(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2024 年 1 月 25 日 (25.01.2024)

WIPO | PCT

(10) 国际公布号
WO 2024/017110 A1

(51) 国际专利分类号:
G10L 21/0208 (2013.01) *G10L 21/0232* (2013.01)
G10L 21/0216 (2013.01) *G10L 25/30* (2013.01)
G10L 21/0224 (2013.01)

(21) 国际申请号: PCT/CN2023/106951

(22) 国际申请日: 2023 年 7 月 12 日 (12.07.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202210864010.4 2022年7月21日 (21.07.2022) CN

(71) 申请人: 广州市百果园信息技术有限公司 (**GUANGZHOU BAIGUOYUAN INFORMATION TECHNOLOGY CO., LTD.**) [CN/CN]; 中国广东省广州市番禺区南村镇万博二路79号201-203房、401-410房、801-811房、1001-1011房、1101-1111房, Guangdong 511402 (CN)。

(72) 发明人: 魏善义 (**WEI, Shanyi**); 中国广东省广州市番禺区南村镇万博二路79号201-203房、401-410房、801-811房、1001-1011房、1101-1111房, Guangdong 511402 (CN)。 刘梁 (**LIU, Liang**); 中国广东省广州市番禺区南村镇万博二路79号201-203房、401-410房、801-811房、1001-1011房、1101-1111房, Guangdong 511402 (CN)。

(74) 代理人: 北京品源专利代理有限公司 (**BEYOND ATTORNEYS AT LAW**); 中国北京市海淀区莲花池东路39号西金大厦6层, Beijing 100036 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(54) **Title:** VOICE NOISE REDUCTION METHOD, MODEL TRAINING METHOD, APPARATUS, DEVICE, MEDIUM, AND PRODUCT

(54) 发明名称: 语音降噪方法、模型训练方法、装置、设备、介质及产品

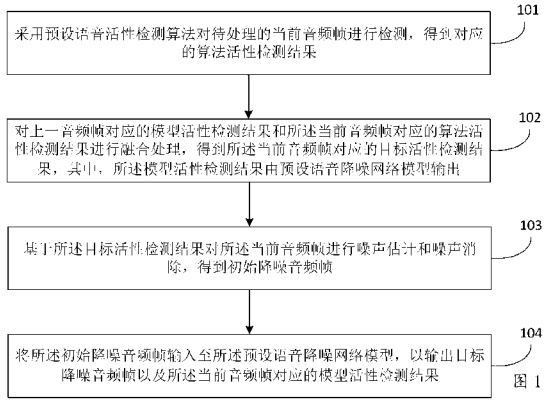


图 1

101 Use a preset voice activity detection algorithm to detect a current audio frame to be processed, and obtain a corresponding algorithm activity detection result

102 Merge a model activity detection result corresponding to the previous audio frame with an algorithm activity detection result corresponding to the current audio frame, and obtain a target activity detection result corresponding to the current audio frame, wherein the model activity detection result is outputted by a preset voice noise reduction network model

103 Based on the target activity detection result, perform noise estimation and noise elimination on the current audio frame, and obtain an initial noise reduction audio frame

104 Input the initial noise reduction audio frame into the preset voice noise reduction network model so as to output a target noise reduction audio frame and also a model activity detection result corresponding to the current audio frame

(57) **Abstract:** A voice noise reduction method, a model training method, an apparatus, a device, a medium, and a product. The voice noise reduction method comprises: using a preset voice activity detection algorithm to detect a current audio frame to be processed, and obtaining a corresponding algorithm activity detection result [101]; merging a model activity detection result corresponding to the previous audio frame with an algorithm activity detection result corresponding to the current audio frame, obtaining a target activity detection result corresponding to the current audio frame, wherein the model activity detection result is outputted by a preset voice noise reduction network model [102]; based on the target activity detection result, performing noise estimation and noise elimination on the current audio frame, obtaining an initial noise reduction audio frame [103]; and inputting the initial noise reduction audio frame into the preset voice noise reduction network model so as to output a target noise reduction audio frame and also a model activity detection result corresponding to the current audio frame [104]. The use of the aforementioned solution can enhance a voice noise reduction effect, and the stability and robustness of the voice noise reduction solution are improved.

WO 2024/017110 A1

[见续页]

(84) 指定国(除另有指明，要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:
— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种语音降噪方法、模型训练方法、装置、设备、介质及产品。其中，该语音降噪方法包括：采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果[101]；对上一音频帧对应的模型活性检测结果和当前音频帧对应的算法活性检测结果进行融合处理，得到当前音频帧对应的目标活性检测结果，模型活性检测结果由预设语音降噪网络模型输出[102]；基于目标活性检测结果对当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧[103]；将初始降噪音频帧输入至预设语音降噪网络模型，以输出目标降噪音频帧以及当前音频帧对应的模型活性检测结果[104]。通过采用上述方案，可以提升语音降噪效果，并提高语音降噪方案的稳定性和鲁棒性。

语音降噪方法、模型训练方法、装置、设备、介质及产品

本公开要求在2022年7月21日提交中国专利局、申请号为202210864010.4的中国专利的优先权，以上申请的全部内容通过引用结合在本申请中。

技术领域

本申请涉及音频处理技术领域，例如涉及语音降噪方法、模型训练方法、装置、设备、介质及产品。

背景技术

随着多媒体技术的飞速发展，各种会议、社交以及娱乐类应用程序层出不穷，其中涉及语音通话、音视频直播以及多人会议等诸多场景，而语音质量是衡量应用性能的重要指标。

终端设备的麦克风所采集的语音通常带有一定程度的噪声，通过语音降噪算法可抑制语音中所携带的噪声，从而提高语音的可懂度和话音质量。

目前，语音降噪方案大致可以分为两大类：传统降噪方案和人工智能（Artificial Intelligence，AI）降噪方案。传统降噪方案是以信号处理的方式实现语音降噪，无法消除非稳态的噪声，也即对突发噪声的降噪能力较弱；AI降噪方案对稳态噪声和非稳态噪声都具有较好的降噪能力，但该方案为数据驱动方案，非常依赖于训练样本，在模型训练过程中如果存在未考虑到的场景（例如信噪比很低的情况），那么在实际应用中遇到此场景可能导致不可估计的信号输出，甚至系统崩溃。

发明内容

本申请实施例提供了语音降噪方法、模型训练方法、装置、设备、介质及产品，可以将传统降噪方案和AI降噪方案进行有效结合，提升语音降噪效果。

根据本申请的一方面，提供了一种语音降噪方法，该方法包括：

采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果；

对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，其中，

所述模型活性检测结果由预设语音降噪网络模型输出；

基于所述目标活性检测结果对所述当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧；

将所述初始降噪音频帧输入至所述预设语音降噪网络模型，以输出目标降噪音频帧以及所述当前音频帧对应的模型活性检测结果。

根据本申请的另一方面，提供了一种模型训练方法，包括：

采用预设语音活性检测算法对当前样本音频帧进行检测，得到对应的样本算法活性检测结果，其中，所述当前样本音频帧关联有活性检测标签和纯净音频帧；

对上一样本音频帧对应的样本模型活性检测结果和所述当前样本音频帧对应的样本算法活性检测结果进行融合处理，得到所述当前样本音频帧对应的目标样本活性检测结果，其中，所述样本模型活性检测结果由语音降噪网络模型输出；

基于所述目标活性样本检测结果对所述当前样本音频帧进行噪声估计和噪声消除，得到初始降噪样本音频帧；

将所述初始降噪样本音频帧输入至所述语音降噪网络模型，以输出目标样本降噪音频帧以及所述当前样本音频帧对应的样本模型活性检测结果；

根据所述目标样本降噪音频帧和所述纯净音频帧确定第一损失关系，根据所述样本模型活性检测结果和所述活性检测标签确定第二损失关系，并基于所述第一损失关系和所述第二损失关系对所述语音降噪网络模型进行训练。

根据本申请的另一方面，提供了一种语音降噪装置，该装置包括：

语音活性检测模块，设置为采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果；

检测结果融合模块，设置为对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，其中，所述模型活性检测结果由预设语音降噪网络模型输出；

降噪处理模块，设置为基于所述目标活性检测结果对所述当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧；

模型输入模块，设置为将所述初始降噪音频帧输入至所述预设语音降噪网络模型，以输出目标降噪音频帧以及所述当前音频帧对应的模型活性检测结果。

根据本申请的另一方面，提供了一种模型训练装置，包括：

语音检测模块，设置为采用预设语音活性检测算法对待处理的当前样本音频帧进行检测，得到对应的样本算法活性检测结果，其中，所述当前样本音频帧关联有活性检测标签和干净音频帧；

融合模块，设置为对上一样本音频帧对应的样本模型活性检测结果和所述当前样本音频帧对应的样本算法活性检测结果进行融合处理，得到所述当前样本音频帧对应的目标样本活性检测结果，其中，所述样本模型活性检测结果由语音降噪网络模型输出；

噪声消除模块，设置为基于所述目标活性样本检测结果对所述当前样本音频帧进行噪声估计和噪声消除，得到初始降噪样本音频帧；

网络模型输入模块，设置为将所述初始降噪样本音频帧输入至所述语音降噪网络模型，以输出目标样本降噪音频帧以及所述当前样本音频帧对应的样本模型活性检测结果；

网络模型训练模块，设置为根据所述目标样本降噪音频帧和所述干净音频帧确定第一损失关系，根据所述样本模型活性检测结果和所述活性检测标签确定第二损失关系，并基于所述第一损失关系和所述第二损失关系对所述语音降噪网络模型进行训练。

根据本申请的另一方面，提供了一种电子设备，所述电子设备包括：

至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的计算机程序，所述计算机程序被所述至少一个处理器执行，以使所述至少一个处理器能够执行本申请任一实施例所述的语音降噪方法和/或模型训练方法。

根据本申请的另一方面，提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序用于使处理器执行时实现本申请任一实施例所述的语音降噪方法和/或模型训练方法。

根据本申请的另一方面，提供了一种计算机程序产品，所述计算机程序产品包括计算机程序，所述计算机程序在被处理器执行时实现本申请任一实施例所述的语音降噪方法和/或模型训练方法。

本申请实施例中提供的语音降噪方案，采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果，对上一音频帧对应的模型活性检测结果和当前音频帧对应的算法活性检测结果进行融合处理，得到当前音频帧对应的目标活性检测结果，模型活性检测结果由预设语音降噪网

络模型输出，基于目标活性检测结果对当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧，将初始降噪音频帧输入至预设语音降噪网络模型，以输出目标降噪音频帧以及当前音频帧对应的模型活性检测结果。通过采用上述方案，预设语音降噪网络模型能够输出模型活性检测结果，在采用传统语音降噪算法对当前音频帧进行处理时，可以对上一音频帧的模型活性检测结果和传统语音降噪算法得到的算法活性检测结果进行结合，使传统降噪算法可以获得更多的活性检测信息，更加合理准确地确定语音活性检测结果，基于该结果进行噪声估计和噪声消除，可以更好的保护语音以及更多的消除噪声，得到信噪比更高的传统降噪结果，再将传统降噪结果作为预设语音降噪网络模型的输入，得到效果更好的降噪音频帧，降低了预设语音降噪网络模型处理恶劣数据的可能性，传统降噪算法和 AI 降噪方法相互促进，对各种噪声具有较好的降噪能力，可以提升语音降噪效果，并提高整体的语音降噪方案的稳定性和鲁棒性。

附图说明

下面将对实施例描述中所需要使用的附图作介绍，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请实施例提供的一种语音降噪方法的流程示意图；

图 2 为本申请实施例提供的又一种语音降噪方法的流程示意图；

图 3 为本申请实施例提供的一种语音降噪方法的推理流程示意图；

图 4 为本申请实施例提供的一种模型训练方法的流程示意图；

图 5 为本申请实施例提供的一种模型训练方法的训练过程示意图；

图 6 为本申请实施例提供的一种语音降噪装置的结构框图；

图 7 为本申请实施例提供的一种模型训练装置的结构框图；

图 8 为本申请实施例提供的一种电子设备的结构框图。

具体实施方式

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

需要说明的是，本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

图1为本申请实施例提供的一种语音降噪方法的流程示意图，本实施例可适用于对语音进行降噪的情况，例如可以适用于如语音通话、音视频直播以及多人会议等各种场景。该方法可以由语音降噪装置执行，该语音降噪装置可以采用硬件和/或软件的形式实现，该语音降噪装置可配置于语音降噪设备等电子设备中。所述电子设备可以为手机、智能手表、平板电脑以及个人数字助理等移动设备；也可为台式计算机等其他设备。如图1所示，该方法包括：

步骤101、采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果。

示例性的，待处理的当前音频帧可以理解为当前需要进行语音降噪处理的音频帧，当前音频帧可以包含于音频文件或音频流中。可选的，当前音频帧可以是音频文件或音频流中的原始音频帧，也可以是对原始音频帧进行预处理后得到的音频帧。

本申请实施例中，语音降噪方案整体可以理解为一个语音降噪系统，当前音频帧可以理解为语音降噪系统的输入信号。语音降噪方案中可包含传统语音降噪算法和AI语音降噪模型。

其中，传统语音降噪算法的类型例如可以是网络即时通信（Web Real-Time Communication，webRTC）中的自适应噪音抑制（Adaptive Noise Suppression，ANS）算法、线性滤波法、谱减法、统计模型算法或子空间算法等。传统语音降噪算法中主要包括语音活性检测（Voice Activity Detection，VAD）估计、噪声估计和噪声消除三大部分。语音活性检测，又称语音端点检测或语音边界检测，可以从声音信号流里识别长时间的静音期。本申请实施例中的预设语音活性检测算法，可以是任意传统语音降噪算法中的语音活性检测算法。

其中，本申请中的预设语音降噪网络模型，可以是AI语音降噪模型，可包括如RNNoise模型、或双路信号变换长短期记忆人工神经网络的实时噪声抑制（Dual-Signal Transformation LSTM Network for Real-Time Noise Suppression，DTLN）降噪模型等。预设语音降噪网络模型中包括两个分支，一个分支用于输

出降噪语音（可简称为降噪分支），另一个分支用于输出语音活性检测结果（可简称为检测分支）。对于已包含检测分支的AI语音降噪模型来说，可以保持原有模型结构；对于未包含检测分支的AI语音降噪模型来说，可以在主干网络基础上，添加检测分支，检测分支的网络结构例如可以包括卷积层和/或全连接层等。

其中，RNNoise是一种采用音频特征抽取+深度神经网络结合的降噪方案。

示例性的，为了便于区分不同来源的语音活性检测结果，采用预设语音活性检测算法对待处理的当前音频帧进行检测后，可将得到的检测结果记为算法活性检测结果，将预设语音降噪网络模型输出的活性检测结果记为模型活性检测结果。

步骤102、对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，其中，所述模型活性检测结果由预设语音降噪网络模型输出。

示例性的，上一音频帧可以理解为当前音频帧之前的最近一个音频帧，也即，上一音频帧位于当前音频帧之前且两者帧序号相邻。在对上一音频帧进行语音降噪处理时，预设语音降噪网络模型可以输出上一音频帧对应的降噪音频帧和模型活性检测结果，可对该模型活性检测结果进行缓存，以用于对当前音频帧的降噪处理。

本申请实施例中，在对当前音频帧进行处理时，可以综合上一音频帧对应的模型活性检测结果和当前音频帧对应的算法活性检测结果，来确定用于供传统语音降噪算法中的噪声估计和噪声消除所使用的活性检测结果（目标活性检测结果）。相比于单纯采用传统语音降噪算法来进行语音活性检测来说，使传统降噪算法可以获得更多的VAD信息，从而得到更准确的噪声估计，可以更好的保护语音并更准确地消除噪声，可提升传统降噪算法的输出信噪比（Signal to Noise Ratio，SNR）。

步骤103、基于所述目标活性检测结果对所述当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧。

示例性的，在得到目标活性检测结果后，可以利用传统语音降噪算法中的噪声估计算法和噪声消除算法，对当前音频帧进行相应处理，将处理后得到的音频帧记为初始降噪音频帧。

步骤104、将所述初始降噪音频帧输入至所述预设语音降噪网络模型，以输出目标降噪音频帧以及所述当前音频帧对应的模型活性检测结果。

示例性的，在得到初始降噪音频帧后，可以直接将初始降噪音频帧作为预

设语音降噪网络模型的输入，也可以根据预设语音降噪网络模型的特点对初始降噪音频帧进行转换，例如转换为预设维度的信号，预设维度例如可以是频域、时域或其他维度域。

本申请实施例中提供的语音降噪方法，采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果，对上一音频帧对应的模型活性检测结果和当前音频帧对应的算法活性检测结果进行融合处理，得到当前音频帧对应的目标活性检测结果，模型活性检测结果由预设语音降噪网络模型输出，基于目标活性检测结果对当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧，将初始降噪音频帧输入至预设语音降噪网络模型，以输出目标降噪音频帧以及当前音频帧对应的模型活性检测结果。通过采用上述方案，预设语音降噪网络模型能够输出模型活性检测结果，在采用传统语音降噪算法对当前音频帧进行处理时，可以对上一音频帧的模型活性检测结果和传统语音降噪算法得到的算法活性检测结果进行结合，使传统降噪算法可以获得更多的活性检测信息，更加合理准确地确定语音活性检测结果，基于该结果进行噪声估计和噪声消除，可以更好的保护语音、更多的消除噪声，得到信噪比更高的传统降噪结果，再将传统降噪结果作为预设语音降噪网络模型的输入，得到效果更好的降噪音频帧，降低了预设语音降噪网络模型处理恶劣数据的可能性，传统降噪算法和AI降噪方法相互促进，对各种噪声具有较好的降噪能力，提高方案整体的稳定性和鲁棒性。

本申请实施例中，语音活性检测可以是帧级别的，也可以是频点级别的，检测结果可以用一个或多个概率值来表示。

在一些实施例中，所述算法活性检测结果包括对应音频帧中存在语音的第一概率值，所述模型活性检测结果包括对应音频帧中存在语音的第二概率值。其中，所述对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，包括：采用预设计算方式，对上一音频帧对应的模型活性检测结果中的第一概率值，和所述当前音频帧对应的算法活性检测结果中的第二概率值进行计算，得到第三概率值，根据所述第三概率值确定所述当前音频帧对应的目标活性检测结果。这样设置，对于帧级别的语音活性检测，可以准确地确定目标活性检测结果。

其中，第一概率值用于表示采用预设语音活性检测算法对对应音频帧进行检测后，得到对应音频帧中包含语音的概率，这里的对应音频帧可以是任意音频帧，可以是当前音频帧，也可以是上一音频帧，不同音频帧对应的第一概率值可以不同；第二概率值用于表示由预设语音降噪网络模型输出的，对应音频

帧中包含语音的概率，这里的对应音频帧也可以是任意音频帧，不同音频帧对应的第二概率值可以不同。

示例性的，当前音频帧对应的算法活性检测结果中的第一概率值，可以用于表示采用预设语音活性检测算法对当前音频帧（假设记为A）进行检测后，得到的当前音频帧中包含语音的概率，可记为Pa。上一音频帧对应的模型活性检测结果中的第二概率值可以用于表示在对上一音频帧（假设记为B）进行语音降噪处理时，预设语音降噪网络模型所预测的上一音频帧中包含语音的概率，可记为Pb。采用预设计算方式对Pa和Pb进行计算，得到第三概率值，可记为Pc。示例性的，可将第三概率值作为当前音频帧对应的目标活性检测结果。

示例性的，所述预设计算方式为取最大值、取最小值、计算平均值、求和、计算加权和、以及计算加权平均值中的一种。以取最大值为例，Pc=max（Pa，Pb）。

在一些实施例中，所述算法活性检测结果包括对应音频帧中，预设数量的频点中每个频点存在语音的第四概率值；所述模型活性检测结果包括对应音频帧中，所述预设数量的频点中每个频点存在语音的第五概率值；其中，所述对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，包括：针对所述预设数量的频点中的每个频点，采用预设计算方式，对上一音频帧对应的模型活性检测结果中的单个频点的第五概率值，和所述当前音频帧对应的算法活性检测结果中的对应的所述单个频点的第四概率值进行计算，得到第六概率值；根据所述预设数量的第六概率值，确定所述当前音频帧对应的目标活性检测结果。这样设置，采用频点级别的语音活性检测，可以更加精准地确定目标活性检测结果。

示例性的，预设数量（记为n）可以根据实际需求设定，例如可以根据预处理阶段时的快速傅里叶变换所采用的点数确定，例如n为256。当前音频帧对应的第四概率值可以用于表示采用预设语音活性检测算法对当前音频帧（假设记为A）进行检测后，得到的当前音频帧中的预设数量的频点中每个频点包含语音的概率，可记为PA[n]，PA[n]可理解为包含n个元素（n位）的向量，每个元素的取值在0至1之间，一个元素的取值用于表示对应的频点中包含语音的概率。上一音频帧对应的第五概率值可以用于表示在对上一音频帧（假设记为B）进行语音降噪处理时，预设语音降噪网络模型所预测的上一音频帧中的预设数量的频点中每个频点包含语音的概率，可记为PB[n]。采用预设计算方式对PA[n]和PB[n]进行计算，得到预设数量的第六概率值，例如可记为PC[n]。示例性的，可将包含第六概率值的向量作为当前音频帧对应的目标活性检测结果。

示例性的，所述预设计算方式为取最大值、取最小值、计算平均值、求和、计算加权和、以及计算加权平均值中的一种。以取最大值为例，PC[n]=max（PA[n]，PB[n]）。例如，对于当前音频帧中的第一个频点，对应的第四概率值和第五概率值中的最大值，成为当前音频帧中的第一个频点对应的第六概率值，后续频点以此类推。

在一些实施例中，所述将所述初始降噪音频帧输入至所述预设语音降噪网络模型，包括：对所述初始降噪音频帧进行预设特征维度的特征提取，得到目标输入信号；将所述目标输入信号输入至所述预设语音降噪网络模型，或者，将所述目标输入信号和所述初始降噪音频帧输入至所述预设语音降噪网络模型。这样设置，可以有针对性地进行特征提取，提高预设语音降噪网络模型的预测准确度和精度。

可选的，预设特征维度包括显性特征维度，可以是基频特征，如基音频率（Pitch），还可以是每通道能量归一化（Per-channel energy normalization，PCEN）特征、或梅尔频率倒谱系数（Mel Frequency Cepstrum Coefficient, MFCC）特征等等。预设特征维度可根据预设语音降噪网络模型的网络结构或特点等来确定。

图2为本申请实施例提供的又一种语音降噪方法的流程示意图，该方法在上述各可选实施例基础上进行优化，图3为本申请实施例提供的一种语音降噪方法的推理流程示意图，可结合图2和图3对本申请实施例方案进行理解。其中，如图2所示，该方法可包括：

步骤201、获取原始音频帧，对原始音频帧进行预处理，得到待处理的当前音频帧。

示例性的，原始音频帧包含于音频文件或音频流中，例如，可以是语音通话场景中的音频流。为了保证通话质量，需要对通话音频进行降噪。预处理可以包括如分帧、加窗以及傅里叶变换等处理。经过预处理后的带噪语音帧即为待处理的当前音频帧，作为预设传统降噪算法的输入信号（记为S0）。

步骤202、采用预设传统降噪算法中的预设语音活性检测算法，对待处理的当前音频帧进行检测，得到对应的算法活性检测结果。

示例性的，预设传统降噪算法可以是ANS算法。利用ANS算法中的VAD估计功能模块对应的预设语音活性检测算法，对S0进行检测，假设为频点级别的检测，可以得到256个频点的语音存在概率Pf[256]，也即S0对应的算法活性检测结果。

步骤203、判断当前音频帧是否存在上一个音频帧，若是，则执行步骤204；否则，执行步骤206。

示例性的，对于第一个音频帧来说，不存在上一个音频帧，因此，可以不需要获取上一音频帧的模型活性检测结果，执行步骤206，基于当前音频帧对应的算法活性检测结果进行噪声估计和噪声消除。

步骤204、获取上一音频帧对应的模型活性检测结果，对所获取的模型活性检测结果和当前音频帧对应的算法活性检测结果进行融合处理，得到当前音频帧对应的目标活性检测结果。

示例性的，上一音频帧对应的模型活性检测结果由基于人工智能的预设语音降噪网络模型输出，可以是上一音频帧中256个频点的语音存在概率PF[256]，可以采用取最大值的方式得到融合的VAD估计结果（目标活性检测结果）：P[256]=max(Pf[256]，PF[256])。

步骤205、基于目标活性检测结果，利用所述预设传统降噪算法对当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧，执行步骤207。

示例性的，预设传统降噪算法根据P[256]，实现噪声估计和噪声消除，得到经过传统降噪处理的语音信号S1，也即初始降噪音频帧。

步骤206、基于当前音频帧对应的算法活性检测结果，利用所述预设传统降噪算法对当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧。

示例性的，预设传统降噪算法根据Pf[256]，实现噪声估计和噪声消除，得到经过传统降噪处理的语音信号S1，也即初始降噪音频帧。

步骤207、对初始降噪语音进行预设特征维度的特征提取，得到目标输入信号。

示例性的，S1作为预设语音降噪网络模型的输入信号，可以是频域、时域或其他维度域的信号，根据预设语音降噪网络模型的模型设计不同，可能存在一步显性的特征提取计算，如基音频率特征，将提取到的特征信息记为目标输入信号S2。

步骤208、将目标输入信号和/或初始降噪音频帧输入至预设语音降噪网络模型，以输出目标降噪音频帧以及当前音频帧对应的模型活性检测结果。

可选的，可以将S1或S2作为模型输入，还可以将S1和S2均作为模型输入，输入至预设语音降噪网络模型中进行推理计算，得出输出信号。输出信号包含两部分，第一部分是语音降噪方法的最终降噪语音的输出S3，第二部分是模型的VAD输出PF[256]，供传统语音降噪算法在处理下一音频帧时使用。

步骤209、判断是否存在待处理的原始音频帧，若是，则返回执行步骤201；否则，结束流程。

示例性的，若语音通话结束，所有原始音频帧已得到降噪处理，此时可以结束流程，若仍存在未降噪的原始音频帧，则可返回执行步骤201，继续进行降噪处理。

本申请实施例提供的语音降噪方法，通过基于人工智能的预设语音降噪网络模型向传统降噪算法进行信息反馈的方式，使传统降噪算法可以获得更多的VAD信息，传统降噪和AI降噪的VAD估计均采用频点级别，可以得到更精准的噪声估计，使得传统降噪算法可以更好的保护语音、更多的消除噪声，提升传统降噪的输出信噪比，高信噪比的初始降噪语音信号经过特征提取后，可以丰富预设语音降噪网络模型的输入，在降低预设语音降噪网络模型处理恶劣数据的可能性的同时，提升模型的语音降噪效果，提升语音降噪性能。

图4为本申请实施例提供的一种模型训练方法的流程示意图，图5为本申请实施例提供的一种模型训练方法的训练过程示意图，可结合图4和图4对本申请实施例进行理解。本实施例可适用于对基于人工智能的语音降噪网络模型进行训练的情况，该模型可以适用于如语音通话、音视频直播以及多人会议等各种场景。该方法可以由模型训练装置执行，该装置可以采用硬件和/或软件的形式实现，该装置可配置于模型训练设备等电子设备中。所述电子设备可以为手机、智能手表、平板电脑以及个人数字助理等移动设备；也可为台式计算机等其他设备。采用本申请实施例训练得到的语音降噪网络模型可以应用于本申请中任意实施例提供的语音降噪方法。

如图4所示，该方法包括：

步骤401、采用预设语音活性检测算法对当前样本音频帧进行检测，得到对应的样本算法活性检测结果，其中，所述当前样本音频帧关联有活性检测标签和纯净音频帧。

示例性的，可将纯净（干净）语音数据集合和噪声数据集按照预设混合规则混合成为带噪语音数据，预设混合规则例如可以基于信噪比或房间声学冲激响应（Room Impulse Response，RIR）来设定。可选的，将混合得到的带噪语音数据集和纯净语音数据集一起作为模型的训练集。当前样本音频帧可以是训练集中的音频帧。当前样本音频帧可以携带活性检测标签，该标签可以通过人工标注的方式添加。以帧级别为例，若包含语音，标签可以为1，若不包含语音，标签可以为0；以频点级别为例，标签可以是包含预设数量的元素的向量，每个元素的取值为1或0，对应频点若包含语音，取值为1，对应频点若不包含语音，取值为0。

步骤402、对上一样本音频帧对应的样本模型活性检测结果和所述当前样本音频帧对应的样本算法活性检测结果进行融合处理，得到所述当前样本音频帧

对应的目标样本活性检测结果，其中，所述样本模型活性检测结果由语音降噪网络模型输出。

示例性的，本步骤中的活性检测结果融合过程可以与本申请实施例提供的语音降噪方法中的融合过程类似，如可以是频点级融合或帧级别融合等，还可采用类似的预设计算方式对相应的频率值进行融合，具体细节可参考本文相关内容，此处不再赘述。

步骤403、基于所述目标活性样本检测结果对所述当前样本音频帧进行噪声估计和噪声消除，得到初始降噪样本音频帧。

步骤404、将所述初始降噪样本音频帧输入至所述语音降噪网络模型，以输出目标样本降噪音频帧以及所述当前样本音频帧对应的样本模型活性检测结果。

步骤405、根据所述目标样本降噪音频帧和所述纯净音频帧确定第一损失关系，根据所述样本模型活性检测结果和所述活性检测标签确定第二损失关系，并基于所述第一损失关系和所述第二损失关系对所述语音降噪网络模型进行训练。

示例性的，损失关系可以用于表征两种数据之间的差异，可以用损失值表示，例如可以采用损失函数来计算。第一损失关系用于表征目标样本降噪音频帧和纯净音频帧之间的差异，第二损失关系用于表征样本模型活性检测结果和活性检测标签之间的差异，其中，用于计算第一损失关系的第一损失函数，以及用于计算第二损失关系的第二损失函数的函数类型可根据实际需求进行设置。

示例性的，可基于所述第一损失关系和所述第二损失关系计算目标损失关系，计算方式例如可以是加权求和等。

示例性的，根据目标损失关系对语音降噪网络模型进行训练，在训练过程中，可以以最小化目标损失关系为目标，利用反向传播等训练手段不断优化语音降噪网络模型中的权重参数值，直到满足预设训练截止条件。训练截止条件可根据实际需求进行设置，例如可以基于迭代次数、损失值收敛程度、或模型准确率等设定。

本申请实施例提供的模型训练方法，在训练过程中，将传统降噪算法和语音降噪网络模型作为一个整体，可避免传统降噪算法串联单独训练的语音降噪网络模型所带来的数据失配风险，训练后得到的模型，可以用于语音降噪，并对各种噪声具有较好的降噪能力，提升降噪效果。

可选的，所述样本算法活性检测结果包括对应样本音频帧中存在语音的第一样本概率值，所述样本模型活性检测结果包括对应样本音频帧中存在语音的第二样本概率值；

其中，所述对上一样本音频帧对应的样本模型活性检测结果和所述当前样本音频帧对应的样本算法活性检测结果进行融合处理，得到所述当前样本音频帧对应的目标样本活性检测结果，包括：采用预设计算方式，对上一样本音频帧对应的样本模型活性检测结果中的第二样本概率值，和所述当前样本音频帧对应的样本算法活性检测结果中的第一样本概率值进行计算，得到第三样本概率值，根据所述第三样本概率值确定所述当前样本音频帧对应的目标样本活性检测结果。

可选的，所述样本算法活性检测结果包括对应音频帧中，预设数量的频点中每个频点存在语音的第四样本概率值；所述模型活性检测结果包括对应音频帧中，所述预设数量的频点中每个频点存在语音的第五样本概率值；

其中，所述对上一样本音频帧对应的样本模型活性检测结果和所述当前样本音频帧对应的样本算法活性检测结果进行融合处理，得到所述当前样本音频帧对应的目标样本活性检测结果，包括：针对所述预设数量的频点中的每个频点，采用预设计算方式，对上一样本音频帧对应的样本模型活性检测结果中的单个频点的第五样本概率值，和所述当前样本音频帧对应的样本算法活性检测结果中的对应的所述单个频点的第四样本概率值进行计算，得到第六样本概率值；根据所述预设数量的第六样本概率值，确定所述当前样本音频帧对应的目标样本活性检测结果。

可选的，所述将所述初始降噪样本音频帧输入至所述语音降噪网络模型，包括：对所述初始降噪样本音频帧进行预设特征维度的特征提取，得到目标输入信号；将所述目标输入信号输入至所述语音降噪网络模型，或者，将所述目标输入信号和所述初始降噪样本音频帧输入至所述语音降噪网络模型。

图6为本申请实施例提供的一种语音降噪装置的结构框图，该装置可由软件和/或硬件实现，一般可集成在语音降噪设备等电子设备中，可通过执行语音降噪方法来进行语音降噪。如图6所示，该装置包括：

语音活性检测模块601，设置为采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果；

检测结果融合模块602，设置为对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，其中，所述模型活性检测结果由预设语音降噪网络模型输出；

降噪处理模块603，设置为基于所述目标活性检测结果对所述当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧；

模型输入模块604，设置为将所述初始降噪音频帧输入至所述预设语音降噪网络模型，以输出目标降噪音频帧以及所述当前音频帧对应的模型活性检测结果。

本申请实施例提供的语音降噪装置，采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果，对上一音频帧对应的模型活性检测结果和当前音频帧对应的算法活性检测结果进行融合处理，得到当前音频帧对应的目标活性检测结果，模型活性检测结果由预设语音降噪网络模型输出，基于目标活性检测结果对当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧，将初始降噪音频帧输入至预设语音降噪网络模型，以输出目标降噪音频帧以及当前音频帧对应的模型活性检测结果。通过采用上述方案，预设语音降噪网络模型能够输出模型活性检测结果，在采用传统语音降噪算法对当前音频帧进行处理时，可以对上一音频帧的模型活性检测结果和传统语音降噪算法得到的算法活性检测结果进行结合，使传统降噪算法可以获得更多的活性检测信息，更加合理准确地确定语音活性检测结果，基于该结果进行噪声估计和噪声消除，可以更好的保护语音、更多的消除噪声，得到信噪比更高的传统降噪结果，再将传统降噪结果作为预设语音降噪网络模型的输入，得到效果更好的降噪音频帧，降低了预设语音降噪网络模型处理恶劣数据的可能性，传统降噪算法和AI降噪方法相互促进，对各种噪声具有较好的降噪能力，提高方案整体的稳定性和鲁棒性。

可选的，所述算法活性检测结果包括对应音频帧中存在语音的第一概率值，所述模型活性检测结果包括对应音频帧中存在语音的第二概率值；

其中，所述检测结果融合模块602设置为通过以下方式对所述模型活性检测结果和所述算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果：

采用预设计算方式，对上一音频帧对应的模型活性检测结果中的第二概率值，和所述当前音频帧对应的算法活性检测结果中的第一概率值进行计算，得到第三概率值，根据所述第三概率值确定所述当前音频帧对应的目标活性检测结果。

可选的，所述算法活性检测结果包括对应音频帧中，预设数量的频点中每个频点存在语音的第四概率值；所述模型活性检测结果包括对应音频帧中，所述预设数量的频点中每个频点存在语音的第五概率值；

其中，所述检测结果融合模块602还设置为通过以下方式对所述模型活性检测结果和所述算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果：

针对所述预设数量的频点中的每个频点，采用预设计算方式，对上一音频帧对应的模型活性检测结果中的单个频点的第五概率值，和所述当前音频帧对应的算法活性检测结果中的对应的所述单个频点的第四概率值进行计算，得到第六概率值；根据所述预设数量的第六概率值，确定所述当前音频帧对应的目标活性检测结果。

可选的，所述预设计算方式为取最大值、取最小值、计算平均值、求和、计算加权和、以及计算加权平均值中的一种。

可选的，所述模型输入模块，包括：

特征提取单元，设置为对所述初始降噪语音进行预设特征维度的特征提取，得到目标输入信号；

信号输入单元，设置为将所述目标输入信号输入至所述预设语音降噪网络模型，或者，将所述目标输入信号和所述初始降噪音频帧输入至所述预设语音降噪网络模型，以输出目标降噪音频帧以及所述当前音频帧对应的模型活性检测结果。

图7为本申请实施例提供的一种模型训练装置的结构框图，该装置可由软件和/或硬件实现，一般可集成在模型训练设备等电子设备中，可通过执行模型训练方法来进行模型训练。如图7所示，该装置包括：

语音检测模块701，设置为采用预设语音活性检测算法对待处理的当前样本音频帧进行检测，得到对应的样本算法活性检测结果，其中，所述当前样本音频帧关联有活性检测标签和干净音频帧；

融合模块702，设置为对上一样本音频帧对应的样本模型活性检测结果和所述当前样本音频帧对应的样本算法活性检测结果进行融合处理，得到所述当前样本音频帧对应的目标样本活性检测结果，其中，所述样本模型活性检测结果由语音降噪网络模型输出；

噪声消除模块703，设置为基于所述目标活性样本检测结果对所述当前样本音频帧进行噪声估计和噪声消除，得到初始降噪样本音频帧；

网络模型输入模块704，设置为将所述初始降噪样本音频帧输入至所述语音降噪网络模型，以输出目标样本降噪音频帧以及所述当前样本音频帧对应的样本模型活性检测结果；

网络模型训练模块705，设置为根据所述目标样本降噪音频帧和所述干净音频帧确定第一损失关系，根据所述样本模型活性检测结果和所述活性检测标签确定第二损失关系，并基于所述第一损失关系和所述第二损失关系对所述语音降噪网络模型进行训练。

本申请实施例提供的模型训练装置，在训练过程中，将传统降噪算法和语音降噪网络模型作为一个整体，可避免传统降噪算法串联单独训练的语音降噪网络模型所带来的数据失配风险，训练后得到的模型，可以用于语音降噪，并对各种噪声具有较好的降噪能力，提升降噪效果。

本申请实施例提供了一种电子设备，该电子设备中可集成本申请实施例提供的语音降噪装置和/或模型训练装置。图8为本申请实施例提供的一种电子设备的结构框图。电子设备800包括处理器801，以及与处理器801通信连接的存储器802，其中，存储器802存储有可被处理器801执行的计算机程序，计算机程序被处理器801执行，以使所述处理器801能够执行本申请任一实施例所述的语音降噪方法和/或模型训练方法。其中，处理器的数量可以是一个或多个，图8中以一个处理器为例。

本申请实施例还提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序用于使处理器执行时实现本申请任一实施例所述的语音降噪方法和/或模型训练方法。

本申请实施例还提供一种计算机程序产品，所述计算机程序产品包括计算机程序，所述计算机程序在被处理器执行时实现如本申请实施例提供的语音降噪方法和/或模型训练方法。

上述实施例中提供的语音降噪装置、模型训练装置、电子设备、存储介质及产品可执行本申请相应实施例所提供的语音降噪方法或模型训练方法，具备执行该方法相应的功能模块和有益效果。未在上述实施例中详尽描述的技术细节，可参见本申请任意实施例所提供的语音降噪方法或模型训练方法。

1. 一种语音降噪方法，包括：

采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果；

对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，其中，所述模型活性检测结果由预设语音降噪网络模型输出；

基于所述目标活性检测结果对所述当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧；

将所述初始降噪音频帧输入至所述预设语音降噪网络模型，以输出目标降噪音频帧以及所述当前音频帧对应的模型活性检测结果。

2. 根据权利要求 1 所述的方法，其中，所述算法活性检测结果包括对应音频帧中存在语音的第一概率值，所述模型活性检测结果包括对应音频帧中存在语音的第二概率值；

所述对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，包括：

采用预设计算方式，对上一音频帧对应的模型活性检测结果中的第二概率值，和所述当前音频帧对应的算法活性检测结果中的第一概率值进行计算，得到第三概率值，根据所述第三概率值确定所述当前音频帧对应的目标活性检测结果。

3. 根据权利要求 1 所述的方法，其中，所述算法活性检测结果包括对应音频帧中，预设数量的频点中每个频点存在语音的第四概率值；所述模型活性检测结果包括对应音频帧中，所述预设数量的频点中每个频点存在语音的第五概率值；

所述对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，包括：

针对所述预设数量的频点中的每个频点，采用预设计算方式，对上一音频帧对应的模型活性检测结果中的单个频点的第五概率值，和所述当前音频帧对应的算法活性检测结果中的对应的所述单个频点的第四概率值进行计算，得到第六概率值；

根据所述预设数量的第六概率值，确定所述当前音频帧对应的目标活性检

测结果。

4. 根据权利要求 2 或 3 所述的方法，其中，所述预设计算方式为取最大值、取最小值、计算平均值、求和、计算加权和、以及计算加权平均值中的一种。

5. 根据权利要求 1 所述的方法，其中，所述将所述初始降噪音频帧输入至所述预设语音降噪网络模型，包括：

对所述初始降噪音频帧进行预设特征维度的特征提取，得到目标输入信号；

将所述目标输入信号输入至所述预设语音降噪网络模型，或者，将所述目标输入信号和所述初始降噪音频帧输入至所述预设语音降噪网络模型。

6. 一种模型训练方法，包括：

采用预设语音活性检测算法对当前样本音频帧进行检测，得到对应的样本算法活性检测结果，其中，所述当前样本音频帧关联有活性检测标签和纯净音频帧；

对上一样本音频帧对应的样本模型活性检测结果和所述当前样本音频帧对应的样本算法活性检测结果进行融合处理，得到所述当前样本音频帧对应的目标样本活性检测结果，其中，所述样本模型活性检测结果由语音降噪网络模型输出；

基于所述目标活性样本检测结果对所述当前样本音频帧进行噪声估计和噪声消除，得到初始降噪样本音频帧；

将所述初始降噪样本音频帧输入至所述语音降噪网络模型，以输出目标样本降噪音频帧以及所述当前样本音频帧对应的样本模型活性检测结果；

根据所述目标样本降噪音频帧和所述纯净音频帧确定第一损失关系，根据所述样本模型活性检测结果和所述活性检测标签确定第二损失关系，并基于所述第一损失关系和所述第二损失关系对所述语音降噪网络模型进行训练。

7. 一种语音降噪装置，包括：

语音活性检测模块，设置为采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果；

检测结果融合模块，设置为对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，其中，所述模型活性检测结果由预设语音降噪网络模型输出；

降噪处理模块，设置为基于所述目标活性检测结果对所述当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧；

模型输入模块，设置为将所述初始降噪音频帧输入至所述预设语音降噪网络模型，以输出目标降噪音频帧以及所述当前音频帧对应的模型活性检测结果。

8. 一种模型训练装置，包括：

语音检测模块，设置为采用预设语音活性检测算法对待处理的当前样本音频帧进行检测，得到对应的样本算法活性检测结果，其中，所述当前样本音频帧关联有活性检测标签和干净音频帧；

融合模块，设置为对上一样本音频帧对应的样本模型活性检测结果和所述当前样本音频帧对应的样本算法活性检测结果进行融合处理，得到所述当前样本音频帧对应的目标样本活性检测结果，其中，所述样本模型活性检测结果由语音降噪网络模型输出；

噪声消除模块，设置为基于所述目标活性样本检测结果对所述当前样本音频帧进行噪声估计和噪声消除，得到初始降噪样本音频帧；

网络模型输入模块，设置为将所述初始降噪样本音频帧输入至所述语音降噪网络模型，以输出目标样本降噪音频帧以及所述当前样本音频帧对应的样本模型活性检测结果；

网络模型训练模块，设置为根据所述目标样本降噪音频帧和所述干净音频帧确定第一损失关系，根据所述样本模型活性检测结果和所述活性检测标签确定第二损失关系，并基于所述第一损失关系和所述第二损失关系对所述语音降噪网络模型进行训练。

9. 一种电子设备，所述电子设备包括：

至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的计算机程序，所述计算机程序被所述至少一个处理器执行，以使所述至少一个处理器能够执行权利要求 1-5 任一项所述的语音降噪方法和/或权利要求 6 所述的模型训练方法。

10. 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序用于使处理器执行时实现权利要求 1-5 任一项所述的语音降噪方法和/或权利要求 6 所述的模型训练方法。

11. 一种计算机程序产品，所述计算机程序产品包括计算机程序，所述计算机程序在被处理器执行时实现权利要求 1-5 任一项所述的语音降噪方法和/或权利要求 6 所述的模型训练方法。

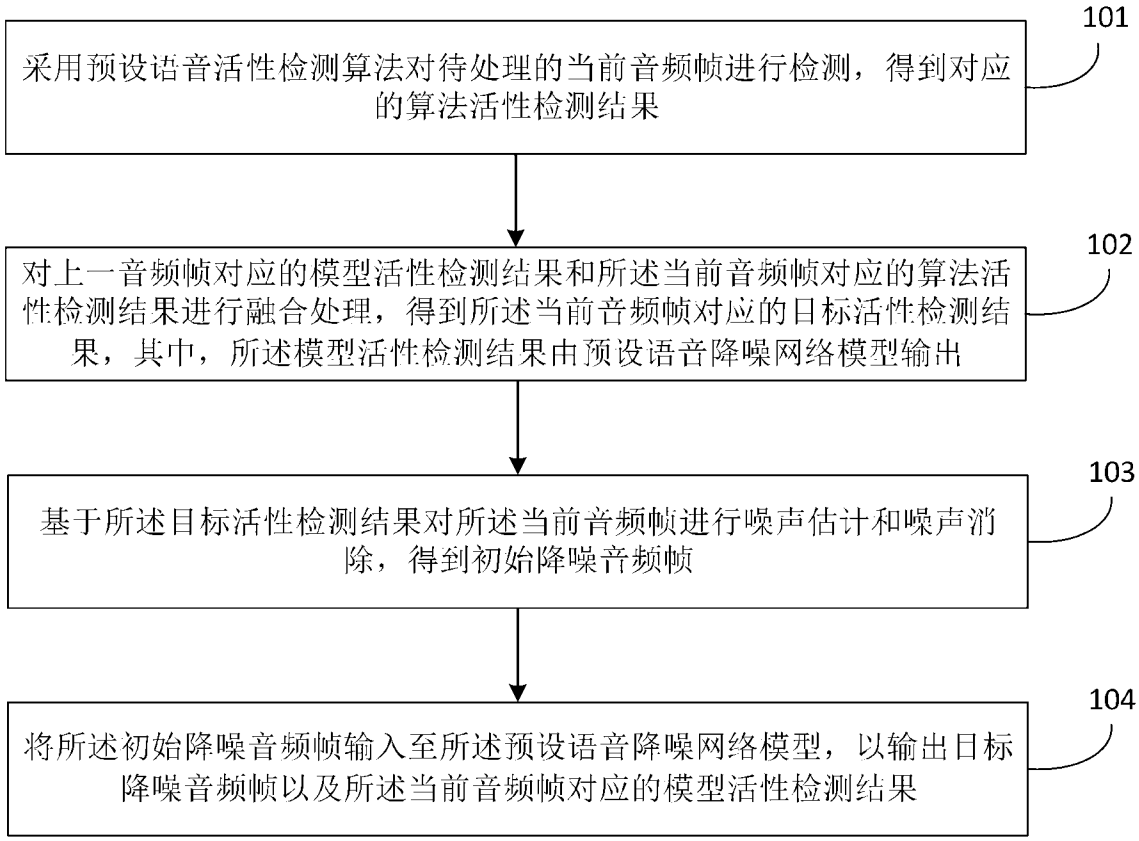

101
采用预设语音活性检测算法对待处理的当前音频帧进行检测，得到对应的算法活性检测结果
102
对上一音频帧对应的模型活性检测结果和所述当前音频帧对应的算法活性检测结果进行融合处理，得到所述当前音频帧对应的目标活性检测结果，其中，所述模型活性检测结果由预设语音降噪网络模型输出
103
基于所述目标活性检测结果对所述当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧
104
将所述初始降噪音频帧输入至所述预设语音降噪网络模型，以输出目标降噪音频帧以及所述当前音频帧对应的模型活性检测结果

图 1

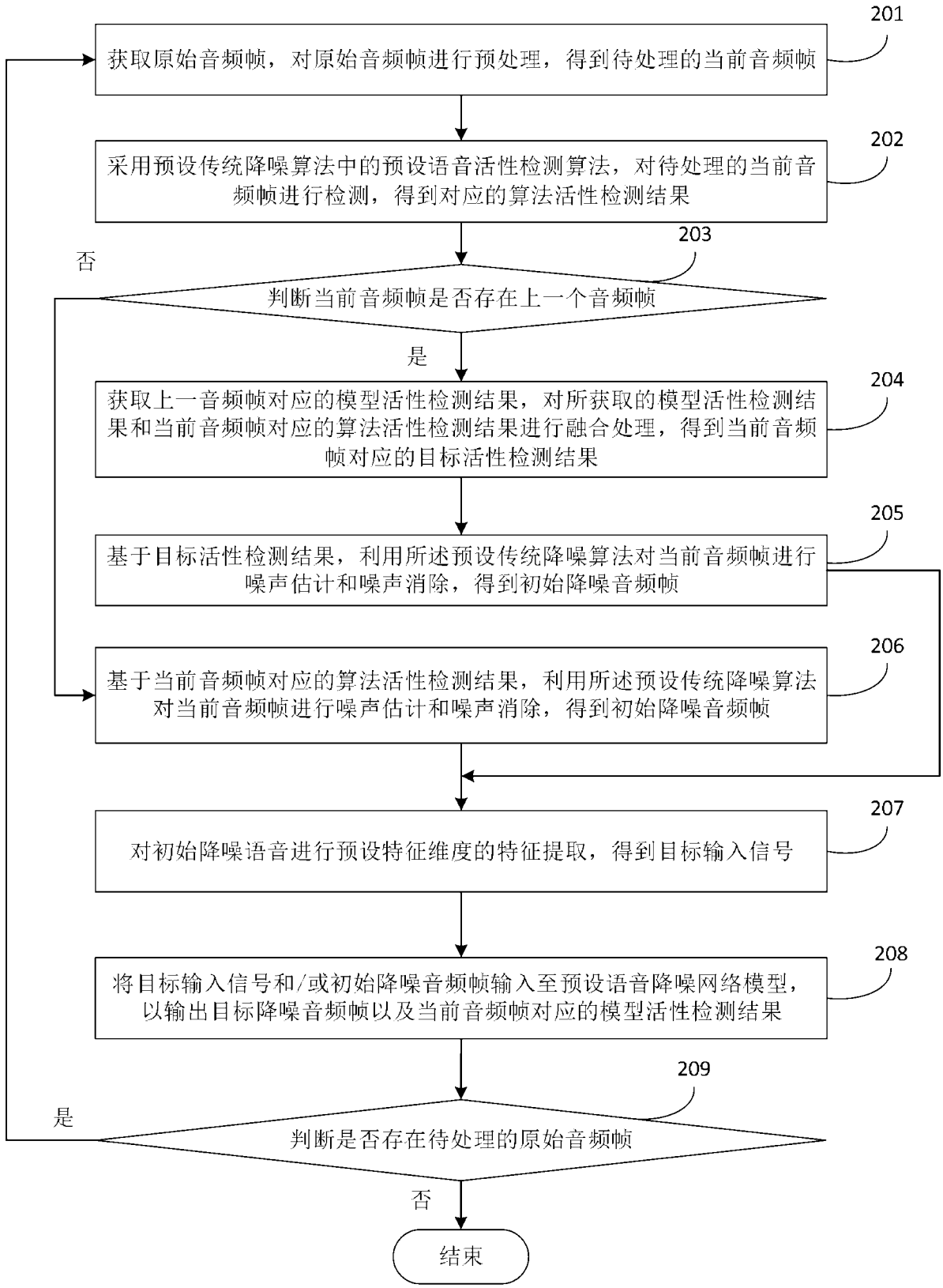

201
获取原始音频帧，对原始音频帧进行预处理，得到待处理的当前音频帧
202
采用预设传统降噪算法中的预设语音活性检测算法，对待处理的当前音频帧进行检测，得到对应的算法活性检测结果
203
否
判断当前音频帧是否存在上一个音频帧
是
204
获取上一音频帧对应的模型活性检测结果，对所获取的模型活性检测结果和当前音频帧对应的算法活性检测结果进行融合处理，得到当前音频帧对应的目标活性检测结果
205
基于目标活性检测结果，利用所述预设传统降噪算法对当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧
206
基于当前音频帧对应的算法活性检测结果，利用所述预设传统降噪算法对当前音频帧进行噪声估计和噪声消除，得到初始降噪音频帧
207
对初始降噪语音进行预设特征维度的特征提取，得到目标输入信号
208
将目标输入信号和/或初始降噪音频帧输入至预设语音降噪网络模型，以输出目标降噪音频帧以及当前音频帧对应的模型活性检测结果
209
是
判断是否存在待处理的原始音频帧
否
结束

图 2

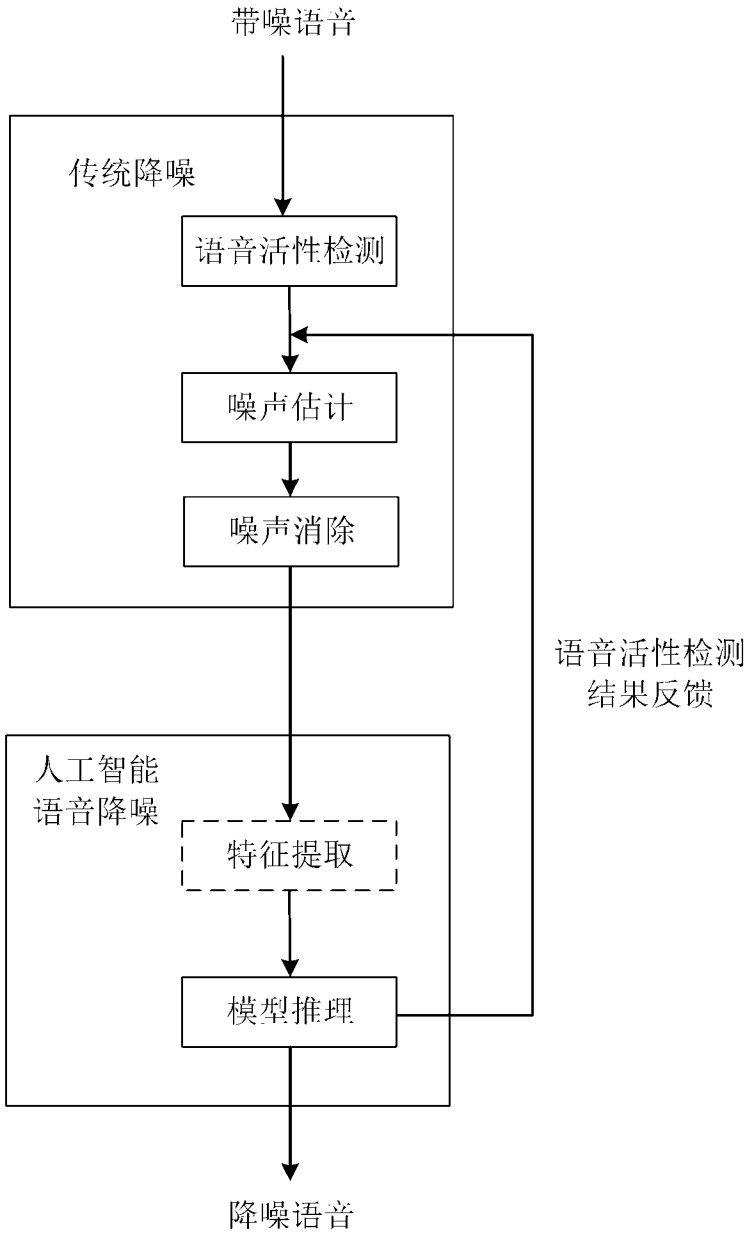

带噪语音
传统降噪
语音活性检测
噪声估计
噪声消除
语音活性检测
结果反馈
人工智能
语音降噪
特征提取
模型推理
降噪语音

图 3

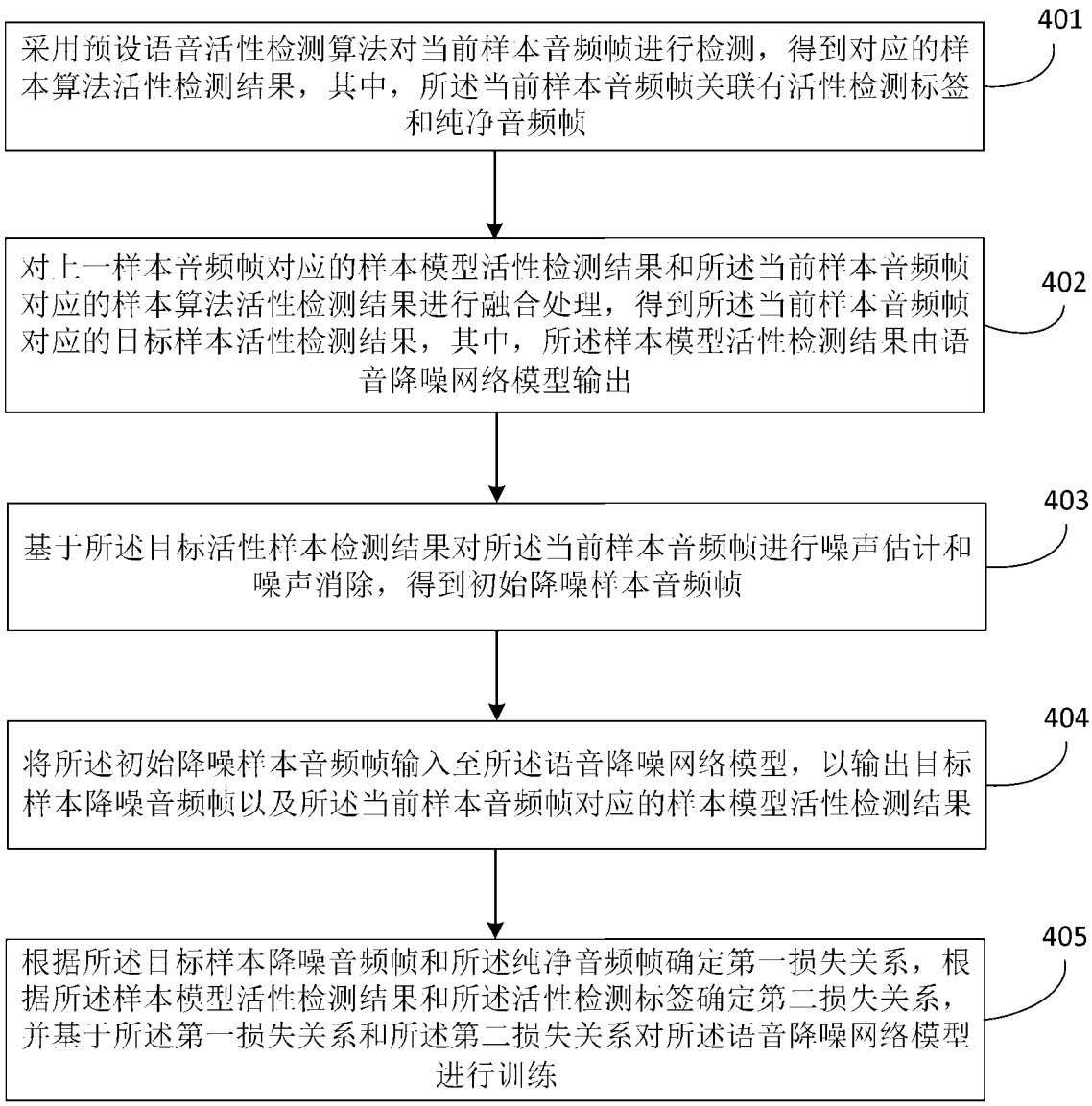

401
采用预设语音活性检测算法对当前样本音频帧进行检测，得到对应的样本算法活性检测结果，其中，所述当前样本音频帧关联有活性检测标签和纯净音频帧
402
对上一样本音频帧对应的样本模型活性检测结果和所述当前样本音频帧对应的样本算法活性检测结果进行融合处理，得到所述当前样本音频帧对应的目标样本活性检测结果，其中，所述样本模型活性检测结果由语音降噪网络模型输出
403
基于所述目标活性样本检测结果对所述当前样本音频帧进行噪声估计和噪声消除，得到初始降噪样本音频帧
404
将所述初始降噪样本音频帧输入至所述语音降噪网络模型，以输出目标样本降噪音频帧以及所述当前样本音频帧对应的样本模型活性检测结果
405
根据所述目标样本降噪音频帧和所述纯净音频帧确定第一损失关系，根据所述样本模型活性检测结果和所述活性检测标签确定第二损失关系，并基于所述第一损失关系和所述第二损失关系对所述语音降噪网络模型进行训练

图 4

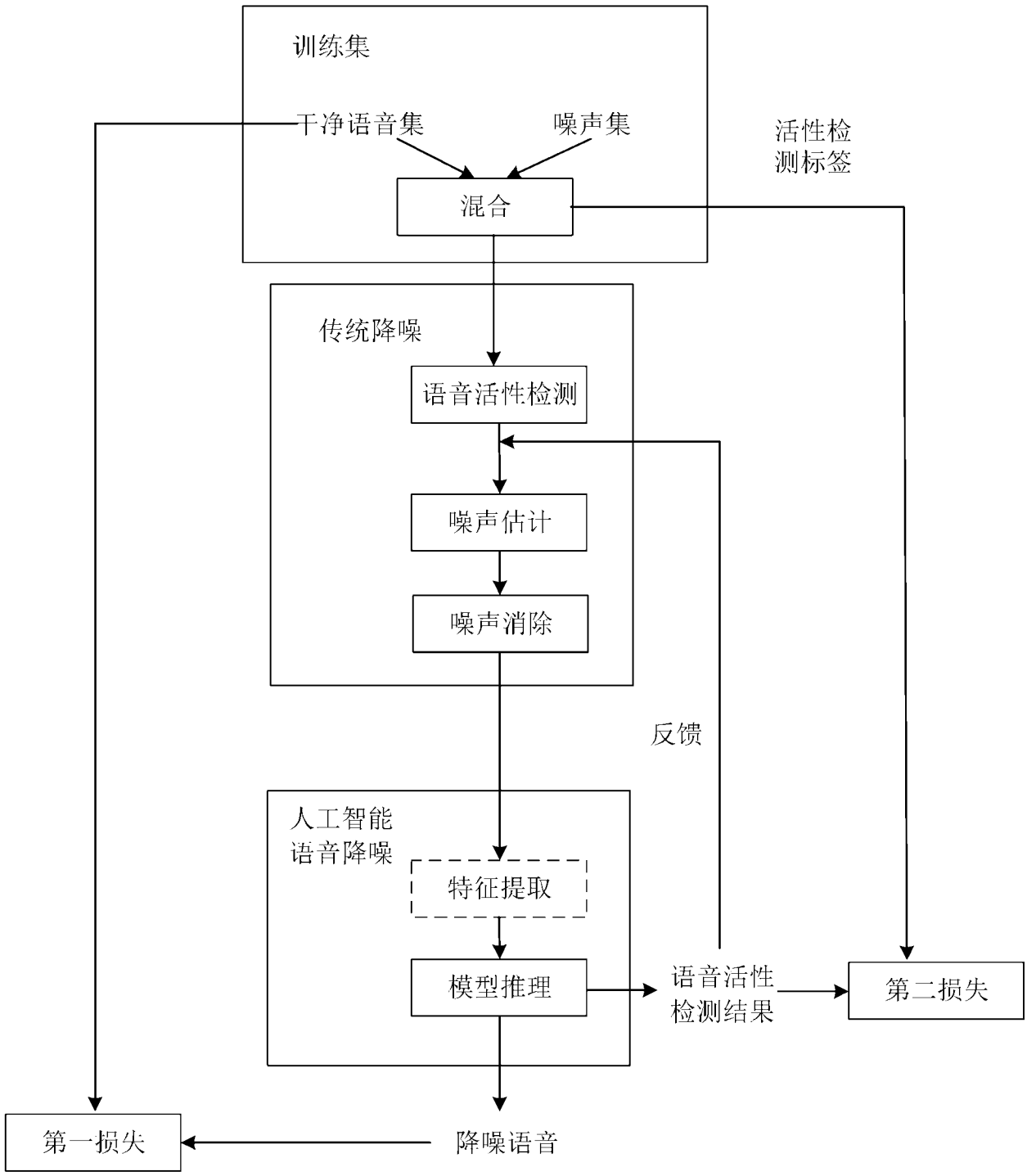

训练集
干净语音集
噪声集
混合
活性检测标签
传统降噪
语音活性检测
噪声估计
噪声消除
反馈
人工智能语音降噪
特征提取
模型推理
语音活性检测结果
第二损失
第一损失
降噪语音

图 5

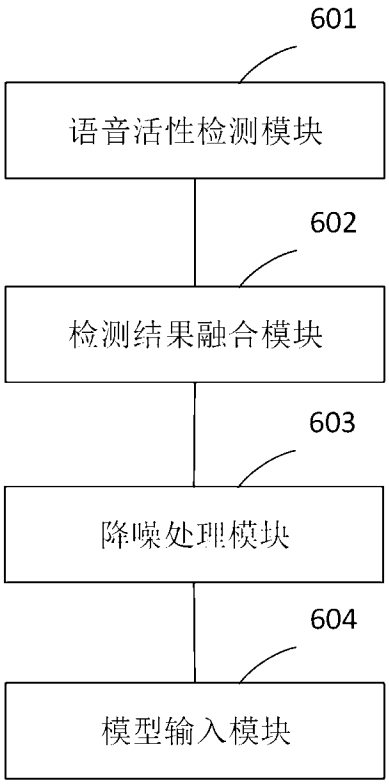

601
语音活性检测模块
602
检测结果融合模块
603
降噪处理模块
604
模型输入模块

图 6

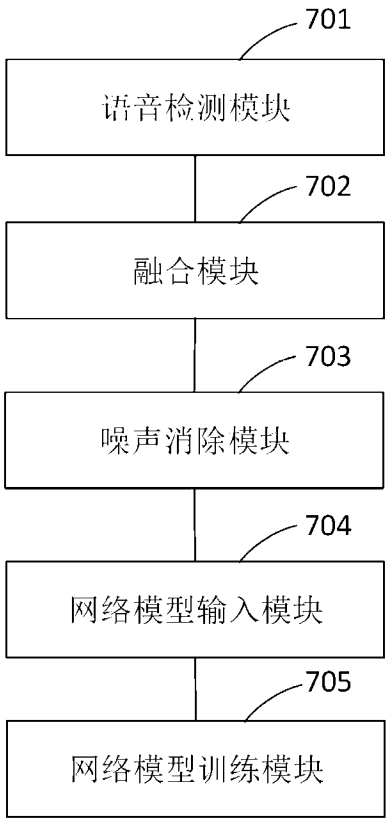

701
语音检测模块
702
融合模块
703
噪声消除模块
704
网络模型输入模块
705
网络模型训练模块

图 7

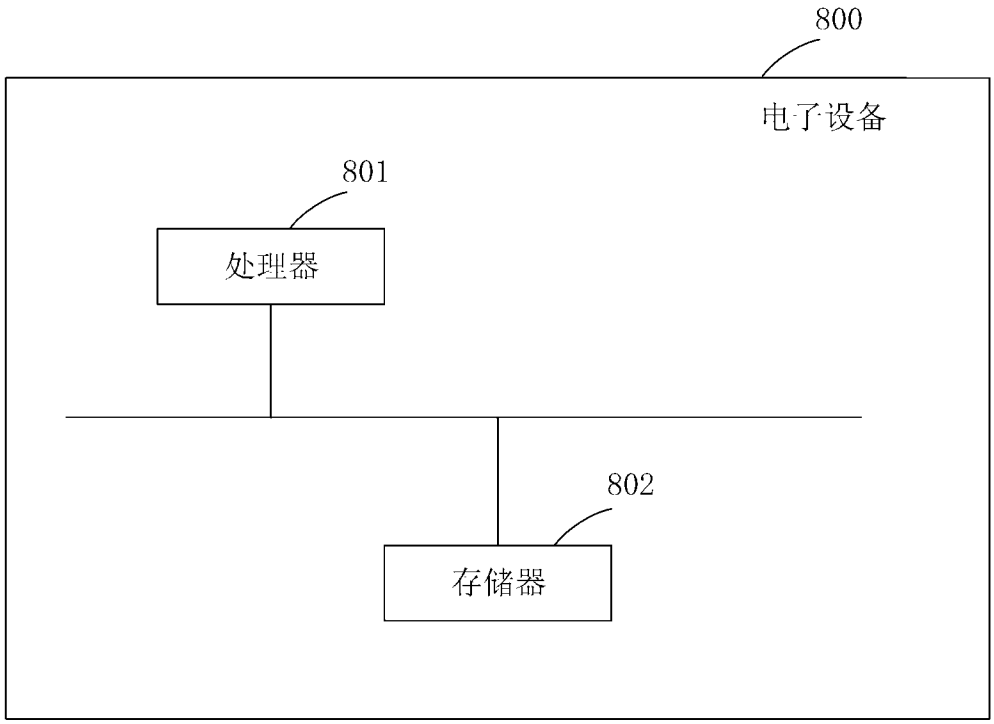


图 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/106951

A. CLASSIFICATION OF SUBJECT MATTER

G10L21/0208(2013.01)i; G10L21/0216(2013.01)i; G10L21/0224(2013.01)i; G10L21/0232(2013.01)i; G10L25/30(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC:G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI, CNTXT, WPABS, ENTXT, CNKI: 语音, 检测, 降噪, 去噪, 噪音, 噪音, 估计, 消除, 前一, 上一, 之前, 活性检测, 网络模型, VAD, voice activity detection, adaptive noise suppression, ANS, speech, voice, detection, noise reduction, de-noise, noise, estimation, cancellation, previous, last, before, network models, RNN, DNN, CNN

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 115273880 A (BAIGUOYUAN TECHNOLOGY (SINGAPORE) PTE LTD.) 01 November 2022 (2022-11-01) claims 1-11	1-11
A	CN 114255778 A (GUANGZHOU HUANCHENG CULTURE MEDIA CO., LTD.) 29 March 2022 (2022-03-29) description, paragraphs [0058]-[0111]	1-11
A	CN 108428456 A (ZHEJIANG KAICHI ELECTRONIC TECHNOLOGY CO., LTD.) 21 August 2018 (2018-08-21) entire document	1-11
A	CN 114495969 A (NANJING FENGHUO TIANDI COMMUNICATION TECHNOLOGY CO., LTD.) 13 May 2022 (2022-05-13) entire document	1-11
A	CN 114596870 A (GUANGZHOU BOGUAN INFORMATION SCIENCE & TECHNOLOGY CO., LTD.) 07 June 2022 (2022-06-07) entire document	1-11

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 September 2023	**20 September 2023**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088**	Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2020286501 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 10 September 2020 (2020-09-10) entire document	1-11
A	WO 2017218386 A1 (MED-EL ELEKTROMEDIZINISCHE GERAETE GMBH) 21 December 2017 (2017-12-21) entire document	1-11

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/106951

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 115273880 A	01 November 2022	None	
CN 114255778 A	29 March 2022	None	
CN 108428456 A	21 August 2018	None	
CN 114495969 A	13 May 2022	None	
CN 114596870 A	07 June 2022	None	
US 2020286501 A1	10 September 2020	EP 3692529 A1	12 August 2020
		EP 3692529 B1	24 May 2023
		WO 2019072395 A1	18 April 2019
WO 2017218386 A1	21 December 2017	EP 3469586 A1	17 April 2019
		EP 3469586 B1	04 August 2021
		US 2019124454 A1	25 April 2019
		US 10785581 B2	22 September 2020
		AU 2017286519 A1	06 December 2018
		AU 2017286519 B2	07 May 2020

国际检索报告

国际申请号
PCT/CN2023/106951

A. 主题的分类

G10L21/0208(2013.01)i; G10L21/0216(2013.01)i; G10L21/0224(2013.01)i; G10L21/0232(2013.01)i; G10L25/30(2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC:G10L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

DWPI, CNTXT, WPABS, ENTXT, CNKI: 语音，检测，降噪，去噪，噪音，噪音，估计，消除，前一，上一，之前，活性检测，网络模型，VAD, voice activity detection, adaptive noise suppression, ANS, speech, voice, detection, noise reduction, de-noise, noise, estimation, cancellation, previous, last, before, network models, RNN,DNN,CNN

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 115273880 A (百果园技术（新加坡）有限公司) 2022年11月1日 (2022-11-01) 权利要求1-11	1-11
A	CN 114255778 A (广州欢城文化传媒有限公司) 2022年3月29日 (2022-03-29) 说明书[0058]-[0111]段	1-11
A	CN 108428456 A (浙江凯池电子科技有限公司) 2018年8月21日 (2018-08-21) 全文	1-11
A	CN 114495969 A (南京烽火天地通信科技有限公司) 2022年5月13日 (2022-05-13) 全文	1-11
A	CN 114596870 A (广州博冠信息科技有限公司) 2022年6月7日 (2022-06-07) 全文	1-11
A	US 2020286501 A1 (HUAWEI TECH. CO., LTD.) 2020年9月10日 (2020-09-10) 全文	1-11

☑ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"D" 申请人在国际申请中引证的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2023年9月13日	2023年9月20日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	受权官员 王玥 电话号码 (+86) 010-53962545

国际检索报告

国际申请号
PCT/CN2023/106951

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	WO 2017218386 A1 (MED-EL ELEKTROMEDIZINISCHE GERAETE G.M.B.H.) 2017年12月21日 (2017-12-21) 全文	1-11

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/106951

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	115273880	A	2022年11月1日	无			
CN	114255778	A	2022年3月29日	无			
CN	108428456	A	2018年8月21日	无			
CN	114495969	A	2022年5月13日	无			
CN	114596870	A	2022年6月7日	无			
US	2020286501	A1	2020年9月10日	EP	3692529	A1	2020年8月12日
				EP	3692529	B1	2023年5月24日
				WO	2019072395	A1	2019年4月18日
WO	2017218386	A1	2017年12月21日	EP	3469586	A1	2019年4月17日
				EP	3469586	B1	2021年8月4日
				US	2019124454	A1	2019年4月25日
				US	10785581	B2	2020年9月22日
				AU	2017286519	A1	2018年12月6日
				AU	2017286519	B2	2020年5月7日