此不做赘述。

另一种实施例中，所述量化编码模块 103 具体的可以包括：

- 5 聚类模块，可以用于将每个所述特征描述符等分成 L 个子向量，对所述 L 个子向量分别进行聚类，并设置所述子向量聚类后的索引编号， $L \geq 2$ ；

映射模块，可以用于将每个所述特征描述符的 L 个子向量分别用与所述子向量距离最近的所述聚类的索引编号表示，形成量化编码后的特征描述符。

- 10 在对所述子向量进行聚类时，每个子向量可以用于 8 位二进制表示，这样可以将每个特征描述符转换成长度为 L 、维度更低的特征向量。

图 5 是本申请提供的一种量化编码模块的模块结构示意图，如图 5 所示，在另一种实施例中，所述量化编码模块 103 具体的可以包括：

索引树构建模块 1031，可以用于建立高度为 $(L+1)$ 的 $K-d$ 树；

- 15 预置规则模块 1032，可以用于根据预先设置规则为所述 $K-d$ 树的非叶子节点划分索引维度和与所述索引维度相对应的划分值，并建立与所述划分值进行比较的结果路径指向；

存储模块 1033，可以用于从所述 $K-d$ 树的根节点开始，将与非叶子节点的索引维度相对应的特征描述符的值与所述非叶子节点的划分值进行比较，并基于比较的结果与所述结果路径指向将所述特征描述符存储至所述 $K-d$ 树的叶子节点中。

- 20 所述建立的结果路径指向可以包括将特征描述符与当前非叶子节点划分值进行比较，如果前者大于后者，则进入当前非叶子节点的左子树继续进行比较，否则进入当前非叶子节点的右子树继续比较。当然，也可以设置所述前者小于所述后者时进入当前非叶子节点的左子树进行比较，具体的可以根据需求自行设定。

优选的实施例中，所述预置规则模块 1032 为非叶子节点划分索引维度具体的可以包括：

- 25 为非叶子节点划分的索引维度值 S 为随机生成的取值范围为 $1 \leq S \leq L$ 的整数，并且当前非叶子节点的索引维度值 S 为从所述 $K-d$ 树根节点到所述当前非叶子节点的路径上未划分过的索引维度值。

图 6 是本申请提供的一种第二搜索模块的模块结构示意图，如图 6 所示，上述所述语音搜索装置的第二搜索模块 106，具体的可以包括：

- 30 距离计算模块 1061，可以用于计算所述目标候选集中的特征描述符与所述待识别语音的特征描述符的欧氏距离；

筛选模块 1062，可以用于在所述目标候选集中选取与所述待识别语音的特征描述符欧氏

距离最小的前 R 个特征描述符作为搜索结果集, $R \geq 1$;

目标语音模块 1063, 可以用于获取与搜索结果集中的特征描述符相对应的目标语音。

另一种实施方式中, 所述描述符模块 102 具体的可以包括:

码本训练模块, 可以用于对所述提取的语音特征通过 k-means 聚类方法获取所述目标语音的码本;

语音特征集合模块, 可以用于获取语音特征集合, 并计算所述语音特征集合与所述码本中距离最近的码字所有残差向量之和; 所述语音特征集合可以例如 $\{x_1, \dots, x_p\}$, 其中每个 x 可以表示对应一个目标语音的语音特征。

归一化处理模块, 可以用于对所述码字的残差向量之和进行归一化, 生成所述目标语音的特征描述符。

本申请所述的一种语音信息搜索装置可以应用于多种设备终端或者服务器中。例如日常生活中常用的智能移动终端可以获取待识别语音信息, 所述智能移动终端可以将所述待识别语音信息发送至服务器, 服务器可以利用本申请所述的语音信息搜索方法和装置中的实施例进行语音搜索, 获取相应的语音搜索结果, 然后可以再根据搜索得到的结果做进一步处理。

因此, 本申请还提供一种语音信息搜索服务器, 所述服务器被设置成可以包括:

第一处理单元, 用于获取目标语音, 生成所述目标语音的特征描述符; 还用于对所述特征描述符进行量化编码;

存储单元, 用于分别存储所述量化编码后的特征描述符中路径相同的特征描述符;

第二处理单元, 用于获取待识别语音的特征描述符; 还用于在所述存储的特征描述符中查找与所述待识别语音相匹配的特征描述符的目标语音, 获取候选集; 还用于根据预定规则在所述候选集中选取所述待识别语音的搜索结果。

当然, 进一步的所述服务器可以将获取的待识别语音的搜索结果返回至发送待识别语音的客户端, 或者结合所述服务器的或者其他服务器的功能模块进行其他处理。本实施例提供的语音搜索服务器, 结合特征描述符和索引树, 优化语音索引方法, 提高了服务器语音搜索速度。

本申请提供的一种语音信息搜索方法、装置及服务器, 可以将目标语音按照规则分别存储到相应的候选集中, 并建立相应的索引机制, 在获取待识别语音后可以快速的通过索引获取目标候选集, 进一步的获取搜索结果, 从而将传统的发音模板的语音识别转换成语音特征的搜索, 并且通过量化编码和 K-d 树优化索引, 提高了搜索速度, 优化查询效率。

尽管本申请内容中提到包括信息传输、数据变换和数据树形结构等之类的描述, 但是,

本申请并不局限于必须是完全符合标准通信协议或数据处理标准的情况。某些协议或者标准的基础上略加修改后的传输机制或者数据处理标准也可以实行上述本申请各实施例的方案。当然，即使不采用上述通用的协议或数据处理标准，而是采用私有协议或数据处理标准，只要符合本申请上述各实施例的信息交互和信息判断反馈方式，仍然可以实现相同的申请，在此不再赘述。

上述实施例阐明的单元、装置或模块，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。为了描述的方便，描述以上装置时以功能分为各种模块分别描述。当然，在实施本申请时可以把各模块的功能在同一个或多个软件和/或硬件中实现，也可以将实现同一功能的模块由多个子模块或子单元的组合实现。

本领域技术人员也知道，除了以纯计算机可读程序代码方式实现控制器以外，完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件，而对其内部包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构、类等等。也可以在分布式计算环境中实践本申请，在这些分布式计算环境中，通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

通过以上的实施方式的描述可知，本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在存储介质中，如ROM/RAM、磁碟、光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，移动终端，服务器，或者网络设备等）执行本申请各个实施例或者实施例的某些部分所述的方法。

本说明书中的各个实施例采用递进的方式描述，各个实施例之间相同或相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。本申请可用于众多通用或专用的计算机系统环境或配置中。例如：个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的电子设备、网络PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。

虽然通过实施例描绘了本申请，本领域普通技术人员知道，本申请有许多变形和变化而不脱离本申请的精神，希望所附的权利要求包括这些变形和变化而不脱离本申请的精神。

权 利 要 求 书

1、一种语音信息搜索方法，其特征在于，所述方法包括：

提取语音信息库中目标语音的语音特征，生成所述目标语音的特征描述符；

5 对所述特征描述符进行量化编码生成量化编码后的特征描述符，并存储所述特征描述符；

获取待识别语音的特征描述符，在所述存储的所述特征描述符中，查找与所述待识别语音的特征描述符相匹配的特征描述符对应的目标语音，将查找到的目标语音作为所述待识别语音对应的目标候选集；

10 根据预定规则在所述目标候选集中选取所述待识别语音的搜索结果。

2、如权利要求 1 所述的一种语音信息搜索方法，其特征在于，所述对所述特征描述符进行量化编码生成量化编码后的特征描述符包括：

将每个所述特征描述符等分成 L 份子向量，对所述 L 个子向量分别进行聚类，并设置所述子向量聚类后的索引编号， $L \geq 2$ ；

15 将每个所述特征描述符的 L 个子向量分别用与所述子向量距离最近的所述聚类的索引编号表示，生成量化编码后的特征描述符。

3、如权利要求 2 所述的一种语音信息搜索方法，其特征在于，所述存储所述特征描述符包括：

建立高度为 $(L+1)$ 的 K -d 树；

20 为所述 K -d 树的非叶子节点划分索引维度和与所述索引维度相对应的划分值；建立与所述划分值进行比较的结果路径指向；

从所述 K -d 树的根节点开始，将与非叶子节点的索引维度相对应的特征描述符的值与所述非叶子节点的划分值进行比较，并基于比较的结果与所述结果路径指向将所述特征描述符存储至所述 K -d 树的叶子节点中。

25 4、如权利要求 3 所述的一种语音信息搜索方法，其特征在于，所述为非叶子节点划分索引维度包括：

为非叶子节点划分的索引维度值 S 为随机生成的取值范围为 $1 \leq S \leq L$ 的整数，并且当前非叶子节点的索引维度值 S 为从所述 K -d 树根节点到所述当前非叶子节点的路径上未划分过的索引维度值。

30 5、如权利要求 1 所述的一种语音信息搜索方法，其特征在于，所述根据预定规则在所述目标候选集中选取所述待识别语音的搜索结果包括：

在所述目标候选集中选取与所述待识别语音的特征描述符欧氏距离最小的前R个特征描述符作为搜索结果集，以所述搜索结果集所对应的目标语音作为所述待识别语音的搜索结果， $R \geq 1$ 。

5 6、如权利要求 1 所述的一种语音信息搜索方法，其特征在于，所述生成所述目标语音的特征描述符包括：

对所述提取的语音特征通过 k-means 聚类方法获取所述目标语音的码本；

获取所述目标语音的语音特征集合，计算所述语音特征集合与所述码本中距离最近的码字所有残差向量之和；

对所述码字的残差向量之和进行归一化，生成所述目标语音的特征描述符。

10 7、如权利要求 1 所述的一种语音信息搜索方法，其特征在于，所述提取语音信息库中目标语音的语音特征包括：

对所述目标语音进行预处理；

计算所述预处理后的目标语音的能量谱；

对所述能量谱进行 Mel 滤波，计算所述 Mel 滤波后的能量谱的对数；

15 对所述能量谱的对数进行 DCT 变换得到 MFCC 系数，获取所述目标语音的语音特征。

8、如权利要求 7 所述的一种语音信息搜索方法，其特征在于，所述对所述目标语音进行预处理包括：对所述目标语音进行语音格式转换、预加重、分帧、加窗处理。

9、如权利要求 8 所述的一种语音信息搜索方法，其特征在于，所述方法还包括：

20 计算所述 MFCC 系数的一阶和二阶差分系数，将所述一阶和二阶差分系数添加至所述语音特征中。

10、一种语音信息搜索装置，其特征在于，所述装置包括：

信息获取模块，用于获取目标语音，并提取所述目标语音的语音特征；

描述符模块，用于基于所述目标语音的语音特征生成所述目标语音的特征描述符；

25 量化编码模块，用于对所述特征描述符进行量化编码，生成量化编码后的特征描述符，并存储所述特征描述符；

识别信息模块，用于获取待识别语音的特征描述符；

第一搜索模块，用于在所述存储的所述特征描述符中，查找与所述待识别语音的特征描述符相匹配的特征描述符对应的目标语音，将查找到的目标语音作为所述待识别语音对应的目标候选集；

30 第二搜索模块，用于根据预定规则在所述目标候选集中选取所述待识别语音的搜索结果。

11、如权利要求 10 所述的一种语音信息搜索装置，其特征在于，所述量化编码模块包括：

聚类模块，用于将每个所述特征描述符等分成 L 个子向量，对所述 L 个子向量分别进行聚类，并设置所述子向量聚类后的索引编号， $L \geq 2$ ；

5 映射模块，用于将每个所述特征描述符的 L 个子向量分别用与所述子向量距离最近的所述聚类的索引编号表示，形成量化编码后的特征描述符。

12、如权利要求 10 所述的一种语音信息搜索装置，其特征在于，所述量化编码模块包括：

索引树构建模块，用于建立高度为 $(L+1)$ 的 K -d 树；

10 预置规则模块，用于根据预先设置规则为所述 K -d 树的非叶子节点划分索引维度和与所述索引维度相对应的划分值，并建立与所述划分值进行比较的结果路径指向；

存储模块，用于从所述 K -d 树的根节点开始，将与非叶子节点的索引维度相对应的特征描述符的值与所述非叶子节点的划分值进行比较，并基于比较的结果与所述结果路径指向将所述特征描述符存储至所述 K -d 树的叶子节点中。

15 13、如权利要求 12 所述的一种语音信息搜索装置，其特征在于，所述预置规则模块为非叶子节点划分索引维度包括：

为非叶子节点划分的索引维度值 S 为随机生成的取值范围为 $1 \leq S \leq L$ 的整数，并且当前非叶子节点的索引维度值 S 为从所述 K -d 树根节点到所述当前非叶子节点的路径上未划分过的索引维度值。

20 14、如权利要求 10 所述的一种语音信息搜索装置，其特征在于，所述第二搜索模块包括：

距离计算模块，用于计算所述目标候选集中的特征描述符与所述待识别语音的特征描述符的欧氏距离；

25 筛选模块，用于在所述目标候选集中选取与所述待识别语音的特征描述符欧氏距离最小的前 R 个特征描述符作为搜索结果集， $R \geq 1$ ；

目标语音模块，用于获取与搜索结果集中的特征描述符相对应的目标语音。

15、如权利要求 10 所述的一种语音信息搜索装置，其特征在于，所述描述符模块包括：

码本训练模块，用于对所述提取的语音特征通过 k -means 聚类方法获取所述目标语音的码本；

30 语音特征集合模块，获取语音特征集合，并计算所述语音特征集合与所述码本中距离最近的码字所有残差向量之和；

归一化处理模块，用于对所述码字的残差向量之和进行归一化，生成所述目标语音的特征描述符。

16、一种语音信息搜索服务器，其特征在于，所述服务器被设置成，包括：

5 第一处理单元，用于获取目标语音，生成所述目标语音的特征描述符；还用于对所述特征描述符进行量化编码；

存储单元，用于分别存储所述量化编码后的特征描述符中路径相同的特征描述符；

第二处理单元，用于获取待识别语音的特征描述符；还用于在所述存储的特征描述符中查找与所述待识别语音相匹配的特征描述符的目标语音，获取候选集；还用于根据预定规则在所述候选集中选取所述待识别语音的搜索结果。

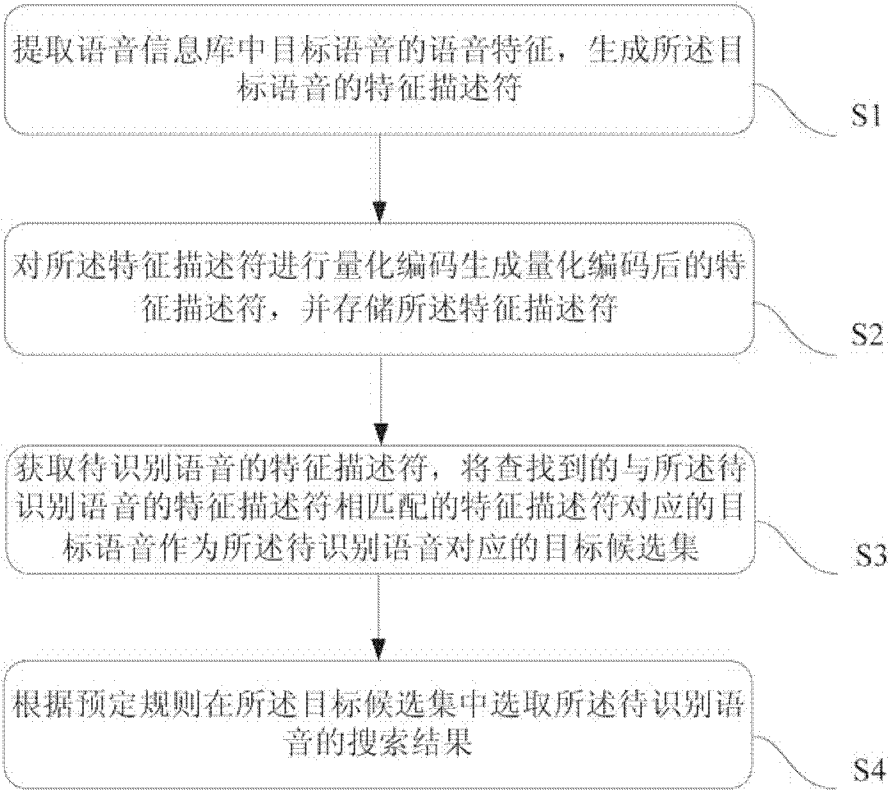


图 1

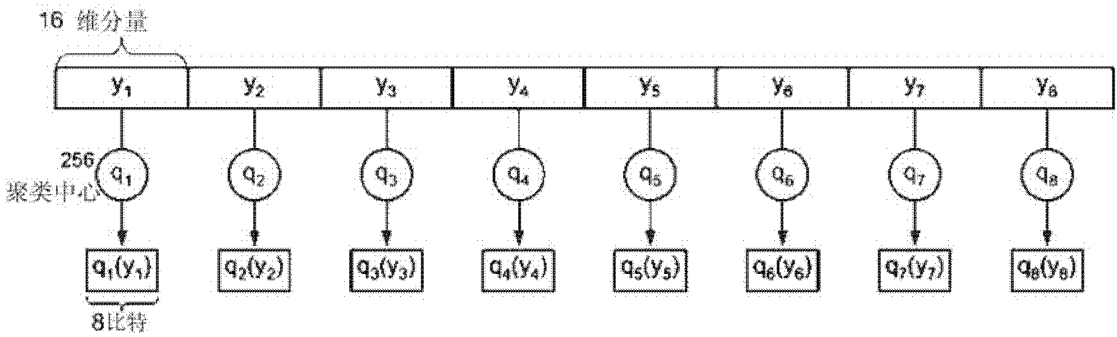


图 2

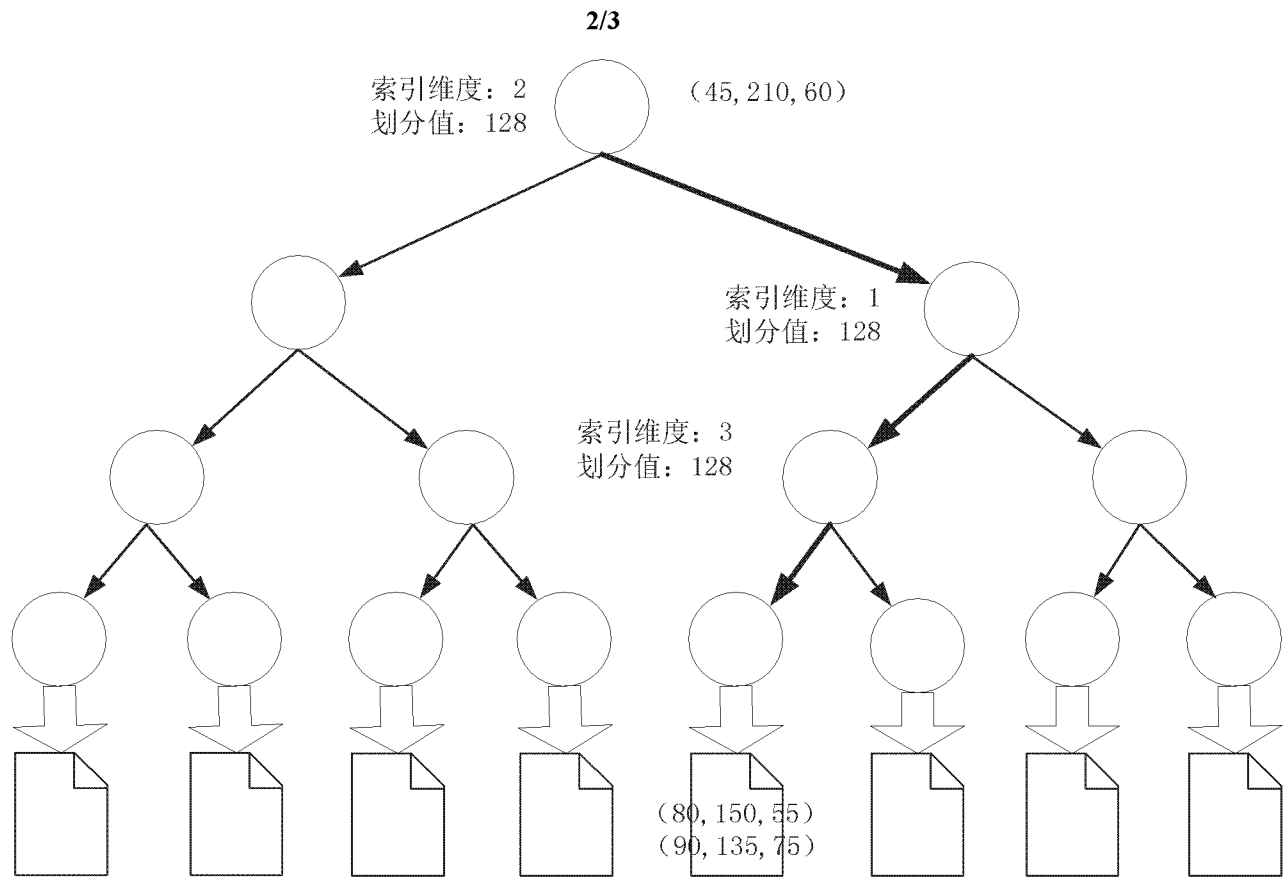


图 3

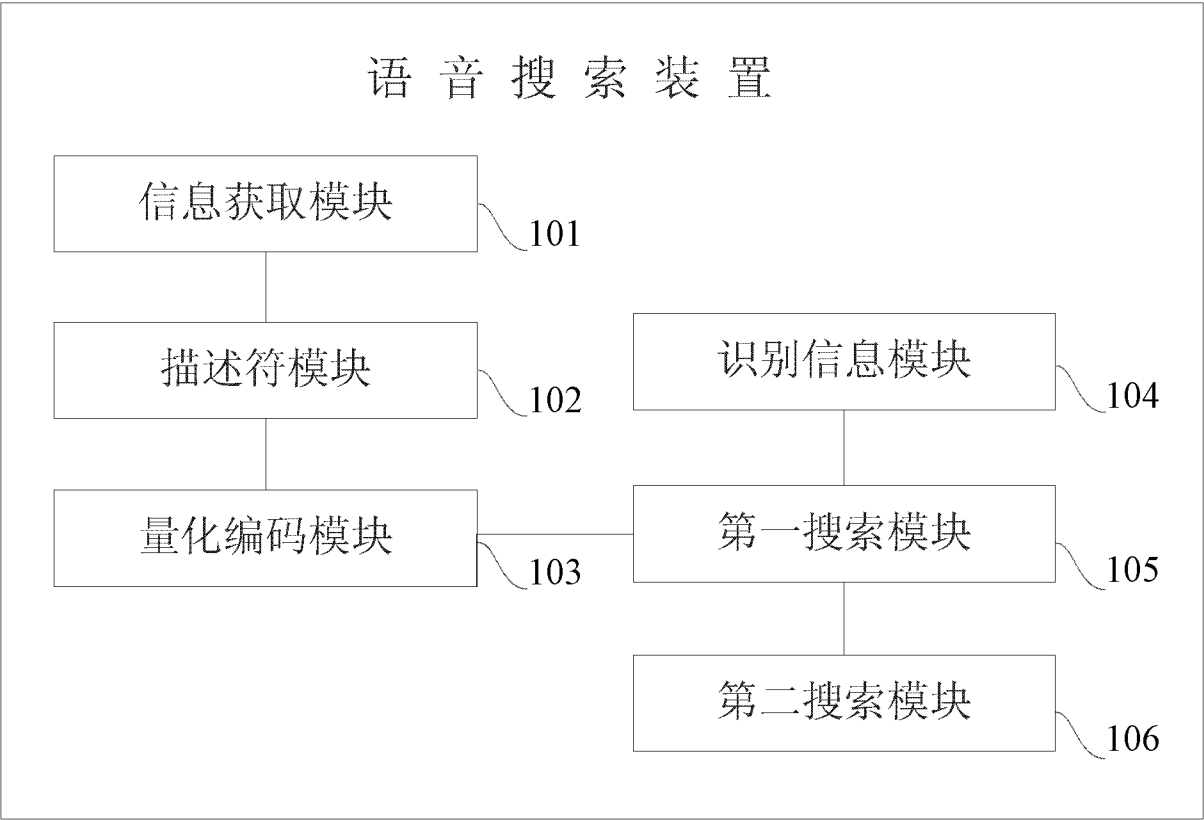


图 4

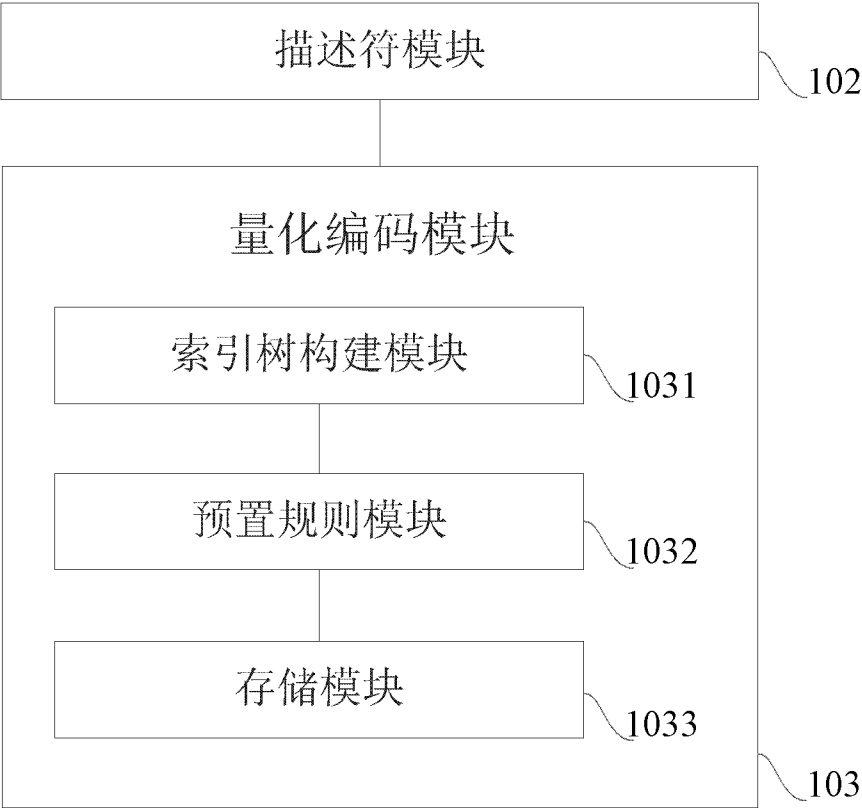


图 5

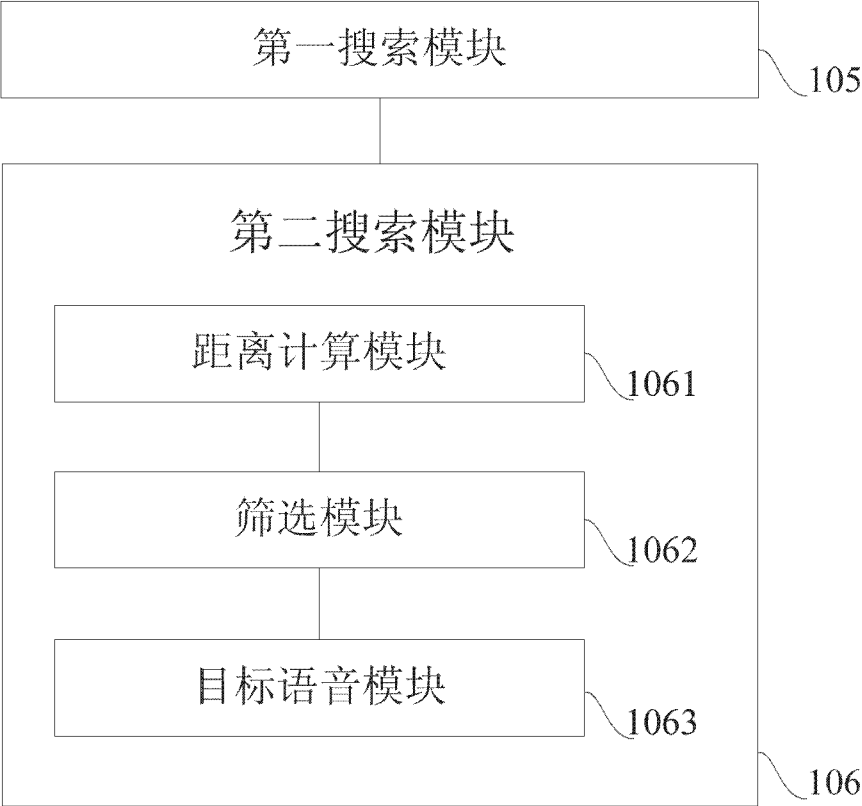


图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/071164

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/08 (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

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)

CNKI, CNPAT, WPI, EPODOC: speech, voice, audio, search+, recogni+, character?, descript+, vector?, compar+, match+, cod+,
k-d w tree

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 101154379 A (SHARP K.K) 02 April 2008 (02.04.2008) description, page 7, line 16 to page 9, line 11	1-16
Y	CN 1455388 A (INSTITUTE OF ACOUSTICS, CHINESE ACADEMY OF SCIENCES et al.) 12 November 2003 (12.11.2003) description, page 5, line 17 to line 24, line 26 to page 6, line 2, page 8, line 10 and line 11	1-16
Y	CN 101030369 A (TSINGHUA UNIVERSITY) 05 September 2007 (05.09.2007) description, page 6, line 3 from the bottom to page 10, line 13	7-9

☒ 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 April 2016	Date of mailing of the international search report 27 April 2016
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 WANG, Wei Telephone No. (86-10) 62413695

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/CN2016/071164

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103310790 A (FUTAIHUA INDUSTRY (SHENZHEN) CO., LTD. et al.) 18 September 2013 (18.09.2013) the whole document	1-16
A	EP 0838803 A2 (NEC CORPORATION) 29 April 1998 (29.04.1998) the whole document	1-16

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/071164

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101154379 A	02 April 2008	WO 2008044582 A1	17 April 2008
		EP 2070079 A1	17 June 2009
		JP 2010504553 A	12 February 2010
		US 2010094626 A1	15 April 2010
CN 1455388 A	12 November 2003	None	
CN 101030369 A	05 September 2007	None	
CN 103310790 A	18 September 2013	US 2013238317 A1	12 September 2013
		TW 201337911 A	16 September 2013
EP 0838803 A2	29 April 1998	JP 3039623 B2	08 May 2000
		US 5953699 A	14 September 1999
		DE 69715343 T2	05 June 2003

A. 主题的分类 G10L 15/08 (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) G10L 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)) CNKI, CNPAT, WPI, EPDOC: 语音, 音频, 搜索, 检索, 查询, 识别, 特征, 特征值, 描述, 矢量, 向量, 匹配, 量化编码, k-d 树, speech, voice, audio, search+, recogni+, character?, descript+, vector?, compar+, match+, cod+, k-d w tree		
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	CN 101154379 A (夏普株式会社) 2008年 4月 2日 (2008 - 04 - 02) 说明书第7页第16行-第9页第11行	1-16
Y	CN 1455388 A (中国科学院声学研究所 等) 2003年 11月 12日 (2003 - 11 - 12) 说明书第5页第17-24行, 第26行-第6页第2行, 第8页第10-11行	1-16
Y	CN 101030369 A (清华大学) 2007年 9月 5日 (2007 - 09 - 05) 说明书第6页倒数第3行-第10页第13行	7-9
A	CN 103310790 A (富泰华工业深圳有限公司 等) 2013年 9月 18日 (2013 - 09 - 18) 全文	1-16
A	EP 0838803 A2 (NEC CORPORATION) 1998年 4月 29日 (1998 - 04 - 29) 全文	1-16
☐ 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 2016年 4月 7日		国际检索报告邮寄日期 2016年 4月 27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		授权官员 王伟 电话号码 (86-10) 62413695

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/071164

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	101154379	A	2008年 4月 2日	WO	2008044582	A1	2008年 4月 17日
				EP	2070079	A1	2009年 6月 17日
				JP	2010504553	A	2010年 2月 12日
				US	2010094626	A1	2010年 4月 15日
CN	1455388	A	2003年 11月 12日	无			
CN	101030369	A	2007年 9月 5日	无			
CN	103310790	A	2013年 9月 18日	US	2013238317	A1	2013年 9月 12日
				TW	201337911	A	2013年 9月 16日
EP	0838803	A2	1998年 4月 29日	JP	3039623	B2	2000年 5月 8日
				US	5953699	A	1999年 9月 14日
				DE	69715343	T2	2003年 6月 5日