



(12) 发明专利申请

(10) 申请公布号 CN 116057627 A

(43) 申请公布日 2023. 05. 02

(21) 申请号 202180062450.1

(22) 申请日 2021.07.13

(30) 优先权数据

20020321.4 2020.07.13 EP

(85) PCT国际申请进入国家阶段日

2023.03.10

(86) PCT国际申请的申请数据

PCT/EP2021/025257 2021.07.13

(87) PCT国际申请的公布数据

W02022/012777 EN 2022.01.20

(71) 申请人 祖恩惴姆股份有限公司

地址 瑞士巴塞尔

(72) 发明人 P·F·英格里托 A·L·普拉达斯

(74) 专利代理机构 北京汇知杰知识产权代理有限公司 11587

专利代理师 李洁 魏文浩

(51) Int.Cl.

G10L 25/63 (2006.01)

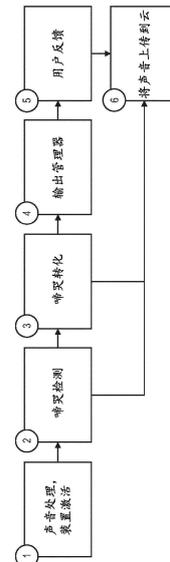
权利要求书3页 说明书46页 附图15页

(54) 发明名称

提供用于自动婴幼儿啼哭评估的数据的计算机实施的方法

(57) 摘要

提议了一种提供用于自动婴幼儿啼哭评估的数据的计算机实施的方法,所述方法包括以下步骤:声学监测婴幼儿并且提供对应的声音数据流,在所述声音数据流中检测啼哭,响应于检测到啼哭从声音数据中选择啼哭相关的数据,确定用于个性化啼哭评估的个人婴幼儿数据,根据个人婴幼儿数据准备用于评估的评估级,以及将啼哭相关的数据馈送到根据个人婴幼儿数据准备的所述啼哭评估级。此外,还提议了一种自动婴幼儿啼哭评估装置。



1. 一种提供用于自动个性化婴幼儿啼哭评估的数据的计算机实施的方法，包括以下步骤：
确定用于个性化啼哭评估的个人婴幼儿数据并且准备与其对应的信息；
声学监测具有背景噪声的环境中的婴幼儿并且提供对应的声音数据样本流，至少考虑到声音的时间和/或谱模式来检测声学监测声音数据样本流中的婴幼儿啼哭，
优选地，在具有考虑是否已经观察到超过阈值的声级的前一检测步骤的多步骤/多级啼哭识别中，
为了进一步评估，选择检测到的啼哭相关的部分，
以及
提供
选择的检测到的啼哭相关的部分
连同与所述个人婴幼儿数据对应的信息给评估级用于进一步评估，其中
根据与所述个人婴幼儿数据对应的所述信息以个性化方式评估选择的部分，使得能够执行对所述啼哭相关的部分的评估
包括
通过与已知的对应于不同类别的啼哭原因的模式进行比较，以产生每个啼哭相关的部分属于所述不同类别中的一个相应的类别的多个概率的方式，对选择的啼哭相关的部分的相继的评价
通过对啼哭相关的部分的相继的评价来建立这样的多个概率的序列，
以及分析所述概率的序列
其中所述评价和进一步评估级中的至少一个根据与所述个人婴幼儿数据对应的所述信息被个性化。
2. 一种提供用于自动个性化婴幼儿啼哭评估的数据并且以个性化方式评估所述数据的计算机实施的方法，包括以下步骤：
确定用于个性化啼哭评估的个人婴幼儿数据并且准备与其对应的信息；
声学监测具有背景噪声的环境中的婴幼儿并且提供对应的声音数据样本流，至少考虑到声音的时间和/或谱模式来检测声学监测声音数据样本流中的婴幼儿啼哭，
优选地，在具有考虑是否已经观察到超过阈值的声级的前一检测步骤的多步骤/多级啼哭识别中，
为了进一步评估，选择检测到的啼哭相关的部分，
以及
提供
选择的检测到的啼哭相关的部分
连同与所述个人婴幼儿数据对应的信息
给评估级用于进一步评估，其中根据与所述个人婴幼儿数据对应的所述信息以个性化方式评估选择的部分，

对所述啼哭相关的部分的评估包括

通过与已知的对应于不同类别的啼哭原因的模式进行比较,以产生每个啼哭相关的部分属于所述不同类别中的一个相应的类别的多个概率的方式,对选择的啼哭相关的部分的相继的评价

通过对啼哭相关的部分的相继的评价来建立这样的多个概率的序列,
以及分析所述概率的序列

其中所述评价和进一步评估级中的至少一个根据与所述个人婴幼儿数据对应的所述信息被个性化。

3. 根据权利要求1或2所述的方法,其中建立声音数据窗的序列,为每个窗并且在每个窗中建立类似谱图的表示,在所述窗中识别啼哭模式,并且选择与所述啼哭模式相关的数据用于进一步评估,优选地使用在时间上重叠的窗。

4. 根据权利要求3所述的方法,其中使用用于识别所述声音数据的类似谱图的表示中的所述啼哭模式的卷积神经网络来实现搜索啼哭模式。

5. 根据权利要求4所述的方法,其中,用于啼哭模式检测和隔离的所述卷积神经网络是非个性化的。

6. 根据权利要求5所述的方法,包括至少临时地存储声音数据,使得能够基于在所述声级超过阈值之前至少部分地获得的声音数据来建立时间和/或谱模式以用于啼哭相关的部分的搜索。

7. 根据前一项权利要求所述的计算机实施的方法,使用类别使得能够实现对“婴幼儿疲倦”、“婴幼儿饥饿”、“婴幼儿需要安慰”、“婴幼儿需要拍嗝”、“婴幼儿疼痛”中的至少一个、优选地至少两个并且特别优选的所有状况的评估。

8. 根据前一项权利要求所述的计算机实施的方法,包括

上传声音相关的数据连同婴幼儿数据信息,所述婴幼儿数据信息至少与年龄、性别、大小、体重、种族渊源、单胞胎/双胞胎/三胞胎、当前医疗状态、已知的医疗先决条件、特别是已知的当前疾病和/或发烧、父母和/或护理者的语言中的多个相关,和/或

上传与一个或多个先前的评估的准确性相关的婴幼儿数据信息。

9. 根据前一项权利要求所述的计算机实施的方法,其中考虑到声音的时间和/或谱模式来检测声学监测声音数据样本流中的婴幼儿啼哭被实现作为具有考虑是否已经观察到超过阈值的声级的前一检测步骤的多步骤/多级啼哭识别的一部分,其中优选地,考虑是否已经观察到超过阈值的声级的步骤基于以下中的至少一个:

当前声级超过阈值,

当前声级超过平均背景噪声一给定的裕度,

一个或多个频带中的当前声级超过阈值,

一个或多个频带中的当前声级超过对应的平均背景噪声一给定的裕度、声音的时间模式,

声级的时间模式和/或谱模式偏离突然的响的非啼哭噪声的时间和/或谱模式,

非声学提示,特别是从所述婴幼儿的视频监视数据、移动检测器和/或呼吸检测器得出的

和/或这样的比较被本地实现,特别是当在远离所述婴幼儿的数据处理装置——特别

是云服务器——上实现使用卷积神经网络来识别声音数据的类似谱图的表示中的所述啼哭模式的步骤时。

10. 根据权利要求9所述的计算机实施的方法,包括以下步骤:

本地检测来自声学监测的婴幼儿的声音是否超过阈值,

以及

响应于检测到所述声音超过阈值,将数据上传到在集中式自动啼哭模式检测中使用的服务器装置中。

11. 一种用于评估根据前述权利要求中任一项所述的方法的自动婴幼儿啼哭评估装置,包括:

传声器,用于连续地声学监测婴幼儿,

数字转换级,用于将监测声音流转换成数字数据流,

存储器级,用于存储个人婴幼儿数据信息,

通信级,用于将数据传输到集中式服务器装置,

其中

设置用于在所述数字数据流中识别啼哭的发作的啼哭识别级

并且

所述通信级被适配

以从所述集中式服务器装置接收与婴幼儿啼哭的个性化评估相关的数据。

12. 根据前一项权利要求所述的自动婴幼儿啼哭评估装置,还包括反馈装置,用于获得与一个或多个先前的评估的准确性相关的反馈信息,并且其中所述通信级被适配用于将反馈信息传输到集中式服务器装置。

13. 根据前一项权利要求所述的自动婴幼儿啼哭评估装置,还包括本地评估级,所述本地评估级被适配以根据从所述集中式服务器装置接收的与婴幼儿啼哭的个性化评估相关的数据来评估婴幼儿啼哭。

14. 根据权利要求12-14中任一项所述的自动婴幼儿啼哭评估装置,包括:

定时器,以及

评价级,该评价级:

在婴幼儿啼哭的评估之前,

评价

从所述集中式服务器装置接收的

并且

与婴幼儿啼哭的个性化评估相关的个人婴幼儿数据信息的当前年龄

和/或

年龄或数据的有效性

所述婴幼儿啼哭评估装置被适配以根据所述评价输出婴幼儿啼哭评估。

提供用于自动婴幼儿啼哭评估的数据的计算机实施的方法

[0001] 本发明涉及婴幼儿(baby)啼哭。

[0002] 每当新生婴幼儿由于或多或少严重的原因诸如饥饿、呼气困难、疲倦、需要更换尿布、具有某种形式的疼痛等而经历任何不舒服,新生婴幼儿确实会啼哭以寻求帮助。父母不仅不得不注意到婴幼儿正在啼哭,而且他们还不得不基于他们的经验、他们对来自婴幼儿的通常有限的信号的理解以及最终基于他们的直觉来查明他们的婴幼儿正在啼哭的当前原因。

[0003] 这可能由于两个简单原因而给父母带来压力。一方面,每当婴幼儿啼哭,必须及时听到婴幼儿;另一方面,父母需要识别(identify)原因,这对于具有他们的第一个新生儿的父母来说是一个特别的问题,而更有经验的父母将理解婴幼儿啼哭的方式经常指示需要注意的问题。

[0004] 已经提议靠近摇篮放置音频发射器以用于将音频声音发射到靠近父母的接收器——这解决了第一个问题,但是识别婴幼儿正在啼哭的原因的第二个问题仍然是用简单的发射器/接收器组合。鉴于此,为了以自动方式识别婴幼儿正在啼哭的原因,已经提出许多提议。例如,已经提议,使用既作为发射器又作为接收器的智能电话,并且在智能电话之一上安装婴幼儿啼哭评估应用程序(app),从而帮助识别婴幼儿正在啼哭的原因。即使在以此方式提供适当的硬件的情况下,识别婴幼儿正在啼哭的原因的问题仍然存在,因为识别婴幼儿啼哭的原因需要合适的应用程序。

[0005] 在科学文献中,已经据提议与这样的识别的方式相关的多个提议。

[0006] 在Charles C.Onu的论文“Harnessing Infant Cry for swift, cost-effective Diagnosis of Perinatal Asphyxia in low-resource settings”中,已经提议,围产期窒息——其在发展中国家是婴儿(infant)死亡率的前三个原因之一——可以通过模式辨识(recognition)系统来辨识,该模式辨识系统对已知的窒息婴儿和正常婴儿的啼哭中的模式进行建模。提议的是,对啼哭进行采样,并且每个啼哭样本被传递通过数个信号处理级,在所述信号处理级结束时提取表示MEL频率倒谱(Cepstrum)的系数的特征向量。然后,辨识过程包括以下步骤:音频采样、特征提取、均值归一化、使用交叉验证训练和测试。确保所使用的特征向量都具有相同的长度和采样率。

[0007] 在Charles Udeogu、Eyenimi Ndiomu、Urbain Kengni、Doina Precup、Guilherme M.Sant’anna、Edward Alikor和Peace Opar的在“美国加利福尼亚州长滩第31届神经信息处理系统会议(NIPS 2017)”中发表的论文“Ubenwa:Cry-based Diagnosis of Birth Asphyxia”中,作者提议,啼哭输入样本被分割、被预处理、特征被提取并且多节段分类被确定;然后,作出关于啼哭原因的决策。

[0008] 在Charles C.Onu、Jonathan Lebenso、William L.Hamilton和Doina Precup的论文“Neural Transfer Learning for Cry-based Diagnosis of Perinatal Asphyxia”中,据陈述,由窒息影响的新生儿的啼哭模式中存在显著变更。作者假设,从成人话音学习的模型参数可以用作用于对婴儿话音训练模型的更好(与随机相比)的初始化。他们还陈述,啼哭和呼吸的生理互联性早就被意识到,并且啼哭以呼吸肌肉的功能为先决条件;此外,啼哭

生成和呼吸被陈述为都由大脑的相同区域协调。作者提议一种模型并且评价了该模型在不同噪声情形——诸如儿童玩耍的声音、狗吠声和警笛声——下的鲁棒性。他们还评价了每个模型对使音频数据的长度变化的响应,并且陈述了真实世界的诊断系统必须能够对尽可能多的可用数据起作用。

[0009] 在J.Saraswathy、M.Hariharan、Wan Khairunizama、J.Sarojini、N.Thiyagar、Y.Sazali和Shafriza Nisha的、在Biocybernetics and Biomedical Engineering 38 (2018) 634-645中发表的论文“Time-frequency analysis in infant cry classification using quadratic time frequency distributions”中,作者提议,对婴儿啼哭的研究可能产生用于辨别婴儿的状况(诸如器质性紊乱、喂养管理、睡眠管理、孕妇健康和感觉运动整合状况)的自动工具。它们指的是诸如音调(pitch)信息、噪声浓度、谱(spectral)能量特征、基于谐波分析的属性、线性预测倒谱系数和MEL频率倒谱系数的参数。作者陈述,婴儿啼哭信号的表示可以使用基于时间-频率的技术,即小波包变换、短时傅里叶变换(STFT)和经验模式分解(EMD)。作者还陈述,在联合t-f分析中,信号的时域表示和频域表示可以被组合成t-f谱能量密度函数,从而引向对多分量信号的特性的清晰探索。提议t-f谱能量含量可用于得出可以表征啼哭信号的不同模式的突出特征,从而强调基于t-f分析的方法在使用多分量信号的分类和检测中的重要性,特别是用于高效地辨别不同的啼哭发声。

[0010] 在M.A.Tugtekin Turan和Engine Erzin的、在Interspeech 2018(2018年9月2--6日,海得拉巴)中公开的论文“Monitoring Infant’s Emotional Cry in Domestic Environments using the Capsule Network Architecture”中,作者提议采用来自表示婴幼儿啼哭的音频信号的短节段(segment)的谱图表示作为到特定深度学习拓扑的输入。为了实现准确的性能,作者应用高通FIR滤波器来去除信号上的话音声音和其他低频噪声。他们声称,婴幼儿啼哭声音不具有完全连续的特性;因此,在应用语音(voice)活动检测算法之前,具有不同大小或持续时间的类似脉冲的序列被分割。

[0011] 在Carlos Alberto Reyes-García、Sandra E.Barajas、Estepan Tlelo-Cuautle和Orion Fausto Reyes-Galaviz的论文“A Hybrid System for Automatic Infant Cry Recognition II”中,作者提议使用遗传算法,并且还提议自动婴儿啼哭辨识与自动话音辨识过程非常类似。

[0012] 在马来亚工程大学的生物医学工程系的Rodney Petrus Balandong R的2013年5月的综述“Acoustic Analysis of Baby Cry”中,据陈述,存在数种用以获得啼哭样本的方法。

[0013] 在Saraswathy Jeyaraman Hariharan Muthsamy、Wan Khairunizam、Sarojini Jeyaramaan、Thiyagar Nadarajaw和Sazali Yaa-cob5&Shafriza Nisha的“A review: Survey on automatic Infant Cry Analysis and Classification”(健康与技术 <https://doi.org/10.1007/s12553-018-0243-5>)中,作者陈述,自动婴儿啼哭分类过程是类似于自动话音辨识的模式辨识问题。他们报告,消除或分割是在婴儿啼哭分类分析中众所周知的预处理技术之一,因为沉默间隔通常携带较少信息,但是增加计算成本。作者还提到不同的啼哭类型,诸如在更换尿布时、在喂食之前、在平静时、在儿科评价期间的自发性啼哭,以及具有病理状况,诸如腔静脉血栓形成、脑膜炎、腹膜炎、窒息、舌系带、IUGR小头畸

形、法洛四联症、高胆红素血症、腹裂、IUGR窒息、牛蛋白过敏、心脏综合征、X染色体。

[0014] 根据M.D.Renanti等人的、在Journal of theoretical and applied Information Technology (I-ESS 1817-31 95)中公布的论文“Infant Cries Identification by using Codebook as Feature Matching, and MFCC as Feature Extraction”中,如果仅在声音信号开始时和结束时从声音数据流删去沉默是不利的。

[0015] 在Stavros Ntalampiras的“Audio Pattern Recognition of Baby Crying Sound Events”(Journal of the Audio Engineering Society,第63卷,第5期,2015年5月)中,提议了一种在五个不同状态之间进行区别的方法,即(a)饥饿、(b)不舒服(需要更换)、(c)需要拍嗝(hurp)、(d)在疼痛中和(e)需要睡眠。据陈述,所涉及的音频信号的周期性性质是负担。作者考虑了数组声学参数,诸如感知线性预测参数、Mel频率倒谱系数、感知小波包、基于Teager能量算子(TEO)的特征、时间调制特征。讨论了诸如支持向量机、多层感知等的多种方法来辨别啼哭。

[0016] 在Rodica Ileana Tuduce、Mircea Sorin Rus、Horia Cucu和Corneliu Burileanu的论文“Automated Baby Cry Classification on aHospital-acquired Baby Cry Database”中,据提议,一种能够在不同种类的婴幼儿啼哭之间进行区别的婴幼儿啼哭辨识系统将在父母自己学习进行这样的区别时帮助父母区别他们的特定婴幼儿的需要。作者检查了多个分类器,但是观察到大多数分类器对真实记录的婴幼儿啼哭的表现低于对从仔细选择的样本提取的啼哭的表现。

[0017] 在Rami Cohen和Yizar Lavner的论文“Infant cry analysis and detection”(2012年IEEE第27届以色列电气和电子工程师大会)中,提议了一种算法,该算法包括三个主要级,即语音活动检测器级、分类级和用于验证分类级的后处理级,以减少负面错误。此算法据陈述基于不同时间尺度上的三个决策级别:即帧级别,其中每个帧(几十毫秒)基于其谱特性被分类为“啼哭”或“不啼哭”;几百毫秒的部段(section);以及数秒的节段,在其中根据它们包含的“啼哭”部段的数目来获得最终决策。多时间尺度分析和决策级别据说旨在提供具有非常高的检测率的分类器,同时保持低的误报率。作者认为,使用婴儿啼哭记录以及其他自然声音(诸如汽车引擎、喇叭声音和语音)的性能评价展示了在存在噪声时的高检测率和鲁棒性。

[0018] 在Shubham Asthana、Naman Varma和Vinay Kumar Mittal的论文“An Investigation into Classification of Infant Cries using Modified Signal Processing Methods”中,据提议,婴儿啼哭是发音、收缩性沉默、咳嗽、哽咽和中断的组合。

[0019] 在专利文档中也已经提议了方法和装置。

[0020] 从CN 103530979A,已知一种用于医院的远程婴幼儿啼哭警报装置,该装置包括婴幼儿啼哭检测模块、警报计划模块、警报接收模块和警报模块,其中一些部分通过电线连接,而其他部分以无线方式连接。

[0021] 从CN 104347066A,已知一种“基于深度神经网络的婴儿啼哭声音辨识方法和系统”。据提议,根据记录的啼哭来区别病理状况和非病理状况。

[0022] 从CN 106653001A,已知一种婴儿哭声辨认方法和系统。据陈述,一个主要问题是仅能够给出一个啼哭理由。提议了一种用于辨识婴儿啼哭的原因的方法,并且据陈述,在此上下文中,可以提取和分析多个以下特征:平均啼哭持续时间、啼哭持续时间方差、平均啼

哭能量、啼哭能量方差、音调频率、音调频率的平均值、音调频率的最大值、音调频率的最小值、音调频率的动态范围、音调频率的平均变化率、第一共振峰频率、第一共振峰频率平均变化率、第一共振峰频率的平均值、第一共振峰频率的最大值、第一共振峰频率的最小值、第一共振峰频率动态范围、第二共振峰频率、第二共振峰频率平均变化率、第二共振峰频率平均值、第二共振峰频率最大值、第二共振峰频率最小值、第二共振峰频率的动态范围、Me1频率倒谱参数、翻转的Me1频率倒谱参数。关于预处理步骤,据提议,对啼哭信号执行降噪以抑制背景噪声,并且使用自动检测算法来去除具有特别嘈杂噪声的数据片段(fragment),从而提高提取到后续特征中的啼哭信号的信噪比。应理解,根据CN 106653001A提取的特征以及提取它们的方式也可以被用在本发明的上下文中。因此,所引用的文档通过引用完全并入本文。

[0023] 从CN 106653059A,已知一种婴儿啼哭自动辨识方法和其系统。据提议,为了识别婴幼儿正在啼哭的原因,婴幼儿的年龄和在啼哭时的啼哭时间可以帮助确定啼哭的病理原因的概率。关于啼哭时间间隔,明确提及最后一次哺乳时间。还据陈述,在记录婴幼儿啼哭声音的同时,执行对捕获婴幼儿的面部的视频的视频的图像分析可以是有帮助的。应注意,通过非实验室条件下的不专业的记录,判断的准确性将下降,从而给出不准确的啼哭原因或误导缺乏经验的父母。明确提及将已知方法实施为智能电话上的应用程序。

[0024] 从CN 107591162A,已知一种基于模式匹配的啼哭辨识方法和智能护理系统。据陈述,年轻父母在其家外花费越来越多的时间,但是雇临时照顾婴幼儿者是昂贵的;因此,婴幼儿啼哭可能未被及时处理。考虑到智能家居,婴幼儿护理功能被提议以解决此问题。

[0025] 从GB 2234840A,已知一种自动婴幼儿啼哭检测,其在检测到婴幼儿正在啼哭时自动产生声音。声音持续一足以确保婴幼儿被哄入睡的时间。此后,啼哭检测器被关闭声音一足以确保父母不忽视真正的悲伤的啼哭的时间。

[0026] US 2008/000 3550A1提议通过以可再现的音频形式存储婴儿声音来教导新父母特定啼哭的含义。存储介质可以是DVD。

[0027] 从KR 2008 003 5549A,已知一种用于向移动电话通知婴幼儿的啼哭的系统,其中当检测到啼哭声音时,自动呼叫母亲的移动电话。

[0028] 从KR 2010 000 466A,已知一种儿科诊断设备,其能够通过儿童的啼哭对儿童儿科肺炎和儿科肺炎进行早期诊断。

[0029] 从KR 2011 0113359A,已知用于使用频率和连续模式检测婴幼儿的啼哭声音的方法和设备。

[0030] 从US 2013/031 7815A1,还已知一种用于分析与婴幼儿啼哭相关联的数字声音音频信号的方法和系统。据提议,通过输入通过在预训练的人工神经网络中处理数字音频信号而确定的时间-频率特性来确定婴幼儿的特定需要。

[0031] 从US 2014/004 4269A1,已知一种智能周围声音监测系统。据提议,所述系统监测周围声音环境,并且将其与预设声音进行比较,例如关于频率特征(signature)、振幅和持续时间,以检测重要的或关键的背景声音,诸如警报、喇叭、定向口头通信、啼哭的婴幼儿、门铃、电话等。据陈述,所述系统对于经由屏蔽周围声音的头戴式受话器收听音乐人是有用的。

[0032] 在US 2019/180772A1中,据提议,音频捕获装置可以在长期或短期时段内存储音

频数据,并且音频捕获装置可以以无线方式传输音频。还据陈述,诸如智能电话的移动终端可以被用来记录和显示啼哭声音,并且在不利环境(诸如嘈杂环境)中,自动判断的准确性将在一定程度上降低。据陈述,通过在终端屏幕上显示啼哭的多个原因,系统将具有更好的错误容忍性。据陈述,可以使用深度神经网络来实施分类器。还据提议,执行分割并且识别用于每个节段的源。此外,该文档考虑以周计的年龄与典型啼哭时间之间的关系。此外,据提议,分割音频流的过程可以涉及机器学习算法,以将音频数据的数据集自动解析为有标签的时间节段,所述时间节段例如将待被评估的婴幼儿与其他儿童、环境噪声或沉默区分开来。然而,任何这样的个性化被提议仅用于啼哭识别。此外,据陈述,可以为多个年龄群体创建发声、啼哭和固定信号/植物性睡眠声音模型,例如,群体各自包括以2个月年龄间隔的婴幼儿。

[0033] 从US 2016/036 4963A1,已知一种用于检测用于智能电话装置的音频事件的方法和系统。据提议,当电子装置获得音频数据时,音频数据被分裂成多个声音分量,每个声音分量与频带的一个相应的频率相关联并且包括一系列时间窗。提议电子装置然后从这些声音分量提取特征向量,并且对提取的特征向量进行分类。以此方式,智能电话装置应能够区别不同的音频事件。

[0034] 从US 2017/017 8667A1,已知用于使用声学特征的时间特性来进行鲁棒的啼哭检测的技术。据提议,将声音数据分裂成帧,然后确定用于每个帧的声学特征向量,并且基于与帧对应的随时间变化的每个声学特征来确定参数。然后基于所述参数确定声音是否与预先定义的声音匹配。提到婴幼儿监测器的使用和婴幼儿啼哭的识别。据陈述,从数据集生成少量参数对于识别期望的声音是有用的,因为这将是使用机器学习技术(诸如神经网络)的一个重要方面。据陈述,已知的声音识别装置可以被体现在计算机、智能电话、膝上型计算机、摄像机装置、消费电子装置或其他装置中。

[0035] 从CN 107657963A,已知一种啼哭识别和啼哭辨识方法,该方法适合于辨识婴儿啼哭的原因,并且收集不同的啼哭样本和根据不同的婴儿来对应啼哭原因,以便为良好的啼哭辨识提供比较。据陈述,通常婴幼儿啼哭具有比纯背景噪声更高的音量和更高的能量。据陈述,可以提供用于存储至少一个啼哭样本的啼哭数据库,并且在识别啼哭的原因的装置的使用期间在已经识别该原因之后,附加的啼哭样本可以被存储在该数据库中。还据提议,在不可以基于数据库中的声音样本确定啼哭的原因的情况下,将附加的啼哭信息存储在数据库中。

[0036] 从CN 107886953A,已知一种基于面部表情和话音辨识的婴儿啼哭语音转化系统。据提议,使用啼哭微处理器以通过学习记忆和反馈自检功能来持续训练和优化样本啼哭数据库中的样本特征数据。提议鉴于强度大于阈值,确定声音节段是否对应于婴幼儿啼哭。

[0037] 从CN 109243493A,已知一种基于改进的长期和短期记忆网络的婴幼儿啼哭情绪辨识方法。在此上下文,必须训练长时间和短时间记忆网络。

[0038] 从CN 110085216A,已知一种婴幼儿啼哭检测方法和装置。该文档陈述,在用于婴幼儿信号啼哭检测的检测技术中存在不足,包括支持向量机学习算法,该算法对婴幼儿啼哭和其他声音具有低区分精度并且对声音的检测不够准确。提议执行感知线性预测系数的特征提取,并且获取与训练它的样本中的话音数据对应的话音特征。将提供至少两种语音类型,并且提议婴幼儿啼哭声音的声学模型考虑每个帧的后验概率以对应于特定的语音

类型。

[0039] 从CN 1564 2458A,已知一种婴幼儿啼哭检测方法,该方法依赖于与多个存储的样本的比较。

[0040] 如可以看到的,存在多种识别婴幼儿啼哭的原因的方法,并且也可以区别多个不同的状况。因此,关于啼哭识别的方法、特别是关于机器学习方法、以及此外关于可以通过分析啼哭声音识别婴幼儿啼哭的不同原因,上文所引用的文档整体被附于本文。

[0041] 然而,虽然过去已经进行了大量研究以根据啼哭本身来识别婴幼儿正在啼哭的原因,并且虽然据提议可以区别多个不同的状况,但是由实际装置所获得的结果仍然需要被改进。在这方面,应注意,已知的是某些状况对啼哭特性具有大的影响,使得不同的婴幼儿在类似的情况下将以不同的方式啼哭。

[0042] 在这方面,在Dror Lederman的硕士论文“Automatic Classification of Infant's Cry”中,新生儿的生理学与它们的啼哭的音频特征相关,并且比较了针对足月新生儿与早产新生儿的静止啼哭的直方图。其他比较尤其包括子宫内可卡因暴露的婴儿的啼哭与未暴露的婴儿的啼哭,以及具有诸如代谢紊乱或染色体异常的紊乱的婴儿的啼哭。作者陈述,当处理啼哭信号时,自动分割的准确性不像语音/单词分割那样关键,其中不准确的分割可能导致重要信息的丢失。作者还陈述,已知年龄在对啼哭信号的分析中是关键参数,并且已经发现如果婴儿发育——尤其是在前几个月期间——则包括基本频率和共振峰的啼哭特征显著变化。

[0043] 在KR20030077489A中,据强调,婴儿快速成长,并且诸如种族、性别等的啼哭特性可以被分类为不同的孩童群体。据陈述,大批量生产的机器不能够分析啼哭婴儿的个体特性。提议使用本地互联网(internet)终端用于从啼哭的婴幼儿获取声音数据,并且利用互联网服务器用于对声音数据进行分析。提及数据可以被存储以供将来在婴儿啼哭的研究中使用。还提议一种用于提供即时状况分析服务的服务方法以及一种用于提供即时状况分析服务的服务方法,其中婴儿人群的细节可以被存储在数据库中。然而,虽然关于婴幼儿啼哭的原因的决策可以基于大型数据库,但是缺点是必须提供到服务器的连接,并且因此,在没有连接的情况下,啼哭表征是不可能的。

[0044] 从KR 2005 0023812A,已知一种使用无线因特网(Internet)连接的用于分析婴儿啼哭的系统。提议提供一种服务器管理系统,该服务器管理系统管理无线因特网服务系统,该无线因特网服务系统进而为无线因特网终端提供无线因特网终端婴儿语音应用程序。据陈述,个性化的声音数据库可以被配置,并且婴儿声音装置应用程序所需的信息可以被修改,使得用户总是能够接收根据最近的研究的对啼哭的准确分析。然而,未提及如何最好地扩大数据库,也未陈述如何以特别高效的方式实现对婴儿声音装置应用程序的修改。

[0045] 从KR 2012 0107382A,已知另一种用于分析婴儿的啼哭的装置。据陈述,如果婴幼儿啼哭声音频率分布信息已经在一预定时段内被辨识最小的次数,则可以统计地处理啼哭频率分布信息,以便调整和优化放置装置的位置处的特定婴幼儿的啼哭声音。据提议,成人使用该装置可以对婴幼儿正在啼哭的原因发声,并且此发声被辨识,使得如果确认在婴幼儿正在啼哭期间或之后一特定时间段内辨识到用户发声,则可以处理发声内容,以便与和婴幼儿啼哭相关的服务功能相关联。这样的发声可能是“38,5°”或“尿布不潮湿”。

[0046] 从CN 109658953A,已知一种婴幼儿啼哭辨识方法和装置。据陈述,可以提供云服

务器,音频特征向量和收集的音频数据节段可以被发送到该云服务器。当装置连接到服务器时,云服务器可以向该装置发送识别模型的最新版本,并且如果识别模型不是最新版本,则该装置可以比较并且发送其自己的识别模型到云服务器。此外,在没有到云服务器的网络连接可用的情况下,可以通过本地存储的神经网络模型来识别音频特征向量。

[0047] 因此,在过去已经提议,以自动方式识别婴幼儿正在啼哭的原因。然而,尽管在过去已经提议个性化可以帮助确定婴幼儿正在啼哭的原因,但是由自动方法提议的评估通常不被认为足够可靠。鉴于此,允许对自动啼哭评估技术的改进将是有帮助的。

[0048] 本发明的目的是提供用于工业应用的新颖事物。

[0049] 此目的通过独立权利要求中所要求保护的主体实现。在从属权利要求中描述了一些优选的实施方案。

[0050] 根据第一总体构想,提议了一种提供用于自动婴幼儿啼哭评估的数据的计算机实施的方法,所述方法包括以下步骤:声学监测婴幼儿并且提供对应的声音数据流,在所述声音数据流中自动检测啼哭,响应于检测到啼哭而从声音数据自动选择啼哭数据,根据选择的啼哭数据确定允许啼哭评估的参数,建立用于个性化啼哭评估的个人婴幼儿数据,根据个人婴幼儿数据准备用于评估的评估级,以及将所述参数馈送到根据个人婴幼儿数据准备的所述啼哭评估级。

[0051] 本发明的发明人已经理解,对于婴幼儿啼哭的个人评估,需要在评估中使用的高质量的啼哭数据。在提供的用于自动婴幼儿啼哭评估的数据具有不足的质量的情况下,以其他方式可能也不能够充分地实现个性化的效果,并且评估的质量(例如,如根据正确评估的百分比推断的)不能够相比于非个性化评估增加或显著增加。相反,在数据的质量足够高的情况下,个性化通常不仅提高可靠性。此外,个性化通常仅需要在非常晚的级处被影响;特别地,尽管评估的个性化,但是通常可以对于所有婴幼儿使用相同的参数集。这简化了评估。

[0052] 尽管如此,即使一旦已经选择了正确的声音输入数据,通过对于所有婴幼儿使用相同的参数确定级可以获得非常好的结果,也将可以根据建立的个人婴幼儿数据来确定不同的参数集。

[0053] 个人婴幼儿数据可以以多种不同的方式被建立,但是将明显的是,在评估之前以个性化方式从父母或其他护理者请求个人婴幼儿数据是最优选的方式并且是最容易实施的。还应理解,仅在用于执行所述方法和用于稍后更新一些输入的装置的初始化期间,才需要从父母或其他护理者请求对应的输入。虽然通过从父母和/或护理者请求输入来建立个人婴幼儿数据被认为是最可靠并且最简单的方式,但是也将可以通过啼哭分析来识别数据中的至少一些;例如,可以评价来自同一婴幼儿的单个啼哭或多个先前啼哭,以然后得出个性化,诸如婴幼儿的最可能的年龄、体重、大小或性别。

[0054] 通过声学监测婴幼儿并且从声音数据流自动选择相关的啼哭数据来确保啼哭数据的高质量。选择的数据可以与声音数据流分离,换言之,它们可以被提取,或可以被标记为啼哭的一部分或潜在地为啼哭的一部分;在不完全清楚声音数据是否属于啼哭的情况下,例如,因为婴幼儿由于周围环境的长期的响的噪声而开始啼哭,对应的数据可能被标记为“潜在地为啼哭的一部分”。这样的标记可能不同于在给出声音数据属于啼哭的较高置信水平的情况下的标记。

[0055] 在这方面,应理解,通常并且优选地以连续的方式进行监测婴幼儿,使得在延长的时间期间从婴幼儿的附近记录声音。这与例如父母仅在他们注意到婴幼儿正在啼哭时才触发声音数据的收集的情形相比,具有多种优点。在延长的时间内监测婴幼儿可以获得既包括啼哭时段又包括非啼哭时段的声音数据。这进而简化了对典型背景行为的考虑。应理解,声学背景特性将关于声音的级别、关于噪声的谱分布以及关于例如由于狗吠、喇叭吹响、摔门、兄弟姐妹啼哭等引起的显著的背景噪声的长度和发生而改变。理解这样的背景行为毫无疑问地帮助从声音流选择数据作为啼哭数据,并且因此帮助改进提供的用于个性化评估的数据的质量。

[0056] 例如,在空调系统在特定频带中生成噪声的情况下,在婴幼儿啼哭评估的一些实施方式中,在确定用以描述婴幼儿啼哭的参数中应忽略这样的频带。通过以连续的方式监测婴幼儿,通过查看在婴幼儿不啼哭的时段中获得的声音数据,可以注意到在特定频带中存在这样的噪声。因此,在依赖于消除特定频带中的噪声的婴幼儿啼哭评估的相应的实施方式中,可以确定应忽略对应的频带,并且对应的信息可以被添加到选择的声音数据。这与简单地滤除掉受噪声影响的频率相比是更优选的,因为那样,虽然已经发现剩余的频带对于婴幼儿啼哭通常是相关的,但是对于特定情况它们将不被考虑。此外,这并不意味着特定的声音数据流必须经受(计算密集的)带通滤波;将足够的是,将对应的信息正向馈送到参数确定级,使得不是分配表示例如相应的频带中的谱强度的值,而是这样的值可以被指定为“不可用”(N/A)。应理解,在某些频带将被忽略的情况下,使用例如不同的滤波器参数,用于啼哭检测的不同算法可能变得必要。还将理解,替代地和/或附加地,可以通过在婴幼儿开始啼哭之后的某一时段内选择帧来选择啼哭数据。然而,还存在优选地不滤除掉特别易于产生噪声的任何特定频带的实施方案。应理解,在数据的量未显著减少的实施方案中,例如因为没有诸如共振峰相关的参数、音调频率相关的参数、第一性能的最大值等的特定参数(特别是比较在下文中列出的参数),而是与婴幼儿啼哭相关的更完整的信息被馈送到卷积神经网络等中,噪声的不利影响显著不那么明显。已经认识到,同时,评估的质量提高。现在,应理解,在卷积神经网络等中处理更完整的信息将需要更大的计算工作量;然而,通过省略任何滤波步骤和/或通过允许独立于存在的特定噪声属性来实现相同的处理,至少部分地补偿了此附加的计算工作量。因此,发明人已经认识到,从总体角度来看,在分析婴幼儿啼哭时将更完整的信息馈送到基于人工智能的评估级比花费相当大的计算工作量用于降低馈送到评估级的数据的复杂性更有益。然而,可以做的是确保待被评估的这样的更完整的信息实际上对于评估婴幼儿啼哭是相关的,这在通常在啼哭中发现的模式已经在声音数据中被识别并且从其隔离的情况下可以容易被假设。因此,在一个优选的实施方案中,选择用于声音评估的正确声音输入数据包括:数据流中的声音相关的模式的识别,并且优选地将这样的声音相关的数据与非声音相关的数据隔离。

[0057] 通常,婴幼儿将啼哭延长的时段,但是也将存在未记录到响的啼哭的短时间,例如因为婴幼儿需要呼吸。与这些短时间相关的信息不需要被完全丢弃。特别地,在参数被确定用于进一步评估的情况下,这些短时间优选地不从声音数据流切除,因为它们也可能含有有用的信息。应理解,在某些情况下,在婴幼儿开始啼哭之后未记录到非常响的声音的这样的时间的长度可能给出在评估婴幼儿啼哭的原因中重要的线索。因此,如果情况是这样,至少包括声音数据的长度的指示可以是有帮助的。在其他情况下,可能有用的是至少确定期

间啼哭相关的模式在声音数据中已经发生的时间或时间标记。

[0058] 然而,在啼哭参数被确定的情况下,将甚至更优选的是,根据更长的、不间断的时段确定啼哭参数,因为以此方式,可以从啼哭的重复发作获得线索,即使婴幼儿在重复发作期间可能不是特别响的。

[0059] 在考虑到较长的不间断的啼哭时段的情况下,可以通过切断啼哭前噪声和/或啼哭后噪声的延长的时段来隔离啼哭。此外,应理解,未接收到足够照顾的婴幼儿能够在非常长的时段内啼哭。因此,应理解,即使啼哭仍然继续,也优选地评估婴幼儿啼哭的原因。在这样的情况下,假使父母或护理者不会足够迅速地作出反应,可以重复评估;如果在这样的延长啼哭时段期间获得的评估应变化,可以作出对不同的评估之间的最佳评估的评价。应理解,在延长的时段期间,评估可能变化,因为婴幼儿啼哭的原因在延长的时段期间缓慢地变化,例如因为先前感觉疼痛的婴幼儿缓慢地变得疲倦。

[0060] 因此,当提供或取得用于自动婴幼儿啼哭评估的数据时,选择或提取或识别啼哭数据可能与识别应被分析的时间和/或应(或不应)被分析的频带或频率相关。关于省略频带,应明确提及的是,用以避免奈奎斯特(Nyquist)混叠的带通滤波不被认为频率的“省略”。相反,在必须提到省略频率的情况下,应理解,省略的频率将低于采样频率,并且通常,对数字数据实现省略。因此,可以通过忽略在最低可处理频率以上并且在最高可处理频率以下的某些频带来省略频率。然而,应注意,虽然特定滤波不是至关重要的,特别是对于基于类似谱图的表示检测啼哭模式并且因此声音处理可以被保持到最小的实施方案,但是可以优选的是将声级归一化,例如以使得归一化的最大声级对于每个窗是相同的方式。应注意,一旦在一个窗中已经识别和隔离一个啼哭模式,就存在如下可能性:在该窗中出现并且用作用于使声级归一化的参考的最大声级不构成啼哭模式的一部分。这将是例如门被非常响地砰地关上从而导致婴幼儿啼哭的情况。因此,在优选的是使用具有归一化声级的啼哭模式用于后续啼哭转化或使用另一个术语——啼哭评估的情况下,可以重新归一化啼哭模式。此外,应注意,在啼哭转化的某些实施方式中,特别是使用卷积或神经网络用于转化的某些实施方式中,优选的是具有标准化的长度的啼哭模式。因此,可以添加表示沉默的数据,例如通过将声音数据延长对应的沉默时段,或通过向表示沉默的谱图添加一个区域,例如通过是完全黑色的。应注意,当使用类似谱图的表示作为输入被馈送到其中的机器学习模型——诸如基于卷积或神经网络——时,在啼哭模式转化中使用归一化和标准化的长度是特别优选的。还应注意,尽管在整个说明书和权利要求书中多次提到声音数据的类似谱图的表示,换言之,分割的窗和/或啼哭模式,但是使用声音数据的线性类似谱图的表示不是必要的;相反,可以使用声音数据的非线性类似谱图的表示,特别是类似梅尔谱图(melspectrogram)的表示和/或类似锁谱图(lock spectrogram)的表示。声音数据的这些非线性类似谱图的表示既可以被用于啼哭模式转化,也可以被用于啼哭模式识别和隔离。

[0061] 如之前所陈述的,以个性化遵循检测啼哭、选择啼哭数据或根据选择啼哭数据确定参数的在先步骤的方式实施个性化评估不是必要的。这进而是有利的,因为个性化的计算和/或组织工作量被保持到最小;此外,在个性化评估将不可能的情况下,例如因为对于具有特定个人数据——诸如性别、年龄、大小、体重、医疗先决条件等——的婴幼儿,同龄群体仍然太小,至少可以实现不受不足的特定数据损害的非个性化评估。应注意,也可以选择“类似的”同龄群体和/或同龄群体的数目可以是更小的直到数据库已经充分增长。关于个

性化,这样的个性化可以被实施为对每个个体婴幼儿使用有区别的并且不同的参数的私有化,或可以被实施为确定具有非常类似的啼哭模式的婴幼儿的同龄群体或聚类的聚类化。应理解,通过特别地关于仅从一个特定婴幼儿获得的婴幼儿啼哭进行训练和建模,私有化是可能的;然而,私有化也可以通过首先确定更通用的模型来实现,例如基于来自具有非常类似的啼哭模式和/或非常类似的个人数据(诸如体重、年龄、大小和性别)的婴幼儿的同龄群体或聚类的啼哭,通过略微适配滤波器参数使得它们更适合特定婴幼儿。这被称为迁移学习,并且应理解,在提供用于婴幼儿啼哭评估的数据的本申请中提议的特定方式在通过迁移学习来个性化婴幼儿啼哭评估中是特别有帮助的。

[0062] 正如上文所陈述的,在Charles C.Onu、Jonathan Lebenso、William L.Hamilton和Doina Precup的论文“Neural Transfer Learning for Cry-based Diagnosis of Perinatal Asphyxia”中,据提议,从成人语音学习的模型参数可以作为用于对婴幼儿语音训练模型的更好的(与随机的相比)初始化,此申请人未意识到通过根据更通用的婴幼儿啼哭评估模型进行迁移学习来使婴幼儿啼哭评估个性化的任何尝试,特别是不以通过数据库条目的聚类化来获得迁移学习基于的初始模型的方式,特别地没有根据评估的婴幼儿啼哭的原因精细聚类区分比例如6、8、10、15、20个不同的数据库条目的聚类更多的原因区分。

[0063] 还应理解,本发明的方法帮助生成具有来自不同婴幼儿的啼哭的数据库,使得聚类可以使用比过去已知的更精细的区别,例如以不大于500g、400g、300g、200g或100g的体重间隔;不大于5cm、4cm、2cm或1cm的大小间隔;不大于8周、6周、4周、2周的年龄间隔对婴幼儿进行分组。显然,也可以选择其间的任何间隔。应理解,甚至大于针对重量、大小或年龄的指示最大值的间隔也将导致个性化,该个性化是相当合理的理由并且因此未充分利用通过本发明可获得的高质量啼哭数据,而针对体重和大小指示的下限反映通常在私人住宅中观察到的测量结果的不准确性,使得更精细的个人个性化将不会是太有帮助的。在聚类将完全地或部分地基于相应的个人婴幼儿数据的情况下,还可以考虑附加的参数,诸如以细致的0.1°C步长(step)、0.2°C步长或0.3°C步长的婴幼儿的当前温度或已知的医疗条件。

[0064] 声音数据将被采样,例如以4kHz、8kHz或10kHz、16kHz的采样频率;采样频率通常是根据婴幼儿啼哭的频率含量、在监测婴幼儿时使用的传声器的频率响应和/或根据可用的计算能力和/或可用于将声音数据上传到在自动婴幼儿啼哭评估中使用的云和/或服务器的带宽来确定的。带宽可以适合于例如可用于将声音数据上传到云的带宽。然而,虽然可以在8kHz以上的频率范围内找到相关的啼哭信息,但根据使用的传声器以及根据它们的方向性二者,在现场记录这些频率通常是困难的,因为即使在传声器在高频率下足够灵敏的情况下,使用的传声器灵敏度的极性模式可能是不利的;这对于较高频率变得更重要。因此,在不限制本发明的情况下,对于大量的用户来说,多达8-10kHz的采样频率给出不能够与使用较高频率获得的那些结果区别开来的结果。来自传声器的声音信号将被预处理,诸如被放大、被低通和/或带通滤波并且被数字化。为了进一步处理和/或为了将声音数据传达到服务器、云服务器等,优选的是限定包括多个样本的帧,特别是固定数目的样本(诸如64个样本、128个样本或256个样本)。虽然没有必要使用固定帧或根本不使用固定帧,但是在下文中,由于使用帧降低了计算复杂性,因此经常必须提到帧。

[0065] 关于从啼哭数据确定的参数,可以确定以下参数中的一个或多个:

[0066] 当前啼哭事件期间的平均啼哭能量,在特定数目的连贯的和/或特别是在2、4、8、

16或32个帧中的帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——内的啼哭能量的滑动平均值；在一个事件期间的在间歇之间的啼哭持续时间方差；

[0067] 特别是在2、4、8、16或32个帧2、4、8、16或32个帧上和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——内的啼哭能量方差；当前音调频率；在2、4、8、16或32个帧2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——上平均的音调频率；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的音调频率的最大值；

[0068] 在啼哭事件期间在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的滑动最大音调频率的变化；

[0069] 在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的音调频率的最小值；在啼哭事件期间在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的滑动最小音调频率的变化；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的音调频率的动态范围；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的频率的音调平均变化率；

[0070] 在啼哭事件中或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧中、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的的第一共振峰频率(应注意，在本发明的上下文中，术语共振峰可以涉及人类声道的谱整形结果；此外，当谈到共振峰和/或通过谐振增强的谐波部分时，可以参考该谱中的峰或局部最大值)；

[0071] 在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间平均的第一共振峰频率的平均变化率；

[0072] 在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间滑动平均的第一共振峰频率的滑动平均变化率；

[0073] 在啼哭数据的2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间平均的第一共振峰频率的平均值；

[0074] 在啼哭数据的2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——内的第一共振频率的最大值；在啼哭数据的2、4、8、16或32个帧中、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的的第一共振频率的最小值；

[0075] 在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的的第一谐振峰频率动态范围；

[0076] 在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的的第二共振峰频率；

[0077] 在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率平均变化率；

[0078] 在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率平均值；

[0079] 在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率最大值；

[0080] 在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率最小值；

[0081] 在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二谐振；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的峰频率动态范围；

[0082] Mel频率倒谱参数(应注意,倒谱是以下数学运算序列的结果:a-信号从时域到频域的变换-b-谱振幅的对数-c-到倒频率域的变换,其中最终的独立变量——倒频率——实际上具有时间尺度),在整个啼哭事件和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间确定该参数;和/或

[0083] 倒置的Mel频率倒谱参数。

[0084] 应理解,虽然可以对于每个啼哭确定上文列出的参数或所述参数中的一些,用于将预先计算的参数馈送到神经网络中,但是这将不是绝对必要的。特别地,可以将包含所有相关信息的记录的啼哭声音的表示馈送到机器学习模型中;在该情况下,机器本身“评价”哪些参数实际上是相关的。这样的表示中的一个实施例将是用于声音的梅尔谱图。

[0085] 应注意,在上文提到特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——的情况下,也可以提到任何其他固定时段——诸如7秒或28秒——一直到明确提及的相应的时间。这对于在下文中提到的其他参数也成立。然而,应理解,相应的时间和/或帧数目是有利的,因为多达5秒的较短时间与确定啼哭数据中的非常适合转化啼哭的典型模式相关。如果信息中的一些被周围噪声淹没,则多达15秒的中等长度是有帮助的,而在同时共存多个这样的原因——诸如婴幼儿疼痛、饥饿和困倦——的情况下,多达30秒的较长时段对于识别婴幼儿啼哭的主要原因是有帮助的。

[0086] 应理解,在评估中,甚至在个性化评估中,不需要所有上文提及的参数。相比之下,产生非常良好的结果的评估可以仅依赖于上文提及的参数中的数个。这尤其是如此,因为参数中的一些将涉及稍微冗余的信息,例如,在整个啼哭事件期间的声级的平均值,2、4、8、16、32或64个帧上的声级的滑动平均值等。在个人评估中使用诸如卷积神经网络的技术的情况下,对于由于相同的年龄、种族渊源、性别、体重、大小等而具有类似的啼哭模式的婴幼儿的不同群体,列出的不同的参数可以是最好的。尽管如此,即使那样,通常也可以发现用于多种不同的婴幼儿的共同的参数集,使得需要被确定的参数的总数目可以是相当低的,并且仍然可以为个人评估提供有用的参数集。对于在由于噪声而不能被使用一些频带的情况下的参数同样成立。这帮助将与个性化相关的计算和组织工作量保持在最小值处,特别是在用于个性化的一个或多个相关描述符指代体重、身体大小或年龄的情况下;在个性化评估级的更新到期或过期的情况下或在必须执行非个性化评估的情况下,例如,因为为

了啼哭数据的个性化评估而通常被寻址的云服务器当前是不可用的,使用对于一定范围的体重、一定范围的身体大小和/或一定范围的年龄最好的参数集可以帮助仍然获得非常好的评估结果。

[0087] 应强调,在上文中,已经提及了对滑动参数的使用。诸如使用滑动参数或交叉关联技术的技术是特别优选的,因为以该方式,可以减少足够精确地确定啼哭的确切发作的影响。

[0088] 然而,如上文提及的,根据声音数据确定用于后续评估的低维参数不是绝对必要的(应注意,必须提到低维参数,因为通过使用有限数目的参数,与评估例如以16比特和48千赫记录的500ms的声音数据的组块(chunk)的情形相比,维度明显降低)。另一个可能性不是计算低维参数,而是初步识别通常可以与婴幼儿啼哭相关联的模式,将这些隔离并且然后将隔离的模式馈送到后续评估级。特别简单地实施的一个可能性是为给定长度的组块生成谱图,以及然后在那些谱图中查找通常与婴幼儿啼哭相关联的模式。优点是,生成谱图是简单的并且可以以低计算工作量来完成;在那些谱图中搜索通常与婴幼儿啼哭相关的模式然后通过常见的图像分析技术来容易地实现。因此,实施必要的步骤是特别容易的。在此上下文中,优选的是,识别和隔离具有至少0.25秒的长度的啼哭模式;优选地,被识别和被隔离的最小啼哭模式长度甚至更长,特别是至少0.3秒或0.4秒。在一个实际实施方式中,尽管通过使用监测的声音数据的类似谱图的表示在典型的现场安装中观察到任何背景噪声,特别是通过实施用于在监测的声音数据的5秒的窗的声级归一化类似的梅尔谱图表示中搜索具有至少0.4秒持续时间的啼哭模式的对象定位方法,具有0.4秒的最小长度的啼哭模式可以被可靠地识别。同时,模式的长度也可以被限制,例如,切除或完全排除具有长于2秒的长度的识别的和隔离的啼哭模式;将可以接受长于2秒的啼哭模式。然而,每个啼哭被分割成多个不同的啼哭模式,所述啼哭模式通常短于2秒。使用标准化的长度的啼哭模式用于其后续转化将意味着,识别的和隔离的啼哭模式的一个主要部分将必须在后续转化的一方被延长例如黑色肩部(black shoulder),并且可以理解这可以损害啼哭评估的精度。因此,使用不长于4秒——优选地不长于3秒并且特别地不长于2秒——的啼哭模式是非常优选的。使用较短的啼哭模式也帮助解决隐私的问题,特别是在仅短的啼哭模式或其表示而不是连续的声音流被上传到云中的情况下。

[0089] 即使当通过添加“沉默”的时段或对应的谱图模式来归一化啼哭模式的长度用于转化时,原始长度仍然可以被馈送到在婴幼儿啼哭的评估中使用的卷积神经网络或机器学习模型中。例如,在一个实施方案中,首先计算每个啼哭模式表示婴幼儿啼哭的特定原因的概率,并且然后根据收集的概率评估婴幼儿啼哭的总体原因,包括考虑的每个模式的长度也可以是有用的。

[0090] 在一个优选的实施方案中,提议连续地声学监测婴幼儿并且至少临时地存储啼哭前声音数据,直到已经发现后续声音数据不是啼哭相关的。在这方面,应注意,这对于根据声音数据确定低维参数以用于后续婴幼儿啼哭评估的情况并且对于简单搜索可以与婴幼儿啼哭相关联的模式并且执行这些模式的隔离的情况成立。例如,在婴幼儿啼哭音数据中搜索的典型模式具有例如大于2秒的音符的长度的情况下,对应于5秒的时段的多个帧可以被分组在一起,并且以使得例如通过确定谱图来在其中找到模式是特别简单的方式被处理。显然,这意味着,在5秒的时段内获取的数据首先需要被存储以用于这样的后续分析。此

外,应注意,由于啼哭模式绝不应被期望仅在一给定的时段(诸如5秒)内开始和结束,因此使用的步幅应是较低的,使得在时段之间存在重叠。步幅优选地应使得搜索到的每个模式完全位于一个序列的窗中的至少一个内。用短步幅,这容易保证。虽然这可以导致一些模式被识别为两个后续时段的情形,但是易于丢弃出现两次的模式;更详细地,应理解,时间戳技术可以被容易地用于丢弃由于时段的重叠而发现两次的这样的啼哭模式。应注意,当通过基于声音数据的类似谱图的表示、特别是基于声音数据中的男性类似谱图的表示的对象定位算法来实施啼哭模式检测和/或啼哭模式隔离时,可以出现如下情形:一个模式未被完全包括在获得类似谱图的表示之前基础声音数据被分割成的一个窗内。这易于注意,因为识别的模式将延伸到该窗的边界。即使在一系列重叠窗中的每个中以稍微不同的方式隔离同一个啼哭模式,但是非常可能的是一个窗中的一个不完整的啼哭模式将被再次发现在另一个窗中,在该另一个窗中啼哭模式然后是完整的。因此,优选的是,丢弃延伸一直到该窗的边界的任何模式,因为在转化不完整的啼哭模式的后续步骤中,通常仅获得低精度。即使在一个随后的窗中不可以识别不完整的啼哭模式的情况下,这几乎不对啼哭转化产生任何负面影响,因为在典型的情况下,多个啼哭模式将被识别并且被观察,使得省略不完整的模式将不具有任何显著的缺点。

[0091] 存储啼哭前数据大大减少了总体计算工作量,因为啼哭检测可以更好地与啼哭评估分离,而不降低啼哭评估准确性。在这方面,应注意,可以简化和/或可以在多步骤过程中进行啼哭的识别。通常,婴幼儿的啼哭将比任何背景噪声显著更响。优选地,这样的声级增加被用作啼哭检测的第一步。即使在云中执行啼哭检测的情况下,本地实施这样的第一步是优选的,因为它相当显著地减少了待被传送到云的数据的量,从而减少了啼哭评估所消耗的总能量。应理解,考虑到声级本地实施啼哭检测的第一步可以被非常容易地实施,例如使用简单的比较器。因此,不需要具有特别强大的本地处理器或微控制器。

[0092] 因此,第一个重要标准是一个样本或帧的绝对声级。不是使用绝对声级,还可以使用在一个给定的——相当短的——时段内增加声级,使得自动实现对允许的周围噪声的适配。应理解,背景/啼哭辨别可以依赖于人工智能/神经网络滤波技术,并且如果情况如此,可以并且优选地将使用与在实际啼哭评估中使用的滤波器不同的滤波器。

[0093] 如上文所陈述的,可以用极低的计算工作量完成声级的对应测试,因为这仅需要将当前声级(换言之,当前声音数据的二进制值)与预定义的或学习的阈值进行比较。然而,背景噪声(诸如狗吠、摔门声等)也可以导致显著高的声级。因此,一旦通过将当前声级与阈值进行比较或通过检测音量的突然增大而已经检测到显著高的声级,仍应确定突然较高的声级是否与突然的响背景噪声或与婴幼儿啼哭相关联。为此,记录啼哭前声音数据是有帮助的,因为这允许评价在已经超过阈值的那些数据之前立即记录的声音。相比于连续地检查多个状况——所述状况仅在组合时以足够高的概率指示婴幼儿正在啼哭,存储这样的啼哭前声音数据用于后续评价需要显著较少的能量。应理解,啼哭前声音数据不需要被存储特别长的时间,使得小存储器通常就足够。在此小存储器中,新数据可以被循环覆写在最旧的数据之上。还将理解,指示婴幼儿正在啼哭的线索也可以从非声学数据——例如从婴幼儿的视频监视——得出,指示移动或指示婴幼儿的表情对于啼哭的婴幼儿是典型的。

[0094] 设置阈值的一个优选的可能性是不断地测量数据流(诸如样本或帧)的连贯的片段的噪声级,例如通过确定一个帧的平均值。应注意,这将甚至可以使用模拟实施方式。此

外,代替在每个片段期间使用平均值,并且考虑到在每个片段内声级可能变化,这些变化的声级中的最小值可以被确定为背景水平。此背景水平可以被单独地或根据每个片段的多个背景水平来考虑,并且可以确定新的总体背景水平,诸如滑动平均背景水平。然后,可以根据相应的背景水平——例如仅考虑比过去的背景水平高至少 x dB的样本,其中 X 例如是6、12或18——来确定需要被超过以便假设婴幼儿啼哭的阈值。然而,应理解,在背景水平特别高的情况下,假设婴幼儿正在更响地啼哭是不合理的,因此需要超过的阈值(或实施例中的“ x ”)可能是总体声级的函数,因为不应期望婴幼儿由于背景噪声而特别响地啼哭。因此,如果周围环境具有更大背景噪声, X 通常变得更小。在此上下文中,应理解,婴幼儿啼哭的实际声级将取决于传声器与婴幼儿的距离以及婴幼儿本身;然而,通常将可以在距婴幼儿1至2米的距离处放置记录传声器;此外,尽管有一些变化,但是婴幼儿啼哭的总体声级可以被假设位于一个特定的有用范围内,特别是考虑到即使使用非昂贵的数字模拟转换器也常规地可实现的声级的分辨率。在背景噪声是极其响的情况下,在声音数据流中连续地搜索婴幼儿啼哭可能是明智的。这是合理的,因为在不可以将声级用作第一线索的情况下,应评估其他参数,诸如频率含量/谱图等。应注意,尽管婴幼儿啼哭识别的以上第一级是优选的,但是将存在其他可能性,例如使用固定阈值,以及使用最初确定的阈值或考虑到在与婴幼儿啼哭的具体原因等正相关的声音事件期间的声级而定期确定的阈值。

[0095] 在一个优选的实施方案中,还提议了,基于以下中的至少一个并且优选地至少两个、特别是至少三个来检测婴幼儿啼哭,特别是检测在连续的声学监测流中婴幼儿啼哭的发作:当前声级超过阈值,当前声级超过平均背景噪声一给定的裕度,一个或多个频带中的当前声级超过阈值,一个或多个频带中或一个或多个频率处的当前声级超过对应的平均背景噪声一给定的裕度,声音的时间模式,包括不仅来自时域而且来自频域的声学特征的模型。换句话说,可以建立声音流的时间模式和/或谱模式,并且可以考虑相应的模式作出决定。应注意,可以以使得可以使用常规图像分析技术来作出关于模式的决定的方式处理声音数据。在通过与平均背景噪声进行比较来检测啼哭的情况下,例如可以在前面的五秒、10秒、20秒、30秒或一分钟内对背景噪声进行平均。

[0096] 应理解,可以建立必须共同满足的多个状况以考虑婴幼儿正在啼哭。例如,如果声能的谱分布对应于婴幼儿啼哭的声能的典型谱分布并且如果它是足够长的,仅响的噪声被认为啼哭。由于对于不同状况计算工作量将是不同的,具有多步骤/多级啼哭识别是合理的,其中识别步骤需要连续地运行的最少的计算工作量,并且仅在连续地运行识别步骤指示已经发现需要更详细分析的声音模式的情况下,才执行剩余的和/或附加的识别步骤。应理解,以此方式,能量消耗可以是相对低的,这在完全或至少相对于电池操作的装置上的初始部分执行所述方法的情况下是特别有利的。此外,应理解,即使在应用时被认为缓慢的处理装置——诸如DSP、FPGA、微控制器、微处理器上,任何啼哭识别步骤也可以被足够快速地执行。因此,尽管多步骤啼哭识别方法,但是等待时间(latency)将是可忽略的。换句话说,这不将在啼哭评估中引起明显或显著的延迟。无论如何,典型的等待时间是如此之低,以至于它将是容易可接受的,因为不存在待被建立的直接通信,例如在成人之间的视频或电话通信中。

[0097] 鉴于此,提议了,在一个优选的实施方案中,基于以下的至少一个来检测婴幼儿啼哭,特别是检测在连续的声学监测流中婴幼儿啼哭的发作:当前声级超过阈值,当前声级超

过平均背景噪声—给定的裕度,一个或多个频带中的当前声级超过阈值,一个或多个频带中的当前声级超过对应的平均背景噪声—给定的裕度,声音的时间模式和/或谱模式,优选地通过多步骤/多级啼哭识别来决定声音数据流中是否存在婴幼儿啼哭,其中需要较少计算工作量的一个识别步骤或多个识别步骤连续地运行,并且仅在连续地运行一个或多个识别步骤指示已经发现需要更详细分析的声音模式的情况下,才执行剩余的和/或附加的识别步骤。

[0098] 应理解,在已经识别婴幼儿啼哭或更精确地说婴幼儿啼哭的发作或声称的婴幼儿啼哭的情况下,存在数个可能性。首先,将可能在检测到啼哭之后评估多个帧或指定的时段中的每个以确定婴幼儿是否仍然在啼哭;这可以例如独立于当前声级来完成;以此方式,来自婴幼儿需要呼吸的那些时段的数据也将被分析,因为呼吸声音也可能为婴幼儿啼哭的原因给出重要线索。(应注意,在其内执行啼哭模式搜索的一个时段或窗中的样本的数目可能不同于在帧中被分组在一起的样本的数目,以例如获得可以被容易地传送到云服务器的文件;例如,为了文件的传送,应以允许简单错误校正的方式将一定数目的样本分组在一起,但是此数目通常应是较小的并且分组在一起的多个样本形成一个窗。

[0099] 在从声音数据提取少量参数以用于评估和/或用于检测声音流中的啼哭的一个实施方案中,可以使用计数器,该计数器计数一直到在已经检测到啼哭事件或声称的啼哭事件之后需要被获取并且被分析的帧的最小数目。此外,在分析声音流中的时段以确定是否已经发现模式的情况下,这样的时段通常将由多个帧组成并且因此也应提供计数器。

[0100] 应注意,优选的是监测婴幼儿是否(仍然)啼哭,使得在进一步传入的声音数据中,还应与计算与婴幼儿啼哭的原因的个性化评估相关的参数并行地,和/或与识别和隔离通常与婴幼儿啼哭相关联的模式并行地,搜索响的噪声。应注意,为了检测啼哭是否继续,可能没有必要要求记录的一个帧的平均声级显著响于先前的帧的声级,但是声级不应下降到一个给定的最小值以下。因此,一旦进一步分析声称的啼哭,就可以实施类似滞后的行为。

[0101] 如果这样做了,无论何时证实婴幼儿仍然正在啼哭,计数器可以被重置。此方法确保,在父母和/或护理者在已经停止啼哭之前没有对婴幼儿的啼哭作出反应的情况下,啼哭的结束阶段被完全记录。这对于证实已经安静下来的婴幼儿是否应被单独留下可以是有帮助的。然而,还将可以仅考虑已经证实在记录帧中的样本期间婴幼儿正在啼哭的那些帧。实现这一点的一个特别有用的方式是,在声音流中搜索通常与婴幼儿啼哭对应的模式。这样的搜索可以例如通过获得声音小段(snippet)或时段的二维表示来实现,例如通过确定指示随时间的声音时段(诸如5秒声音时段)的频率含量的谱图。然后通过使用已知与婴幼儿啼哭相关的模式训练人工智能模型来实现搜索通常与声音流中的婴幼儿啼哭对应的此模式,使得模型的输出将是谱图中与婴幼儿啼哭最好地对应的那些部分。例如,可以选择具有与婴幼儿啼哭相关的可能性大于50%、60%、70%、75%、80%、90%或95%的声音流的一部分。此外,没有必要隔离声音流的谱图(或声谱图(sonogram)、声纹或录音电报(voicegram))或类似谱图的表示的实际部分;相反,识别啼哭模式的起始时间将是足够的——并且,在评估中考虑具有不同长度的啼哭模式的情况下,还将识别啼哭模式的时间。这在识别的模式类似评估然后在评估级中被实现的情况下特别有帮助,因为可能期望的是使用不同于用于模式转变的分辨率的用于模式识别和隔离的分辨率,特别地,对于模式转变,频率分辨率和时间分辨率二者可以是更低的。在相当一小部分的声音流因与婴幼儿

啼哭不相关而被抛弃的情况下,可以产生计算工作量的减少。相比之下,在声音流的一个相当大的部分被选择用于转化的情况下,从计算角度来看,使用与用于稍后的模式转化的时间和/或频率分辨率相同的时间和/或频率分辨率来搜索啼哭模式可以是更有用的。然而,应注意,在其中实现搜索啼哭相关的模式的窗或时段应具有相当大的重叠,这进而意味着,啼哭相关的模式——特别是短的啼哭相关的模式——可以在前一时段的结束时和后一时段的开始时都可以被发现。在确定使用与用于模式转化的时间和/或频率分辨率不同的时间和/或频率分辨率来实现搜索啼哭模式是否有利时,这应被纳入考虑。

[0102] 在任何情况下,优选的是使用时间戳,使得可以确定啼哭的任何中断的长度——例如因为婴幼儿正在喘气或吸气,使得其可以被用于改进啼哭模式的转化或评估。

[0103] 应理解,对一般的背景进行采样和/或分析可以是有利的,所述背景特别是在其中观察到平均声级的背景,并且因此短的、脉冲式的响的声音——诸如摔门声或狗吠声——将不会不利地影响对背景的分析。背景分析可以用于确定对于啼哭评估最有用的参数,并且也可以帮助确定用于第一啼哭检测级的阈值。还应理解,由于背景噪声或因为监测传声器必须被放置得离婴幼儿太远而在实际环境中不能够恰当的测量在极其安静的环境中将最好地评估婴幼儿啼哭的原因的某些参数。在这样的情况下,原本可获得的相关信息将被埋在噪声之下并且应选择用于评估婴幼儿啼哭的原因的其他参数。

[0104] 由此,可以看到,与声学背景相关的信息在确定最好参数中和/或用于啼哭识别中是非常有帮助的,特别是因为为了个性化评估,应提供非常高质量的数据。应理解,考虑到由于传声器的放置和/或传声器的特性引起的变化,在可以使用用于实施所述方法的各种不同的装置(例如不同的智能电话)的情况下,参数的选择应考虑到参数的“稳定性”。

[0105] 对于将随着婴幼儿和监测传声器之间的距离改变的参数,诸如啼哭的总声级,这可以是容易理解的;然而,其他因素也具有影响,所述因素是诸如婴幼儿是否被放置在摇篮里、在房间里是否关闭窗帘并且因此更高的频率受到更高的吸收、传声器灵敏度极性模式看起来像什么——例如心形、超心型指向(hypercardioid)、超心型(supercardioid)、亚心形(subcardioid)或单极性——以及其如何朝向婴幼儿定向等。应理解,这不仅影响到从声音流得出的一些低维参数,而且还可以不利地影响使用谱图或类似谱图转化的声音流中的啼哭模式的识别和隔离。因此,在AI模型时,非常优选的是依赖不只一个设备/传声器获得的数据,而是使用包括用多种不同的装置获得的声音样本的训练集。具体地,为了建立训练集,可以用多个装置同时记录相同的声音。这也是有帮助的,因为这些多个装置将不被完全同步,使得与同一个婴幼儿啼哭相关的声音模式在它们各自的5秒时段内将具有不同的起始时间。此外,应理解,即使放置在完全相同的位置处用于记录完全相同的声音,记录的声音也将根据装置改变。这些变化有很多原因,例如关于传声器灵敏度、传声器响应、在数字化之前用于调节模拟信号的放大器的响应等的变化。应理解,甚至将可以使用多个装置来合成训练集,所述多个装置记录先前记录的(优选地使用高质量的传声器装置先前记录的)多个婴幼儿啼哭的回放。

[0106] 特别是在需要确定神经网络滤波器用于基于仅少量参数对啼哭进行个性化评估的情况下是有帮助的,以还考虑声学背景的行为。因此,将一些非啼哭背景声音模式上传到服务器可以是有利的,使得典型的背景模式也可以被考虑到,特别是对神经网络滤波器/神经网络滤波器参数的评价和确定。应理解,虽然这里通常参考本申请的神经网络滤波器或

神经网络参数,但是也可以参考分类、分类模型等;这将不被认为是在以描述这样的技术使用的措辞所实施的技术和方法中的差异。

[0107] 特别优选的是,如果啼哭数据——根据其确定允许啼哭评估的参数——包括自啼哭事件的发作起的声音数据,特别是自啼哭的最初两秒起的声音数据,优选地自啼哭的前两秒起的声音数据,特别优选地自啼哭的最初500毫秒起的声音数据。以自动方式进行婴幼儿正在啼哭的确定的情况是容易可能的;考虑到啼哭事件的发作在评估中可以是有帮助的,因为啼哭本身可以增加不舒服,例如因为由于婴幼儿的需要长时间内保持应答的事实而感到附加的压力,和/或如果啼哭继续延长的时间则啼哭本身会使婴幼儿筋疲力尽。此外,如上文所指示的,在考虑参数的变化——诸如共振峰的第一频率的变化——的情况下,最初的变化可以包含特别有价值的信息。应注意,在使用卷积或神经网络或其他人工智能方法从声音流时段隔离啼哭模式的情况下,这经常是在使用非常简单的电路系统(例如模拟或数字比较器)进行第一声级评估之后完成的。在这样的情况下,还优选的是,评价响的噪声的实际发作;因此,应存储声音数据,优选地使得使用比较器可检测到的响的噪声将靠近在其中执行搜索正确模式的窗的结束,例如4、5、6、7、8、9或10秒的窗。

[0108] 应注意,在许多情况下,有利的是,经常更改评估婴幼儿啼哭的方式,例如通过考虑到婴幼儿的成长和发育,经常更改用于评估婴幼儿啼哭的神经网络滤波器中的滤波器系数。这可以是有利的,例如在出生后很早的时候,因为新生婴幼儿的声音特性快速变化;此外,可能是快速变化的医疗条件(诸如升高温度)强烈影响到将评估数据的方式,使得为了有用的个性化,神经网络滤波器的滤波器系数也应经常变化。根据个性化的确切实施方式,本地实施个性化评估步骤的执行可能是不可行的,因为这将需要用于不同的滤波器系数的相当大的存储器,和/或因为当前合适的滤波器系数必须被识别并且被下载;因此,评估有时将优选地在集中式服务器上和/或云中执行。

[0109] 因此,在一个优选的实施方式中,还提议,实施从本地声学监测的婴幼儿获得的声音中进行啼哭检测的全部或一些(例如至少一个)步骤,并且将数据上传到在集中式自动婴幼儿啼哭评估中使用的(云)服务器装置,特别是上传用于在云中评估婴幼儿啼哭的数据。即使啼哭检测的简单的第一步——诸如阈值比较器——将帮助减少需要被上传的数据流,从而节省能量和带宽。应注意,甚至在其中在云服务器中执行对婴幼儿啼哭或其大部分的实际评估的情况下,这样的啼哭检测的第一简单的本地步骤因此是优选的。还应注意,没有必要将声音上传到分散式远程云服务器。例如,可能存在如下情况:靠近婴幼儿放置的装置具有极其有限的计算能力,并且被用来通知父母婴幼儿正在啼哭的装置是智能电话,该装置如今至少具有至少计算能力来实现评估。在这样的情况下,不是将声音上传到远程设置的云服务器,而是也可以将来自靠近婴幼儿放置的装置的声音上传到智能电话用于进一步评估,从而减少父母可能具有的任何隐私问题。

[0110] 然而应注意,虽然显然优选的是至少在某种可能程度上本地确定当前是否记录了啼哭事件以节省带宽,否则需要连续地将声音数据上传到啼哭识别级,但是通过考虑到婴幼儿的特性中的至少一些——特别是与个性化最相关的特性(诸如性别、年龄、体重和大小)——已经可以获得改进的结果,换言之,甚至不需要一个考虑其他个人数据——诸如当前温度、当前医疗条件、自婴幼儿被喂食起的时间等——的完全个性化。由于年龄、体重和大小仅缓慢变化,因此个性化评估也可以本地实施,特别是对于声音数据的上传被损害的

情形。应理解,即使在这样的情形下,非常优选的是在记录婴幼儿声音的装置和(云)服务器之间建立连接,使得用于个性化婴幼儿啼哭评估的神经网络滤波器可以被经常地更新。此外,在连接的时段期间,本地收集的数据可以被上传到服务器,并且可以下载新的滤波器或可执行指令以根据记录的声音数据评估婴儿啼哭的原因。

[0111] 应理解,为了改进婴幼儿啼哭的个性化评估,优选的是,将与婴幼儿啼哭的声学监测相关的数据和/或与允许啼哭评估的选择的啼哭数据相关的参数上传到云和/或中央服务器。在本地可用的计算能力足以实现啼哭识别的情况下,至少在带宽不足的情况下,可以优选的是不上传任何数据;但是,由于当上传大量的模式时声音样本的数据库增大,优选地提供来自确认或不同意先前的评估的反馈的标签,并且由于可以使用标记的样本的扩大的数据库来重新训练模型,因此非常优选的是尽可能多地上传与啼哭模式相关的样本。

[0112] 因此,为了改进个性化评估,啼哭中的至少一些和/或从啼哭得出的参数可以与相应的个人婴幼儿数据一起被存储在服务器上,以允许根据存储在服务器上的信息进行个性化评估。应理解,在例如已经为特定装置获取订阅的新的滤波器的情况下,以及在父母或护理者已经最初指示婴幼儿的年龄和另一些细节的情况下,仅传输装置的ID可以是足够的。然而,由于其他参数——诸如婴幼儿的体重和大小——也应经常地被更新,优选的是,要求父母和/或护理者定期录入对应的信息。应理解,录入这样的信息可以尤其使用诸如运行合适的应用程序的智能电话的单独的装置和/或通过使用本地或集中式语音辨识允许用户经由语音向装置录入对应的信息来完成。

[0113] 如从上文可以看到的,在一个优选的实施方案中提议,所述方法包括从集中式服务器下载允许本地个性化婴幼儿啼哭评估的信息的步骤。考虑到婴幼儿成长和变得越来越大,在一些时间之后,个性化滤波器将不再给出最有利的结果。因此,将本地个性化婴幼儿啼哭评估的使用限制到特定的时间是可能的并且有用的。一旦这样的时段已经过去,可以发出个性化不再是可靠的警告,和/或可以使用标准的、非个性化的滤波器,和/或可以向用户发出要求续订滤波器而不是指示婴幼儿啼哭的原因的信息。在如下特别的情况下:在该情况下,父母或护理者已经订阅了滤波器的定期更新并且这样的滤波器已经在延长的时段内未被更新,例如因为到集中式服务器的连接已经被损害和/或被阻止,在总体上停止使用个性化的滤波器之前和/或在仅以非个性化方式实现评估之前的某个时间可以生成警告。此外,明显的是,个性化评估也可以在云服务器中实现。

[0114] 如上文所指示的,在一个优选的实施方案中,提议了,在啼哭的发作之前获取的声音数据(也)被用来确定声学背景和/或确定用户婴幼儿啼哭评估的附加的参数。关于确定用于婴幼儿啼哭评估的附加的参数,可能出现如下情形:不能够以足够高的概率确定啼哭的确切发作,例如由于与响的声学背景的一致。然而应注意,相比于仅通过依赖于从声音流得出的仅少量参数,通过依赖于谱图等用于模式识别,可以更显著可靠地隔离啼哭相关的声音数据。因此,啼哭时间或啼哭模式的基于谱图的识别和隔离对抗声学干扰明显更鲁棒,并且应理解,这非常显著地帮助获得更好的啼哭转化结果。

[0115] 应理解,然而,啼哭的确切发作可能不可以足够高的概率确定,并且对附加的(优选地先前的)帧的评价可能有助于评估。这可以优选地通过以下方式完成:通过评估滑动参数和/或通过交叉关联技术或通过分析时间段,例如用于在响的噪声超过阈值之前识别和隔离声音流中的啼哭模式的1、2或3个标准化的时间段。此外,在响的噪声之后检测到婴幼

儿啼哭的情况下,婴幼儿更可能需要被安慰;因此,这样的事件在评估中可以是有益的,即使它们不被认为是需要从声音数据中减去或滤除掉的背景模式。

[0116] 关于被认为已经检测到啼哭的发作而不需要评价在啼哭的发作之前获取的声音数据的最小概率,首先应理解,在多级啼哭检测的典型情况下,可以确定这样的可能性或概率,并且如果这样的概率高于例如70%、80%、90%、95%或99%,则作出已经检测到啼哭的假设,认为概率足够高的确切阈值尤其取决于背景噪声的模式和/或多级啼哭检测的质量。

[0117] 考虑到申请人已经达到的当前标准,在第一时间内在一个帧中已经检测到啼哭并且因此已经检测到啼哭的发作的概率容易超过99%的概率。然而,可以设置较低的阈值,诸如97%、95%、90%或80%。应注意,即使已经实现了准确确定啼哭的发作的极高概率,仍然可以将先前的帧与在(非常可能的)啼哭的发作之后记录的那些帧一起馈送到婴幼儿啼哭评估级。这在其中在评估中使用诸如交叉关联的技术的情况下可以是特别有帮助的。甚至可以根据概率来确定应被馈送到啼哭评估中的啼哭的假设发作之前的帧的数目,例如通过公式(以%计的 $100 - \text{概率}$) $\times A$ ——其中a是0,5或1或1,5或之间的任何数字——来确定先前帧的数目;显然,从这些公式获得的先前的帧的数目四舍五入到下一个较大的整数。

[0118] 应理解,上文的提供用于评估的数据的方法在现场环境中——换言之在声音或音频实验室之外——是特别有帮助的。在现场,合适地准备允许增加啼哭评估的准确性的数据是特别重要的。例如,在典型的实验室设置中,声音将具有干净的、低噪声的背景并且啼哭可以被清晰地记录。相比之下,在典型的现场环境中,背景噪声将明显更高,啼哭的声音音量将变化得更强,并且由于不太最佳的传声器定位,记录例如相对于高频内容不那么“清晰”。这些差异通常导致现场中的准确性显著低于实验室环境中的准确性。然而,通过使用交叉关联技术和/或滑动平均值,尽管存在显著的噪声,在现场识别啼哭的发作和/或识别啼哭本身的准确性变得绝对比得上在实验室环境中可获得的准确性。此外,应注意,关于转化的最终精度,通过使用与声音时段的类似谱图的变换相关的类似图像处理的人工智能方法来识别和隔离啼哭模式,可以获得显著更好的结果。

[0119] 尽管如此,应理解,在实验室和在现场二者中可获得的并且确定的绝对准确性可能仍然取决于例如使用的样本、实际评估的质量(例如由神经网络滤波器表示的)、记录的长度或对“准确性”的度量(measure)的定义和数学确定。因此,通过不同的方法确定的准确性不容易被比较。通常,准确性将被定义成使得方法在实验室中给出高于90%的准确性。此准确性在现场通常将显著地下降。

[0120] 尽管如此,使用同样的方法,如果向评估级提供适当的数据,例如允许考虑滑动平均值和/或交叉关联的数据,或在声音时段的类似谱图的表示中搜索模式,则在现场的总体准确性不必再从实验室中的 $>90\%$ 下降到在现场的不到80%。应注意,类似谱图的表示可以是标准的谱图或可以不同,因为频率的分辨率在音频谱上是不同的,和/或对于不同的频率可以具有不同的动态范围。

[0121] 从上文,应已经理解,优选地,参数以允许使用神经网络、卷积神经网络和/或其他已知的人工智能技术评估啼哭的方式被馈送到啼哭评估级。应注意,通常这样的技术依赖于多个所谓的“层”,并且如果至少一个这样的层是个性化的则可以实现个性化;然而,还将可以通过使不止一个层个性化来使啼哭评估个性化,换言之通过根据与婴幼儿相关的个人数据为一个或多个卷积层选择例如不同的滤波器参数。在多个层被个性化的情况下,可

以使每个层完全个性化,例如通过根据性别、年龄和体重选择不同的滤波器参数;相比之下,还将可以在第一个个性化层中仅根据性别选择不同的滤波器参数,在第二个个性化层中仅根据体重选择不同的滤波器参数,并且在第三个个性化层中仅根据年龄选择不同的滤波器参数(其中第一、第二和第三个个性化层不一定按该顺序处理数据)。对于本领域技术人员来说明了的是,当多个层被个性化时,还将可以使它们完全地或部分地个性化,例如使用例如1、2、3或更多个人参数使第一层个性化并且使用2、3、4或更多个人参数使第二层个性化,所述参数可以或可以不与用于使第一层个性化的个人参数部分地重叠。

[0122] 在如下计算机实施的多级啼哭评估方法中:所述方法具有例如基于对象定位方法的用于啼哭检测和啼哭模式隔离的一个或多个级,该一个或多个级基于对原始声音数据进行分割的窗的重叠序列的类似(梅尔)谱图的表示来搜索至少0,4秒、并且优选地短于3秒的长度的啼哭模式;所述方法还具有用于对隔离的啼哭模式进行分类的一个或多个级,该一个或多个级优选地还基于啼哭模式的二维表示,例如与用于啼哭模式检测和隔离的表示相比,具有不同的时间和/或频率分辨率的表示,和/或通过以下方式被归一化:通过将每个啼哭模式中的最大声级设置为特定值和/或通过短补将短的啼哭模式补充一直到期望的长度,并且确定被隔离并且被认为属于多个不同类别(或婴幼儿啼哭的原因)中的一个的每个啼哭模式的概率;以及另一个“收集器”级,其评估根据啼哭模式的序列确定的概率的序列,可以实施每个级与其他级分离,例如以如下方式:第一级靠近婴幼儿被执行,第二级在云中被执行并且第三级在父母或护理者的智能电话上被执行;在这样的情况下,容易看到,一个级——优选地最后的收集器级和/或为每个啼哭模式分配概率的级——的个性化将足以获得婴幼儿啼哭的总体个性化评估。然而,不需要使不同的级分离并且将很可能将计算机实施的多级啼哭评估方法以如下方式来实施,使得不同的级以及接口连接的(interfacing)步骤——诸如重新归一化长度和声级——以如下方式被执行:用户不会注意到不同的级,而是将整个评估看作唯一的过程。这可以通过将用于在同一个地点处执行的各级连接在一起来实现,例如通过将整个声音流上传到云中并且仅接收来自云的对啼哭的评估。尽管如此,这仍然可以被认为多级啼哭评估。应明确地注意,提供啼哭模式检测和隔离,啼哭模式序列的啼哭模式概率评估,以及对从评估概率获得的概率的个性化自动集体评估被认为是创造性的,特别是当使用对象定位用于啼哭模式检测和隔离时。

[0123] 还应理解,在一个优选的实施方式中,参数和/或记录的声音的数据流将连同婴幼儿数据信息被上传。应理解,在要加强现有婴幼儿啼哭数据库的情形下上传整个声音记录是优选的,而在其他情况下,仅上传提取的参数是优选的,因为需要传输较少的数据,从而特别是在数据传输带宽是低的情况下更快的响应。在此上下文中,应理解,在对数据进行人工智能评价时,重要的步骤中的一个降维。例如,如果考虑到声音数据的组块包括128个连贯的16比特样本的64个帧,则初始空间是 $(64 \times 128 \times 16 =)$ 131072维。为了处理此,可以确定诸如上文列出的参数,诸如平均声级、第一共振峰频率的变化等;(作为一个替代方案,可以使用谱图源类似谱图的表示)。现在,从上文可以看到,在期望基于一些参数而不是二维的、类似谱图的表示来评估婴幼儿啼哭的情况下,存在可以在描述和评估婴幼儿啼哭中使用的大量不同的参数;此大量不同的参数通常通过仅选择最相关的参数而进一步减少。

[0124] 在婴幼儿啼哭的个性化评估中,识别在来自其他婴幼儿的声音数据中也发现的模式,并且搜索最好地描述这些模式的参数集。在使用基于类似谱图的表示的技术的情况下,

在啼哭模式的类似谱图的表示中的这样的识别可以基于类似图像的分析——其将待被评估的啼哭的类似谱图的表示与已知的啼哭相关的啼哭模式进行比较。已知的啼哭模式可以从大的数据库选择,使得仅来自具有类似的年龄、性别、体重、医疗条件等的婴幼儿的啼哭被分组;在这样的情况下,用于模式的比较的滤波器将具有明显的个性化,因为每个类似(同龄)群体的啼哭将与属于不同同龄群体的婴幼儿的啼哭稍微不同,即使婴幼儿因相同的原因。然而,当数据库未大到足以建立多种同龄群体时,也将可以在第一步骤中对所有婴幼儿使用完全相同的滤波器参数来评估由一个或多个层记录的原始声音流中识别和隔离的啼哭模式。由于啼哭将包括多个啼哭模式,因此在此第一步骤中,将产生婴幼儿啼哭的多个可能的原因,其中原因或原因的可能性被分配给每个啼哭模式。然后,可以在另一个步骤中评估婴幼儿啼哭的多个这样的可能的原因或可能性,其中此步骤然后被个性化。应注意,通过以这样的方式处理数据,用于个性化的个性化滤波器系数的数目显著低于在使将啼哭的原因的合理的可能性分配给每个单个啼哭模式的第一步骤个性化时。

[0125] 应注意,在为了个性化而建立婴幼儿的不同(“同龄”)群体的情况下,可能出现以下情形:在所述情形中,不同的参数集对于每个不同的婴幼儿群体可以是最好的。期望的是,减少用以确定参数的计算工作量并且因此选择足以进行个人评估的小参数集。然而,如果仅将有限数目的参数或甚至更差地仅将减少的参数集传输到服务器,而不是传输完整的语音流或不是传输与啼哭模式或其类似谱图的表示相关的语音流的至少小部分,则对新颖模式的识别可能被损害。因此,至少为了从附加的数据识别新颖的模式,优选的是传输完整的语音数据——或完整的提取/分离的啼哭数据——而不是仅传输从其提取的参数。

[0126] 在一个优选的实施方案中,本发明的计算机实施的方法包括上传参数和/或记录的声音的数据流和/或其片段和/或与这样的片段相关的信息(诸识别的和隔离的啼哭模式)连同婴幼儿数据信息,特别是与年龄、性别、大小、体重、种族渊源、单胞胎/双胞胎/三胞胎、当前医疗状态、已知医疗先决条件、特别是已知的当前疾病和/或发烧、父母和/或护理者的语言中的至少一个、优选地至少两个、三个或四个相关的婴幼儿数据信息。

[0127] 应理解,诸如出生日期、单胞胎/双胞胎/三胞胎的信息不需要在每次从本地装置向服务器或云传输数据时被传输。然而,由于需要婴幼儿数据信息中的一些,至少可以传输允许识别本地装置并且可与对应的必要婴幼儿数据相关联的信息,例如本地使用的装置的ID;这样的情况,实际的婴幼儿数据可以在个性化评估之前被独立地传输,被存储在云或服务服务器上并且根据传输的信息(诸如本地使用的装置的ID)被检索。应理解,在此上下文中,如果父母使用应用程序、网站表格或等注册婴幼儿或装置,输入相应的婴幼儿数据,这是足够的。

[0128] 此外,如果与一个或多个先前的评估的准确性相关的(反馈)信息被上传到服务器,这是优选的。这可能帮助重新校准机器学习模型中使用的滤波器(或分类)和/或消除先前的错误。再次,在不需要对当前啼哭事件进行评估的情况下,与先前的评估的准确性相关的信息可以被上传。优选的是,传输与一个或多个先前的评估的准确性相关的信息连同与啼哭事件相关的数据,诸如啼哭事件ID、装置id+时间标签等;此外,这可以例如是实际的自动评估、父母或护理者对评估的反馈以及附加的信息——诸如对应的婴幼儿啼哭参数和/或声音原始数据——的组合,特别是在评估被判断为不良的情况下;代替原始数据,先前评估的啼哭事件的时间戳也可以被传输,所述先前评估的啼哭事件优选地是如下先前评估

的啼哭事件:对于所述啼哭事件,数据已经被传输并且一直被存储在集中式服务器(诸如云服务器)上。是优选重新传输根据声音数据确定的声音数据或参数,还是只传输ID或时间标签,将尤其取决于服务器上的存储空间。也可以仅仅传输指示评估总体正确的频率的统计数据,或指示诸如“婴幼儿希望被安慰”或“婴幼儿需要拍嗝”的具体评估是正确或错误的频率的统计数据。使用关于评估的统计信息允许向不同的用户(尽管与同一同龄群体中的婴幼儿相关)提供不同的评估算法/滤波器或使用不同的滤波器和/或算法的评估结果,并且然后以统计方式评价不同的评估。这在用户的群体是足够大的情况下是特别有帮助的。

[0129] 应理解,不同的信道和/或不同的时间可以被用来传输不同种类的数据。

[0130] 提供的用于个性化评估的数据优选地根据提取的啼哭数据来确定,使得评估允许区别“婴幼儿疲倦”、“婴幼儿饥饿”、“婴幼儿不舒服并且需要照顾”、“婴幼儿需要拍嗝”、“婴幼儿疼痛”中的至少一个。在传送参数的情况下,这样的参数被优选地选择并且被提供使得可以区别和识别至少两个、特别是至少三个并且特别是所有不同的状况。可以估计的是,一旦足够大的数据库是可用的,可以很好地识别某些医疗状况,诸如“婴幼儿具有反流(reflux)”、“婴幼儿具有胀气”、“婴幼儿具有中耳炎”,或更具体的不舒服的原因,诸如“婴幼儿太热”、“婴幼儿太冷”、“婴幼儿无聊”。在此上下文中,应理解,在本发明中提议的提供用于个性化评估的数据的方法也非常有助于扩大婴幼儿啼哭的现有数据库,因此有助于改进婴幼儿啼哭评估。因此,通过适当地实施本发明,可以扩大啼哭的数据库以允许在短时间内评估的高度精细的个性化。

[0131] 虽然上文所描述的方法可以使用多种装置和/或系统来实施,但是特别寻求保护的是一种自动婴幼儿啼哭评估装置,所述自动婴幼儿啼哭评估装置包括:传声器,用于连续地声学监测婴幼儿;数字转换级,用于将监测声音流转换成数字数据流;存储器级,用于存储个人婴幼儿数据信息;通信级,用于将数据传输到集中式服务器装置和用于指示评估的结果的指示装置,诸如用于声学指示评估的结果的扬声器装置、显示器和/或到显示器的接口;设置用于识别数字数据流中啼哭的发作的啼哭识别级,并且其中所述通信级被适配用于向集中式服务器装置传输与用于根据个人婴幼儿数据信息而进行评估的啼哭相关的数据,并且从所述集中式服务器装置接收与婴幼儿啼哭的个性化评估相关的数据。啼哭评估装置可以如下方式实现:用于声学或视觉指示评估的结果的显示器和扬声器装置被布置在与包括传声器的装置分离的装置中。

[0132] 应理解,一个或多个级,特别是用于识别数字数据流中啼哭的发作的啼哭识别级可以通过硬件和软件的组合来实施。此外,从集中式服务器装置,可以接收用于对自动识别的啼哭的本地评估的个性化的滤波器,或,在考虑到声音数据获得的一些或全部参数被传输到集中式或云服务器用于评估的情况下,可以接收评估的结果。

[0133] 在一个优选的实施方案中,提议了,所述自动婴幼儿啼哭评估装置包括反馈装置,用于获得与一个或多个先前的评估的准确性相关的反馈信息,并且通信(或I/O)级被适配用于将反馈信息传输到集中式服务器装置。在一个优选的实施方案中,所述反馈装置被集成到用于声学监测婴幼儿的装置中;这帮助确保反馈的高质量。

[0134] 此外,在一个优选的实施方案中,所述自动婴幼儿啼哭评估装置将包括本地评估级,所述本地评估级被适配以根据从所述集中式服务器装置接收的与婴幼儿啼哭的个性化评估相关的数据来评估婴幼儿啼哭。本地评估可以是允许在没有声音数据可以被传输到集

中式服务器装置的情况下进行评估的辅助评估级,或可以是主要的或唯一的评估级,其中生成向父母和/或护理者指示的所有评估。

[0135] 上文已经陈述的是,婴幼儿啼哭数据的个性化评估取决于诸如婴幼儿的年龄、大小和体重的因素,所述因素将随着婴幼儿年龄的增长而显著变化;这导致个性化评估可能变得过时的事实。为了防止使用过时的滤波器尝试个性化评估,应进行对应的检查。因此,优选的是,如果自动婴幼儿啼哭评估装置包括定时器和评价级,该评价级:在婴幼儿啼哭的评估之前,评价从所述集中式服务器装置接收的并且与婴幼儿啼哭的个性化评估相关的(滤波器/算法)数据的:个人婴幼儿数据信息的当前年龄和/或年龄或有效性,所述婴幼儿啼哭评估装置被适配以根据所述评价输出婴幼儿啼哭评估。

[0136] 现在将通过实施例的方式并且参考附图来描述本发明。在附图中,

[0137] 图1a示出了婴幼儿啼哭的评估中的步骤的序列,其中这些步骤中的一些实施本发明的一个实施方案;

[0138] 图1b示出了啼哭检测/数据预处理的一部分;

[0139] 图2示出了可以被用来指示婴幼儿的当前需要的多个符号;

[0140] 图3a-图3e示出了从表示指示不同需要的婴幼儿啼哭的多个音频记录提取的三维谱图——时间沿X轴增加,频率沿Y轴增加并且强度沿Z轴增加。单位是任意的但是对所有部分是相同的。

[0141] 图4a-图4f示出了谱图的比较,以使对于不同的啼哭、对于多个频率强度随时间的变化可视化;

[0142] 更详细地,

[0143] 图4a与来自不同饥饿婴幼儿的啼哭相关

[0144] 图4b与来自同一饥饿婴幼儿的不同啼哭相关

[0145] 图4c与来自不同疼痛婴幼儿的啼哭相关

[0146] 图4d与来自处于疼痛中的同一婴幼儿的不同啼哭相关

[0147] 图4e与来自需要拍嗝的不同婴幼儿的啼哭相关

[0148] 图4f与来自需要拍嗝的同一婴幼儿的不同啼哭相关。

[0149] 图5a-图5f示出了不同啼哭的聚类,其中

[0150] 图5a示出了啼哭的总体聚类

[0151] 图5b示出了总体聚类中的饥饿啼哭

[0152] 图5c示出了总体聚类中的“困倦”啼哭

[0153] 图5d示出了总体聚类中的“需要拍嗝”啼哭

[0154] 图5e:总体聚类中的不舒服啼哭

[0155] 图5f示出了总体聚类中的疼痛啼哭。

[0156] (应理解,二维图形中的聚类的区分并不像考虑附加的区别参数那样完整;然而,变得明显的是,即使在所示出的二维图形中,聚类开始出现)。

[0157] 图6从两个不同的视角示出了T-SNE降维梅尔谱图(melspectrogram)的三维表示。

[0158] 图7示出了K-均值聚类,其中用于5个不同标签的每个聚类的形心被绘制为白色叉形记号,并且被划分为不同的单元(cell,小区)。

[0159] 图1例示了在婴幼儿啼哭评估中有用的步骤,为了婴幼儿啼哭评估,执行一种提供

用于自动婴幼儿啼哭评估的数据的计算机实施的方法,所述提供用于自动婴幼儿啼哭评估的数据的计算机实施的方法包括以下步骤:声学监测婴幼儿并且提供对应的声音数据流,在所述声音数据流中检测啼哭或啼哭的一部分,响应于检测到啼哭或啼哭的一部分而从声音数据中选择数据,确定用于个性化啼哭评估的个人婴幼儿数据,根据个人婴幼儿数据准备用于评估的评估级,处理用于啼哭评估的选择的数据,以及将经处理的信息馈送到根据个人婴幼儿数据准备的用于个性化评估的啼哭评估级。

[0160] 在这方面,图1表明,为了婴幼儿啼哭的评估,首先合适的声音处理或预处理装置被激活、被放置得足够靠近待被监测的婴幼儿并且被接通。

[0161] 在一个优选的实施方案中,靠近婴幼儿实现声音预处理,并且然后,只要到中央服务器的连接是可用的,经预处理的声音数据连同个人婴幼儿数据信息被上传到集中式服务器,该集中式服务器可以是云服务器。在这样的优选的实施方案中,声音预处理装置将是自动婴幼儿啼哭评估装置(未示出)的一部分,该自动婴幼儿啼哭评估装置包括用于连续地声学监测婴幼儿的传声器,用于将监测声音流转换成数字数据流的数字转换级,用于存储个人婴幼儿数据信息的存储器级,用于将数据传输到集中式服务器装置的通信级,其中设置用于在所述数字数据流中识别啼哭的发作的啼哭识别级别,并且所述通信级被适配以从所述集中式服务器装置接收与婴幼儿啼哭的个性化评估相关的数据。

[0162] 应理解,基本上,典型的智能电话可以被用作预处理装置,因为典型的智能电话将包括电池、传声器和合适的传声器信号转换电路系统、处理单元和无线I/O连接。在通过使用智能电话来实施预处理装置的情况下,可以安装合适的应用程序以实施功能和处理级,使得可以在智能电话上执行所有必要的预处理(以及,在适用情况下,预处理和评估二者);然而,由于不是所有的父母和/或护理者都具有闲置的智能电话,并且由于一些应用程序(例如医院的应用程序和儿科站)需要相当大量的预处理装置,优选的是,将必要的硬件集成到独立的封装中或集成到其他婴幼儿监测装置——诸如用于婴幼儿监视的视频摄像机或监测婴幼儿是否呼吸的传感器装置——中。应注意,可以使用靠近婴幼儿的非智能电话装置,从该非智能电话装置声音数据通过诸如蓝牙和/或Wi-Fi的短程通信被传送到父母或护理者的智能电话装置,在此实现附加的(预)处理,使得经预处理的声音相关的数据可以被上传到集中式服务器。在这样的装置中,声音数据将仅在婴幼儿附近检测到特别响的噪声的情况下才必须被传送到智能电话装置;然而,许多父母希望从他们的婴幼儿接收连续的声音流,并且在该情况下,显然可以将来自婴幼儿的连续的声音流发送到父母或护理者的智能电话、膝上型计算机、平板计算机等,以实现在那里对婴幼儿啼哭的评估必要的声音流的任何预处理,包括检测声音流的与啼哭相关的或可以被假定与具有某个非零概率的啼哭相关的那些部分。

[0163] 现在将描述优选的集成独立装置(未示出)。独立装置将包括电力源(power source)、传声器、处理单元、存储器、无线I/O连接和输入/输出装置,并且优选地包括定时器。应理解,这样的装置可以以将特别快启动的方式被构建,使得将不发生接通和实际操作之间的显著延迟。

[0164] 电力源可以是电池,例如可再充电电池,或可以是将被插入到电力插座中的电力供应装置(power supply)。

[0165] 传声器可以是在150Hz-3000Hz范围内灵敏的任何传声器;应理解,更广泛的范围

是优选的,例如,范围从100Hz或甚至80Hz作为下限并且将多达3500Hz、优选地多达4000Hz布置作为上限。应理解,现代传声器将容易地记录此频率范围;然而还应理解,谱灵敏度的变化可以不利地影响对婴幼儿啼哭的评估,因为这样的谱灵敏度的变化可以导致某些频率被克制或被过度强调的声音数据。虽然在允许使用智能电话作为独立装置的情况下,这是一个特别的问题,因为来自各种不同的制造商的不同智能电话可能具有广泛变化的谱灵敏度,但是该问题不太明显并且使用一个或数个型号的独立装置将预期更好的结果,因为在该情况下,可以使用相同的传声器型号。甚至可以校准传声器并且在装置上安装校准数据,使得可以针对给定装置的实际(谱)灵敏度对记录的任何声音进行校正。然而,人们应意识到,谱灵敏度的变化也可能由环境的变化导致,例如,因为在婴幼儿周围放置更多或更少的吸收材料,从而导致特别是对高频的更高或更低的吸收。因此,传声器的总体灵敏度应使得放置在大约0,25米到1,5米之间的距离处,非常响的婴幼儿啼哭应给出靠近但不超过最大数字信号强度的数字信号。在一个优选的实施方案中,装置的灵敏度将被手动或自动设置。传声器的极性模式将使得装置的定向将不显著地影响总体灵敏度和/或谱灵敏度;因此,单极性模式是优选的。传声器信号将被放大,优选地被带通滤波并且被转换成数字信号音频信号。应理解,根据奈奎斯特定理,模拟数字转换的采样频率将高到足以避免混叠问题。因此,在传声器灵敏度多达4000Hz作为上限的情况下,8kHz的采样频率被认为是最小值。此外,在使用8kHz采样频率的情况下,使用适当的模拟(带通或低通)滤波器在4kHz处切断模拟信号是有用的。在一个典型的实施方式中,模拟数字转换将产生至少12比特的输出信号,并且优选地产生14比特的输出信号。由于通常将不可避免地存在一些背景噪声,因此更高的动态分辨率通常将不改善评估。

[0166] I/O连接形成通信级的一部分,用于将信息传输到在附近的父母或护理者以及将数据传输到集中式服务器装置。可以选择不同的连接,一方面与父母或护理者通信并且另一方面与集中式服务器通信;例如,短程无线协议——诸如蓝牙、蓝牙LE、Zigbee等——可以被用来将信息传输到护理者,而广域无线协议——诸如与互联网接入点通信的G4G5 GSM UMTS或WiFi——可以被用来与集中式服务器通信。在此上下文中,应理解,仅有限量的信息需要从靠近婴幼儿的装置传输到父母或护理者。例如,可以传输指示装置正在正确地工作或指示诸如“电池电量低”的另一个状态的常规的装置心跳信号;此外,在婴幼儿啼哭的情况下,可以独立于对啼哭的实际评估传输啼哭指示,并且应指示一旦可用的啼哭评估。普通技术人员将理解,这可以通过传输非常少量的比特来完成,并且因此,带宽和能量消耗二者可以是相当低的。然而,在一个优选的实施方案中,父母可以具有决定任何声音的传输是否是优选的可能性。在一些情况下,父母希望具有对婴幼儿的永久声学监视。

[0167] 然而,应理解,甚至没有必要将实际评估传输到父母或护理者,因为操作中的一个模式将是仅通知父母或护理者婴幼儿正在啼哭,使得护理者移动到然后指示实际评估的装置。相反,当将数据传输到集中式服务器时,通常应传输来自当前待被评估的啼哭的啼哭数据连同从多个啼哭收集的个性化婴幼儿信息和/或数据。由于经常难以安抚婴幼儿,所以即使婴幼儿啼哭的原因是已知的,可以预期的是,可以在诸如数分钟的延长的时间段内收集这样的啼哭数据,从而导致显著更大量的数据将被传输。因此,具有到服务器的宽频带连接是有用的。虽然将大量的数据传输到位于远离婴幼儿的房间的看护者父母不是绝对必要的,因此不需要使用宽频带连接,但是应理解,不需要使用诸如蓝牙LE、Zigbee等的低能量

协议。相反,也可以使用(宽频带)I/O——诸如Wi-Fi——用于与父母或看护者通信。

[0168] 优选的独立装置的输入-输出装置一方面用于将个性化婴幼儿数据信息——诸如婴幼儿的年龄、体重、大小、事实和/或当前或长久的医疗先决条件——输入到装置。当数据可以被录入并且被传输到独立装置以便被存储时,当输入装置与智能电话、膝上型计算机、平板计算机、PC等连接使用时,输入装置可以使用上文所描述的I/O连接来实施。然而,一种甚至更优选的输入个性化婴幼儿数据信息的方式将是使用传声器和附加的话音辨识;在选择此用以输入个性化婴幼儿信息的方式的情况下,按钮等可以被设置,使得可以通过按下按钮来请求进入个性化婴幼儿信息输入模式。应注意,没有必要使话音辨识级实施在装置本身上,但是父母或护理者的话音数据可以被上传到云用于在那里进行话音处理,从而发送回个性化信息和/或与根据话音相比根据其可以更容易确定个性化信息的信息,例如文本文件。应理解,与个性化信息相关的话音辨识可以使用网络中已经可用的服务来实施。

[0169] 甚至将可以通过使用集成的扬声器询问特定的输入信息来引导用户,并且优选地经由机器合成的语音来确认如通过扬声器理解的输入。这对于应被定期更新的个性化婴幼儿数据信息(诸如婴幼儿的体重或与升高温度相关的信息)是特别有帮助的,因为使用传声器,父母或护理者可以毫不费力地并且快速地完成个性化信息的更新。在期望使婴幼儿信息个性化的情况下,应使用不同的装置和无线或有线连接来录入婴幼儿数据信息,有线连接诸如USB——其无论如何可以被用于供应电力——然而优选地设置的唯一输入装置可以是用于确认或拒绝评估的确认/拒绝按钮,因此提供关于评估的质量的反馈。此外,在某些情况(诸如使用独立装置和儿科站)下,如果医疗状况可以被录入作为个性化信息,可能是优选的。这是有利的,因为在儿科站中,来自具有不寻常医疗状况的婴幼儿的啼哭将是更丰富的,从而使数据库更快地增长。

[0170] 关于提供反馈的必要性,应理解,虽然没有必要允许在每一个啼哭时每一个装置的反馈,但是这样做仍然是非常优选的,因为提供反馈帮助扩大可用的样本的数据库,并且因此帮助改进评估;此外,在提供合适的反馈的情况下,大量的“标记的”样本是可用的。应理解,在诸如神经网络滤波器的技术被用于评估和/或用于检测或多或少的嘈杂背景中的啼哭时,需要样本来训练模型以确定合适的滤波器。现在,在提供反馈的情况下,用父母或护理者的反馈标记的可用样本的数据库——无论自动评估是否正确——将显著大于其他情况,并且特别是一旦已经部署了足够数目的装置就可以快速增长。此外,具有来自具有不同年龄、不同性别、不同大小、不同体重等的多个婴幼儿的样本的足够大的数据库允许提供被更高程度地个性化的评估。此外,更大的数据库对于识别新颖事物可以是有帮助的;因此,非常期望向父母或护理者询问反馈并且将反馈提供给集中式服务器,并且优选地以允许将反馈与个性化信息和与啼哭相关的声音数据组合的方式。然而,在某些情况下,在声音数据先前已经被评估并且保持存储在服务器上直到接收到反馈的情况下,没有必要传输或重新传输整个声音数据。

[0171] 关于个性化的评估,对婴幼儿啼哭的当前理解是,对于多达某一年龄(诸如4至6个月)的新生婴幼儿,来自不同国家、种族渊源或“种族”的婴幼儿之间不存在大差别。相反,根据申请人的目前理解,啼哭的差异可以归因于较小和较大的婴幼儿、新生的婴幼儿和年龄较大的(elder)婴幼儿之间的生理差异,其中婴幼儿的医疗状况也具有显著影响。应理解,可以可能更清楚地区别不同的啼哭和/或区别婴幼儿啼哭的更大量的原因。从对现有技术

的以上讨论应理解,某些医疗状况可能更改婴幼儿啼哭的方式,使得从对声音数据的分析可以获得重要的医疗提示。此外,还应理解,为数据库提供新的样本也被实现以提供用于自动婴幼儿啼哭评估的数据,并且根据准备用于扩大样本数据库的样本的方式,可以构成根据本发明的计算机实施的方法。

[0172] 优选的独立装置的存储器将被用来存储用于处理单元——诸如独立盒的微控制器、CPU、DSP和/或FPGA——的可执行指令;然后,个性化信息将被存储在独立装置中,装置ID、用于上传到服务器的声音数据/反馈数据以及用于本地啼哭识别和个性化的本地啼哭评估的经滤波的数据应被存储。此外,存储器将允许对最近的声音数据的缓冲,使得一旦在声音数据中检测到啼哭,紧接在啼哭之前——例如在20秒和0.5秒之间的时段之前——的声音数据也是可得的,优选地至少5秒、特别地至少10秒并且具体地至少15秒的时段之前,其中5秒的窗被用来在声音流中搜索啼哭模式。紧接啼哭之前的数据的长度可以根据成本和合适的缓冲存储器的可用性以及根据预期的噪声水平来确定。在预期或允许环境特别嘈杂的情况下,存储背景/周围噪声的样本也是有帮助的,例如,以识别特别嘈杂或特别安静的频带。应理解,不同类型的存储器——诸如ROM(例如EEPROM存储器)、RAM存储器、闪存等——可以被用于所指示的特定不同目的。此外,应理解,考虑到预期的用途和两次传输声音数据样本到集中式服务器之间允许的时段,以及考虑到至少要本地存储某一时间的数据的种类,可以容易地估计必要的存储器的大小。这可能只是反馈数据、从啼哭数据得出的参数、自前一次上传以来识别的所有啼哭的最初(原始)声音数据、不同水平下背景噪声的样本——例如具有特别响的非啼哭背景噪声、或具有经常观察到的频率含量的非啼哭背景噪声;应注意,后者意味着对背景行为的本地统计分析。

[0173] 提供的存储器的大小也将取决于考虑的婴幼儿啼哭的长度。如上文所指示的,可以使用在评估之前经训练的AI/CNN-滤波器本地和/或以集中式的方式来实现评估,以将啼哭与背景噪声区别开来。考虑到婴幼儿啼哭的典型长度,这样的滤波可以是非常精确的,当然,前提是使待被评估的整个啼哭可用于评估级。通常,当然,在婴幼儿结束啼哭之前需要评估。因此,在确定存储啼哭以用于稍后上传和/或用于缓冲数据所需的存储器的大小时,应考虑通常评价的婴幼儿啼哭的小部分。应注意,在一个优选的实施方式——其在实际实施中已经产生了非常高的正确检测率——中,长度小于5秒、特别地在1.5秒和4.5秒之间、具体地大约2秒、3秒或4秒的啼哭模式被从声音流隔离以用于对啼哭的评估。由于优选地分析多个这样的隔离的啼哭模式,婴幼儿可能在延长的时段内啼哭,除非父母或护理者成功地告诫(sue)婴幼儿,优选的是,存储至少10个、优选地至少20个并且特别地至少30个上文所指示的长度的啼哭隔离模式。应注意,即使具有例如对应于CD质量的采样率,通常也将只需要在0.5MB到8MB之间来实施非常有用的存储器。

[0174] 如上文所指示的,独立盒将具有某种数据处理可能性,例如微控制器、CPU、DSP和/或FPGA以及用以存储分别用于这些装置的指令和/或配置的存储器。这些指令除其他之外将优选地包括至少本地实现啼哭检测的指令。这允许选择那些与啼哭相关的数据,因此与仅在集中式服务器上实现整个评估的情况相比,显著地减少需要被传输到集中式服务器的数据的量。应注意,即使基于仅用于检测特别响的噪声的声音强度的第一级评估也已经导致待被传输的数据的非常显著的减少,因为每个婴幼儿将具有婴幼儿不啼哭的延长的时段。

[0175] 本地评估所必需的处理能力可以被容易地估计。在这方面,应注意,虽然处理单元优选地应能够实现本地啼哭检测,但是考虑到例如存储器和处理约束以及关于更新的频率的约束,在集中式服务器上可能的多达一定程度的完全个性化可能没有必要或不可能在本地装置上。然而,应理解,某些个性化在本地也将是可能的。

[0176] 因此,优选的处理单元通常将被配置为使得自动婴幼儿啼哭评估装置包括本地评估级,该本地评估级被适配以根据从集中式服务器装置接收的并且与对婴幼儿啼哭的个性化评估相关的滤波器或评估指令数据来评估婴幼儿啼哭。

[0177] 应理解,相应地,本地评估不需要对每一个啼哭都有效,而是可以被限制到发现集中式服务器不可访问或仅以特别低的数据传输率可访问的情况。这样的状况可以容易地由如所描述的独立盒的I/O级来确定。

[0178] 本地评估级可以用作不具有个性化的辅助评估级,但是通常也将被个性化,尽管比服务器上可能的程度要低,个性化的程度取决于例如滤波器的可用性和/或本地可用的处理能力。然而,考虑到在典型的应用中,仅相当低的分辨率的音频数据需要被分析并且被处理,通常本地可用的处理能力是足够的,从而甚至允许诸如傅里叶滤波、交叉关联等的处理步骤,而没有处理装置上的过度的负担。

[0179] 因此,在足够的宽频带连接不可用于将数据上传到服务器的情况下,在优选的实施方案中,假如在装置上实施合适的评估级,则可以实现对啼哭的本地评估。通过本地分析声音数据,父母不需要具有永久的因特网接入,这在旅行时或在父母不合理地担心WI-FI辐射可能伤害他们的婴幼儿时可以是有利的。无论如何,在一个优选的实施方案中,可能的并且优选的是,仅一旦已经检测到啼哭才激活无线传输。这减少电池消耗,并且解决父母害怕电磁辐射源靠近婴幼儿的担忧。

[0180] 独立盒通常还将包括定时器。定时器可以是常规的时钟,但是应理解,通常需要计数至少多天,以便判断个性化的评估是否仍然有效,或考虑到自升高温度事件以来已经流逝的时间,决定与婴幼儿的升高温度相关的先前个性化是否仍然应被认为有效。此外,可以测量自上一次个性化以来的时间,并且如果数据通常是过时的,则可以发出警告,例如因为自上一次个性化数据输入以来,健康婴幼儿在正常状况下应具有超过10%的体重增加。

[0181] 因此,本地独立盒可以包括定时器和评价级,该评价级在对婴幼儿啼哭的评估之前评价个人婴幼儿数据信息的当前年龄和/或年龄和/或从集中式服务器装置接收的和/或与对婴幼儿啼哭的个性化评估相关的数据的有效性,并且婴幼儿啼哭评估装置将被适配以根据该评价来输出婴幼儿啼哭评估。

[0182] 定时器对于推断个性化数据是有帮助的。例如,装置在第一次使用时被初始化,并且然后可以对于每个声音样本确定自第一次使用时初始化以来的时间。虽然特别优选的是,在初始化时,录入婴幼儿的年龄,然而这不是绝对必要的,因为可以以非个性化方式评估啼哭直到父母有时间录入所有信息。然后,在记录特定声音样本的时间时婴幼儿的年龄可以被确定并且可以被录入到数据库。稍后可以基于此信息容易地计算婴幼儿的年龄。然而,另外的信息——诸如大小——可以被推断。例如,如果将初始大小与婴幼儿的年龄和性别一起被录入,可以确定在此时间婴幼儿是具有相同年龄、相同性别的同龄群体婴幼儿的平均大小还是在相同年龄、相同性别的同龄群体婴幼儿的特定的给定百分位数以上或以下。可以假设婴幼儿在某一时间内保持在该百分位数范围内,并且可以实现对大小的推断。

[0183] 滤波器更新可以作为用户必须支付的服务来提供,例如经由订阅。一旦订阅到期,用户可以使用具有最终滤波器、通用滤波器的装置或作为不具有任何啼哭检测能力的简单婴幼儿电话的装置。由于订阅将被限制到特定时间,一旦装置被出售——这通常将是在婴幼儿长大相当多之后的情况,订阅期通常将到期,并且将支付新的订阅。此外,原因可以被允许或启用,例如通过传输对应的重置代码。

[0184] 装置还将包括输出装置用于输出婴幼儿啼哭评估的结果。该输出装置可以是如图2中所示出的屏幕或LED点亮符号,这在除了提供LED的原因之外将不指示附加的不同原因的情况下是特别有帮助的。附加地或替代地,输出装置可以是或包括用于与用户的智能电话通信的扬声器和/或I/O,使得可以在遥远位置——例如在远离婴幼儿的不同房间中、在父母或护理者所在地方的智能电话的屏幕上——提供即时评估。

[0185] 使用先前所描述的装置,可以以下方式评估啼哭,并且为此可以提供数据用于使用以下方法的自动个性化婴幼儿啼哭评估,如可以理解的,所述以下方法将是计算机实施的方法。

[0186] 首先,本地装置被激活,即,被接通并且被放置在婴幼儿摇篮附近。一旦该装置已经启动,就检查是否能够以足够的带宽到达集中式服务器。需要被上传到集中式服务器的任何数据,例如先前采样的啼哭数据连同本地评估和/或对先前的评估的反馈都被上传到集中式服务器。然后,通过经由I/O通信接口传输合适的的数据来建立与靠近父母——例如靠近智能电话——的远程站的通信。检查针对个性化评估的当前订阅是否仍然是有效的或应被续订。如果当前订阅不再是有效的,警告信息被传输到远程站。如果当前订阅仍然是有效的,则声音采样开始,并且向远程站传输消息,指示本地装置现在正在“听取你的婴幼儿”。

[0187] 然后实现声音采样,使得传声器被设置在活动模式下,并且对来自传声器的电信号的合适放大被设置成使得信号远高于装置的电子噪声本底,而在响声音事件期间不过载黑色信号(black)。此外,电输入信号以4kHz截止被滤波。经滤波并且经放大的电模拟信号然后以8kHz的采样频率和14比特的动态分辨率被转换成数字信号。只要本地装置保持激活,这就连续地并且自动地进行。

[0188] 数字信号经过自动多级啼哭检测。在这里所讨论的实施方案中,为了检测啼哭,首先,数字音频流中的样本被分组为每个有128个样本的帧,因此所述帧具有16ms的长度。所述帧被写入在所描述的实施方案中存储1024个帧的帧环形缓冲器,循环地将最新的帧存储在当前最旧的帧先前被存储的存储器位置。然而,应注意,不是使用每个有128个样本的帧,而是帧中的样本的数目可以是不同的。虽然使用具有少于128个样本的帧允许对不相关的数据进行更精细的切片或切断,但是具有更多样本的帧更容易由低功率CPU处理。应注意,以8kHz采样率获得的1024个128个样本的帧对应于 $1024 \times 128 \text{ 个样本} \times (1/8000) \text{ 秒} = 16.38 \text{ 秒}$ 。因此,在其中搜索啼哭模式的5秒的窗的情况下,可以在响的噪声之前的3个窗中搜索啼哭模式或啼哭的发作。

[0189] 然后,作为啼哭评估的初始步骤,对于每一个帧,确定128个样本的数字值的均方根,以便提供对当前平均帧声级的估计。对当前平均帧声级的估计也被存储。根据当前平均帧声级,确定阈值,该阈值必须被新的帧的声级超过,以便满足可能已经检测到啼哭的第一标准。应注意,阈值可以以自适应的方式被确定并且不需要是恒定的,不考虑当前平均帧声级。

[0190] 如果检测到平均帧声级未超过前一个帧的平均帧声级一对应于阈值的量,则确定未检测到啼哭并且分析下一个帧。

[0191] 应理解,第一啼哭检测级可以以不同的方式来实施,例如使用前面的平均帧声级的平均值或前面的多个帧(诸如,例如4个、8个或16个前面的帧)中的平均帧声级的最小值。这里,应选择该最小值,因为该最小值是对环境至少有多嘈杂的良好估计。因此,在这里所描述的优选的实施方案中,啼哭检测第一级滤波掉低于并且自适应阈值的聲音。

[0192] 如果检测到新的帧的平均帧声级超过前一个帧的平均帧声级一对应于至少一个阈值的量,则满足关于可能已经检测到啼哭的第一标准,并且帧缓冲器中的1024个帧被保存到另一个存储器位置。在一个实施方案中,声称的啼哭相关的数据将保持受保护,直到已经由附加的啼哭检测级决定尽管满足第一标准但是不存在啼哭,或以其他方式,直到数据可以被选择为与啼哭相关。这允许稍后考虑乍一看似乎不相关的声音数据,尽管实际上已经包括与啼哭的发作相关的声音数据。在另一个实施方案中,与声称的啼哭相关的信息经过啼哭模式识别和隔离步骤。为此,可以定义重叠的窗,例如,5秒长度和3秒步幅的窗。可以非本地实现啼哭模式分析,换言之在例如集中式云服务器上;在该情况下,其他存储器位置将在云服务器上。在其他情况下,特别是在本地提供的处理能力足够高的情况下,将可以本地搜索啼哭模式,甚至不将存储的帧存储在另一个存储器位置,考虑到如果处理足够快,可以在先前存储的1023个帧中直接检测声称的啼哭模式。然后,只有隔离的啼哭模式必须被传送到云服务器或另一个本地处理级。应理解,通过首先定义具有待被检测的啼哭模式的长度的倍数的窗,通过定义所定义的窗的类似谱图的表示,以及通过使用人工智能技术——特别是卷积神经网络在类似谱图的表示中搜索啼哭模式,可以影响啼哭模式识别隔离。应理解,对于啼哭模式的这样的搜索,使用以个性化方式训练的模型不是绝对必要的,因为婴幼儿啼哭模式对于大量各种婴幼儿将具有非常类似的特征。这在本地搜索啼哭模式中是有帮助的,因为以个性化方式搜索啼哭模式不是绝对必要的。在以上文本中,已经陈述,搜索到的啼哭模式应优选地具有最小长度;0,25秒或0,3秒或优选地0,4秒的这样的最小长度是有利的,因为对于考虑的较长的模式,啼哭模式和非啼哭时段之间的区别是显著更可靠的,即使当未实施啼哭隔离和啼哭检测的个性化时。换句话说,通过搜索较长的啼哭模式,啼哭检测的个性化变得可有可无,特别是当本身已知的对象定位算法被用于声音数据的类似谱图的表示中的啼哭模式检测时。

[0193] 然而,将啼哭检测的第二级作为啼哭模式检测和隔离来实施不是绝对必要的——因此,可以通过除搜索啼哭模式以外的手段来实现更严格的分析,以确定由新的帧的平均帧声级判断的声音强度的突然增加——超过前一个帧的平均帧声级对应于至少一个阈值的量——实际上是否是由于婴幼儿啼哭。对于这样的不具有啼哭模式识别和隔离的更严格的分析,必须理解,婴幼儿在延长的时段内啼哭,使得仅与延长的高声音强度相关的声音数据应被完全考虑。因此,超过一特定强度的后续帧——例如因为如上文所定义的均方根平均帧强度超过当前最小噪声——自适应阈值水平——被复制到啼哭检测帧缓冲器以用于进一步分析。另一方面,在考虑到声音数据的类似谱图、使用对象定位方法来实施啼哭模式检测和隔离的情况下,仅在声级超过给定阈值的情况下才开始这样的啼哭模式检测和隔离不是绝对必要的。可以基于类似谱图的表示中的对象定位来连续地执行啼哭模式检测和隔离方法。因此,可以作出是否实施用于特别响的声级的在前检查的决定,例如考虑到能量消

耗,例如当根据电池操作本地装置时,或考虑到当在云中实施啼哭模式检测和隔离时用于将声音数据上传到集中式装置(诸如云服务器)的可用数据带宽。

[0194] 如果发现在一特定时段内需要被复制到啼哭检测帧缓冲器中用于进行进一步分析的后续帧的数目太低,可以安全地假设婴幼儿未啼哭并且可以删除存储在缓冲器中的数据。否则,将执行附加的测试。因此,这里,通过计数在一给定时段期间输入到缓冲器内的帧,实施另一个啼哭检测级。这是一种拒绝将噪声作为声称的啼哭的优选方式,尽管在一些实施方式中,将不使用这样的附加的拒绝。

[0195] 应强调,由于仅那些具有足够高的平均声级的帧被复制到缓冲器中,缓冲器将不表示一个完整的帧序列,因为一个或多个中间的帧可能具有较低的平均声级,并且因此将不被复制到啼哭检测帧缓冲器中。应理解,此方法不同于典型的实验室设置,在实验室设置中,不具有背景噪声的声音可用于分析,并且因此,考虑到估计的背景环境噪声水平,不需要抛弃帧。应注意,在作为啼哭检测本身的一部分来识别和隔离啼哭模式的实施方案中,显然,最初记录的声音流的部分也被省略。

[0196] 然而,在一些实施方案和实施方式中,在潜在的啼哭的发作之后,没有帧被省略。在不省略帧的一个实施方式中,可以容易地使用交叉关联/滑动平均技术。删除具有低声级的帧的优点是待被处理和分析的数据的总量是较低的;不删除帧的优点之一是可以获得更高的精度/准确性,特别是在使用滑动/交叉关联技术的情况下。在此上下文中,应理解,一旦已经证实在声音数据中存在啼哭,分析完整的序列以评估啼哭的原因通常是有帮助的,因此,无论如何都应存储包括满足第一啼哭检测级标准的第一帧之后的所有帧的完整的序列。显然,当从在原始记录的声音流中定义的窗识别和隔离啼哭模式时,包括所有帧的完整的序列也被评估;应注意,作为啼哭检测的一部分,识别和隔离啼哭模式在对啼哭的评估中产生特别良好的结果。

[0197] 还应理解,在包括所有啼哭前数据的循环缓冲器足够大的情况下,为完整的序列提供单独的缓冲器是非必要的;在该情况下,完整的序列将简单地存储在循环帧环形缓冲器中。

[0198] 还应强调,超过给定的——并且在适用情况下自适应的——阈值的帧的数目不被用来识别和拒绝短事件,缓冲器仍然可以被关闭,例如因为一些连贯的帧被识别为不相关,例如因为由于低的声级。在该情况下,显然,缓冲器将不被完全填充。

[0199] 为了证实特别响的帧是否构成婴幼儿啼哭的一部分,然后使用用先前已经识别的婴幼儿啼哭训练的神经网络滤波器。应注意,此神经网络滤波器将不同于在对啼哭的评估(或“转化”)中使用的神经网络滤波器。如上文所解释的,一种使用神经网络滤波器的方式是定义多个帧的类似谱图的表示,并且然后在此表示中搜索啼哭模式。因此,对于啼哭检测神经网络滤波器,婴幼儿为什么正在啼哭不是重要的,并且个性化对于提高啼哭检测的准确性也不是绝对必要的,尽管应理解,在某些情况下,谱图的声能和某些频率范围可以给出重要的线索,所述线索既根据不同婴幼儿又根据不同环境而不同。此外,考虑到如果使用诸如卷积神经网络的人工智能技术,使啼哭检测和/或啼哭评估个性化是相当计算密集的,将任何个性化限制到仍然获得有利结果所必需的最小值被认为是有利的。

[0200] 例如,在某些环境中,婴幼儿啼哭得特别响的频带可能也经历更强的背景噪声,从而致使它们不太适合于啼哭检测。遗憾的是,背景噪声模式可能比对婴幼儿的个性化啼哭

评估变化得甚至更快,例如因为监测婴幼儿的位置经常变化、因为窗根据天气状况而被打开或关闭所以背景噪声变化等。因此,在最优选的实施方案中,啼哭检测本身未被个性化。尽管如此,在实际实施方式中,使用适当地训练的神经网络滤波器作为声级评价之后的级,可以容易地实现具有本领域的99%更好的准确性的啼哭检测。还应注意,包括经过滤的(leached)的啼哭声音的声音窗的类似谱图的表示一个优点是抗噪声的鲁棒性;换句话说,尽管有噪声,但是在使用这样的技术的窗的表示中将非常可靠地识别啼哭模式。

[0201] 然而,在一些情况下,可以优选的是,不隔离在类似谱图的表示中发现的啼哭模式;在这样的情况下,对于使用神经网络滤波器的啼哭检测,要么原始的完整的声音数据——例如缓冲器中的每个帧——可以被直接地输入到合适的神经网络滤波器中,要么参数。上文,已经公开了可以从啼哭数据提取的多个参数。类似的参数可以被确定用于啼哭检测,例如,在声称的啼哭缓冲器内的帧的平均声称的啼哭能量,在缓冲器内的特定数目的连贯的和/或帧上——特别是在2、4、8、16或32个帧中——和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——内的声称的啼哭能量的滑动平均值;在缓冲器内的帧的声称的啼哭持续时间方差;特别是在2、4、8、16或32个帧2、4、8、16或32个帧上和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——内的声称的啼哭能量方差等;

[0202] 当前音调频率;在2、4、8、16或32个帧2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——上的平均音调频率;在声称的啼哭事件期间和/或在缓冲器内的声称的啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的音调频率的最大值;

[0203] 在声称的啼哭事件期间在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的滑动最大音调频率的变化;

[0204] 在根据缓冲的帧的声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的音调频率的最小值;在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的滑动最小音调频率的变化;

[0205] 在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的音调频率的动态范围;在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的频率的音调平均变化率;

[0206] 此外,假设帧将表示啼哭数据,可以确定共振峰相关的参数,诸如在声称的啼哭事件中或在声称的啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧中、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第一共振峰频率;

[0207] 在声称的啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间平均的第一共振峰频率的平均变化率;

[0208] 在声称的啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧上、和/或在特定时间

间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间滑动平均的第一共振峰频率的滑动平均变化率；

[0209] 在声称的啼哭数据的2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间平均的第一共振峰频率的平均值；

[0210] 在声称的啼哭数据的2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——内的第一共振峰频率的最大值；在声称的啼哭数据的2、4、8、16或32个帧中、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第一共振峰频率的最小值；

[0211] 在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率动态范围；

[0212] 在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率；

[0213] 在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率平均变化率；

[0214] 在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率平均值；

[0215] 在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率最大值；

[0216] 在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率最小值；

[0217] 在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振；

[0218] 在声称的啼哭事件期间和/或在声称的啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的峰频率动态范围；此外，再次假设婴幼儿实际上正在啼哭，可以确定Me1频率倒谱参数，在整个声称的啼哭事件和/或在声称的啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间确定该参数；和/或倒置的Me1频率倒谱参数。

[0219] 虽然可以使用这样的参数用于啼哭检测，但是应容易地理解，考虑到上文关于背景所描述的大的方差，通常，通过使用更精细的并且在计算上更密集的参数将无法提高准确性。因此，婴幼儿当前是否正在啼哭的决定可以基于被选择为使得计算工作量保持低同时仍然提供高准确性的啼哭检测的参数。这进而允许本地实现啼哭检测，即使使用多级啼哭检测，其中至少一个级使用神经网络滤波器技术。

[0220] 因此，从上文可以看到，作为啼哭检测级中的一个步骤，缓冲器的内容被进一步分析。通常， n 个复数帧将被存储在缓冲器中，并且生成指示缓冲器中的帧是否与婴幼儿啼哭

相关的输出信号;作为一个替代方案,可以首先识别和隔离啼哭模式,并且然后存储这些啼哭模式用于对婴幼儿啼哭的原因的进一步评估和/或将它们传输到集中式服务器和/或立即处理它们。

[0221] 在依赖于对参数的评价而不依赖于啼哭模式识别隔离的一个最典型的实施方式中,作出缓冲器中的帧是否与婴幼儿啼哭相关的判断,指示声音数据与啼哭和/或与判断的可靠性的程度相关的概率。这样的确定可以在缓冲器内逐帧地进行,但是如果以逐缓冲器的方式进行确定,则处理明显更容易。在这方面,可以一个接一个地分析多个缓冲器,例如因为第一缓冲器已经被完全填充,和/或因为由于两个超过阈值的帧之间的时间距离变得太大因此前一个缓冲器已经被关闭。一旦已经分析了多个缓冲器,可以确定最终输出,因此最终输出是为每个单个帧的每个单个条计算的结果的函数。这可以通过以相等的权重平均一个相应的缓冲器与啼哭相关的概率来完成,也可以以考虑每个概率的可靠性的方式通过一个相应的缓冲器与啼哭相关的概率来完成。在一个优选的实施方式中,在啼哭检测期间被分析的缓冲器的数目被设置为2或3。在一个优选的实施方式中,要求判断检测到啼哭,至少一个缓冲器中的声音数据的概率应超过75%,并且还要求 $N=3$ 个缓冲器的概率的(线性)平均值大于50%。如果所有的标准都满足,则声音数据就被判断为属于婴幼儿啼哭。然而,如果不满足标准,则对随后的缓冲器重复对应的分析,直到在延长的时间内未检测到考虑到最初的第一阈值标准而必须被认为是啼哭的候选者的帧。换句话说,如果啼哭检测分析的最后一级的输出对于缓冲器(n,n+1,n+2)是负的,则只要一些帧指示特别响的声音,就对缓冲器(n+1,n+2,n+3)重复该分析。

[0222] 应理解,如果在最终一级中未检测到啼哭,则啼哭检测帧缓冲器中的数据将被刷新,除非决定将一个或多个啼哭检测帧缓冲器中的背景声音数据上传到服务器用于训练啼哭检测神经网络过滤器和/或用于识别典型或关键的背景模式。

[0223] 这样的将非啼哭数据上传到服务器的决定可以是随机的或缓冲的非啼哭数据被标记用于上传,因为平均概率非常靠近被判断为指代啼哭的概率或因为平均概率是极其低的。在此上下文中应理解,定期将非啼哭数据上传到服务器的目的是为了识别背景行为模式中的新颖事物,并且改进神经网络啼哭检测过滤器。应理解,即使非啼哭数据将被上传,也可以遵守数据保护条例。具体地,只有在拥有本地装置的人同意上传特定的非啼哭模式之后,才将可以允许上传。此外,还可以实现话音检测,以便防止上传与话音相关的声音数据;此外,还将可以在不提到特定装置的情况下上传非啼哭。

[0224] 上文,已经陈述,如果在最终一级未检测到啼哭,则啼哭检测帧缓冲器中的数据将被刷新。在一种类似的方法中,在一给定时段中的啼哭模式的数目太少的情况下,换言之如果婴幼儿仅一直啼哭了极其短的时段(诸如5秒或6秒),即使肯定地被识别为具有与啼哭相关的非常高的可能性,也将可以丢弃啼哭模式。在最终评估依赖于对大量的啼哭模式(例如5至10个啼哭模式)的评估的情况下将尤其如此。

[0225] 否则,换言之当确定检测到啼哭时,首先通过向远程站发送对应的消息来通知父母或护理者。这是有帮助的,因为父母或护理者将需要一些时间到达本地站并且在该时间期间常常能够执行对婴幼儿啼哭的原因的评估。除了婴幼儿啼哭的消息之外或作为婴幼儿啼哭的消息,优选地包括存储在循环缓冲器中的预先发作的声音数据的啼哭的声音数据可以被传输到远程站进行音频重现。应理解,传输啼哭前声音可以帮助引起父母的注意,因为

在远程站处重现的声音像更接近护理者在靠近婴幼儿时将听到的声音。此外，它可以帮助在婴幼儿开始哭泣的时间对婴幼儿的周围环境进行声学评价，这进而在啼哭是由外部影响（诸如兄弟姐妹或宠物进入房间）引起的情况下可以是有帮助的。然而，应理解，一些实施方案将不需要通知父母或护理者婴幼儿正在啼哭，因为父母和护理者无论如何都将保持在婴幼儿的听力范围内，即使他们在不同的房间。

[0226] 然后，在向远程站发送消息之后，作出是否可以由集中式服务器评估啼哭或是否需要本地进行评估的决定。为此，准备数据传输文件，该文件包括啼哭的声音数据流的所有相关帧和/或一直到该文件的准备的所有啼哭模式。在不依赖于本地啼哭模式识别和隔离的一个实施方案中，此文件不仅可以包括在第一帧超过阈值之后已经超过给定阈值的那些帧（并且因此将包括比仅为了啼哭检测而缓冲的更多的帧），而且将包括自第一帧已经超过阈值以来记录的帧的完整的序列。此外，该文件将包括来自循环啼哭前帧缓冲器的、在检测到第一极其高的声级之后被锁定的那些帧。在本地啼哭模式识别和隔离被实现的一个实施方案中，仅隔离的啼哭模式将被传输；在其中仅第一啼哭识别级——诸如阈值比较——被实现、但啼哭模式识别隔离被实现在集中式装置上的一个实施方案中，上传到集中式装置的文件将优选地包括环形缓冲器的全部内容。然后，个性化婴幼儿数据被添加到该文件，诸如体重、性别、年龄、医疗先决条件等。这可以以编码的方式来完成，例如通过包括先前分配的ID。在此情况下，可以对应于分配给装置的特定ID地检索存储在集中式服务器上的集中式数据库中的对应的信息。这是优选的，因为被许多用户认为是机密的数据将不得不较少地被传输，使得避免机密性问题。在一个优选的实施方案中，装置将与服务器协商以使用令牌系统获得上传数据的许可。与服务器协商以获得上传数据的许可将减少服务器上的来自传入的不希望的数据的负载。此外，可以将来自特定用户的传入数据存储在唯一地分配给该特定用户的预定位置，因此提高机密性。应理解，在传输到服务器的总体数据率无关紧要的情况下，将甚至没有必要诸如通过使用比较器来实现第一本地简单评估。

[0227] 应理解，数据文件的确切内容和/或数据文件的确切结构可以变化。此外，尽管不太优选，将可以省略啼哭前数据——或即使婴幼儿还不是特别响也可以在声音表示中识别的啼哭模式——并且，在带宽是一个特别的问题的情况下，可以在帧序列中留下空白，例如抛弃在啼哭的发作之前确定的非常靠近最小背景水平的那些帧。然而，显然，这明显是不太优选的，因为以这样的方式获得的结果通常不那么准确。具体地，使用交叉关联技术的可能性可以被削弱。还应强调，对于正在进行的啼哭，在第一数目的帧已经被传输到服务器使得评估可以开始之后，附加的数据可以被收集并且被传输到服务器，以改进对正在进行的啼哭的评估。

[0228] 现在将描述一种优选的啼哭转化或啼哭评估的方式；应理解，存在多种不同的评估婴幼儿啼哭的方式。还应理解，本申请提议的提供用于评估的数据的方法对于所有或至少非常多的各种这样的不同的评估婴幼儿啼哭的方式。尽管如此，通过描述啼哭评估的典型实施方式，提供用于评估的数据的方法如何被最好地实施将变得更明显。

[0229] 为了理解啼哭转化，应理解，对于婴幼儿的当前特定需要，婴幼儿的啼哭示出明显区别的特征。例如，这可以例如在图3a-图3e中所示出的三维谱图中看到，该三维谱图清楚地示出在同一个婴幼儿具有不同需要时记录的啼哭的谱图之间的差异。在图3中可以清楚地看到的是三维谱图明显不同。应注意，谱图示出了对于多个不同的频率（分别为x轴和y

轴)能量含量(z轴)随时间的推移的三维图。基本上,对于不同的啼哭在图4a和图4b中给出了相同的信息。

[0230] 所示出的模式由于相应的原因是典型的,使得啼哭的显著差异原则上允许将啼哭相互区别开来,或允许考虑到声音来评估婴幼儿啼哭的原因,比较图4a-图4f。然而,应理解,不仅对于来自同一个婴幼儿的不同的啼哭,而且对于由于相同的原因的不同的婴幼儿啼哭,模式看起来不同,差异尤其取决于婴幼儿的年龄、体重、大小等。尽管如此,不同种类的啼哭之间的显著差异仍然可以被识别,特别是当从啼哭隔离特定的参数和/或使用机器学习算法时,比较图4a-图4f。

[0231] 然后,应理解,通常,婴幼儿啼哭的原因并不单一;例如,婴幼儿可能既疲倦又可能需要拍嗝;婴幼儿可能饥饿并且需要安慰等。这将反映在啼哭中并且对应地反映在相应的啼哭模式中;因此,任何给定的啼哭模式可能同时反映婴幼儿正在啼哭的多个模式,并且恰当的评估优选地将其考虑在内。

[0232] 对于以自动方式识别不同的啼哭模式之间的差异的第一种方式,使用足够的参数来描述每个啼哭或将从啼哭获得的特定参数馈送到模型中可以是有帮助的。使用足够的参数,可以定义啼哭的群组(或“云”),其中每个云包括由于不同原因的啼哭。这被示出在图5a-图5e和图6中。应注意,图5中所示出的模式来自通过被称为自组织映射的无监督深度学习技术增强的原始啼哭数据集。在图6中,每个啼哭由多维参数空间中的点来表示。不同类型的点表示婴幼儿啼哭的不同的原因,并且图6清楚地指示,可以将不同类型的点进行分组,并且因此可以将婴幼儿啼哭的不同类型的原因进行分组。即使在实际的啼哭评估基于在原始声音数据流中被识别和隔离的啼哭模式的情况下,学习新颖事物和/或对同龄群体的分组可以依赖于基于特定参数的自组织映射。然后,一旦建立了这样的同龄群体,先前针对每个同龄群体获得的啼哭模式可以被用来训练为每个同龄群体个性化的相应的模型。在另一种改进个性化的方式中,可以为每个啼哭模式确定此啼哭模式与多个不同类别的原因中的每个原因相关的概率;从啼哭模式序列,然后可以根据针对不同类别中的每个类别的概率的序列来提供婴幼儿为什么啼哭的总体评估。具有非常良好的结果的个性化的简化是仅使对此概率的序列的评估个性化。以此方式,计算密集的个性化可以被减少到最小,同时仍然给出优秀的个性化结果。应理解,这简化了个性化,因为可以以非个性化方式实现给每个啼哭模式的概率的分配。

[0233] 啼哭转化或“啼哭评估”的目的是将已经被识别为属于婴幼儿啼哭的声音数据分类为多个不同类别之一。在图3中,示出了由于婴幼儿啼哭的5个不同的原因的啼哭,即“饥饿、不舒服、需要拍嗝、感到疼痛和困倦”。将可以使用这5个不同的原因作为每个婴幼儿啼哭被分类为的类别。然而,这些类别虽然对于年轻父母非常有帮助并且容易区别,但是不应被解释为限制啼哭评估的可能性。相反,可以实施更少的类别,将例如“不舒服”和“疼痛”组合,或可以使用更多的类别,诸如描述例如与“咳嗽”、“打嗝(hiccup)”、“打喷嚏”相关的呼吸模式的类别。此外,将可以使用与啼哭根本不相关的附加的类别,例如“沉默”或“未定义”;使用这样的有意与啼哭不相关的附加的类别帮助滤波我们的潜在的肯定(positive),因为即使具有良好的啼哭检测级,被错误地识别为啼哭的声音数据可能被传送到啼哭评估级。在滤波器中提供一个或多个非啼哭类别帮助减少误报的数目。

[0234] 应理解,这是在数据挖掘中和在数据分析中经常发现的情形,并且因此,人工智能

技术并且特别是神经网络技术——诸如CNN(卷积神经网络)技术——可适用于区别不同的啼哭,如果可以提供合适的训练数据并且发现足够的参数。

[0235] 因此,一些数据必须被提供给本地或集中式啼哭评估级。在两种情况下,可以使用类似的技术,例如人工智能/神经网络滤波技术。此外,在两种情况下,可以以个性化方式评估声音数据,尽管应理解,个性化的程度和/或可以承担的计算工作量分别对于本地和集中式的情况可以是不同的。具体地,在集中式的情况下,通常,可用的处理能力可能是更大的,在一些情况下显著大于在本地装置中的本地评估的情况。因此,在评估中使用的作为输入到神经网络滤波器中的参数的数目可能是更大的,因为计算如上文所列出的那些的参数需要至少一些计算工作量。

[0236] 然后,在用于服务器装置的集中式能力中使用的神经网络滤波器可以比可以在具有较低处理能力的本地站上本地实施的那些滤波器更复杂。还应理解,在最典型的情况下,用于本地评估的滤波器系数将在集中式服务器上被确定,并且从集中式服务器传送到本地装置(应注意,提到了集中式服务器,因为此服务器可以被使声音数据从他们的本地装置传送到服务器的大量的用户使用;这并不排除“服务器”在空间上分布的可能性,就像“云服务器”的情况一样)。在使用集中式服务器的情况下,滤波器的更新/个性化可以是更好的,因为通常,个性化的滤波器将更经常在服务器上被确定,然后被下载到本地装置,并且另外在一些情况下,尽可能仅一部分个性化在本地装置上。例如,当婴幼儿发烧或发烧同时伴有在特定范围内的温度升高时,啼哭可能改变。由于发烧可能经常地并且自发地发生,每次个性化的滤波器系数被更新,对应的滤波器系数集将必须被本地存储在本地装置上。由于可能需要考虑诸如发烧的类似的状况,存储大量各种滤波器系数所需的存储器大小将是非常大的,并且此外,为了更新针对在服务器上可以被区分的每个不同的状况的滤波器系数而将必须从服务器传送到本地装置的数据的量通常将是太大的。因此,考虑到技术困难,本地评估经常必然不如集中式评估精确。

[0237] 尽管如此,甚至对于本地声音评估,也必须向评估级提供声音数据,并且根据声音数据是否被本地评估——考虑到用于特定婴幼儿的特定滤波器已经被下载或经由推送服务被获得,评估在被本地完成时可以被认为是个性化的。对于上传,如果与和ID相关的婴幼儿相关的个性化数据之前已经被存储在服务器上,则声音数据可以与ID组合;此外,完整的个人信息可以被传输;应理解,考虑到以最好的方式保护隐私的愿望,可以根据数据保护条例作出是使用个人信息的频繁传输还是使用与存储在服务器上的个人信息有关的ID的传输的决定。

[0238] 尽管存在可以对在声音评估级中评估声音数据的确切评估方式具有影响的困难——诸如可用的计算能力,但是在目前情况下,如果足够的计算能力是可用的,例如在将声音数据上传到集中式服务器之后,描述可以做什么被认为足够。由此,也可以容易地推断可以如何影响本地评估。例如,在交叉关联技术是太计算密集的情况下,将可以不通过在以逐样本的方式移位输入信号时计算最好对应性来确定评估,而是可以仅考虑在以逐帧的方式移位输入信号时或在移位输入信号并且超过2帧时获得的结果,因此将计算负载减少到二分之一。

[0239] 对于一个实施方案,在下文中还将假设,最初传送到集中式服务器的帧的数目足以实施和执行交叉关联步骤,并且一旦以多碎片形式最初传送的啼哭数据已经被评估,父

母就已经到达本地装置并且可以确认最初作出的任何评估。应注意,对任何评估的确认不需要是立即的;一方面,常常仅一旦婴幼儿响应于由父母采取的行动而停止啼哭父母就将感觉有信心确认或拒绝评估。此外,甚至在评估被立即认为是正确的情况下,父母也应在评价评估之前照顾婴幼儿。因此,任何评估原则上也可以在稍后作出,使用例如运行合适的应用程序的智能电话。尽管如此,在一个实施方案中,父母可以具有将评估或反馈立即输入到装置中的可能性,并且在此实施方案中父母还未判断评估的情况下,可以上传附加的数据并且可以像最初已经传输了较大的文件一样实现啼哭评估。与重复地而不是在较大的文件中传输数据的情况的唯一区别在于,使用数据的第一部分,可以进行并且传输第一评估。然后,如果接收到越来越多的数据,评估可以被校正或确认;在最初的评估通过分析越来越多的数据而未改变并且因此被确认的情况下,用户甚至可能不知道附加的数据被评估,特别是除非评估正确的概率被指示;可以预期,通过提供更多的数据,评估正确的概率将增加。在评估随时间变化的情况下,将可以明确地告知用户评估已经改变,使得用户不认为用户可能已经注意到的最初的评估是小故障(glitch)。应理解,将数据传送到啼哭评估级可以继续,直到用户已经确认啼哭评估和/或直到婴幼儿已经停止啼哭。尽管如此,考虑到上文所述的内容,在本申请中,描述仅根据传送到集中式服务器的第一文件实现评估的情况将是足够的。

[0240] 从上文,应理解,分析长的帧序列是可能的并且有用的。上文,已经提到第一包括例如1024个帧。多个这样的缓冲器可以被打包成一个单个文件,该文件然后被分析以确定婴幼儿啼哭的原因。应理解,虽然尽快检测到婴幼儿正在啼哭是有用的,因为对应的信息应尽快被传达到父母或护理者,但是考虑到父母或护理者无论如何需要的反应时间,评估可以花费更多一些时间。因此,使用大量的帧用于评估通常不构成一个重大问题。因此,在啼哭检测优选地基于不大于3个缓冲器工作的情况下,啼哭评估可以在显著早期的较大量的缓冲器——诸如5、6、7、8或16个缓冲器(其中每个缓冲器保持例如1024个帧)——上来实现。尽管如此,优选的是,如果在啼哭检测之后在少于15秒——优选地不超过10秒——内使第一评估对用户可用。否则,用户可能认为本地装置是无反应的。

[0241] 从上文,应理解,优选的是,使用较大量的缓冲器用于啼哭检测,并且本地装置应优选地具有足够的存储器来存储至少16个缓冲器,优选地在中央服务器上不能够分析检测的啼哭的情况下,更多的缓冲器必须被存储用于稍后加载在本地装置上。

[0242] 一旦足够的数据已经被收集以用于在啼哭评估级中进行评估并且被上传到集中式服务器,就可以开始预处理。在预处理期间,确定对应于个性化信息的滤波器参数集,例如通过按照参数集数据库来参考滤波器,并且根据此滤波器参数集配置神经网络滤波器。

[0243] 然后,声音数据本身例如以逐帧的方式被馈送到个性化的神经网络滤波器中,或描述声音数据的参数被确定并且确定的描述声音数据的参数被输入到个性化的神经网络滤波器中。

[0244] 如上文所指示的,可以被输入到神经网络滤波器中的可以描述声音数据的参数可以包括当前啼哭事件期间的平均啼哭能量,在特定数目的连贯的和/或帧上——特别是在2、4、8、16或32个帧中——和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——内的啼哭能量的滑动平均值;在2、4、8、16或32个帧2、4、8、16或32个帧上和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——内的啼哭能量方差;当前音调频率;

在2、4、8、16或32个帧2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——上平均的音调频率；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的音调频率的最大值；在啼哭事件期间在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的滑动最大音调频率的变化；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的音调频率的最小值；在啼哭事件期间在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的滑动最小音调频率的变化；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的音调频率的动态范围；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的频率的音调平均变化率；在啼哭事件中或在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧中、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第一共振峰频率；在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间平均的第一共振峰频率的平均变化率；在啼哭数据的2、4、8、16或32个帧2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间滑动平均的第一共振峰频率的滑动平均变化率；在啼哭数据的2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间平均的第一共振峰频率的平均值；在啼哭数据的2、4、8、16或32个帧上、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——内的第一共振峰频率的最大值；在啼哭数据的2、4、8、16或32个帧中、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第一共振峰频率的最小值；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第一谐振峰频率动态范围；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率平均变化率；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率平均值；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率最大值；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二共振峰频率最小值；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的第二谐振；在啼哭事件期间和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、五秒、10秒、15秒、30秒——期间的峰频率动态范围；Mel频率倒谱参数，在整个啼哭事件和/或在啼哭数据的2、4、8、16或32个帧期间、和/或在特定时间——诸如一秒、两秒、

五秒、10秒、15秒、30秒——期间确定该参数；和/或倒置的Mel频率倒谱参数。

[0245] 如上文所指示的，在已经在参数列表中提到特定时间的情况下，也可以参考任何其他固定时段，一直到明确提及的相应的时间。此外，提醒阅读者上文已经指示了为什么某些长度是有利的解释。

[0246] 上文还已经反复强调，如果由声音数据表示的模式与由于不同原因的典型模式进行比较，并且由神经网络滤波器表示的以交叉关联的方式被分析，换言之通过考虑啼哭的不同潜在发作，则可以获得优点；考虑啼哭的不同潜在发作可以容易地完成，其中滑动平均或其他滑动参数是逐帧或逐样本地确定的，并且非常对应的参数的序列被使用分别作为到神经网络滤波器的输入。然而，应理解，这样的技术是在计算上更密集的。

[0247] 在另一个实施方案中，在已经在啼哭检测级中隔离了啼哭模式和/或已经确定了啼哭模式发生的时间而不是计算用于与这样的时间相关的帧的特定参数的情况下，可以准备在观察到啼哭模式的时间期间记录的声音的类似谱图的表示；应注意，在可以使用相同的时间和频率分辨率用于啼哭模式识别和隔离的情况下，不需要准备附加的类似谱图的表示，因为可以使用之前隔离的啼哭模式。类似谱图的表示然后可以被馈送到卷积神经网络中，该卷积神经网络输出被馈送到它中的每个啼哭模式属于预定类别中的任何一个的可能性。在一个特别精确的但是在计算上更密集的实施方案中，相应的卷积神经网络滤波器参数将被个性化，换言之，对于每个定义的同龄群体使用不同的滤波器参数集。在其他实施方式中，将可以对于所有同龄群体使用相同方的滤波器参数，以确定每个隔离的啼哭模式属于婴幼儿要啼哭的不同原因中的每个原因的概率，并且然后以个性化方式评估从每个啼哭模式获得的组合原因。应理解，通过使仅一个或数个最终层个性化，可以减少训练模型的计算工作量以及在评估中使用的并且必须被存储的参数的数目。

[0248] 为了清楚起见，应强调，虽然神经网络滤波器可以在啼哭检测的最终一级中使用并且虽然神经网络滤波器实施方式也可以被用于啼哭评估，但是在啼哭检测的最终一级中使用的神经网络滤波器将不同于在啼哭评估中使用的滤波器，分别到用于啼哭检测和啼哭评估的相应的神经网络滤波器中的总体输入也将不同。用于啼哭转化的滤波器可能是更复杂的，使用例如卷积神经网络中的更多层和/或更多输入，例如使用更高的分辨率和/或具有用于个人参数的附加输入和/或更多层。

[0249] 应理解，一旦已经检测到啼哭，甚至就没有绝对必要评估啼哭。如果父母确信在没有来自电子装置的任何帮助的情况下能理解婴幼儿啼哭的原因，则有时通知父母婴幼儿正在啼哭是足够的。因此，在这样的情况下，将不需要转化啼哭，因此节省例如能量。因此，在一些实施方案中，甚至将可以仅在父母明确需要支持的情况下才触发自动啼哭评估。在又一个装置中，将可以根本不允许啼哭转化并且仅使用啼哭检测，以便改进对婴幼儿监视监测器的反应。在这样的装置中，传输可以仅响应于啼哭的检测而被反映和/或声音可以在接收器处产生特别响的声音的方式被传输，例如通过响应于啼哭的检测而改变数字信号的增益。

[0250] 在啼哭评估级中，确定输入到神经网络滤波器中的声音数据属于多个预定义的类别——诸如“饥饿”、“困倦”、“需要拍嗝”等——之一的概率，并且因此获得多个概率，从而给出概率的n元组，其中n元组的分量表示婴幼儿由于与相应的类别相关的原因而啼哭的概率。从帧到帧或从缓冲器到缓冲器进行，每次获得的n元组的分量将变化。因此，根据n元组

的序列,必须计算总体评估。

[0251] 存在多个可能性来确定总体评估。例如,可以计算N-元组的每个分量的平均值,并且选择具有最大平均值并且因此具有总体最高概率的分量作为评估。此平均值可以是线性平均值、均方根平均值等。在一个优选的简单的实施方案中,计算线性平均值。此外,考虑到交叉关联技术可以导致用于非常良好的匹配的非常高的概率,可以比较所有“与”元组上的每个分量的最大值;因为由于采样和噪声使用交叉关联技术可实现的模式匹配可能是不完美的,所以甚至可能优选的是,考虑使例如2、3、4或5个连贯的n元组的每个分量平均的滑动平均值的最大值。如果一个给定分量的最大值未超过用于所有考虑的帧的特定的阈值,则将可以将一些分量完全排除在考虑之外;以此方式,从未给出令人信服的匹配的分量将不导致错误评估。

[0252] 另一个可能性将是由为M个连贯的帧获得的M个连续的N-元组逐步构建NxM矩阵,并且然后将此M跨M的矩阵馈送到另一个神经网络滤波器用于最终评估(或通过实施卷积神经网络的对应的层来使用类似的技术。应理解,虽然一些参考文献已经提到神经网络滤波器,但是本文将不解释这样的滤波器的细节,因为可以设计各种数据处理技术并且特别是实施神经网络滤波器的数据处理技术,特别是参考滤波器层的数目和大小。通常,当然应理解,在本地装置上,通常将实施较少的层和/或较不复杂的层。

[0253] 在一个特别优选的实施方案中,啼哭检测包括(可选的)非常简单的(可选地:第一级)仅检测记录的声级超过阈值;啼哭检测的(第二)级包括在完全覆盖声音流(或可选地,紧接在过多的声级之前和之后的时间)的重叠时间窗中识别和隔离(声称的)啼哭模式,啼哭模式是通过在时间窗的类似谱图的表示中搜索已知的对应于婴幼儿啼哭识别的;使识别的和隔离的识别的啼哭模式经历模式分类,从而根据啼哭模式——特别是根据隔离的啼哭模式的类似谱图的表示——确定婴幼儿啼哭的原因的概率,并且此后根据对于隔离的啼哭模式的序列获得的概率识别婴幼儿啼哭的具体原因,其中模式分类和/或根据概率识别婴幼儿啼哭的具体原因的步骤中的至少一个通过根据个人婴幼儿数据个性化的评估来执行,特别是根据对于每个啼哭模式获得的概率识别婴幼儿啼哭的具体原因的步骤。

[0254] 无论如何实施神经网络滤波器和/或根据给定的输入的输出选择评估的算法的最终决定是什么,应理解,评估的可靠性在很大程度上取决于可用于分析的数据的质量。应理解,诸如交叉关联和/或滑动参数的技术对于更精确的分析是特别有用的,并且特别是在使评估个性化中,并且提供适合于这样的技术的数据对于允许改进的评估是至关重要的。在这方面,应注意,一个实施方案的实际实施方式,其中第一啼哭检测级简单地将当前声级与平均在前背景声级进行比较并且使紧接在响的噪声超过可比较的侵权(tort)阈值之前记录的声音数据和随后获得的声音数据经历基于类似谱图的表示的啼哭模式识别和隔离,在对啼哭的评估中产生特别良好的结果,并且已经被发现是特别鲁棒的,例如,不管声音记录装置相对于声音源的确切位置,不管具体的声音记录装置以及其使用的传声器和/或记录婴幼儿啼哭时存在的任何典型背景噪声。

[0255] 一旦已经获得评估结果,必须生成对应的输出。为此,评估的结果将被馈送到输出级(或“输出管理器”)以生成对应的输出信号,该输出信号可以是听觉信号、视觉信号,例如显示在监测器上的模式或闪烁的LED。

[0256] 应理解,指示输出可以由被适配以改进用户的体验的特定的输出级来实现。例如,

在最初评估被传送到仅与来自一个或两个缓冲器的声音数据相关的输出级的情况下,输出管理器可以抑制最初评估的输出,或如果自通知用户关于啼哭的发作以来的时间已经是短的,则可能抑制最初评估的输出。不立即示出最初评估帮助避免因评估随时间变化而使用户感到困惑。此外,在用户已经设置偏好的情况下,例如指示用户希望使两个最可能的原因与相应的概率一起被指示,输出管理器可以按照要求准备这样的输出。

[0257] 在评估级指示啼哭检测可能由于误报已经被触发的情况下,也可以向用户示出对应的信息和/或可以提出移动婴幼儿的对应的请求。此外,如果最初已经进行了第一评估,特别是以足够高的概率并且基于大量的帧和/或啼哭模式,则可能发生如下情形:与第一评估不同的第二评估也将是合理的,例如因为婴幼儿啼哭的原因实际上已经改变。然而,可以优选的是,在一定时间段(诸如2分钟或3分钟)内防止评估的改变,再次以避免用户感到困惑。

[0258] 一旦婴幼儿在给定的时间——例如30秒、一分钟、2分钟、3分钟、4分钟或5分钟——之后已经停止啼哭,在啼哭已经停止之后,对应的LED的闪烁的显示或听觉信号的输出可以结束,并且代替地可以生成诸如“听取你的婴幼儿”或“跟踪音频流”的标准消息。在一些情况下,向用户示出婴幼儿一直啼哭的原因可以是有帮助的,因为通常,在婴幼儿特别筋疲力尽的情况下,尽管先前的不舒服的原因可能保持有效,但是婴幼儿入睡。因此,显示先前的评估一些时间可以对于父母或护理者是有帮助的。

[0259] 然而,在通常的情况下,预期的是父母或护理者在婴幼儿还在啼哭时就照顾婴幼儿。在那,他们将通常试图安抚婴幼儿,例如通过喂养饥饿的婴幼儿、通过帮助婴幼儿打饱嗝或通过安慰婴幼儿直到其入睡。根据显示的评估以及他们试图根据评估安抚婴幼儿的成功或失败,反馈可以被录入本地装置。如果反馈被传输到集中式服务器——优选地以使得反馈可以与先前的评估的声音数据相关的方式,这是特别有帮助的。应理解,这样的反馈帮助改进啼哭数据库,并且特别是帮助提供用户标记的样本用于改进数据库。应理解,通过将反馈、使反馈与特定的声音数据和先前得出的任何评估相关的信息以及个性化信息上传到集中式服务器,对于集中式服务器的操作者(因为上传帮助扩大数据库)和对于父母(因为具有多个标记的声音啼哭)二者可以获得优点,并且优选地个性化信息还帮助改进个性化,例如通过识别具有类似啼哭模式的其他婴幼儿的同龄群体。这帮助区别婴幼儿群体,即使其他参数(诸如性别、年龄、大小、体重)是相同的。因此,个性化被改进。另外,在父母录入的参数(诸如大小或体重)是过时的或已经被不正确地录入时,考虑到实际啼哭的这样的个性化是有帮助的。

[0260] 另外,应理解,一旦已经识别了其他婴幼儿的同龄群体,对于这样的同龄群体获得的信息就可以被用于被发现具有与一个同龄群体的啼哭模式类似的啼哭模式的特定婴幼儿。例如,在发现给定的婴幼儿的所有啼哭与患有特定罕见疾病的其他婴幼儿的同龄群体的啼哭非常类似的情况下,可以向父母发出对应的警告。应理解,在父母和其他护理者提供反馈的情况下,可以实施用于奖励父母和其他护理者的方法;例如,在已经实施订阅模式的情况下,可以进行退款或延长当前订阅而不要求附加的支付。因此,在一个优选的实施方案中,提供了激励生成手段和/或激励步骤以将反馈上传到集中式服务器。应理解,在与集中式服务器的连接被中断的情况下,相关数据被存储在本地装置中直到在已经传送的数据中建立了连接。

[0261] 应理解,反馈不仅需要与转化的准确性相关,而且还可以提供与啼哭检测的准确性相关的反馈。还应理解,存在多种实施反馈的方式,例如,使用被作为远离本地装置的站的智能电话上的应用程序、按下本地装置上的按钮或对着本地装置的传声器说话以确认或拒绝评估。

[0262] 根据数据库的大小,神经网络滤波器并且因此评估起初可能不像在稍后时间的已经收集了更多样本并且可以区别更多情况时的评估那样具体;因此,可以使用通用啼哭检测滤波器。然而,随着数据库增长,滤波器将变得越来越具体。

[0263] 此后,可以预期的是,一旦已经收集了足够多的样本——例如从将发出的声音不同于较小、更轻的婴幼儿的特别重或大的婴幼儿——就可以作出进一步区别。滤波器的更新可以是自动的,例如每周一次,使滤波器适应于用于特定年龄/同龄群体的平均通用滤波器。此外,虽然根据当前的知识,在非常幼小的新生儿的婴幼儿啼哭中没有发现显著差异,但是应预期的是,婴儿变得年龄越大,啼哭的差异化就将越多,这取决于例如原始国或他定期附带的母语。因此,通过对啼哭检测滤波器的合适的适配,可能可以使用该装置更长时间和/或对年龄较大的孩子获得更精确的结果,这在分析父母不知道的非标准啼哭(诸如与特定疾病相关的那些啼哭)的情况下是特别有利的。

[0264] 在此上下文中,可以假设,婴幼儿在年龄增长时可能以与在其相同年龄/相同性别/相同大小的婴幼儿的同龄群体中的其他、类似的婴幼儿相同的方式长大,并且因此,将经历其话音器官的类似的发育。只要父母未录入矛盾的信息和/或只要通过父母反馈标记的啼哭与同一同龄群体中的婴幼儿的对应的啼哭并非不同,此假设就可以被认为有效,使得通常可以为相应的同龄群体确定滤波器。然而,每次滤波器被更新和/或特定的时间段已经过去,也可以实现对同龄群体的重新评估。

[0265] 上传到集中式服务器的数据将被录入到数据库中,并且数据库中的通过用户反馈标记的声音样本将被重复地使用,以重新训练在啼哭转化中使用的神经网络滤波器,并且只要背景噪声也被传输到数据库,就可以被使用以重新训练在啼哭检测中使用的神经网络滤波器。关于神经网络滤波器的训练、根据新颖事物对数据库的重新训练等,认为这样的技术是本领域众所周知的。这允许提供自适应滤波器,即使在仅上传针对特定孩子的有限量的数据的情况下,例如因为父母出于隐私原因而不希望传输数据。

[0266] 上文,已经提及,所述方法将是适用的并且装置将可用在儿科站中。在儿科站中,本地装置可以从不止一个婴幼儿拾取声音。例如,对于对双胞胎的监视同样成立。在存在本地装置可能从不止一个婴幼儿拾取声音的危险的设置,存在多个可能性。首先,将可能将本地装置连接到多个传声器,其中每个传声器放置得非常靠近婴幼儿中的一个。然后,根据从这些传声器中的每个接收的声音强度,可以作出哪个婴幼儿正在啼哭的决定。在使用多个本地装置,而不是使用连接到单个本地装置电线电缆的多个传声器的情况下,装置可以交换关于记录在每个本地装置处的声音强度的信息,并且决定可以基于交换的信息。应理解,这甚至对于同卵双胞胎起作用。另一个可能性将是检测任何啼哭,并且使用多个不同的个性化来评估每个啼哭,其中每个个性化对应于被监测的婴幼儿中的一个。对于个性化中的每个,可以确定对啼哭的评估是正确的可能性,并且可以发布具有最高可能性的评估。另一个可能性将是指示所有可能的评估,并且让护理者决定哪个婴幼儿正在啼哭,相应地决定哪个评估是相关的。这对于儿科站可以是一个优选的实施方式。因此,装置可以容易地被同

时用于多个婴幼儿。

[0267] 因此,上文所提议的内容尤其并且在不限本申请的情况下是一种提供用于自动婴幼儿啼哭评估的数据的计算机实施的方法,所述方法包括以下步骤:声学监测婴幼儿并且提供对应的声音数据流,在所述声音数据流中检测啼哭,响应于检测到啼哭而从声音数据选择啼哭相关的数据,根据选择的啼哭数据确定允许啼哭评估的参数,确定用于个性化啼哭评估的个人婴幼儿数据,根据个人婴幼儿数据准备用于评估的评估级,以及将所述参数馈送到根据个人婴幼儿数据准备的所述啼哭评估级。应注意,在一个具体的实施方案中,所述参数可以是啼哭模式已经被识别的时间和/或以类似谱图的方式对应于隔离的啼哭模式的声音数据的表示。

[0268] 此外,对于提议的方法,已经公开了,婴幼儿可以被连续声学监测并且啼哭前声音数据被至少临时地存储,直到随后的声音数据已经被发现不是啼哭相关的,例如,因为在其中未识别具有足够高的概率的啼哭模式。还公开了,在如提议的方法中,基于以下中的至少一个来检测婴幼儿啼哭,特别是检测在连续声学监测流中婴幼儿啼哭的发作:当前声级超过阈值,当前声级超过平均背景噪声一给定的裕度,一个或多个频带中的当前声级超过阈值,一个或多个频带中的当前声级超过对应的平均背景噪声一给定的裕度,声音的时间模式,声级的时间模式和/或谱模式偏离突然的响的非啼哭噪声的时间模式和/或谱模式,非声学提示,特别是从婴幼儿的视频监视数据、移动检测器和/或呼吸检测器得出的。

[0269] 还应注意,确定允许啼哭评估的参数所根据的选择的数据可以包括自啼哭事件的发作起的声音数据,特别是自啼哭的最初两秒起的声音数据,优选地自啼哭的最初二秒起的声音数据,特别优选地自啼哭的最初500毫秒起的声音数据。这可以通过从其(如果存在的话)隔离啼哭模式的方式检查包括在声级的增加之前的时间的窗来完成。此外,已经公开了一种计算机实施的方法,所述方法附加地可以包括以下步骤:本地检测从声学监测的婴幼儿获得的声音中的啼哭,并且将数据上传到在集中式自动婴幼儿啼哭评估中使用的服务器装置中,特别是将用于评估婴幼儿啼哭的选择的数据上传到云中。

[0270] 然后,还公开了,所述方法可以包括将与婴幼儿啼哭的声学监测相关的数据和/或与允许啼哭评估的选择的啼哭相关的参数上传到云的步骤,和/或包括将啼哭中的至少一些和/或从啼哭得出的参数连同个人婴幼儿数据存储在服务器上并且以根据存储在服务器上的信息来建立评估的步骤。

[0271] 应理解,如公开的和提议的方法也可以包括从集中式服务器下载允许本地个性化婴幼儿啼哭评估——特别是允许有限时间内的本地个性化婴幼儿啼哭评估——的信息的步骤。应注意,在如公开的和提议的优选的计算机的实施方法中,在啼哭的发作之前获取的监测声音数据被用于确定声学背景和/或确定用于婴幼儿啼哭评估的附加的参数,特别是如果不能以足够高的概率确定确切的发作。

[0272] 然后,还提议了,以允许使用神经网络和/或人工智能技术来评估啼哭的方式将参数馈送到啼哭评估级中,特别是其中馈送到啼哭评估级中的参数是通过迁移学习获得的和/或通过仅对单个婴幼儿的啼哭训练模型获得的。

[0273] 此外,已经公开了,所述方法还可以包括:上传参数和/或记录的声音数据流连同婴幼儿数据信息,特别是与以下中的至少一个相关的婴幼儿数据信息:年龄、性别、大小、体重、种族渊源、单胞胎/双胞胎/三胞胎、当前医疗状态、已知医疗先决条件、特别是已知的当

前疾病和/或发烧、父母和/或护理者的语言；和/或上传与一个或多个先前的评估的准确性相关的婴幼儿数据信息。

[0274] 应注意，还公开了，选择根据选择的啼哭数据中确定的参数，使得允许对“婴幼儿疲倦”、“婴幼儿饥饿”、“婴幼儿需要安慰”、“婴幼儿需要拍嗝”、“婴幼儿疼痛”中的至少一个状况的评估。

[0275] 应注意，还公开了一种自动婴幼儿啼哭评估装置，在一个实施方案中，所述自动婴幼儿啼哭评估装置包括：传声器，用于连续地声学监测婴幼儿；数字转换级，用于将监测声音流转换成数字数据流；存储器级，用于存储个人婴幼儿数据信息；通信级，用于将数据传输到集中式服务器装置，其中设置用于在所述数字数据流中识别啼哭的发作的啼哭识别级并且所述通信级被适配以从所述集中式服务器装置接收与婴幼儿啼哭的个性化评估相关的数据。

[0276] 此外，还公开了，所述自动婴幼儿啼哭评估装置还可以包括反馈装置，用于获得与一个或多个先前的评估的准确性相关的反馈信息，并且其中所述通信级被适配用于将反馈信息传输到集中式服务器装置。据提议，自动婴幼儿啼哭评估装置可以包括本地评估级，所述本地评估级被适配以根据从所述集中式服务器装置接收的与婴幼儿啼哭的个性化评估相关的数据来评估婴幼儿啼哭。应注意，所述自动婴幼儿啼哭评估装置可以包括定时器和评价级，所述评价级：在婴幼儿啼哭的评估之前，评价从所述集中式服务器装置接收的并且与婴幼儿啼哭的个性化的评估相关的数据的：个人婴幼儿数据信息的当前年龄和/或年龄或有效性，所述婴幼儿啼哭评估装置被适配以根据所述评价输出婴幼儿啼哭评估。

[0277] 虽然鉴于如提交的权利要求的引用，某些特征被明确公开为可组合的，但是并不意在将本公开内容仅限制于在如原始提交的权利要求中公开的组合。例如，根据所附权利要求2所述的方法的一个实施方案可以是优选的，其中用于啼哭模式检测和隔离的卷积神经网络是非个性化的，并且其中搜索啼哭模式是使用用于识别啼哭模式的卷积神经网络实现的，而不是基于声音数据的类似谱图的表示。此外，还将可以例如可能使用如权利要求5中所描述的计算机实施的方法，并且上传任何声音相关的数据——诸如识别的并且隔离的啼哭模式表示——连同婴幼儿数据信息，所述婴幼儿数据信息与至少多个年龄、性别、大小、体重、种族渊源、单胞胎/双胞胎/三胞胎、当前医疗状态、已知的医疗先决条件、特别是已知的当前疾病和/或发烧、父母和/或护理者的语言——相关；和/或上传与一个或多个先前的评估的准确性相关的婴幼儿数据信息，即使用于啼哭模式检测和隔离的机器学习模型是个性化的。

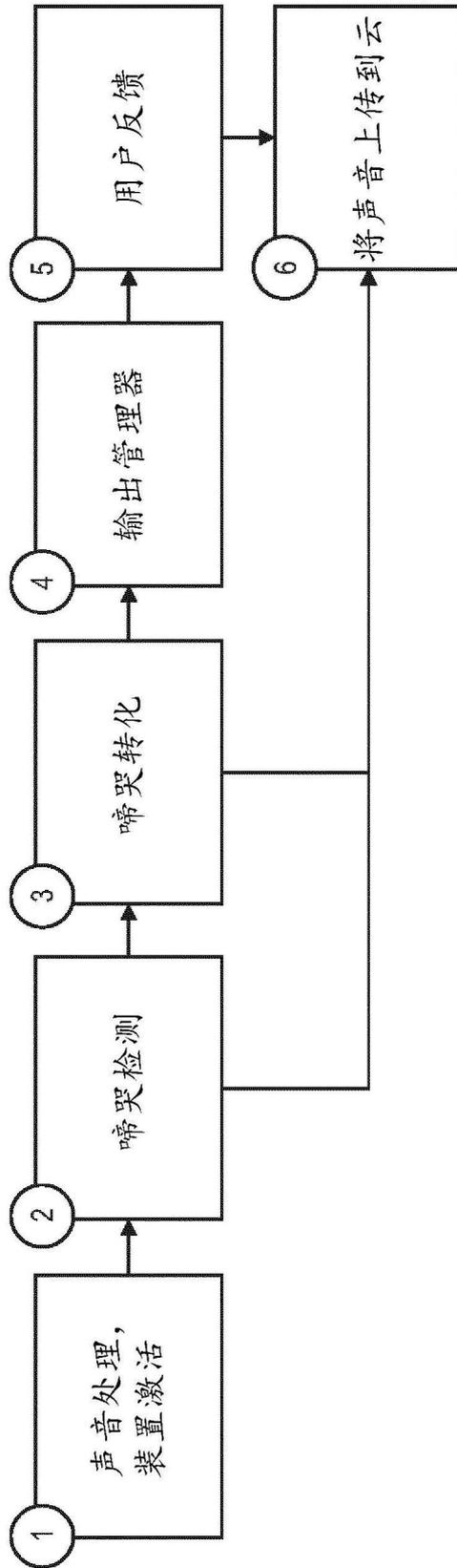


图1a

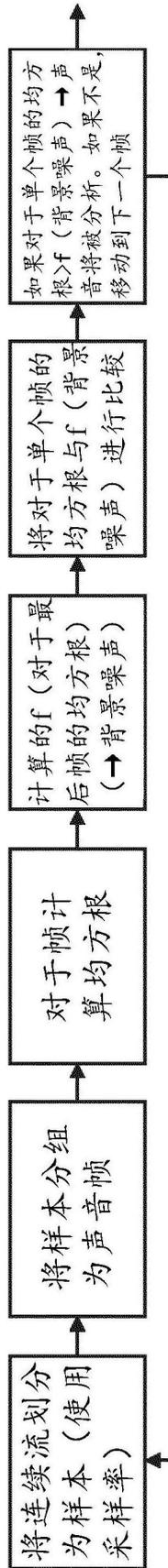
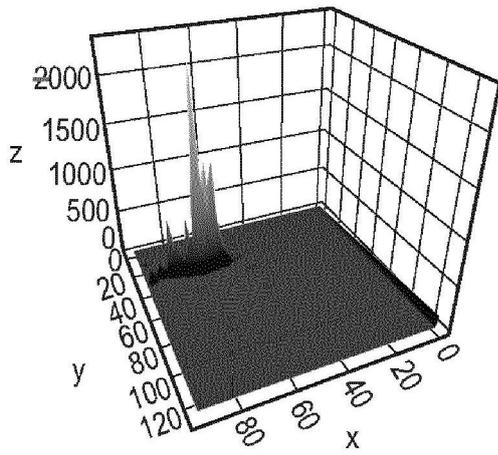


图1b

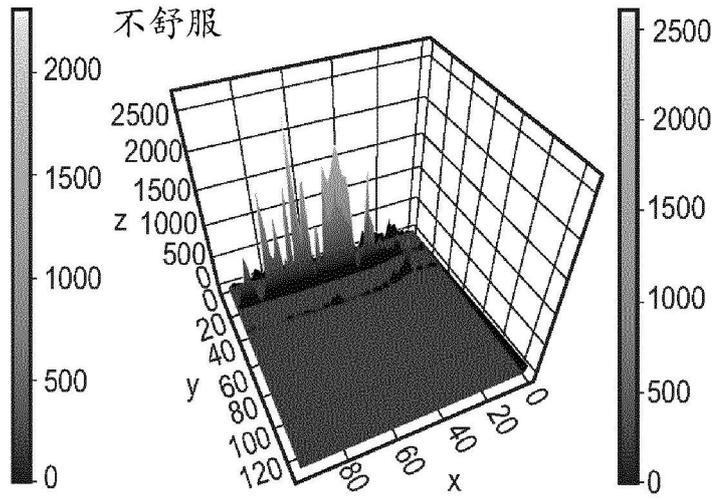


图2

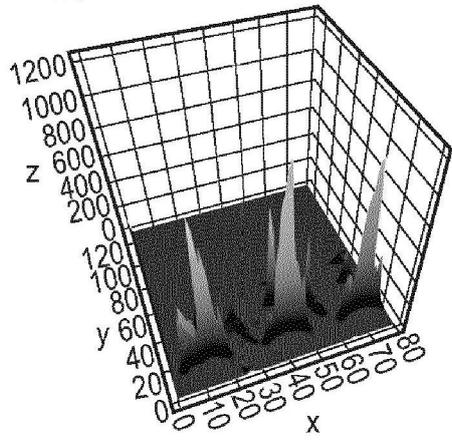
打饱嗝



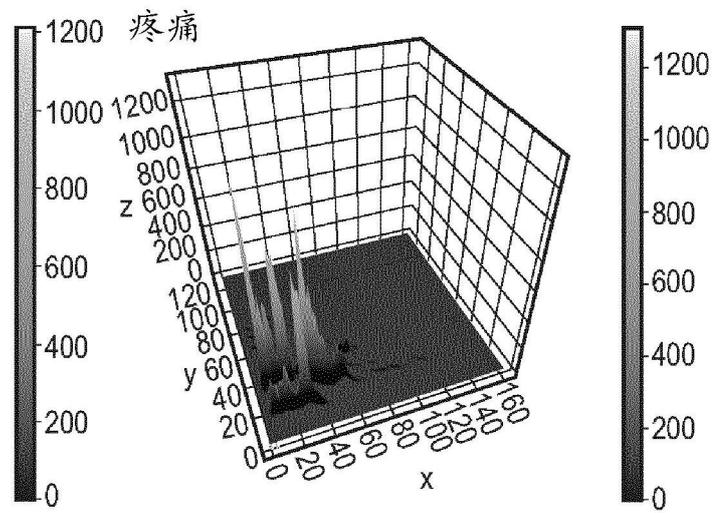
不舒服



饥饿



疼痛



困倦

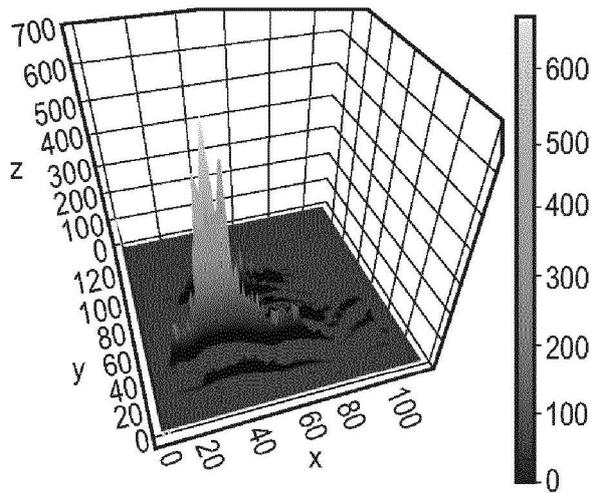


图3a-e: 三维谱图

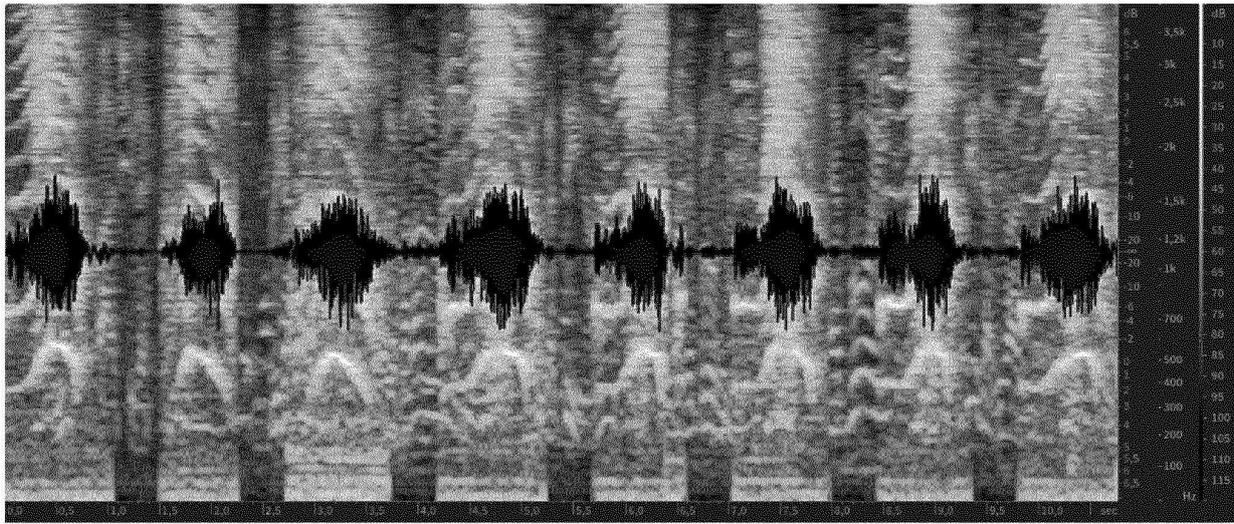


图4a:来自不同饥饿婴幼儿的啼哭

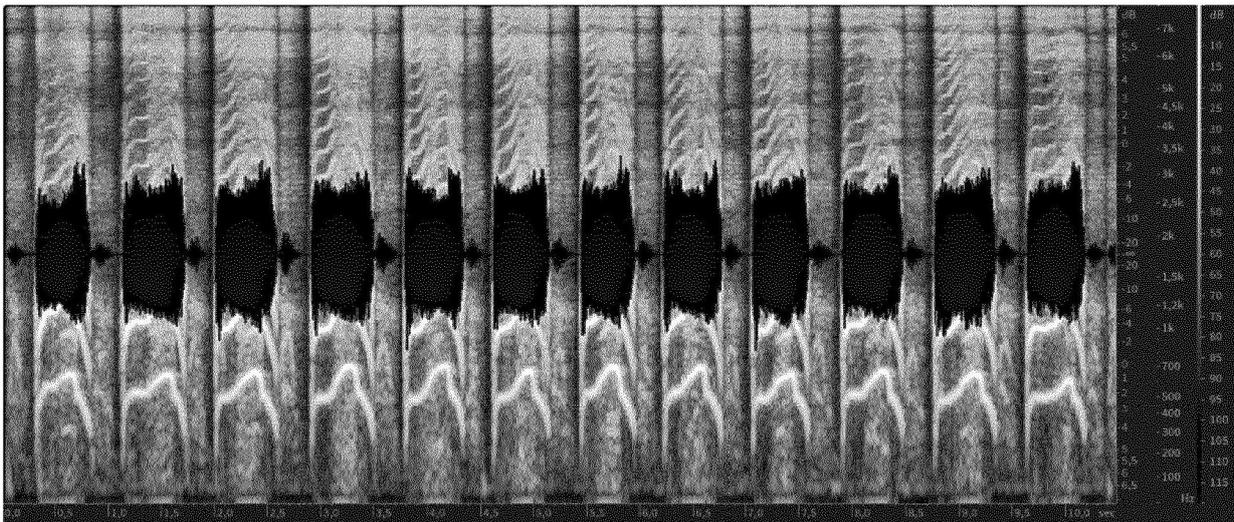


图4b:来自同一饥饿婴幼儿的不同啼哭

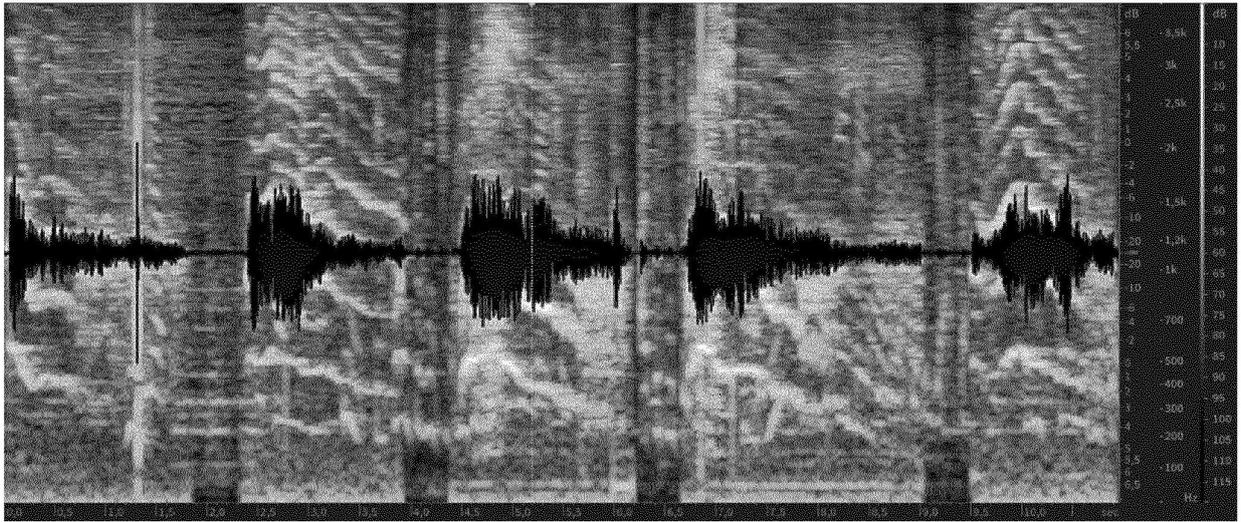


图4c:来自疼痛中的不同婴幼儿的啼哭

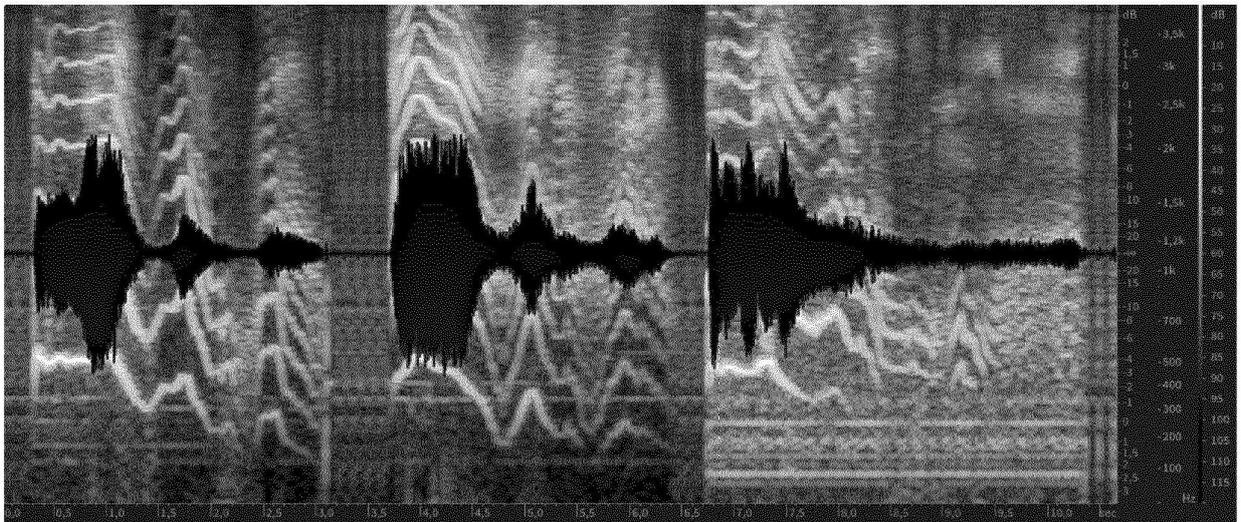


图4d:来自疼痛中的同一婴幼儿的不同啼哭

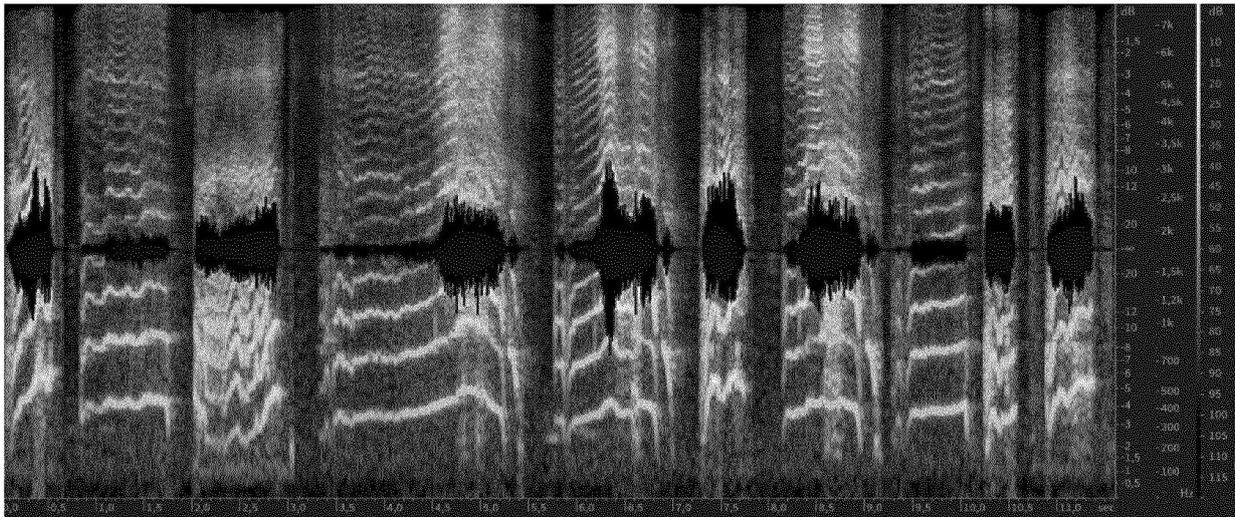


图4e:来自需要拍嗝的不同婴幼儿的啼哭

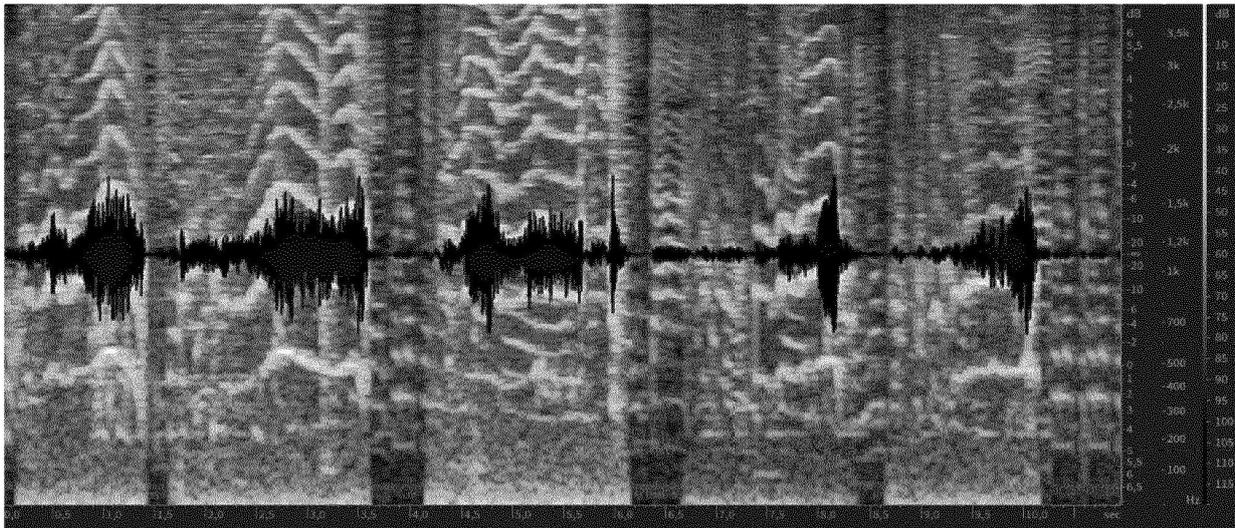


图4f:来自需要拍嗝的同一婴幼儿的不同啼哭

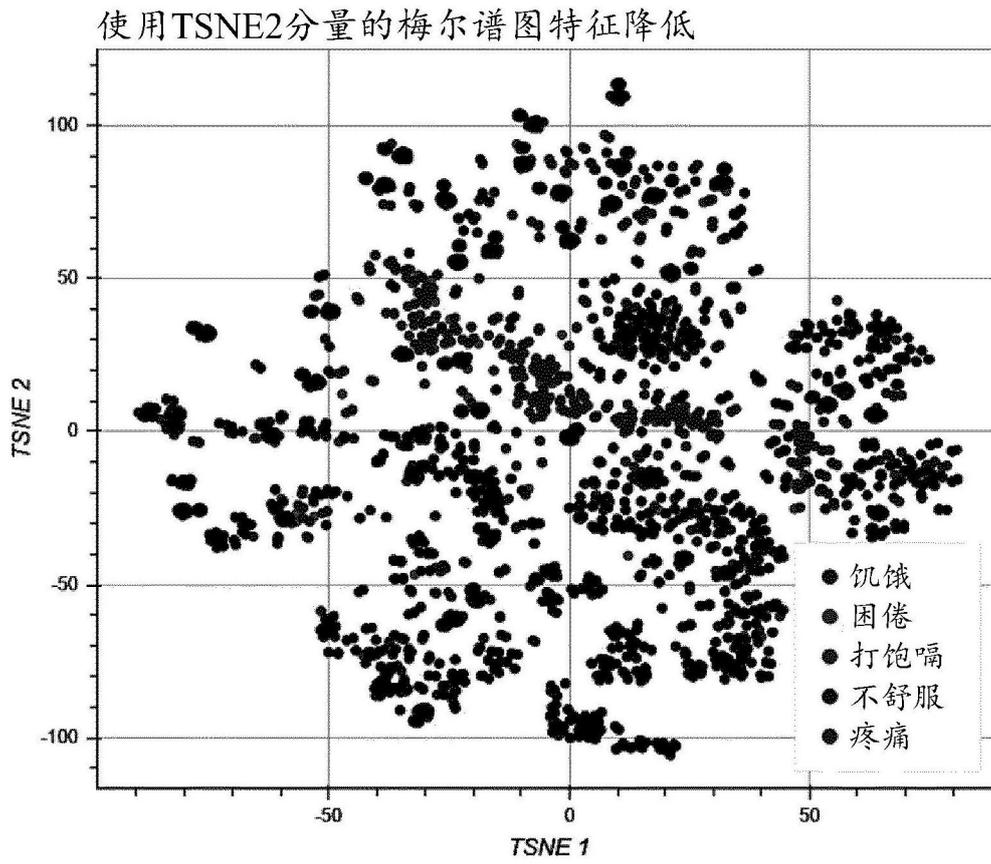


图5a啼哭的总体聚类

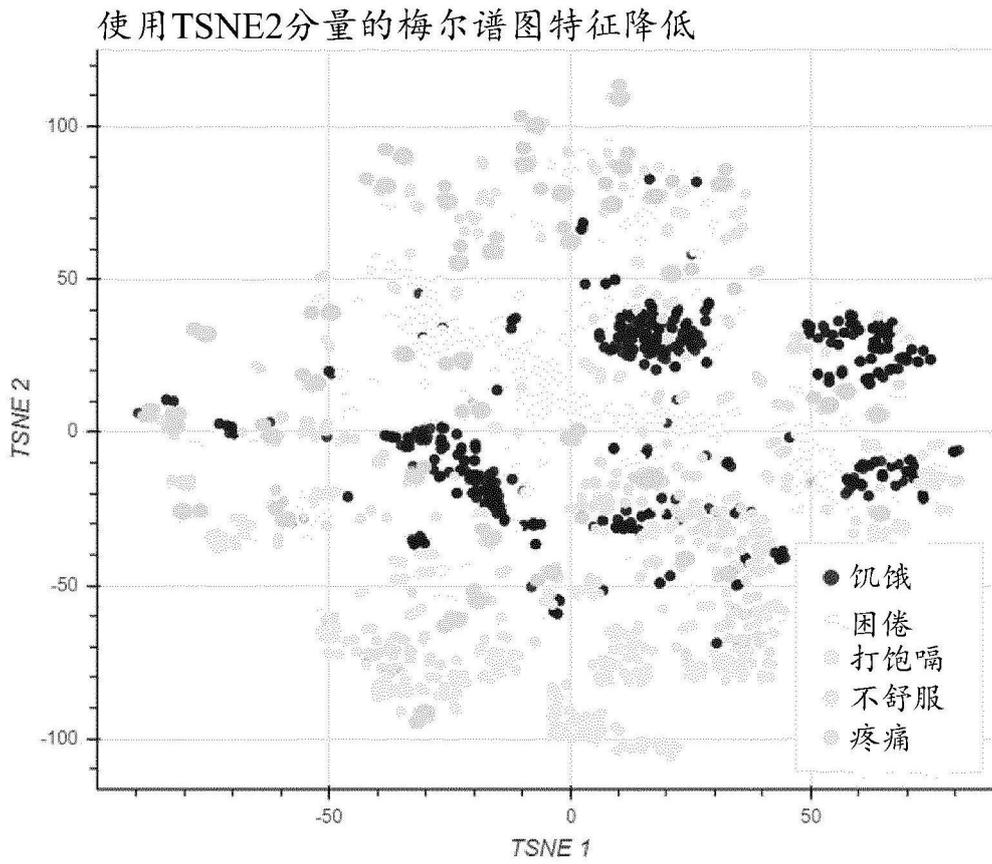


图5b总体聚类中的饥饿啼哭

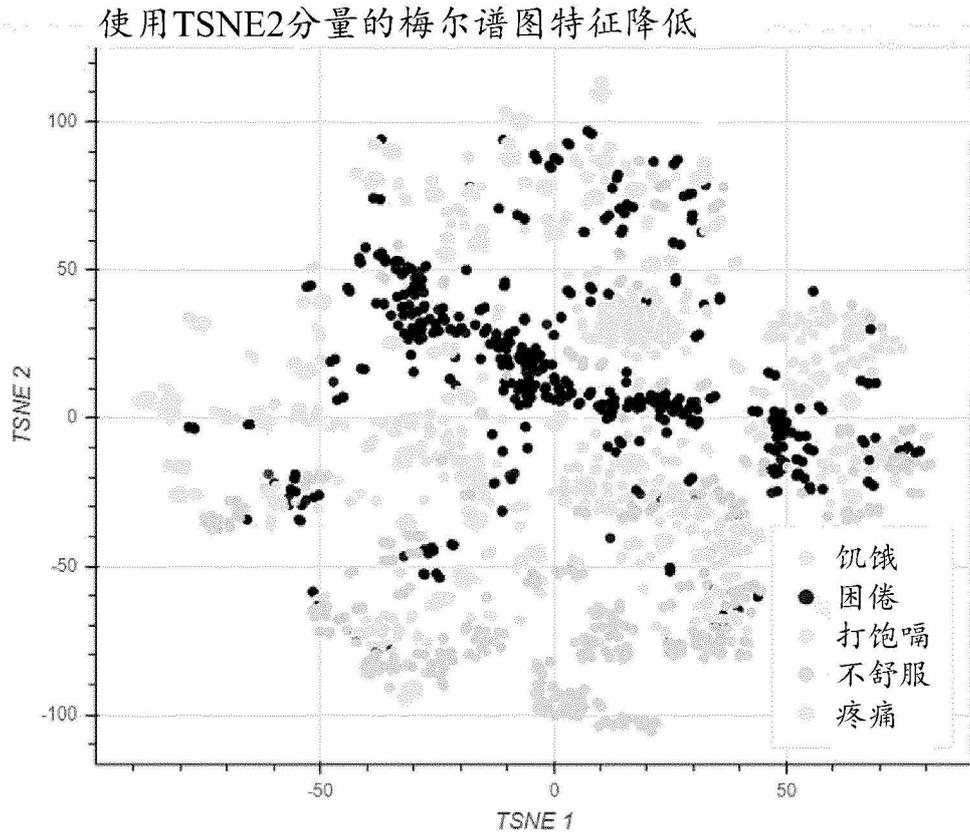


图5c“总体聚类中的困倦啼哭”

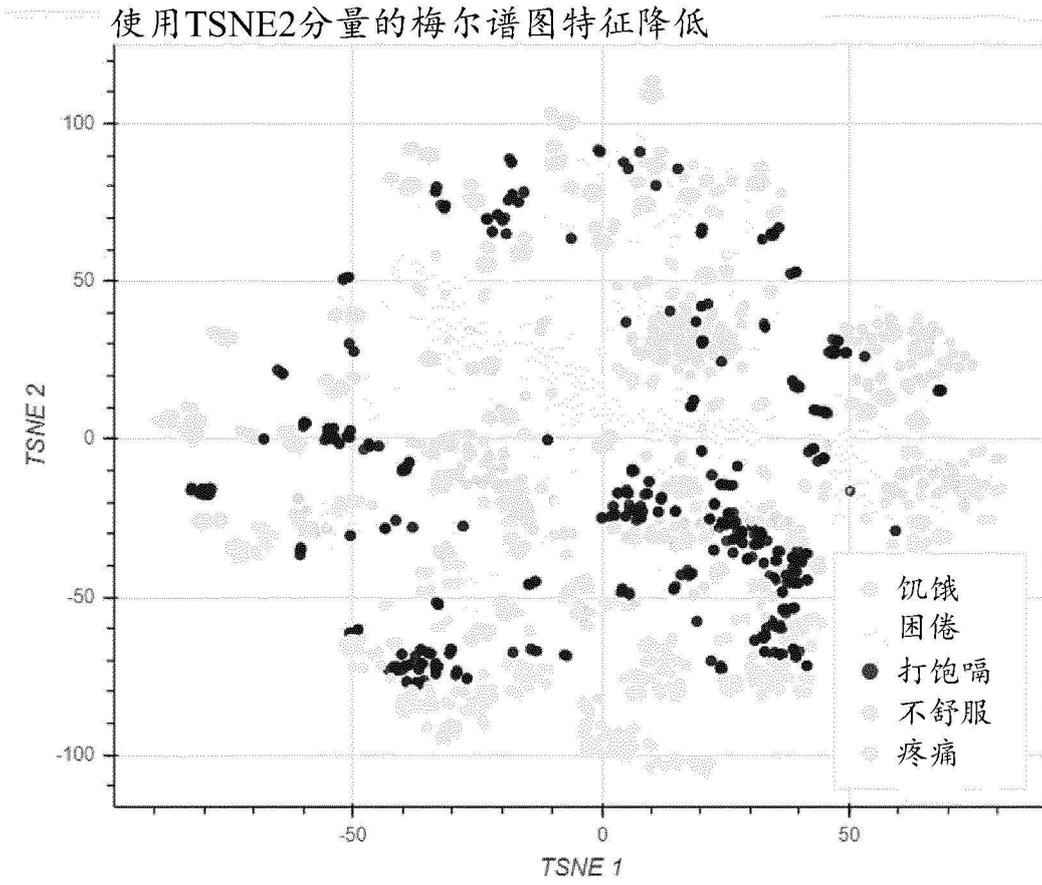


图5d总体聚类中的“需要拍嗝”啼哭

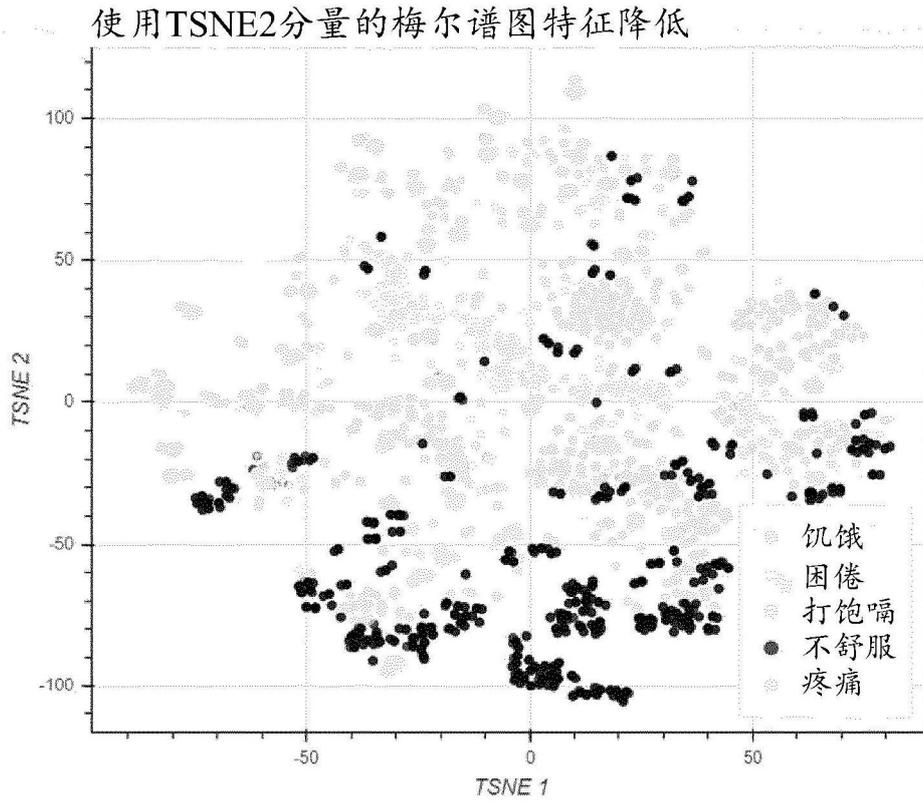


图5e总体聚类中的不舒服啼哭

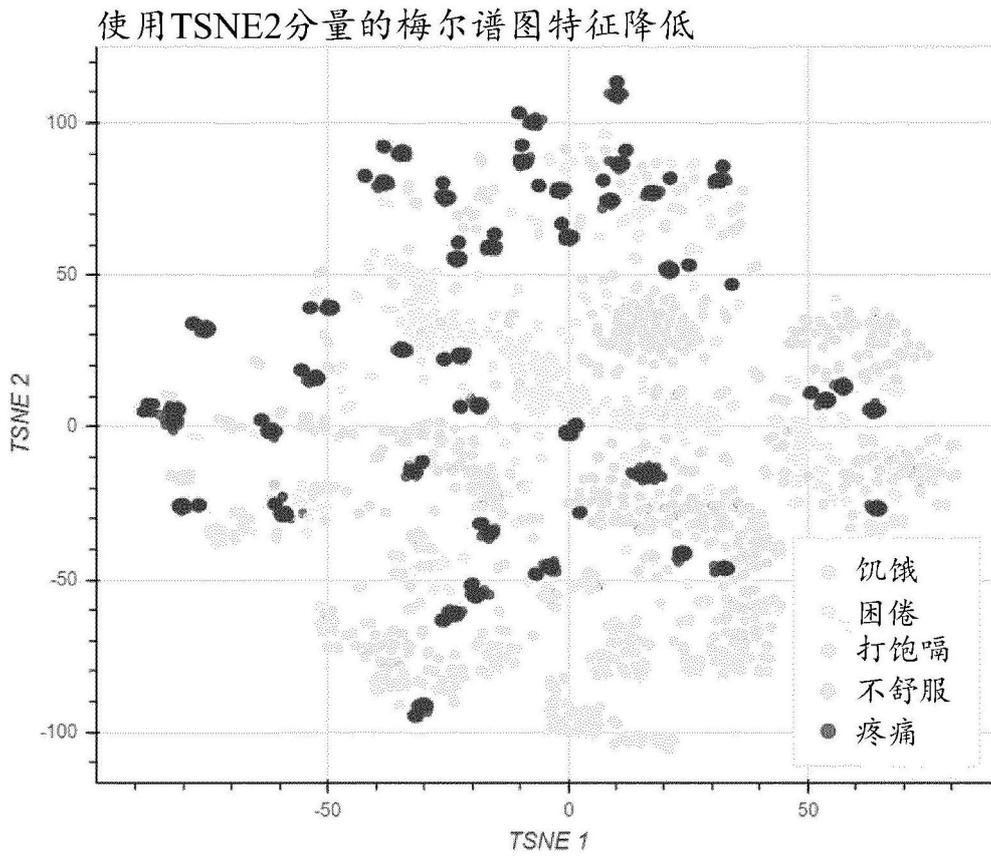
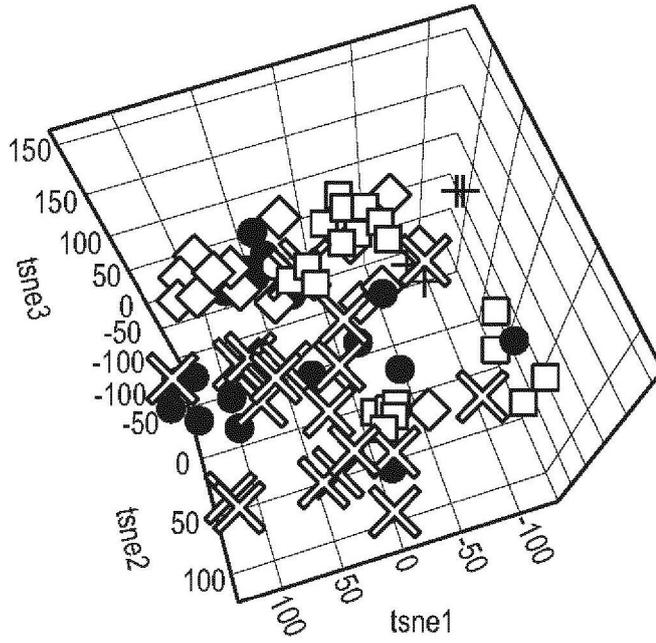


图5f总体聚类中的疼痛啼哭

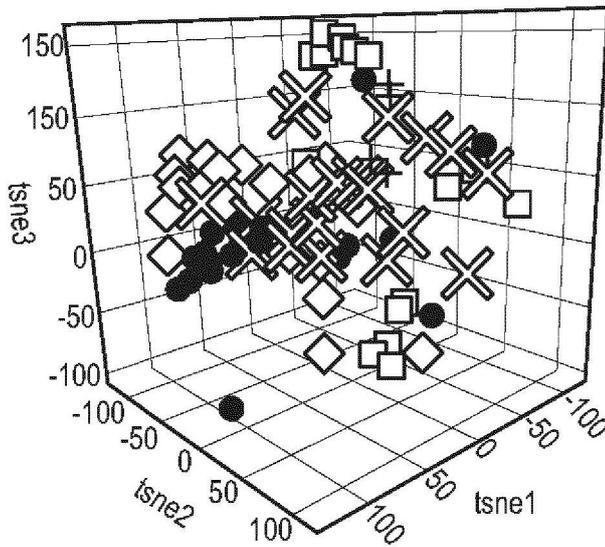
T-SNE 降维梅尔谱图



标签

- 打饱嗝
- ◆ 不舒服
- 饥饿
- × 困倦
- ⊕ 疼痛

T-SNE 降维梅尔谱图



标签

- 打饱嗝
- ◆ 不舒服
- 饥饿
- × 困倦
- ⊕ 疼痛

图6从两个不同角度示出的T-SNE降维梅尔谱图的三维表示

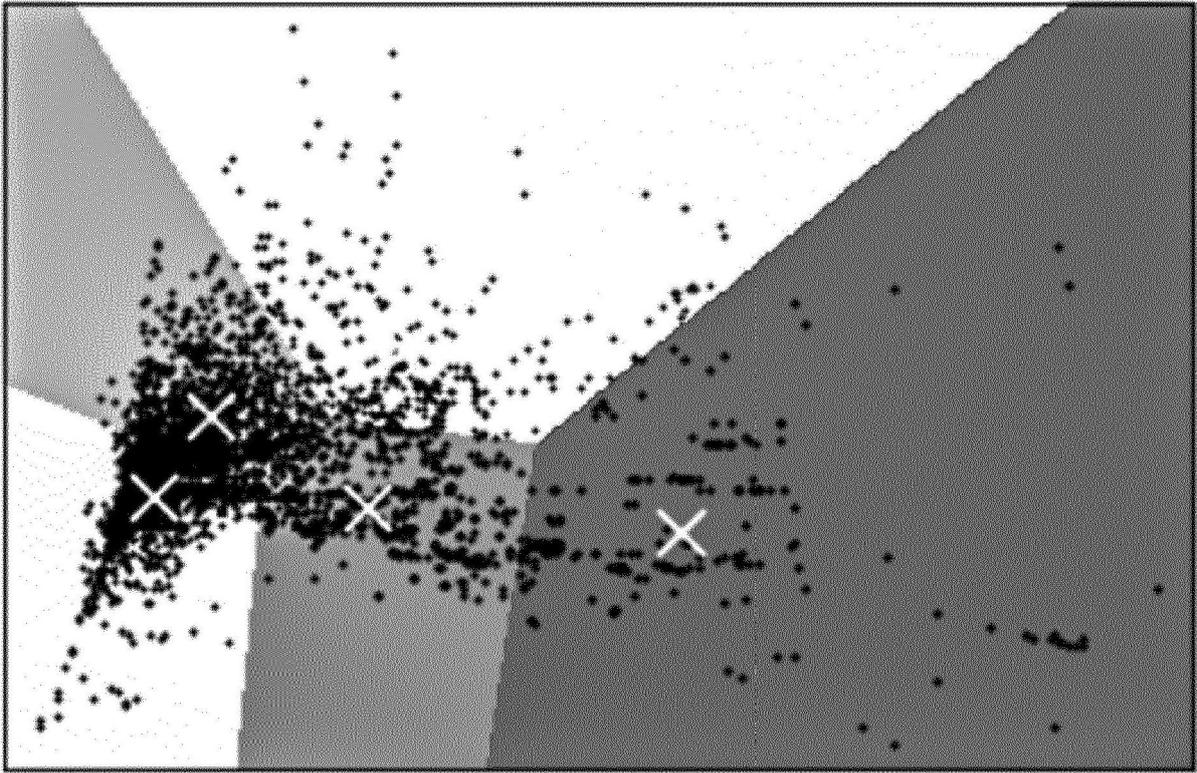


图7K-均值聚类将用于5个不同的啼哭模式中的每个的聚类的形心示出为白色叉形记号,并且划分成不同的单元