(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103824565 B
(45)授权公告日 2017.02.15

(21)申请号 201410066451.5

(22)申请日 2014.02.26

(65)同一申请的已公布的文献号
申请公布号 CN 103824565 A

(43)申请公布日 2014.05.28

(73)专利权人 曾新
地址 410083 湖南省长沙市中南大学信息可视艺术与设计研究中心
专利权人 徐明 王利斌

(72)发明人 曾新 徐明 王利斌

(74)专利代理机构 深圳市恒申知识产权事务所（普通合伙）44312
代理人 陈健

(51)Int.Cl.
G10L 21/06(2013.01)

(56)对比文件
CN 1607575 A,2005.04.20,
CN 102682752 A,2012.09.19,
CN 102956224 A,2013.03.06,
CN 102664016 A,2012.09.12,
CN 101471074 A,2009.07.01,
CN 101093661 A,2007.12.26,
US 2009119097 A1,2009.05.07,
CN 1607575 A,2005.04.20,
US 5038658 A,1991.08.13,
CN 101916250 A,2010.12.15,
徐明等.“一种高效的基于CHMM的哼唱式旋律检索方法”.《第三届全国数字误乐与艺术暨数字家庭交互应用技术与设计学术研讨会论文集》.2007,

审查员 董小东

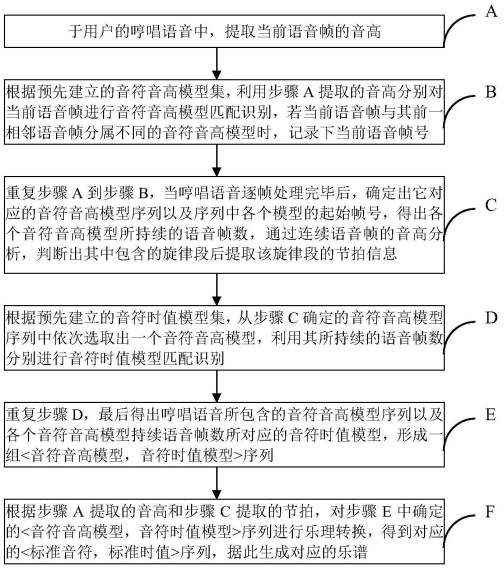
权利要求书6页 说明书8页 附图2页

(54)发明名称

一种基于音符和时值建模的哼唱识谱方法及系统

(57)摘要

本发明适用于计算机应用技术领域，提供了一种基于音符和时值建模的哼唱识谱方法，本发明建立有包括音符音高模型集和音符时值模型集在内的乐理高斯混合模型库，所述乐理高斯混合模型库中的所有模型均事先通过乐理高斯混合模型训练单元进行模型参数训练，并可选用乐理高斯混合模型重估训练单元进行模型参数的重估训练，哼唱识谱时，对采集的用户哼唱语音分别进行音高特征提取、乐理信息解码识别、节拍提取、乐理处理与变换，最后输出成标准乐谱。本发明方法设计的哼唱识谱系统识别率高、稳定性好，还能适应个人的唱歌行为特点，可作为专业人员或音乐爱好者的创作助手和备用工具，具有推广应用价值和产业化前景。



CN 103824565 B

1.一种基于音符和时值建模的哼唱识谱方法,其特征在于,所述方法包括下述步骤:

步骤A,于用户的哼唱语音中,提取当前语音帧的音高;

步骤B,根据预先建立的音符音高模型集,利用步骤A提取的音高分别计算出当前语音帧属于所述音符音高模型集中各个音符音高模型的概率值,根据计算的概率值以及音符音高模型集对当前语音帧进行音符音高模型匹配识别,若当前语音帧与其前一相邻语音帧分属不同的音符音高模型时,记录下当前语音帧号;

步骤C,重复步骤A到步骤B,当哼唱语音依序逐语音帧全部处理完毕后,确定出哼唱语音所对应的音符音高模型序列以及序列中各个音符音高模型的起始语音帧号,计算出所述各个音符音高模型各自所持续的语音帧数,并累积分析语音帧的音高变化情况,判断出其中包含的旋律段后提取该旋律段的节拍信息;

步骤D,根据预先建立的音符时值模型集,从步骤C确定的音符音高模型序列中依次选取出一个音符音高模型,利用其所持续的语音帧数分别计算出选取的音符音高模型属于所述音符时值模型集中各个音符时值模型的概率值,根据计算的概率值以及音符时值模型集对选取的音符音高模型进行音符时值模型匹配识别;

步骤E,重复步骤D,当步骤C中确定的全部音符音高模型序列处理完毕后,得出哼唱语音所包含的各个音符音高模型序列以及各个音符音高模型持续语音帧数所对应的音符时值模型,形成一组<音符音高模型,音符时值模型>序列;

步骤F,根据步骤A提取的音高和步骤C提取的节拍信息,对步骤E确定的哼唱语音的<音符音高模型,音符时值模型>序列进行乐理转换处理,得到对应的<标准音符,标准时值>序列,据此生成对应的乐谱。

2.如权利要求1所述的方法,其特征在于,所述音符音高模型集包含分别为处于低八度、中八度、高八度区段中的各个标准音符以及一个静音所建立的模型,其基于高斯混合模型技术进行建模,采用多个单高斯分布进行混合,通过如下公式对音符音高模型的概率密度输出函数$G_f(x)$进行加权混合计算:

$$G_f(x)=\sum_{j=1}^{M}\alpha_j P_j(x,\mu_j,\Sigma_j),\ \sum_{j=1}^{M}\alpha_j=1$$

其中,M为包含的单高斯分布的个数,α_j为各个单高斯分布的概率密度函数的混合权重,$P_j(x,\mu_j,\Sigma_j)$的定义如下:

$$P(x,\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)|\Sigma|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

其中,T表示矩阵的转置,x为待估算的哼唱语音帧的音高特征列向量,μ为模型期望,Σ为模型方差,μ、Σ均由若干训练样本音符语音帧的音高特征列向量c_j得出,$\mu=\frac{1}{n}\sum_{j=1}^{n}c_j$为均值向量,$\Sigma=\frac{1}{n-1}\sum_{j=1}^{n}(c_j-\mu)(c_j-\mu)^T$是协方差矩阵,n为训练样本的个数;

所述音符时值模型集包含分别为各种标准音符时值基于高斯混合模型技术所建立的模型,采用多个单高斯分布进行混合,通过如下公式对音符时值模型的概率密度输出函数$G_t(x)$进行加权混合计算:

$$G_t(x)=\sum_{j=1}^{M}\alpha_j P_j(x,\mu_j,\Sigma_j),\ \sum_{j=1}^{M}\alpha_j=1$$

其中，M为包含的单一高斯分布的个数，α_j为各个单一高斯分布的概率密度函数的混合权重，$P_j(x,\mu_j,\Sigma_j)$的定义如下：

$$P(x,\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)|\Sigma|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

其中，T表示矩阵的转置，x为待估算的某音符持续哼唱时长所对应的语音帧数，μ为模型期望，Σ为模型方差，μ、Σ均由若干训练样本时值所对应的语音帧数c_j得出，$\mu=\frac{1}{n}\sum_{j=1}^{n}c_j$为均值向量，$\Sigma=\frac{1}{n-1}\sum_{j=1}^{n}(c_j-\mu)(c_j-\mu)^T$是协方差矩阵，n为训练样本的个数。

3.如权利要求1所述的方法，其特征在于，所述音符音高模型集的参数通过如下步骤G1至G3训练得到：

步骤G1，进行音符音高模型高斯混合概率密度输出函数工作参数的初始化，对于每一个音符音高模型，将该音符的国际标准音高作为所述工作参数的初始期望均值；

步骤G2，在步骤G1音符音高模型参数初始化的基础上，利用从哼唱语料中提取出来的该音符的音高作为观察样本值，利用期望最大化算法进行最大似然估计，确定音符音高模型高斯混合概率密度输出函数的各个工作参数；

步骤G3，依次根据步骤G1和G2训练得到的每一个音符音高模型，将哼唱语料中提取出来的所有音高观察样本值划分成两类，一类是属于该音符音高模型的接受域，另一类是不属于该音符音高模型的拒绝域，利用后验概率和似然比分析的方法对所述接受域和拒绝域所包含的观察样本值进行处理以确定该音符音高模型的拒识阈值；

所述音符时值模型集的参数通过如下步骤H1至H3训练得到：

步骤H1，进行音符时值模型高斯混合概率密度输出函数工作参数的初始化，对于每一个音符时值模型，将该音符时值的国际标准时长转化成语音帧数作为所述工作参数的初始期望均值；

步骤H2，在步骤H1音符时值模型参数初始化的基础上，以从哼唱语料中提取出来的该音符的哼唱时长所对应的语音帧数作为观察样本值，利用期望最大化算法进行最大似然估计，确定音符时值模型高斯混合概率密度输出函数的各个工作参数；

步骤H3，依次根据步骤H1和H2训练得到的每一个音符时值模型，将哼唱语料中提取出来的所有时值观察样本值划分成两类，一类是属于该音符时值模型的接受域，另一类是不属于该音符时值模型的拒绝域，利用后验概率和似然比分析的方法对所述接受域和拒绝域所包含的观察样本值进行处理以确定该音符时值模型的拒识阈值。

4.如权利要求1所述的方法，其特征在于，在所述步骤A之前，根据用户的哼唱特征对所述音符音高模型及音符时值模型的高斯混合概率密度输出函数工作参数进行重估，重估步骤如下：

步骤I1，采集用户按照预先设定好的固定哼唱模板逐一进行哼唱的哼唱语音；其中，每一固定哼唱模板个哼唱模板由一组特定的<音符，时值>序列组成；

步骤I2，对步骤I1采集到的哼唱语音逐帧提取音高，根据哼唱模板的乐理知识得到该

用户哼唱各个音符时的个性音高值，作为新的观察样本值，重新利用期望最大化算法进行最大似然估计，分别对音符音高模型集中的各个音符音高模型参数进行重估训练；

步骤I3，对步骤I2逐帧提取到的音高特征进行连续分析，根据哼唱模板的乐理知识得到该用户哼唱各个音符时，相对于标准时值所表现出的个性时长，作为新的观察样本值，重新利用期望最大化算法进行最大似然估计，分别对音符时值模型集中的各个音符时值模型参数进行重估训练；

步骤I4，将通过步骤I2重估训练得到的各个音符音高模型的新参数以及通过步骤I3重估训练得到的各个音符时值模型的新参数，更新到乐理高斯混合模型库，得到反映该用户发音特点的新的乐理高斯混合模型参数。

5.如权利要求1所述的方法，其特征在于，所述步骤B具体包括如下步骤：

步骤B1，根据预先建立的音符音高模型集，对步骤A提取的当前语音帧的音高分别代入所述音符音高模型集中各个音符音高模型的混合概率密度输出函数，计算出所述语音帧属于各个音符音高模型的概率值；

步骤B2，将当前语音帧与所述概率值中最大者所对应的音符音高模型进行匹配，当该最大概率值低于相应音符音高模型的拒识阈值时进行拒识处理；

步骤B3，若匹配结果为当前语音帧与前一语音帧分属不同的音符音高模型时，记录当前语音帧号；

所述步骤D具体包括如下步骤：

步骤D1，根据预先建立的音符时值模型集，逐音符音高模型将其所持续的语音帧数分别代入所述音符时值模型集中各个音符时值模型的概率密度输出函数，计算出对各个音符时值模型的概率值；

步骤D2，将当前音符音高模型与所述概率值中最大者所对应的音符时值模型进行匹配，当该最大概率值低于相应音符时值模型的拒识阈值时进行拒识处理。

6.如权利要求1所述的方法，其特征在于，所述步骤F包括如下步骤：

步骤F1，根据提取的哼唱语音节拍特征，与中速标准歌唱速度下的节拍特征作对比分析，得出哼唱节拍与中速标准节拍之间快慢程度比率，将步骤E中识别出的各音符时值模型均转化成对应的标准时值；

步骤F2，根据步骤C对哼唱语音音高变化情况的分析结果，得出哼唱语音的整体音高特点，对步骤E中识别出的各音符音高模型进行纠正处理，最终将所述各音符音高模型一一转化成对应的标准音符；

步骤F3，根据步骤F1和步骤F2的结果，形成哼唱语音所对应的<音符，时值>序列，按照乐理常识将所述<音符，时值>序列自动转化成五线谱或者简谱。

7.一种基于音符和时值建模的哼唱识谱系统，其特征在于，包括：

哼唱输入采集器，用于采集用户的哼唱语音；

音高提取器，用于从用户的哼唱语音中逐语音帧提取音高；

节拍提取器，用于从音高提取器获取哼唱语音各语音帧的音高，累积分析语音帧的音高变化情况，判断出其中包含的旋律段后提取该旋律段的节拍信息；

乐理信息解码识别器，用于根据预先建立的音符音高模型集，利用提取的音高分别计算出当前语音帧属于所述音符音高模型集中各个音符音高模型的概率值，根据计算的概率

值以及音符音高模型集对当前语音帧进行音符音高模型匹配识别，若当前语音帧与其前一相邻语音帧分属不同的音符音高模型时，记录下当前语音帧号；在按照上述方式依序处理完哼唱语音的所有语音帧后，确定出哼唱语音所对应的音符音高模型序列以及序列中各个音符音高模型的起始语音帧号，计算出所述各个音符音高模型各自所持续的语音帧数，并通过节拍提取器提取哼唱语音包含的节拍信息；根据预先建立的音符时值模型集，从确定的音符音高模型序列中依次选取出一个音符音高模型，利用其所持续的语音帧数分别计算出所述音符音高模型属于所述音符时值模型集中各个音符时值模型的概率值，进行音符时值模型匹配识别；在按照上述方式依序处理完所确定的全部音符音高模型序列后，得出哼唱语音所包含的各个音符音高模型序列以及各个音符音高模型持续语音帧数所对应的音符时值模型，形成一组<音符音高模型，音符时值模型>序列；

乐理处理与变换器，用于根据音高提取器提取的音高和节拍提取器提取的节拍信息，对确定出的哼唱语音的<音符音高模型，音符时值模型>序列进行乐理转换处理，得到对应的<标准音符，标准时值>序列；

标准乐谱生成器，用于根据所述<标准音符，标准时值>序列生成对应的乐谱。

8.如权利要求7所述的系统，其特征在于，所述音符音高模型集包含分别为处于低八度、中八度、高八度区段中的各个标准音符以及一个静音所建立的模型，其基于高斯混合模型技术进行建模，采用多个单高斯分布进行混合，通过如下公式对音符音高模型的概率密度输出函数$G_f(x)$进行加权混合计算：

$$G_f(x)=\sum_{j=1}^{M}\alpha_j P_j(x,\mu_j,\Sigma_j),\ \sum_{j=1}^{M}\alpha_j=1$$

其中，M为包含的单高斯分布的个数，α_j为各个单高斯分布的概率密度函数的混合权重，$P_j(x,\mu_j,\Sigma_j)$的定义如下：

$$P(x,\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)|\Sigma|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

其中，T表示矩阵的转置，x为待估算的哼唱语音帧的音高特征列向量，μ为模型期望，Σ为模型方差，μ、Σ均由若干训练样本音符语音帧的音高特征列向量c_j得出，$\mu=\frac{1}{n}\sum_{j=1}^{n}c_j$为均值向量，$\Sigma=\frac{1}{n-1}\sum_{j=1}^{n}(c_j-\mu)(c_j-\mu)^T$是协方差矩阵，n为训练样本的个数；

所述音符时值模型集包含分别为各种标准音符时值基于高斯混合模型技术所建立的模型采用多个单高斯分布进行混合，通过如下公式对音符时值模型的概率密度输出函数$G_t(x)$进行加权混合计算：

$$G_t(x)=\sum_{j=1}^{M}\alpha_j P_j(x,\mu_j,\Sigma_j),\ \sum_{j=1}^{M}\alpha_j=1$$

其中，M为包含的单一高斯分布的个数，α_j为各个单一高斯分布的概率密度函数的混合权重，$P_j(x,\mu_j,\Sigma_j)$的定义如下：

$$P(x,\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)|\Sigma|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

其中，T表示矩阵的转置，x为待估算的某音符持续哼唱时长所对应的语音帧数，μ为模型期望，Σ为模型方差，μ、Σ均由若干训练样本时值所对应的语音帧数c_j得出，$\mu=\frac{1}{n}\sum_{j=1}^{n}c_j$为均值向量，$\Sigma=\frac{1}{n-1}\sum_{j=1}^{n}(c_j-\mu)(c_j-\mu)^T$是协方差矩阵，n为训练样本的个数。

9.如权利要求7所述的系统，其特征在于，所述系统还包括一个乐理高斯混合模型训练单元，所述乐理高斯混合模型训练单元包括：

音符及时长信息标注器，用于将训练语料库中采集的每一个哼唱样本参照其对应的歌谱标注好其中的音符名称以及该音符时值被哼唱的时长，保存到标注文件中；

音高及时值特征提取器，用于从哼唱语料中，根据标注文件的定义为每个标注好的音符名称提取其对应语音帧的音高，按照音符名称进行分类保存，并根据标注文件的定义为每个标注好的音符时值提取其对应的语音帧数，作为该音符时值的哼唱时长，按照音符时值名称进行分类保存；

先验知识导入器，用于进行音符音高模型及音符时值模型的高斯混合概率密度函数工作参数的初始化，对于每一个音符音高模型，将该音符的国际标准音高作为所述工作参数的初始期望均值，对于每一个音符时值模型，将该音符时值的国际标准时长作为所述工作参数的初始期望均值；

乐理高斯混合模型训练器，用于进行音符音高模型工作参数的训练，对于每一个音符音高模型，在音符音高模型参数初始化的基础上，利用从哼唱语料中提取出来的该音符的音高值作为观察样本值，利用期望最大化算法进行最大似然估计，确定音符音高模型高斯混合概率密度输出函数的各个工作参数，然后依次根据上述方式训练得到的每一个音符音高模型，将哼唱语料中提取出来的所有音高观察样本值划分成两类，一类是属于该音符音高模型的接受域，另一类是不属于该音符音高模型的拒绝域，利用后验概率和似然比分析的方法对所述接受域和拒绝域进行处理以确定该音符音高模型的拒识阈值；还用于进行音符时值模型工作参数的训练，对于每一个音符时值模型，在音符时值模型参数初始化的基础上，利用从哼唱语料中提取出来的该音符的哼唱时长所对应的语音帧数作为观察样本值，利用期望最大化算法进行最大似然估计，确定音符时值模型高斯混合概率密度输出函数的各个工作参数，然后依次对按照上述方式训练得到的每一个音符时值模型，将哼唱语料中提取出来的所有时值观察样本值划分成两类，一类是属于该音符时值模型的接受域，另一类是不属于该音符时值模型的拒绝域，利用后验概率和似然比分析的方法对所述接受域和拒绝域进行处理以确定该音符时值模型的拒识阈值。

10.如权利要求7所述的系统，其特征在于，所述系统还包括一个乐理高斯混合模型重估训练单元，所述乐理高斯混合模型重估训练单元包括：

旋律模板加载器，用于加载预先设定好的若干旋律模板，以便用户按照所述旋律模板中约定的音符及时值序列进行哼唱；

个性哼唱采集器，用于采集用户按照上述旋律模板约定的内容进行哼唱的语音；

音高及时值提取器，用于从通过个性哼唱采集器采集的哼唱语音中，根据旋律模板的定义为每个音符名称提取其对应语音帧的音高，并根据旋律模板的定义为每个音符时值提取其对应的语音帧数；

乐理高斯混合模型重估训练器，用于选取若干旋律片段作为固定哼唱模板，每一个哼唱模板由一组特定的<音符，时值>序列组成，用户按照哼唱模板逐一进行哼唱，采集哼唱语音；然后对采集到的哼唱语音逐帧提取音高，根据哼唱模板的乐理知识得到该用户哼唱各个音符时的个性音高值，作为新的观察样本值，重新利用期望最大化算法进行最大似然估计，分别对音符音高模型集中的各个音符音高模型参数进行重估训练；再对逐帧提取到的音高特征进行连续分析，根据哼唱模板的乐理知识得到该用户哼唱各个音符时，相对于标准时值所表现出的个性时长，作为新的观察样本值，重新利用期望最大化算法进行最大似然估计，分别对音符时值模型集中的各个音符时值模型参数进行重估训练；最后将通过重估训练得到的各个音符音高模型的新参数以及通过重估训练得到的各个音符时值模型的新参数，更新到乐理高斯混合模型库，得到反映该用户发音特点的新的乐理高斯混合模型参数。

11.如权利要求7所述的系统，其特征在于，所述乐理处理与变换器用于根据提取的哼唱语音节拍特征，与中速标准歌唱速度下的节拍特征作对比分析，得出哼唱节拍与中速标准节拍之间快慢程度比率，将识别出的各音符时值模型均转化成对应的标准时值；然后根据对哼唱语音音高变化情况的分析结果，得出哼唱语音的整体音高特点，对识别出的各音符音高模型进行纠正处理，最终将所述各音符音高模型一一转化成对应的标准音符；最后根据转化的标准时值与标准音符，形成哼唱语音所对应的<音符，时值>序列，以使所述标准乐谱生成器按照乐理常识将所述<音符，时值>序列自动转化成五线谱或者简谱。

一种基于音符和时值建模的哼唱识谱方法及系统

技术领域

[0001] 本发明属于计算机应用技术领域，尤其涉及一种基于音符和时值建模的哼唱识谱方法及系统，具体对对音乐的音符和时值进行建模，通过模型训练和解码识别两个关键过程实现哼唱识谱的功能。

背景技术

[0002] 随着计算机与网络技术的发展，人们越来越多地利用数字技术提供音乐服务，如卡啦OK、音乐检索、歌唱评价、哼唱搜歌、音乐合成等，既丰富了人们的娱乐生活，也推动了音乐创作活动的发展。对于资深的音乐爱好者，常常会即兴哼唱出一些旋律，希望能够找到专业软件把这些旋律转化成歌谱保存起来，以用于今后的音乐原创活动。而对于专业的音乐创作人，在生活中随时会突发灵感，唱出自己新构思的旋律，这时也迫切需要具有哼唱识谱功能的专业软件把歌唱语音自动转化成歌谱，以便后续的加工润色工作。

[0003] 与音乐合成、哼唱识别技术相比，哼唱识谱方面的研究工作开展得较少。现有的技术主要是对哼唱录音数据在时域上进行自相关等技术提取基频信号，获取音高数值，然后直接利用单一的音高参数去进行音符切分，通过与标准的音符音高及标准时值进行比对，得出音符及时值序列作为识别结果。

[0004] 然而，上述哼唱识谱方法在实际应用中存在不足，表现在准确性不高。由于噪音的影响，自相关提取基频的技术抗干扰能力差，往往出现倍频或半频的错误，造成音符识别的不准确。歌唱或哼唱过程中，协同发音现象普遍存在，造成音符切分上的困难，多切和遗漏现象严重，影响时值判别的准确性。更为重要的是，对音乐爱好者来说，每个人的发音系统和发音习惯不尽相同，歌唱时音高和时值的把握与国际标准音高和时值有差异。即便是专业人士，也存在这种差异。况且，相邻音符之间的音高还存在一定区域的重叠，也给音符判别过程带来难度。直接以个性化的音高及时长数值去与标准音高和时值进行匹配，软件系统的适应性强很差。

[0005] 总之，现有的哼唱识谱技术存在不足，推广应用存在困难，需要采用新的思路研究精度高、稳定性好、适应性强的方法。

发明内容

[0006] 鉴于上述原因，本发明所要解决的技术问题在于提供一种基于音符和时值建模的哼唱识谱方法，该方法识别率高、稳定性好、适应性广，能够针对多数人的唱歌行为特点保持高识别率和运算性能，具有推广应用价值和产业化前景。

[0007] 本发明是这样实现的，一种基于音符和时值建模的哼唱识谱方法，包括下述步骤：

[0008] 步骤A，于用户的哼唱语音中，提取当前语音帧的音高；

[0009] 步骤B，根据预先建立的音符音高模型集，利用步骤A提取的音高分别计算出当前语音帧属于所述音符音高模型集中各个音符音高模型的概率值，根据计算的概率值以及音符音高模型集对当前语音帧进行音符音高模型匹配识别，若当前语音帧与其前一相邻语音

帧分属不同的音符音高模型时，记录下当前语音帧号；

[0010] 步骤C，重复步骤A到步骤B，当哼唱语音依序逐语音帧全部处理完毕后，确定出哼唱语音所对应的音符音高模型序列以及序列中各个音符音高模型的起始语音帧号，计算出所述各个音符音高模型各自所持续的语音帧数，并累积分析语音帧的音高变化情况，判断出其中包含的旋律段后提取该旋律段的节拍信息；

[0011] 步骤D，根据预先建立的音符时值模型集，从步骤C确定的音符音高模型序列中依次选取出一个音符音高模型，利用其所持续的语音帧数分别计算出所选取的音符音高模型属于所述音符时值模型集中各个音符时值模型的概率值，根据计算的概率值以及音符时值模型集对选取的音符音高模型进行音符时值模型匹配识别；

[0012] 步骤E，重复步骤D，当步骤C中确定的全部音符音高模型序列处理完毕后，得出哼唱语音所包含的各个音符音高模型序列以及各个音符音高模型持续语音帧数所对应的音符时值模型，形成一组<音符音高模型，音符时值模型>序列；

[0013] 步骤F，根据步骤A提取的音高和步骤C提取的节拍信息，对步骤E确定的哼唱语音的<音符音高模型，音符时值模型>序列进行乐理转换处理，得到对应的<标准音符，标准时值>序列，据此生成对应的乐谱。

[0014] 本发明还提供了一种基于音符和时值建模的哼唱识谱系统，包括：

[0015] 哼唱输入采集器，用于采集用户的哼唱语音；

[0016] 音高提取器，用于从用户的哼唱语音中逐语音帧提取音高；

[0017] 节拍提取器，用于从音高提取器获取哼唱语音各语音帧的音高，累积分析语音帧的音高变化情况，判断出其中包含的旋律段后提取该旋律段的节拍信息；

[0018] 乐理信息解码识别器，用于根据预先建立的音符音高模型集，利用提取的音高分别计算出当前语音帧属于所述音符音高模型集中各个音符音高模型的概率值，根据计算的概率值以及音符音高模型集对当前语音帧进行音符音高模型匹配识别，若当前语音帧与其前一相邻语音帧分属不同的音符音高模型时，记录下当前语音帧号；在按照上述方式依序处理完哼唱语音的所有语音帧后，确定出哼唱语音所对应的音符音高模型序列以及序列中各个音符音高模型的起始语音帧号，计算出所述各个音符音高模型各自所持续的语音帧数，并通过节拍提取器提取哼唱语音包含的节拍信息；根据预先建立的音符时值模型集，从确定的音符音高模型序列中依次选取出一个音符音高模型，利用其所持续的语音帧数分别计算出所述音符音高模型属于所述音符时值模型集中各个音符时值模型的概率值，进行音符时值模型匹配识别；在按照上述方式依序处理完所确定的全部音符音高模型序列后，得出哼唱语音所包含的各个音符音高模型序列以及各个音符音高模型持续语音帧数所对应的音符时值模型，形成一组<音符音高模型，音符时值模型>序列；

[0019] 乐理处理与变换器，用于根据音高提取器提取的音高和节拍提取器提取的节拍信息，对确定出的哼唱语音的<音符音高模型，音符时值模型>序列进行乐理转换处理，得到对应的<标准音符，标准时值>序列；

[0020] 标准乐谱生成器，用于根据所述<标准音符，标准时值>序列生成对应的乐谱。

[0021] 本发明与现有技术相比，通过抗噪音的音高特征提取、音符音高模型集及音符时值模型集参数训练、乐理信息解码识别，具有较高的识别率和计算速度，适应性强。实验结果表明，本发明方法设计的哼唱识谱系统抗噪音干扰能力强，能够满足不同歌唱水平人员

的使用需求，能够针对多数人的唱歌行为特点保持高识别率，具有推广应用价值和产业化前景。

附图说明

[0022] 图1是本发明提供的基于音符和时值建模的哼唱识谱方法的实现流程图；

[0023] 图2是本发明提供的基于音符和时值建模的哼唱识谱系统的结构原理图。具体实施方式

[0024] 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

[0025] 本发明采用统计模型对音乐中的音符和时值进行声学建模，采集有代表性的哼唱语料，以高精度抗干扰的音高提取方法计算出音符的观察样本数据，通过迭代的方法训练出稳定的模型参数。提供模型参数重估的方法，允许将个人的哼唱语音作为样本参与模型参数的再训练，使得模型参数反映出个人的发音特点和习惯，使哼唱识谱系统具有很好的适应性。

[0026] 图1示出了本发明提供的基于音符和时值建模的哼唱识谱方法的实现流程，详述如下：

[0027] 在步骤A中，于用户的哼唱语音中，提取当前语音帧的音高。

[0028] 本发明中，采用一种高精度、抗干扰的方法提取哼唱语音帧音高，具体步骤是：首先针对哼唱语音帧在数字信号经典功率谱估计方法的基础上进行自相关运算，快速提取若干基音周期候选值。然后针对这些候选值实施多重后处理方法，具体为：先利用通过预设的峰值阈参数对候选值进行初步的筛选，接着利用通过预设的一次均值参数将语音分为不同的音高段，再使用通过预设的二次均值参数为每个音高段确定合适的频率范围，最后提取出基音周期作为该哼唱语音帧的音高。上述峰值阈参数、一次均值参数、二次均值参数均可以通过实验预先确定。

[0029] 在步骤B中，根据预先建立的音符音高模型集，利用步骤A提取的音高分别计算出当前语音帧属于所述音符音高模型集中各个音符音高模型的概率值，根据计算的概率值以及音符音高模型集对当前语音帧进行音符音高模型匹配识别，若当前语音帧与其前一相邻语音帧分属不同的音符音高模型时，记录下当前语音帧号。

[0030] 本发明中，根据歌谱均由处于不同八度区间的CDEFGAB七个基本音符组成这一乐理常识，并考虑到的大众歌曲的特点和人们的发音规律，主要对处于低八度、中八度、高八度这一区段的各个音符进行建模。实施例中，对国际标准音符中的$CDEFGABC^1D^1E^1F^1G^1A^1B^1C^2D^2E^2F^2G^2A^2B^2$，也就是简谱中$\dot{1}\dot{2}\dot{3}\dot{4}\dot{5}\dot{6}\dot{7}1\dot{2}\dot{3}\dot{4}\dot{5}\dot{6}\dot{7}1\dot{2}\dot{3}\dot{4}\dot{5}\dot{6}\dot{7}$这21个音符进行建模，还增加一个静音模型。针对这些音符模型，基于高斯混合模型技术进行建模，即采用多个单高斯分布进行混合，通过如下公式对音符音高模型的概率密度输出函数Gf(x)进行加权混合计算：

[0031] $$G_f(x)=\sum_{j=1}^{M}\alpha_j \mathrm{P}_j(x,\mu_j,\Sigma_j),\ \sum_{j=1}^{M}\alpha_j=1 \qquad (1)$$

[0032] 其中，M为包含的单高斯分布的个数，α_j为各个单高斯分布的概率密度函数的混合权重，μ为均值向量，Σ是协方差矩阵，$P_j(x,\mu_j,\Sigma_j)$是单高斯分布的概率密度函数，其计算方法如下：

[0033] $$P(x,\mu,\Sigma)=\frac{1}{\sqrt{(2\pi)|\Sigma|}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] \qquad (2)$$

[0034] 其中，T表示矩阵的转置，x为待估算的哼唱语音帧的音高特征列向量，μ为模型期望，Σ为模型方差，μ、Σ均由若干训练样本音符语音帧的音高特征列向量c_j得出，$\mu=\frac{1}{n}\sum_{j=1}^{n}c_j$为均值向量，$\Sigma=\frac{1}{n-1}\sum_{j=1}^{n}(c_j-\mu)(c_j-\mu)^T$是协方差矩阵，n为训练样本的个数。

[0035] 训练音符音高模型参数的具体过程是：

[0036] 首先，进行音符音高模型高斯混合概率密度输出函数工作参数的初始化，对于每一个音符音高模型，将该音符的国际标准音高看作先验知识，作为模型工作参数的初始期望均值，以便加快训练速度，稳定模型参数。

[0037] 然后，进行音符音高模型高斯混合概率密度输出函数工作参数的训练，对于每一个音符音高模型，在音符音高模型参数初始化的基础上，利用从哼唱语料中提取出来的该音符的音高值作为观察样本值，利用期望最大化算法进行最大似然估计，确定音符音高模型高斯混合概率密度输出函数的各个工作参数，即确定模型的期望、方差和混合权重等参数。核心过程就是通过迭代计算，不断更新权值α_j、均值μ_j和方差矩阵Σ_j，满足$\max\sum_{i=1}^{N}\log(\sum_{j=1}^{M}\alpha_j P_j(x,\mu_j,\Sigma_j))$的要求，使期望值最大。

[0038] 最后，进行音符音高模型的拒识阈值参数的训练。依次对按照以上方式训练得到的每一个音符音高模型，将哼唱语料中提取出来的所有音高观察样本值划分成两类，一类是属于该音符音高模型的接受域，另一类是不属于该音符音高模型的拒绝域，利用后验概率和似然比分析的方法确定该音符音高模型的拒识阈值。

[0039] 在事先完成各音符音高模型参数训练的基础上，才可以实施步骤B中的哼唱语音帧匹配识别过程，具体方法是：首先，根据预先建立的音符音高模型集，对步骤A提取的当前语音帧的音高分别代入所述音符音高模型集中各个音符音高模型的混合概率密度输出函数计算出所述语音帧属于各个音符音高模型的概率值；然后，将当前语音帧与所述概率值中最大者所对应的音符音高模型进行匹配，当该最大概率值低于相应音符音高模型的拒识阈值时进行拒识处理；最后，若匹配结果为当前语音帧与前一语音帧分属不同的音符音高模型时，记录当前语音帧号。

[0040] 在步骤C中，重复步骤A到步骤B，当哼唱语音依序逐语音帧全部处理完毕后，确定出哼唱语音所对应的音符音高模型序列以及序列中各个音符音高模型的起始语音帧号，计算出所述各个音符音高模型各自所持续的语音帧数，并累积分析语音帧的音高变化情况，判断出其中包含的旋律段后提取该旋律段的节拍信息。

[0041] 本发明中，通过跟踪分析哼唱语音音高的连续变化情况，判断出旋律段和非旋律段，针对其中的旋律段采用自相关相位-熵序列分析的方法提取其哼唱的节拍速度，供后续

处理过程使用。

[0042] 在步骤D中，根据预先建立的音符时值模型集，从步骤C确定的音符音高模型序列中依次选取出一个音符音高模型，利用其所持续的语音帧数分别计算出所选取的音符音高模型属于所述音符时值模型集中各个音符时值模型的概率值，根据计算的概率值以及音符时值模型集对选取的音符音高模型进行音符时值模型匹配识别。

[0043] 本发明中，根据音符发音时长方面的乐理常识以及人们的歌唱发音规律，主要为标准全音符、二分音符、四分音符、八分音符、十六分音符、三十二分音符、六十四分音符等这些音符歌唱时的标准时值进行建模。实施例中，基于高斯混合模型技术对音符时值进行建模，采用多个单高斯分布进行混合，通过如下公式对音符时值模型的概率密度输出函数Gt(x)进行加权混合计算：

[0044] $$G_t(x)=\sum_{j=1}^{M}\alpha_j \mathrm{P}_j(x,\mu_j,\Sigma_j),\ \sum_{j=1}^{M}\alpha_j=1 \qquad (3)$$

[0045] 其中，M为包含的单一高斯分布的个数，α_j为各个单一高斯分布的概率密度函数的混合权重，μ为均值向量，Σ是协方差矩阵，$\mathrm{P}_j(x,\mu_j,\Sigma_j)$是单高斯分布的概率密度函数，其定义和计算方法见公式(2)。

[0046] 训练音符时值模型参数的具体过程是：

[0047] 首先，进行音符时值模型高斯混合概率密度输出函数工作参数的初始化，对于每一个音符时值模型，将该音符时值的国际标准时长看作先验知识，作为模型工作参数的初始期望均值，以便加快训练速度，稳定模型参数。

[0048] 然后，进行音符时值模型高斯混合概率密度输出函数工作参数的训练，对于每一个音符时值模型，在音符时值模型参数初始化的基础上，利用从哼唱语料中提取出来的该音符的哼唱时长所对应的语音帧数作为观察样本值，利用期望最大化算法进行最大似然估计，确定音符时值模型高斯混合概率密度输出函数的各个工作参数，即确定模型的期望、方差和混合权重等参数。核心过程就是通过迭代计算，不断更新权值α_j、均值μ_j和方差矩阵Σ_j，满足$\max\sum_{i=1}^{N}\log(\sum_{j=1}^{M}\alpha_j \mathrm{P}_j(x,\mu_j,\Sigma_j))$的要求，使期望值最大。

[0049] 最后，进行音符时值模型的拒识阈值参数的训练。依次对按照以上方式训练得到的每一个音符时值模型，将哼唱语料中提取出来的所有时值观察样本值划分成两类，一类是属于该音符时值模型的接受域，另一类是不属于该音符时值模型的拒绝域，利用后验概率和似然比分析的方法确定该音符时值模型的拒识阈值。

[0050] 进一步地，为使哼唱识谱系统能够适应每个用户的个性发音特点和发音习惯，即当用户歌唱时音符的音高、音符的时值与国际标准音高及国际标准时值存在差异时，识谱系统仍具有较为稳定的识别能力，本发明提供根据用户的哼唱特征对音符音高模型及音符时值模型的高斯混合概率密度输出函数工作参数进行重估训练的方法。在所述步骤A之前，用户可以选择利用自己的发音样本对音符音高模型集及音符时值模型集里的模型参数进行重估再训练，从而得到反映该用户自己发音特点的新的乐理高斯混合模型参数。重估的具体步骤如下：

[0051] 首先，设定若干旋律片段作为固定哼唱模板，每一个哼唱模板由一组特定的<音符，时值>序列组成，用户按照哼唱模板逐一进行哼唱，采集哼唱语音；然后，对以上步骤中

采集到的哼唱语音逐帧提取音高，根据哼唱模板的乐理知识得到该用户哼唱各个音符时的个性音高值，作为新的观察样本值，重新利用期望最大化算法进行最大似然估计，分别对音符音高模型集中的各个音符音高模型参数进行重估训练。

[0052] 同时，对以上步骤中逐帧提取到的音高特征进行连续分析，根据哼唱模板的乐理知识得到该用户哼唱各个音符时，相对于标准时值所表现出的个性时长，作为新的观察样本值，重新利用期望最大化算法进行最大似然估计，分别对音符时值模型集中的各个音符时值模型参数进行重估训练。

[0053] 在事先完成各音符时值模型参数训练的基础上，才可以实施步骤D中为哼唱中已经匹配出来的音符模型序列片段进行时值匹配识别过程，具体方法是：首先，根据预先建立的音符时值模型集，利用步骤C中得出的音符模型序列及其它们各自所持续的语音帧数，逐音符音高模型将其所持续的语音帧数分别代入所述音符时值模型集中各个音符时值模型的概率密度输出函数计算出对各个音符时值模型的概率值；然后，将当前语音帧与所述概率值中最大者所对应的音符时值模型进行匹配，当该最大概率值低于相应音符时值模型的拒识阈值时进行拒识处理。

[0054] 步骤E的具体处理过程为：重复步骤D，当步骤C中确定的全部音符音高模型序列处理完毕后，得出哼唱语音所包含的各个音符音高模型序列以及各个音符音高模型持续语音帧数所对应的音符时值模型，形成一组<音符音高模型，音符时值模型>序列。

[0055] 在步骤F中，根据步骤A提取的音高和步骤C提取的节拍信息，对步骤E确定的哼唱语音的<音符音高模型，音符时值模型>序列进行乐理转换处理，得到对应的<标准音符，标准时值>序列，据此生成对应的乐谱。

[0056] 本发明中，针对已经识别出来的哼唱语音所包含的<音符音高模型，音符时值模型>序列，进行乐理转换处理的具体过程为：

[0057] 根据提取的哼唱语音节拍特征，与中速标准歌唱速度下的节拍特征作对比分析，得出哼唱节拍与中速标准节拍之间快慢程度比率，将步骤E中识别出的各音符时值模型均转化成对应的标准时值；根据步骤C对哼唱语音音高变化情况的分析结果，得出哼唱语音的整体音高特点，对步骤E中识别出的各音符音高模型进行纠正处理，最终将所述各音符音高模型一一转化成对应的标准音符；根据以上两步出来的结果，形成哼唱语音所对应的<音符，时值>序列，按照乐理常识以及哼唱语音中提取的与音阶相关的节拍音乐信息将所述<音符，时值>序列自动转化成五线谱或者简谱。生成的五线谱或者简谱可以在计算机屏幕展现出来，并可保存为外部文件。

[0058] 本领域普通技术人员可以理解实现上述各实施例提供的方法中的全部或部分步骤可以通过程序来指令相关的硬件来完成，所述的程序可以存储于一计算机可读取存储介质中，该存储介质可以为ROM/RAM、磁盘、光盘等。

[0059] 图2示出了本发明提供的基于音符和时值建模的哼唱识谱系统的结构原理，为了便于描述，仅示出了与本发明相关的部分。

[0060] 参照图2，该系统至少包括哼唱输入采集器1、节拍提取器2、音高特征提取器3、乐理信息解码识别器4、乐理处理与变换器5、标准乐谱生成器6。其中，哼唱输入采集器1用于采集用户的哼唱语音，音高特征提取器3从用户的哼唱语音中逐语音帧提取音高，节拍提取器2用于音高提取器获3取哼唱语音各语音帧的音高，累积分析语音帧的音高变化情况，判

断出其中包含的旋律段后提取该旋律段的节拍信息。

[0061] 然后，乐理信息解码识别器4根据预先建立的音符音高模型集，利用提取的音高分别计算出当前语音帧属于所述音符音高模型集中各个音符音高模型的概率值，根据计算的概率值以及音符音高模型集对当前语音帧进行音符音高模型匹配识别，若当前语音帧与其前一相邻语音帧分属不同的音符音高模型时，记录下当前语音帧号；在按照上述方式依序处理完哼唱语音的所有语音帧后，确定出哼唱语音所对应的音符音高模型序列以及序列中各个音符音高模型的起始语音帧号，计算出所述各个音符音高模型各自所持续的语音帧数，并通过节拍提取器3提取哼唱语音包含的节拍信息；根据预先建立的音符时值模型集，从确定的音符音高模型序列中依次选取出一个音符音高模型，利用其所持续的语音帧数分别计算出所述音符音高模型属于所述音符时值模型集中各个音符时值模型的概率值，进行音符时值模型匹配识别；在按照上述方式依序处理完所确定的全部音符音高模型序列后，得出哼唱语音所包含的各个音符音高模型序列以及各个音符音高模型持续语音帧数所对应的音符时值模型，形成一组<音符音高模型，音符时值模型>序列。

[0062] 乐理处理与变换器5用于根据音高提取器2提取的音高和节拍提取器3提取的节拍信息，对确定出的哼唱语音的<音符音高模型，音符时值模型>序列进行乐理转换处理，得到对应的<标准音符，标准时值>序列，最后，标准乐谱生成器6根据乐理处理与变换器5处理后得到的哼唱语音所对应的<标准音符，标准时值>序列生成对应的乐谱。

[0063] 本发明中，音符音高模型集和音符时值模型集均包含在乐理高斯混合模型参数库7中。音符音高模型和音符时值模型均基于高斯混合模型技术进行建模，采用多个单高斯分布进行混合，每个单高斯分布的概率密度函数按照公式(2)进行定义和计算，音符音高模型的概率密度输出函数按照公式(1)进行定义以及进行加权混合计算，音符时值模型的概率密度输出函数按照公式(3)进行定义以及进行加权混合计算。

[0064] 与上述任一实施例相结合，本系统还包括一乐理高斯混合模型训练单元8，用于进行音符音高模型工作参数的训练，对于每一个音符音高模型，在音符音高模型参数初始化的基础上，利用从哼唱语料中提取出来的该音符的音高值作为观察样本值，利用期望最大化算法进行最大似然估计，确定音符音高模型高斯混合概率密度输出函数的各个工作参数，然后依次对按照上述方式训练得到的每一个音符音高模型，将哼唱语料中提取出来的所有音高观察样本值划分成两类，一类是属于该音符音高模型的接受域，另一类是不属于该音符音高模型的拒绝域，利用后验概率和似然比分析的方法确定该音符音高模型的拒识阈值；还用于进行音符时值模型工作参数的训练，对于每一个音符时值模型，在音符时值模型参数初始化的基础上，利用从哼唱语料中提取出来的该音符的哼唱时长所对应的语音帧数作为观察样本值，利用期望最大化算法进行最大似然估计，确定音符时值模型高斯混合概率密度输出函数的各个工作参数，然后依次对按照上述方式训练得到的每一个音符时值模型，将哼唱语料中提取出来的所有时值观察样本值划分成两类，一类是属于该音符时值模型的接受域，另一类是不属于该音符时值模型的拒绝域，利用后验概率和似然比分析的方法确定该音符时值模型的拒识阈值。

[0065] 与上述任一实施例相结合，本系统还包括乐理高斯混合模型重估训练单元9，用于采集某哼唱人按照固定哼唱模板的歌谱的个性哼唱信息，进行音高、时值特征的提取，将提取的特征作为新的观察样本值分别对音符音高模型集、音符时值模型集中的各个模型参数

进行再训练，得到反映该哼唱人发音特点的新的乐理高斯混合模型参数。具体方法是：首先，选取若干旋律片段作为固定哼唱模板，每一个哼唱模板由一组特定的<音符，时值>序列组成，用户按照哼唱模板逐一进行哼唱，采集哼唱语音；然后对采集到的哼唱语音逐帧提取音高，根据哼唱模板的乐理知识得到该用户哼唱各个音符时的个性音高值，作为新的观察样本值，重新利用期望最大化算法进行最大似然估计，分别对音符音高模型集中的各个音符音高模型参数进行重估训练；再对逐帧提取到的音高特征进行连续分析，根据哼唱模板的乐理知识得到该用户哼唱各个音符时，相对于标准时值所表现出的个性时长，作为新的观察样本值，重新利用期望最大化算法进行最大似然估计，分别对音符时值模型集中的各个音符时值模型参数进行重估训练；最后将通过重估训练得到的各个音符音高模型的新参数以及通过重估训练得到的各个音符时值模型的新参数，更新到乐理高斯混合模型库，得到反映该用户发音特点的新的乐理高斯混合模型参数。

[0066] 与上述任一实施例相结合，乐理信息解码识别器4根据音符音高模型集，逐帧计算哼唱语音对各音符音高高斯混合模型的匹配度，对匹配度低的语音帧进行拒识，解码出音符模型序列，同时记录下音符发生变化的语音帧号，确定出各个音符模型的起始语音帧号；然后依次取出音符模型序列中每一个音符模型持续的语音帧数去计算对各个音符时值模型的匹配度，取最优的结果作为该音符的时值。最终得出哼唱语音所包含的各个音符模型序列以及各个音符模型持续时长所对应的时值模型，形成一组<音符，时值>序列。

[0067] 与上述任一实施例相结合，乐理处理与变换器5将根据音高差特征识别出的音符与对应的音高绝对值做比较，进行八度处理，并根据旋律的音高变化情况，进行节拍分析，确定可能的节拍信息，得到最终的音符及其时值序列。然后，标准乐谱生成器6根据所述最终的音符及其时值序列以及与音阶相关的节拍音乐信息。

[0068] 综上所述，本发明提出的哼唱识谱技术可以作为专业音乐人员的音乐创作助手，也可以作为业余音乐爱好者的备用工具，促进更大范围的音乐原创活动，也可以设计成声乐教学软件应用在艺术学院、社会培训机构的教学培训活动中，还可以设计成数字娱乐软件应用在唱歌练歌等社会娱乐活动中，弥补自动记谱软件市场的空白，解决目前音乐创作过程的许多不便之处，具有独特的市场前景。

[0069] 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作的任何修改、等同替换和改进等，均应包含在本发明的保护范围之内。

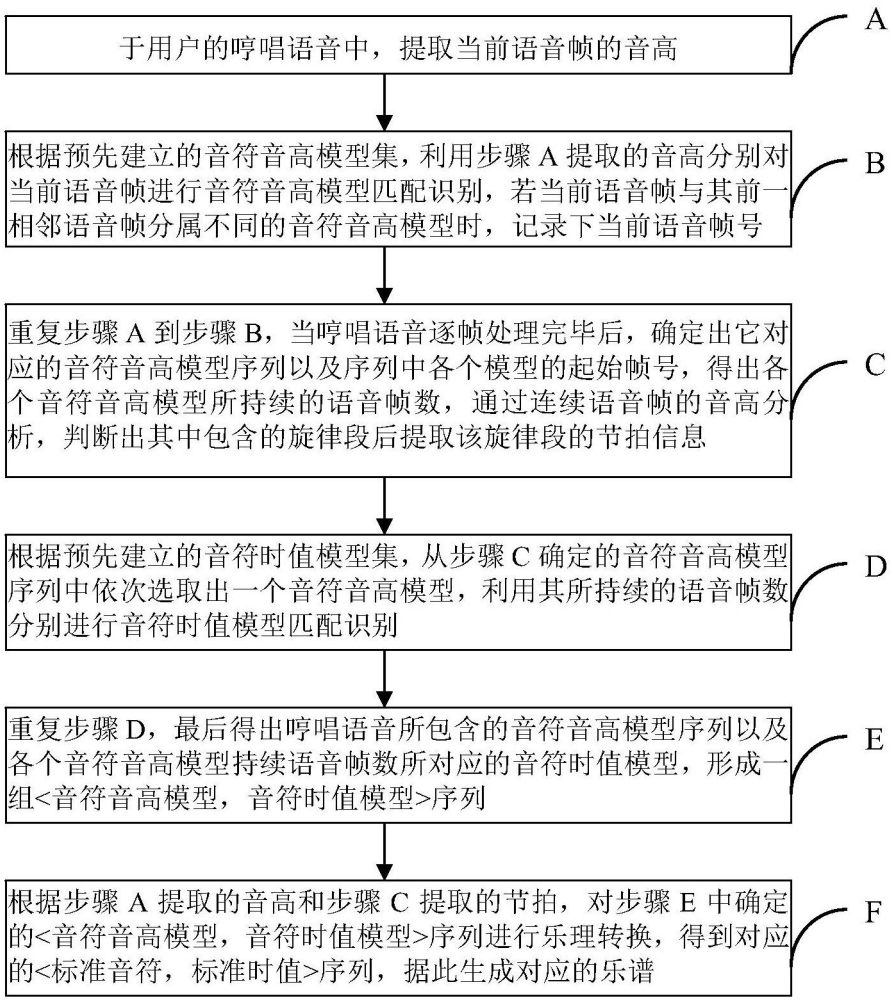

于用户的哼唱语音中，提取当前语音帧的音高
A
根据预先建立的音符音高模型集，利用步骤 A 提取的音高分别对当前语音帧进行音符音高模型匹配识别，若当前语音帧与其前一相邻语音帧分属不同的音符音高模型时，记录下当前语音帧号
B
重复步骤 A 到步骤 B，当哼唱语音逐帧处理完毕后，确定出它对应的音符音高模型序列以及序列中各个模型的起始帧号，得出各个音符音高模型所持续的语音帧数，通过连续语音帧的音高分析，判断出其中包含的旋律段后提取该旋律段的节拍信息
C
根据预先建立的音符时值模型集，从步骤 C 确定的音符音高模型序列中依次选取出一个音符音高模型，利用其所持续的语音帧数分别进行音符时值模型匹配识别
D
重复步骤 D，最后得出哼唱语音所包含的音符音高模型序列以及各个音符音高模型持续语音帧数所对应的音符时值模型，形成一组<音符音高模型，音符时值模型>序列
E
根据步骤 A 提取的音高和步骤 C 提取的节拍，对步骤 E 中确定的<音符音高模型，音符时值模型>序列进行乐理转换，得到对应的<标准音符，标准时值>序列，据此生成对应的乐谱
F

图1

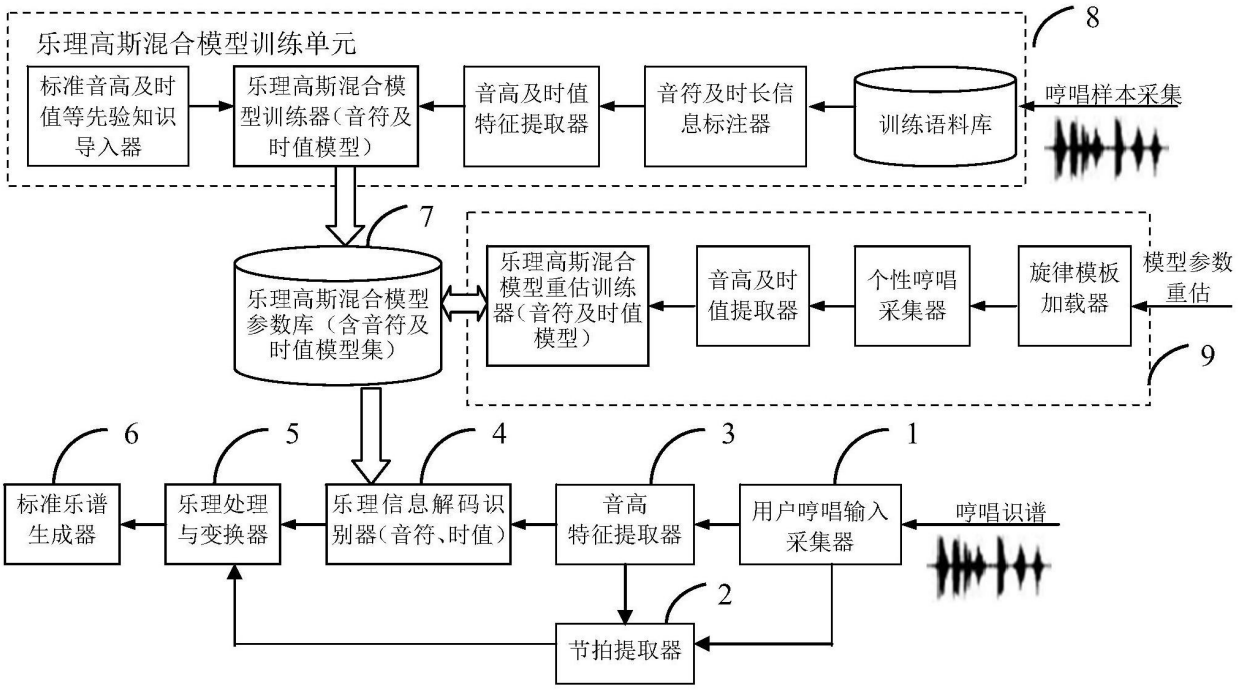

乐理高斯混合模型训练单元
标准音高及时值等先验知识导入器
乐理高斯混合模型训练器(音符及时值模型)
音高及时值特征提取器
音符及时长信息标注器
训练语料库
哼唱样本采集
8
7
乐理高斯混合模型参数库(含音符及时值模型集)
乐理高斯混合模型重估训练器(音符及时值模型)
音高及时值提取器
个性哼唱采集器
旋律模板加载器
模型参数重估
9
6
5
4
3
1
标准乐谱生成器
乐理处理与变换器
乐理信息解码识别器(音符、时值)
音高特征提取器
用户哼唱输入采集器
哼唱识谱
2
节拍提取器

图2