(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105575383 A
(43)申请公布日 2016.05.11

(21)申请号 201510657714.4

(22)申请日 2015.10.13

(30)优先权数据
10-2014-0147474 2014.10.28 KR

(71)申请人 现代摩比斯株式会社
地址 韩国京畿道

(72)发明人 权吾泫

(74)专利代理机构 北京同立钧成知识产权代理有限公司 11205
代理人 杨贝贝 臧建明

(51)Int.Cl.
G10L 13/08(2013.01)
G10L 25/03(2013.01)

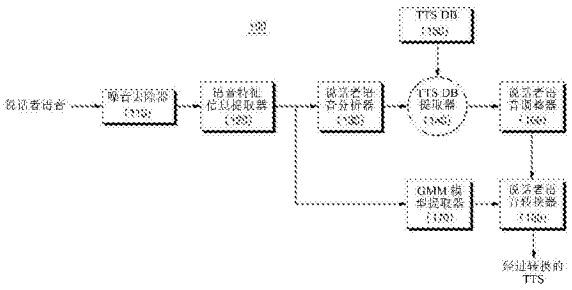
权利要求书2页 说明书12页 附图3页

(54)发明名称

利用用户的语音特征的对象信息语音输出控制装置及方法

(57)摘要

本发明提供一种根据从用户的语音得到的特征信息提供TTS服务的利用用户的语音特征的对象信息语音输出控制装置及方法。本发明的对象信息语音输出控制装置包括:特征信息生成部,其根据用户的语音信息生成所述用户的特征信息;对象信息生成部,其根据所述特征信息,利用文本形式的第一对象信息生成语音形式的第二对象信息;以及,对象信息输出部,其输出所述第二对象信息。本发明的对象信息语音输出控制装置能够构建自然的语音识别系统,能够提供非机械性的亲和、易懂的语音。



CN 105575383 A

1. 一种利用用户的语音特征的对象信息语音输出控制装置，其特征在于，包括：

特征信息生成部，其根据用户的语音信息生成所述用户的特征信息；

对象信息生成部，其根据所述特征信息，利用文本形式的第一对象信息生成语音形式的第二对象信息；以及

对象信息输出部，其输出所述第二对象信息。

2. 根据权利要求 1 所述的利用用户的语音特征的对象信息语音输出控制装置，其特征在于：

所述特征信息生成部从所述语音信息提取共振峰信息、频率信息、线性预测系数信息、频谱包络线信息、能量信息、说话速度信息及对数谱信息中的至少一种信息，并根据所述至少一种信息实时生成所述特征信息。

3. 根据权利要求 1 所述的利用用户的语音特征的对象信息语音输出控制装置，其特征在于：

所述特征信息生成部实时生成所述用户的性别信息、所述用户的年龄信息及所述用户的感情信息中的至少一种信息作为所述特征信息。

4. 根据权利要求 1 所述的利用用户的语音特征的对象信息语音输出控制装置，其特征在于：

所述特征信息生成部从所述语音信息中去除噪音信息后生成所述特征信息。

5. 根据权利要求 1 所述的利用用户的语音特征的对象信息语音输出控制装置，其特征在于：

所述特征信息生成部向所述语音信息适用加权值信息生成所述特征信息，其中，所述加权值信息为通过学习对应于所述语音信息的输入信息与各输入信息的目标信息得到的信息。

6. 根据权利要求 5 所述的利用用户的语音特征的对象信息语音输出控制装置，其特征在于：

所述特征信息生成部利用人工神经网络算法、误差反向传播算法及梯度下降法获取所述加权值信息。

7. 根据权利要求 1 所述的利用用户的语音特征的对象信息语音输出控制装置，其特征在于：

所述对象信息生成部从数据库中提取对应于所述特征信息的基准信息，并根据所述基准信息对所述第一对象信息转换成语音得到的信息进行调整生成所述第二对象信息。

8. 根据权利要求 7 所述的利用用户的语音特征的对象信息语音输出控制装置，其特征在于：

所述对象信息生成部根据从所述基准信息得到的说话速度信息或频率信息，对所述第一对象信息转换成语音得到的信息进行调整生成所述第二对象信息。

9. 根据权利要求 7 所述的利用用户的语音特征的对象信息语音输出控制装置，其特征在于：

所述对象信息生成部根据所述基准信息与从所述特征信息获取的说话者识别信息生成所述第二对象信息。

10. 根据权利要求 9 所述的利用用户的语音特征的对象信息语音输出控制装置，其特

征在于：

所述对象信息生成部根据高斯混合模型获取所述说话者识别信息。

11. 一种利用用户的语音特征的对象信息语音输出控制方法，其特征在于，包括：

根据用户的语音信息生成所述用户的特征信息的步骤；

根据所述特征信息，利用文本形式的第一对象信息生成语音形式的第二对象信息的步骤；以及

输出所述第二对象信息的步骤。

12. 根据权利要求 11 所述的利用用户的语音特征的对象信息语音输出控制方法，其特征在于：

生成所述特征信息的步骤具体是，从所述语音信息提取共振峰信息、频率信息、线性预测系数信息、频谱包络线信息、能量信息、说话速度信息及对数谱信息中的至少一种信息，并根据所述至少一种信息实时生成所述特征信息。

13. 根据权利要求 11 所述的利用用户的语音特征的对象信息语音输出控制方法，其特征在于：

生成所述特征信息的步骤具体是，实时生成所述用户的性别信息、所述用户的年龄信息及所述用户的感情信息中的至少一种信息作为所述特征信息。

14. 根据权利要求 11 所述的利用用户的语音特征的对象信息语音输出控制方法，其特征在于：

生成所述第二对象信息的步骤具体是，从数据库中提取对应于所述特征信息的基准信息，并根据所述基准信息对所述第一对象信息转换成语音得到的信息进行调整生成所述第二对象信息。

15. 根据权利要求 14 所述的利用用户的语音特征的对象信息语音输出控制方法，其特征在于：

生成所述第二对象信息的步骤具体是，根据所述基准信息与从所述特征信息获取的说话者识别信息生成所述第二对象信息。

利用用户的语音特征的对象信息语音输出控制装置及方法

技术领域

[0001] 本发明涉及通过语音输出对象信息的控制装置及方法，尤其涉及一种车辆用语音输出对象信息的控制装置及方法。

背景技术

[0002] 通常，文转声（Text To Speech；以下简称‘TTS’）是将文字或记号转换成语音输出的技术。TTS构建关于音素的发音数据库并将此连接成连续的语音，此时关键是通过调节语音大小、长度、高低等合成自然的语音。

[0003] 即，TTS是将字符串（文章）转换成语音的文字-语音转换装置，大致分为语言处理、生成韵律、波形合成等三个步骤，具体是接收到文本时在语言处理步骤分析接收到的文书的语法结构，根据分析得到的语法结构生成像真人朗读一样的韵律，并根据生成的韵律汇集存储的语音数据库（以下简称‘DB’）的基本单位生成合成音。

[0004] TTS无对象词汇限制，将一般文字形式的信息转换成语音，因此构建系统时应用语音学、语音分析、语音合成及语音识别技术等输出多种自然的语音。

[0005] 但目前提供这种TTS的终端在用语音输出文字消息等情况下，无论对方是谁，都用预先设定的相同语音进行输出，因此无法满足各类用户的需求。

[0006] 韩国公开专利第2011-0032256号公开了一种TTS引导广播装置。但由于该装置只不过是一种单纯地将指定文本转换成语音的装置，因此无法解决上述问题。

发明内容

[0007] 技术问题

[0008] 为解决上述问题，本发明的目的在于提供一种根据从用户的语音获取的特征信息提供TTS(Text To Speech)服务的利用用户的语音特征(characteristic of user voice)的对象信息语音输出控制装置及方法。

[0009] 但本发明的目的不限于以上记载的内容，本领域技术人员可通过以下记载内容明确理解未记载的其他目的。

[0010] 技术方案

[0011] 为达成上述目的，本发明提供一种利用用户的语音特征的对象信息语音输出控制装置，其特征在于，包括：特征信息生成部，其根据用户的语音信息生成所述用户的特征信息；对象信息生成部，其根据所述特征信息，利用文本形式的第一对象信息生成语音形式的第二对象信息；以及，对象信息输出部，其输出所述第二对象信息。

[0012] 优选地，所述特征信息生成部从所述语音信息提取共振峰(Formant)信息、频率(Log f0)信息、线性预测系数(Linear Predictive Coefficient；LPC)信息、频谱包络线(Spectral Envelope)信息、能量信息、说话速度(Pitch Period)信息及对数谱(Log Spectrum)信息中的至少一种信息，并根据所述至少一种信息实时生成所述特征信息。

[0013] 优选地，所述特征信息生成部实时生成所述用户的性别信息、所述用户的年龄信

息及所述用户的感情信息中的至少一种信息作为所述特征信息。

[0014] 优选地，所述特征信息生成部从所述语音信息中去除噪音信息后生成所述特征信息。

[0015] 优选地，所述特征信息生成部向所述语音信息适用加权值信息生成所述特征信息，其中，所述加权值信息为通过学习 (training) 对应于所述语音信息的输入信息与各输入信息的目标信息得到的信息。

[0016] 优选地，所述特征信息生成部利用人工神经网络 (Artificial Neural Network；ANN) 算法、误差反向传播 (Error Back Propagation；EBP) 算法及梯度下降法 (Gradient Descent Method) 获取所述加权值信息。

[0017] 优选地，所述对象信息生成部从数据库中提取对应于所述特征信息的基准信息，并根据所述基准信息对所述第一对象信息转换成语音得到的信息进行调整生成所述第二对象信息。

[0018] 优选地，所述对象信息生成部根据从所述基准信息得到的说话速度 (Pitch Period) 信息或频率 (Log f0) 信息，对所述第一对象信息转换成语音得到的信息进行调整生成所述第二对象信息。

[0019] 优选地，所述对象信息生成部根据所述基准信息与从所述特征信息获取的说话者识别信息生成所述第二对象信息。

[0020] 优选地，所述对象信息生成部根据高斯混合模型 (GMM) 获取所述说话者识别信息。

[0021] 并且，本发明提供一种利用用户的语音特征的对象信息语音输出控制方法，其特征在于，包括：根据用户的语音信息生成所述用户的特征信息的步骤；根据所述特征信息，利用文本形式的第一对象信息生成语音形式的第二对象信息的步骤；以及，输出所述第二对象信息的步骤。

[0022] 优选地，生成所述特征信息的步骤具体是，从所述语音信息提取共振峰 (Formant) 信息、频率 (Log f0) 信息、线性预测系数 (Linear Predictive Coefficient；LPC) 信息、频谱包络线 (Spectral Envelope) 信息、能量信息、说话速度 (Pitch Period) 信息及对数谱 (Log Spectrum) 信息中的至少一种信息，并根据所述至少一种信息实时生成所述特征信息。

[0023] 优选地，生成所述特征信息的步骤具体是，实时生成所述用户的性别信息、所述用户的年龄信息及所述用户的感情信息中的至少一种信息作为所述特征信息。

[0024] 优选地，生成所述特征信息的步骤具体是，从所述语音信息中去除噪音信息后生成所述特征信息。

[0025] 优选地，生成所述特征信息的步骤具体是，向所述语音信息适用加权值信息生成所述特征信息，其中，所述加权值信息为通过学习 (training) 对应于所述语音信息的输入信息与各输入信息的目标信息得到的信息。

[0026] 优选地，生成所述特征信息的步骤具体是，利用人工神经网络 (Artificial Neural Network；ANN) 算法、误差反向传播 (Error Back Propagation；EBP) 算法及梯度下降法 (Gradient Descent Method) 获取所述加权值信息。

[0027] 优选地，生成所述第二对象信息的步骤具体是，从数据库中提取对应于所述特征

信息的基准信息，并根据所述基准信息对所述第一对象信息转换成语音得到的信息进行调整生成所述第二对象信息。

[0028] 优选地，生成所述第二对象信息的步骤具体是，从所述基准信息得到的说话速度(Pitch Period)信息或频率(Log f0)信息，对所述第一对象信息转换成语音得到的信息进行调整生成所述第二对象信息。

[0029] 优选地，生成所述第二对象信息的步骤具体是根据所述基准信息与从所述特征信息获取的说话者识别信息生成所述第二对象信息。

[0030] 优选地，生成所述第二对象信息的步骤具体是根据高斯混合模型(GMM)获取所述说话者识别信息。

[0031] 技术效果

[0032] 本发明根据从用户的语音获取的特征信息提供文转声(Text To Speech，以下简称‘TTS’)服务，从而具有如下效果：

[0033] 第一，从单向方式改成双向方式沟通，从而能够构建自然的语音识别系统。

[0034] 第二，系统提供与驾驶员性别、年龄、爱好等相符的TTS服务，因此车辆的语音识别系统能够提供非机械性的亲和、易懂的语音。

附图说明

[0035] 图1为显示根据本发明一个实施例的车辆用语音引导提供系统的内部构成的概念图；

[0036] 图2及图3为用于说明图1所示车辆用语音引导提供系统中的说话者语音分析器的参考图；

[0037] 图4为显示根据本发明一个实施例的车辆用语音引导提供系统工作方法的流程图。

具体实施方式

[0038] 以下参照附图具体说明本发明的优选实施例。首先，需要注意的是在对各图的构成要素添加附图标记方面，即使相同的构成要素出现在不同的附图上也尽可能添加相同的附图标记。并且在说明本发明时若判断认为对相关公知结构或功能的具体说明可能对本发明的主题造成混淆，则省略相关具体说明。另外，以下将说明本发明的优选实施例，但本发明的技术方案并不限定或限制于此，所属技术领域的技术人员可做多种变形实施。

[0039] 本发明的目的在于分析车辆内驾驶员的语音特征并提供更加自然亲切的语音引导服务。

[0040] 图1为显示根据本发明一个实施例的车辆用语音引导提供系统的内部构成的概念图。

[0041] 车辆用语音引导提供系统100是利用驾驶员的语音，通过与当前驾驶员的语音相似的图案提供语音引导的系统，如图1所示，包括噪音去除器110、语音特征信息提取器120、说话者语音分析器130、文转声数据库提取器(以下简称‘TTS DB提取器’)140、TTS DB(以下简称‘文转声数据库’)150、说话者语音调整器160、高斯混合模型提取器(Gaussian Mixture Model提取器，以下简称‘GMM提取器’)170及说话者语音转换器180。

[0042] 车辆内导航引导语音或语音识别引导语音一般使用生产时既已固定的特定TTS DB。因此，无法充分满足希望按年龄、性别、驾驶员爱好进行语音引导的消费者需求(Needs)。例如，岁数大的老年人可能不太容易听懂朝气蓬勃的二十多岁人员语速较快的语音，而年轻人则认为五十多岁人员的慢速语音枯燥、无个性。

[0043] 本发明的车辆用语音引导提供系统100的目的在于为年轻人、中年人、老年人及男性、女性、性格活泼或温柔的驾驶员提供亲和、易懂的语音品质，而不是提供机械性的TTS引导语音。

[0044] 并且，车辆用语音引导提供系统100的目的在于在双向沟通方式的技术发展下通过语音识别这种说话者识别功能区分驾驶员并首先推荐最适合驾驶员的功能，以适应人工智能趋势。

[0045] 以下参照图1进行具体说明。

[0046] 噪音去除器110的功能是在接收到说话者的语音信息时去除该语音信息中的噪音成分。噪音去除器110通过去除车辆内的噪音获取更加清楚的驾驶员语音。

[0047] 语音特征信息提取器120的功能是从去除噪音成分后的语音信息中提取说话者的语音特征信息。为了分析说话者的年龄、性别、爱好等，语音特征信息提取器120提取个人的语音特征信息。

[0048] 语音特征信息提取器120从语音信息中提取共振峰(Formant)信息、频率(Log f0)信息、线性预测系数(Linear Predictive Coefficient;LPC)信息、频谱包络线(Spectral Envelope)信息、能量(Energy)信息、说话速度(Pitch Period)信息、对数谱(Log Spectrum)信息等语音特征信息。

[0049] 说话者语音分析器130的功能是利用语音特征信息提取器120提取的语音特征信息，对说话者的年龄、性别、爱好等进行分类(Classification)。说话者语音分析器130在区分性别时可采用Log f0信息，Log f0平均值为120Hz～240Hz时可判断为女性，Log f0平均值为0Hz～120Hz时可判断为男性。

[0050] 语音特征信息提取器120提取个人的语音特征信息后，说话者语音分析器130利用人工神经网络(Artificial Neural Network;ANN)算法建模(Modeling)，提取一般化的按年龄、性别、爱好等进行分析的人工神经网络算法的加权值(Weight)信息。说话者语音分析器130可以根据如上提取的一般化的加权值信息(即，利用人工神经网络算法得到的建模结果数据)提取实时输入的驾驶员的语音的特征信息，以此推定说话者的年龄、性别、爱好等。

[0051] 为推定说话者的年龄、性别、爱好等，说话者语音分析器130可以利用年龄分析用神经网络(Neural Network)、性别分析用神经网络、爱好分析用神经网络等人工神经网络算法。

[0052] 以下参照图2及图3进一步说明说话者语音分析器130。

[0053] 图2及图3为用于说明图1所示车辆用语音引导提供系统中的说话者语音分析器的参考图。

[0054] 人工神经网络(Artificial Neural Network;ANN)算法是按神经细胞间的连接关系建模和区分人类大脑的作用的算法。本实施例中，说话者语音分析器130通过依次执行以下两个步骤实现人工神经网络算法。图2为说明适用于本发明的人工神经网络算法的人

工神经网络的神经单元（处理要素）结构的参考图。

[0055] 1. 学习步骤 (Training, Modeling)

[0056] 在学习步骤，说话者语音分析器 130 将大量输入向量与目标向量输入到指定的神经网络中进行图案分类，以获取最佳的加权值 (Weight)220。

[0057] 2. 判别 (Classification)

[0058] 在判别步骤，说话者语音分析器 130 通过学习得到的加权值 220 与输入向量 210 之间的运算式 230 算出输出值 240。说话者语音分析器 130 可以计算加权值 220 与输入向量 210 之间的差值，判别最接近的输出 (Output) 为最终算出结果。运算式 230 中 θ 表示临界值。

[0059] 在利用人工神经网络算法，根据说话者语音特征信息分析说话者的年龄、性别、爱好等时，说话者语音分析器 130 可适用多层感知机 (Multi-Layer Perceptron)，尤其可以适用误差反向传播 (Error Back Propagation ;EBP) 算法。以下参照图 3 进一步进行说明。图 3 为用于显示将适用于本发明的 EBP 算法的结构的参考图。

[0060] 目前与语音相关的感知机理论一直以来用于识别语音（接收到语音时判断语音的内容）或判别人的感情。

[0061] 多层感知机 (multilayer perceptron) 是输入层与输出层之间具有一个以上中间层的神经网络。网络是按照输入层、隐层、输出层方向连接，不存在各层内连接及从输出层到输入层的直接连接的前馈 (Feedforward) 网络。

[0062] 为了将这种多层感知机适用到说话者语音分析器 130，本发明采用 EBP 算法。

[0063] 本发明中，EBP 算法具有位于输入层与输出层之间的一个以上隐层。并且，本发明中 EBP 算法如数学式 1 所示，通过梯度下降法 (gradient-descent method) 向最小化的方向学习代价函数 (Cost function) 值得出所需的加权值，其中所述代价函数是利用一般化的德尔塔 (delta) 定律定义的所需目标值 D_{pj} 与实际输出值 O_{pj} 之间的误差平方和：

[0064] 【数学式 1】

[0065] $$E=\sum_{p}E_{p},\qquad \left(E_{p}=\frac{1}{2}\sum_{j}(D_{pj}-O_{pj})^{2}\right)$$

[0066] 其中，p 表示第 p 学习图案，E_p 表示关于第 p 图案的误差。并且，D_{pj} 表示关于第 p 图案的第 j 要素，O_{pj} 表示实际输出的第 j 要素。

[0067] 说话者语音分析器 130 通过利用以上说明的 EBP 算法，为隐层学习而利用输出层发生的误差计算隐层误差，并将该值逆向传播到输入层，通过重复该过程直至输出层的误差达到目标水平，如上得到最佳的加权值。

[0068] 说话者语音分析器 130 可利用 EBP 算法按如下步骤执行学习 (Training) 步骤。

[0069] 首先，第一步骤初始化加权值 (Weight) 与临界值。

[0070] 然后，第二步骤给出输入向量 (Input Vector) X_p 与目标向量 (Target Vector) d_p。

[0071] 然后，第三步骤利用给出的输入向量计算用于输入到隐层 (Hidden Layer) 第 j 神经单元的输入值。此时可利用数学式 2：

[0072] 【数学式 2】

[0073] $$net_{pj}=\sum_{i=0}^{N-1}W_{ji}X_{pi}-\theta_j$$

[0074] 其中，net_{pj}表示输入到隐层第 j 神经单元的输入值。W_{ji}表示从第 j 神经单元到第 i 神经单元的连接加权值，X_{pi}表示输入向量。并且，θ_j表示临界值。并且，N 表示输入神经单元的个数。

[0075] 然后，第四步骤利用 S 型 (Sigmoid) 函数计算隐层的输出 O_{pj}。

[0076] 然后，第五步骤利用隐层的输出计算用于输入到输出层神经单元 k 的输入值。此时可利用数学式 3：

[0077] 【数学式 3】

[0078] $$net_{pk}=\sum_{j=0}^{L-1}W_{kj}O_{pj}-\theta_k$$

[0079] 其中，net_{pk}表示输入到输出层神经单元 k 的输入值。并且 L 表示隐匿神经单元的个数。

[0080] 然后，第六步骤利用 net_{pk}与 S 型 (Sigmoid) 函数计算输出层的输出 O_{pk}。

[0081] 然后，第七步骤计算输入图案的目标输出与实际输出之间的误差，并将输出层误差和作为学习图案的误差累积。此时可利用数学式 4：

[0082] 【数学式 4】

[0083] $$\delta_{pk}=(d_{pk}-O_{pk})f_k'(net_{pk})=(d_{pk}-O_{pk})O_{pk}(1-O_{pk})$$

[0084] $$E=E+E_p,\quad (E_p=\sum_{k=1}^{M-1}\delta_{pk}^2)$$

[0085] 其中，d_{pk}表示输入图案的目标输出，O_{pk}表示输入图案的实际输出。并且，δ_{pk}表示目标输出与实际输出之间的误差。E 表示输出层误差和，E_p表示学习图案的误差。M 表示输出神经单元的个数。

[0086] 然后，第八步骤利用输出层误差值 d_{pk}、隐层及输出层的加权值 W_{kj}等计算隐层的误差 δ_{pj}。此时可利用数学式 5：

[0087] 【数学式 5】

[0088] $$\delta_{pj}=f_j'(net_{pj})\sum_{k=0}^{M-1}\delta_{pk}W_{kj}=\sum_{k=0}^{M-1}\delta_{pk}W_{kj}O_{pj}(1-O_{pj})$$

[0089] 然后，第九步骤利用在第四步骤及第七步骤求得的隐层神经单元 j 的输出值 O_{pj}与输出层的误差值 δ_{pk}更新输出层的加权值 W_{kj}。此时还调整临界值，假设为与常数值输入相关联的加权值，因此按近似方式适用。此时可利用数学式 6：

[0090] 【数学式 6】

[0091] $$W_{kj}(t+1)=W_{kj}(t)+\eta\,\delta_{pk}O_{pj}$$

[0092] $$\theta_k(t+1)=\theta_k(t)+\beta\,\delta_{pk}$$

[0093] 其中，η 与 β 是增益值，特别地，η 表示学习率，t 表示时刻。$W_{kj}(t)$ 表示时间 t 时从隐匿神经单元 j 到输出神经单元 k 的加权值。

[0094] 然后，第十步骤也像输出层一样更新输入层与隐层的加权值 W_{ji}及临界值 θ_j。此时可利用数学式 7：

[0095] 【数学式 7】

[0096] $W_{ji}(t+1) = W_{ji}(t)+\eta\ \delta_{pj}X_{pi}$

[0097] $\theta_j(t+1) = \theta j(t)+\beta\ \delta_{pj}$

[0098] 然后，第十一步骤分支到第二步骤重复执行直至全部学习所有学习图案。

[0099] 然后，第十二步骤在输出层的误差和 E 为允许值以下或大于最大重复次数时结束，否则转到第二步骤并执行之后的步骤。

[0100] 另外，说话者语音分析器 130 还可以在说话者为多人时，利用多层感知机(multilayer perceptron)根据各说话者的语音特征信息分析各说话者的年龄、性别、爱好等。以下对此进行说明。

[0101] 根据一般噪音过滤方法，语音识别麦克风开启预定时间后发出语音识别用语音，因此将语音识别前进入麦克风的信号判断为车辆内噪音，然后只过滤信号中的该噪音。

[0102] 车辆内具有朝向驾驶员方向的指向性麦克风，但由于将发出语音前的短时间内输入的信号判断为噪音，因此如果发出语音识别用语音的时间点除驾驶员之外还有其他座位人员说话，那么语音相参杂造成语音识别率下降。

[0103] 因此，本发明在车辆内四个座位区域分别设置指向性麦克风，以驾驶员区域的麦克风的输入信号为基准，将其他区域的麦克风信号判别为噪音并过滤。信号处理过程中实时判别驾驶员区域驾驶员的特征，以使多媒体设备提供适合驾驶员的信息。

[0104] 以下对此做进一步说明，以下说明将驾驶座定义为 A 区域，将副驾驶座定义为 B 区域，将驾驶座的后侧与副驾驶座的后侧分别定义为 C 区域与 D 区域。

[0105] 驾驶员启动语音识别功能时，A、B、C、D 区域的麦克风同时开启，通过麦克风接收四个区域的语音信号。由于四个区域的麦克风接收到的除人类语音之外的车辆噪音值是几乎相同的，因此在 A 过滤车辆噪音值。然后分析四个区域的语音。首先分析四个区域的表示性别的语音向量值，若以 A 区域为基准从 B、C、D 区域提取到表示与 A 区域不同性别的向量值，则从 A 区域中过滤相当于该向量值的信号。性别分析结束后按相同方法分析年龄、心情 / 状态等。

[0106] A 区域中最大的必然是驾驶员的语音信号，但还存在 B、C、D 区域的语音信号时，A 区域无法只提取驾驶员的完整语音，因此采用该方法。

[0107] 此时可以利用除相互关系 (CORRELATION)、ICA 技术、波束形成 (BEAM FORMING) 技术之外的其他算法判别信号独立还是具有近似性。

[0108] 可以在通过四个麦克风进行过滤的同时分析说话者的个别特征，可利用获分析个别特征得到的信息过滤噪音，以此提高识别率。

[0109] 车辆一般具有四个座位，车辆内语音识别系统使用者一般是驾驶员，若驾驶员使用语音识别系统的过程中其他座位乘客说话，则多人的语音相叠加，因此语音识别系统无法识别驾驶员的命令。目前一般使用的语音识别系统是在语音识别区间前设置无语音的区间并将该区间的输入识别为噪音，在语音输入区间过滤噪音的结构。

[0110] 本发明是利用多层感知机理论提取语音的特征并识别说话者的特征，根据该数据实时地为说话者提供适合的信息的技术。通过采用多层感知机，①能够根据说话者的特征提供适配信息，或者，②能够识别说话者的位置并提供该位置的说话者所需的功能。以下进一步说明①与②。

[0111] 1. 根据说话者特征提供适配信息

[0112] 利用多层感知机构建系统的情况下，即使多人的语音相叠加也能够提取驾驶员的语音。该方法不仅可以适用于驾驶员，还可以识别其他人员。例如，只提取A区域的语音特征并忽略B、C、D区域的语音信号。

[0113] 多层感知机的大前提是预先形成根据大量DB及反向传播(BACK PROPAGATION)技术进行学习的算法。

[0114] 多层感知机建模具体是，例如分析20～29岁且状态佳的首尔女性的大量语音提取特征（共振峰、基本频率、能量值、LPC值等）并输入到输入端，将20～29岁且状态佳的首尔女性作为输出(OUTPUT)对象的情况下，感知机结构内部经过反向传播(BACK PROPAGATION)过程确定适当的加权(WEIGHT)值。在如上学习多种特征的人的情况下，输入的任何语音都能够在经过学习的结构内找到特征。LPC值是线性预测编码值，是基于人类发声模型的语音编码方式中的一种，具有二十六维向量。

[0115] 输入特定对象的大量语音的共振峰、基本频率、LPC模型的二十六维向量值的情况下，通过反向展开过程向多个目标重复合适的加权值规所定的作业（例如20～29岁且状态佳的首尔女性、30～40岁且状态不佳的庆尚道地区男性…）。

[0116] 在经过该学习过程的情况下，无论任何语音，只要输入到对该语音的特征向量建模的感知机结构即可获知说话者的特征。

[0117] 将即按即通(push to talk，以下简称‘PTT’)作为座位选择基准。若有四个PTT键，则根据位置将相应PTT输入位置的麦克风接收到的语音判断为需要分析的语音，将其余判断为噪音并过滤。根据过滤后的语音进行识别并为说话者提供最佳信息，以说话者向多媒体产品发出命令的情况为例，若想要查找的是餐厅，则首先查找与说话者特征相符的餐厅。

[0118] 整理以上说明内容可导出如下特征。

[0119] 首先，判别PTT位置并提取对应于各语音信号特征的向量。

[0120] 然后，将四种信号的特征向量输入到多层感知机结构。

[0121] 然后，分别提取各语音信号的特征。

[0122] 然后，当具有与基准语音A不同的特征时，将A麦克风信号中的其他特征值判断为噪音并过滤。

[0123] 然后，利用只提取A区域语音得到的数据识别语音，并判别语音的意思。

[0124] 然后，针对A区域的说话者的命令提供最佳信息。

[0125] 2. 识别说话者位置并提供该位置的说话者所需的功能

[0126] 将即按即通(push to talk，以下简称‘PTT’)作为座位选择基准。若有四个PTT键，则根据位置将相应PTT输入位置的麦克风接收到的语音判断为需要分析的语音，将其余判断为噪音并过滤。以空调为例，若D区域的乘坐人员发出关于空调温度的命令，可以使仅D区域的空调装置按命令调节空调档位。

[0127] 以下再次参照图1进行说明。

[0128] TTS DB 150是存储关于年龄的基准特征信息(10～19岁、20～29岁、30～39岁、40～49岁、50～59岁、60～69岁、70岁以上等)、关于性别的基准特征信息（男性、女性等）、关于爱好的基准特征信息（温柔、活泼等）等信息的数据库。

[0129] TTS DB提取器140的功能是从TTS DB 150检测对应于说话者语音分析器130发

现的说话者年龄、性别、爱好等的信息。

[0130] 说话者语音调整器 160 的功能是根据从 TTS DB 150 检测到的信息调整 (tuning) 为了 TTS 服务而要输出的语音。说话者语音调整器 160 可以将从驾驶员的语音获取的说话速度信息 (Pitch Period)、频率的高低的信息 (Log f0) 等适用到要输出的语音进行调整。

[0131] GMM 模型提取器 170 的功能是根据语音特征信息提取器 120 提取的说话者的语音特征信息生成高斯混合模型。

[0132] 说话者语音转换器 180 的功能是向说话者语音调整器 160 调整的语音适用高斯混合模型以进一步转换语音。本发明中，可以提供经过说话者语音调整器 160 调整的语音作为用于 TTS 服务的语音。但本发明不限于此，本发明还可以通过 GMM(Gaussian Mixture Model) 进一步转换说话者的语音，以确保能够实时合理转换说话者的语音特征。

[0133] 以下进一步说明利用高斯混合模型的说话者语音转换器 180。

[0134] $x \in R^n$这一特定随机向量的高斯混合密度 (Gaussian Mixture Density) 可用数学式 8 表示：

[0135] 【数学式 8】

[0136] $$p(x|\lambda)=\sum_{i=0}^{Q}\alpha_i b_i(x),\quad \sum_{i=0}^{Q}\alpha_i=1,\ \alpha_i\geq 0$$

[0137] 其中 p(x|λ) 是成分参数，表示具有平均与离散的高斯函数。Q 表示单高斯密度 (Gaussian Density) 的总个数，α_i表示单高斯密度的加权值。

[0138] $b_i(x)$ 表示多维高斯混合密度 (Gaussian mixture density)。该 $b_i(x)$ 用单高斯密度表示如数学式 9 所示：

[0139] 【数学式 9】

[0140] $$b_i(x)=\frac{1}{(2\pi)^{n/2}|C_i|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\right]$$

[0141] μ_i:nxl mean vector, C_i:nxn cov ariance matrix

[0142] 因此，完成的高斯混合密度 (Gaussian Mixture Density) 由如下三个变量构成：

[0143] $\lambda = \{\alpha i, \mu i, Ci\}, i = 1, \cdots, Q$

[0144] 将 $x \in R^n$定义为 TTS DB 提取器 140 筛选出的语音，将 $y \in R^n$定义为驾驶员的语音，则 $z = (x, y)^T$可以定义为 TTS DB 提取器 140 筛选出的语音与驾驶员语音之间的联合密度 (joint density) 语音。这可以用如下数学式表示：

[0145] 【数学式 10】

[0146] $$p(z|\lambda)=\sum_{i=1}^{Q}\frac{\alpha_i}{(2\pi)^n|C_i|^{1/2}}\exp\left[-\frac{1}{2}(z-\mu_i)^T C_i^{-1}(z-\mu_i)\right]$$

[0147] $$\sum_{i=1}^{Q}\alpha_i=1,\ \alpha_i\geq 0$$

[0148] 因此，说话者语音转换器 180 如数学式 11 所示发现最小化均方误差 (Mean Square Error) 的映射 (Mapping) 函数 F(x)。

[0149] 【数学式 11】

[0150] $\varepsilon_{mse} = E[\| y-F(x) \|^2]$

[0151] E 表示期望值 (Expectation)，F(x) 表示所推定 (estimated) 语音的光谱向量 (Spectral Vector)。

[0152] 利用联合密度推定方法 (Joint Density Estimation Method) 的情况下，F(x) 可定义成如以下数学式 12 所示。此时，可参见 'A.Kain and M.Macon,"Spectral voice conversion for text-to-speech synthesis"Proc.ICASSP,pp.285 ～ 288,1998.'。

[0153] 【数学式 12】

[0154] $$F(x) = E[y|x] = \sum_{i=1}^{Q} h_i(x)\left[\mu_i^y + C_i^{yx} C_i^{xx-1}(x-\mu_i^x)\right]$$

[0155] $$h_i(x) = \frac{\dfrac{\alpha_i}{(2\pi)^{n/2}|C_i^{xx}|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i^x)^T C_i^{xx-1}(x-\mu_i^x)\right]}{\sum_{j=1}^{Q} \dfrac{\alpha_j}{(2\pi)^{n/2}|C_j^{xx}|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_j^x)^T C_j^{xx-1}(x-\mu_j^x)\right]}$$

[0156] $$C_i = \begin{bmatrix} C_i^{xx} & C_i^{xy} \\ C_i^{yx} & C_i^{yy} \end{bmatrix}, \qquad \mu_i = \begin{bmatrix} \mu_i^x \\ \mu_i^y \end{bmatrix}$$

[0157] 以下具体说明参照图 1 至图 3 说明的车辆用语音引导提供系统 100 的工作方法。图 4 为显示根据本发明一个实施例的车辆用语音引导提供系统的工作方法的流程图。

[0158] 步骤 S405 中，驾驶员说出特定命令时，步骤 S410 中，语音特征信息提取器 120 从说话者的语音提取特征信息。

[0159] 然后在步骤 S415 中，说话者语音分析器 130 根据特征信息实时分析性别、年龄、爱好等。

[0160] 然后在步骤 S420 中，TTS DB 提取器 140 从 TTS DB 150 选择对应于各分析结果的信息。

[0161] 然后在步骤 S425 中，说话者语音调整器 160 根据 TTS DB 提取器 140 选择的信息调整经过语音转换的信息。

[0162] 然后在步骤 S430 中，说话者语音转换器 180 将根据从说话者语音得到的 GMM 模型调整后的语音转换成接近驾驶员的实际语音。

[0163] 然后在步骤 S435 中，TTS 输出部（未示出）输出经过说话者语音转换器 180 转换后的语音。

[0164] 以上参照图 1 至图 4 说明了本发明的一个实施形态。以下说明能够从这些实施形态得到的本发明优选形态。

[0165] 根据本发明优选实施例的对象信息语音输出控制装置包括特征信息生成部、对象信息生成部、对象信息输出部、电源部及主控制部。

[0166] 电源部的功能是向构成对象信息语音输出控制装置的各构成供应电源。主控制部的功能是控制构成对象信息语音输出控制装置的各构成的所有工作。对象信息语音输出控制装置适用于车辆的情况下，本实施例不具备电源部与主控制部也无妨。

[0167] 特征信息生成部的功能是根据用户的语音信息生成用户的特征信息。特征信息生成部是对应于图 1 中语音特征信息提取器 120 的概念。

[0168] 特征信息生成部从语音信息提取共振峰 (Formant) 信息、频率 (Log f0) 信息、线性预测系数 (Linear Predictive Coefficient ;LPC) 信息、频谱包络线 (Spectral Envelope) 信息、能量信息、说话速度 (Pitch Period) 信息及对数谱 (Log Spectrum) 信息中的至少一种信息，并根据至少一种信息实时生成特征信息。

[0169] 特征信息生成部可以实时生成特征信息，所述特征信息包括用户的性别信息、用户的年龄信息及用户的感情信息中的至少一种信息。这种特征信息生成部是对应于图 1 的语音特征信息提取器 120 与说话者语音分析器 130 的结合构成的概念。

[0170] 特征信息生成部可以从语音信息中去除噪音信息生成特征信息。这种特征信息生成部是对应于图 1 的噪音去除器 110 与语音特征信息提取器 120 的结合构成的概念。

[0171] 特征信息生成部可以向语音信息适用对应于语音信息的输入信息与通过学习 (training) 各输入信息的目标信息得到的加权值信息生成特征信息。

[0172] 特征信息生成部可利用人工神经网络 (Artificial Neural Network ;ANN) 算法、误差反向传播 (Error Back Propagation ;EBP) 算法及梯度下降法 (Gradient Descent Method) 获取加权值信息。

[0173] 对象信息生成部的功能是根据特征信息，利用文本形式的第一对象信息生成语音形式的第二对象信息。

[0174] 对象信息生成部从数据库中提取对应于特征信息的基准信息，并根据该基准信息调整第一对象信息转换成语音得到的信息生成第二对象信息。这种对象信息生成部是对应于图 1 中 TTS DB 150、TTS DB 提取器 140 及说话者语音调整器 160 的结合构成的概念。

[0175] 对象信息生成部可以根据从基准信息获取的说话速度 (Pitch Period) 信息或频率 (Log f0) 信息调整将第一对象信息转换成语音得到的信息以生成第二对象信息。

[0176] 对象信息生成部可以根据基准信息与从特征信息获取的说话者识别信息生成第二对象信息。这种对象信息生成部是对应于 TTS DB 150、TTS DB 提取器 140、说话者语音调整器 160、GMM 模型提取器 170 及说话者语音转换器 180 的结合构成的概念。

[0177] 对象信息生成部可以根据高斯混合模型 (GMM) 获取说话者识别信息。

[0178] 以下说明对象信息语音输出控制装置的工作方法。

[0179] 首先，特征信息生成部根据用户的语音信息生成用户的特征信息。

[0180] 然后，对象信息生成部根据特征信息，利用文本形式的第一对象信息生成语音形式的第二对象信息。

[0181] 然后，对象信息输出部输出第二对象信息。

[0182] 以上记载了构成本发明实施例的所有构成要素结合成一体或结合工作，但本发明并不限定于这些实施例。即在本发明的目的范围内，其所有构成要素中一个以上可选择性结合工作。并且，其所有构成要素可分别为一个独立的硬件，但也可以选择性地组合各构成要素的一部分或全部，通过具有用于执行一个或多个硬件组合实现的部分或全部功能的程序模块的计算机程序来实现。并且，这种计算机程序可存储于 USB 存储器、CD 磁盘、闪存盘 (Flash Memory) 等计算机可读记录介质 (Computer Readable Media)，由计算机读取并执行，实现本发明的实施例。计算机程序记录介质可包括磁性记录介质、光记录介质、载波

(Carrier Wave)介质等。

[0183] 并且,包括技术或科学用语在内的所有用语在具体说明中无另行定义的情况下,表示和本发明所属技术领域的普通技术人员的通常理解相同的意思。通常使用的词典定义的用语,应解释为与相关技术的文章脉络的意思相一致的意思,若本发明中无明确定义,不得解释为理想或过度性的意思。

[0184] 最后应说明的是:以上各实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照上述各实施例对本发明进行了具体的说明,本领域的普通技术人员应当理解:其依然可以对上述各实施例所记载的技术方案进行修改,或者对其中部分或者全部技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

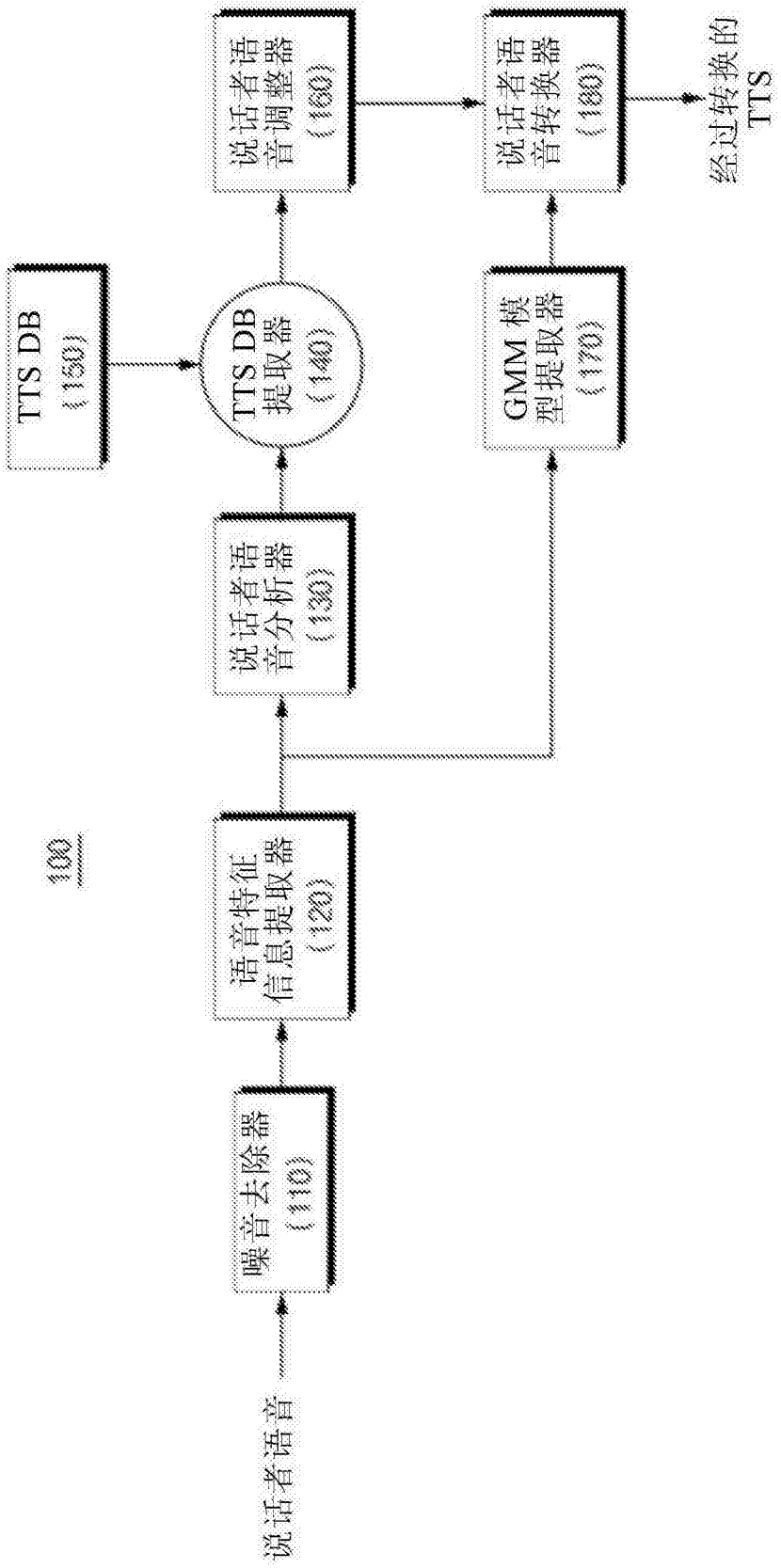

说话者语音
噪音去除器（110）
语音特征信息提取器（120）
说话者语音分析器（130）
TTS DB提取器（140）
TTS DB（150）
说话者语音调整器（160）
GMM模型提取器（170）
说话者语音转换器（180）
经过转换的TTS
100

图 1

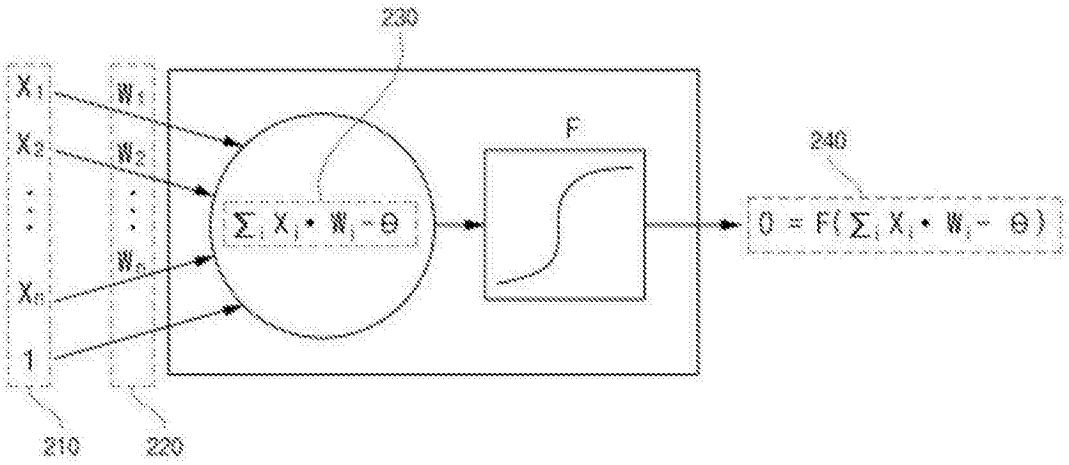

230
F
240
$O = F(\sum_i X_i \cdot W_i - \theta)$
X_1
X_2
X_n
1
W_1
W_2
W_n
$\sum_i X_i \cdot W_i - \theta$
210
220

图 2

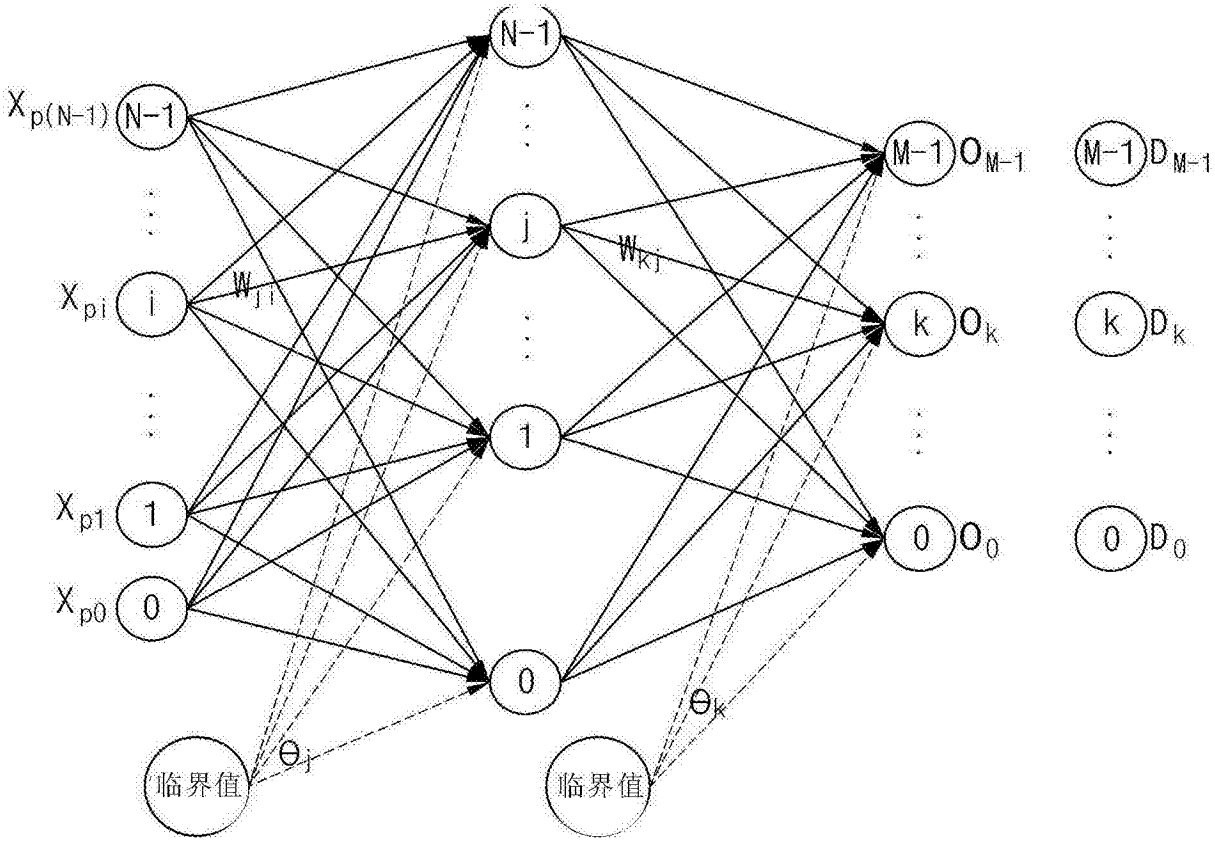

N-1
j
1
0
$X_{p(N-1)}$
X_{pi}
X_{p1}
X_{p0}
i
W_{ji}
W_{kj}
M-1 O_{M-1}
k O_k
0 O_0
M-1 D_{M-1}
k D_k
0 D_0
θ_j
θ_k
临界值
临界值

图 3

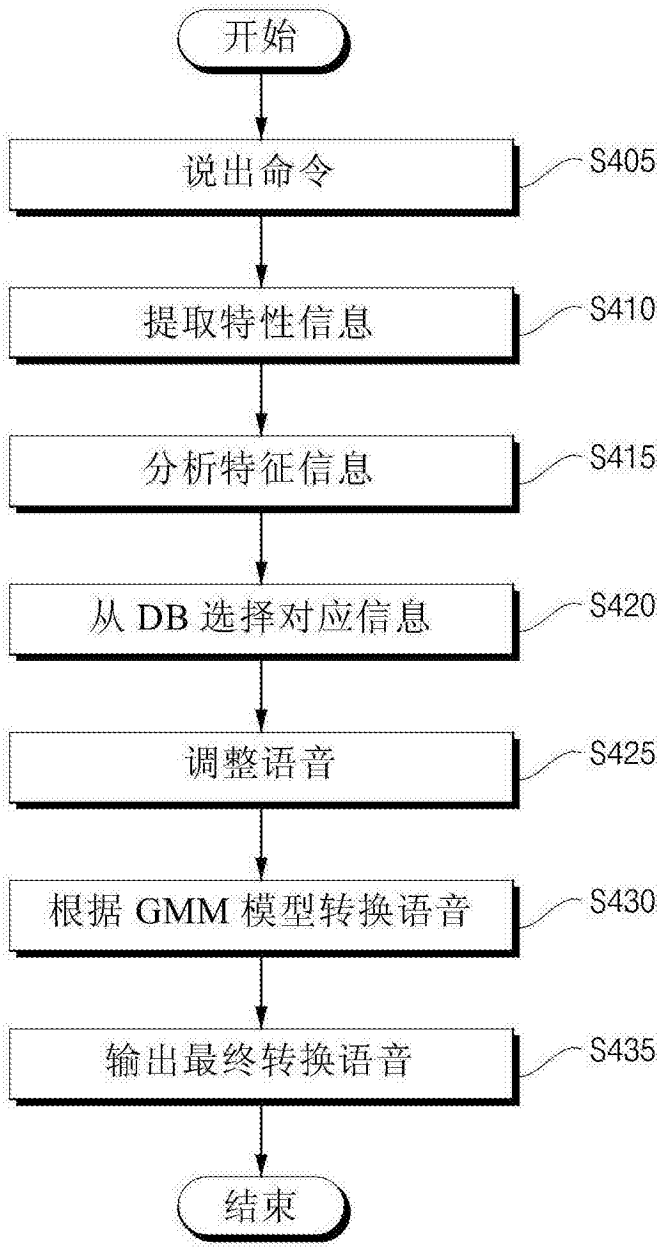

开始
说出命令
S405
提取特性信息
S410
分析特征信息
S415
从 DB 选择对应信息
S420
调整语音
S425
根据 GMM 模型转换语音
S430
输出最终转换语音
S435
结束

图 4