

Rockchip Sound Event Detection开发文档

文件标识：RK-KF-SF-959

发布版本：V1.5.1

日期：2023-11-03

文件密级：☐绝密 ☐秘密 ☐内部资料 ☒公开

免责声明

本文档按“现状”提供，瑞芯微电子股份有限公司（“本公司”，下同）不对本文档的任何陈述、信息和内容的准确性、可靠性、完整性、适销性、特定目的性和非侵权性提供任何明示或暗示的声明或保证。本文档仅作为使用指导的参考。

由于产品版本升级或其他原因，本文档将可能在未经任何通知的情况下，不定期进行更新或修改。

商标声明

“Rockchip”、“瑞芯微”、“瑞芯”均为本公司的注册商标，归本公司所有。

本文档可能提及的其他所有注册商标或商标，由其各自拥有者所有。

版权所有 © 2022 瑞芯微电子股份有限公司

超越合理使用范畴，非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

瑞芯微电子股份有限公司

Rockchip Electronics Co., Ltd.

地址：福建省福州市铜盘路软件园A区18号

网址：www.rock-chips.com

客户服务电话：+86-4007-700-590

客户服务传真：+86-591-83951833

客户服务邮箱：fae@rock-chips.com

产品版本

芯片名称	内核版本
全系列	通用

读者对象

本文档（本指南）主要适用于以下工程师：

技术支持工程师

软件开发工程师

修订记录

版本号	作者	修改日期	修改说明
V1.0.0	廖华平、江迪	2022-07-23	初始版本
V1.0.1	廖华平、郑兴	2022-08-15	整理文档格式
V1.0.2	廖华平	2022-08-20	更新接口，加入蜂鸣器检测
V1.1.0	赖陈潇	2022-12-15	更新接口，加入AGC及玻璃破碎声检测
V1.2.0	赖陈潇	2023-01-16	添加初始化结果输出接口
V1.3.0	赖陈潇	2023-03-09	添加fir滤波器接口
V1.4.0	赖陈潇	2023-08-17	修改aed流程，将fir置于agc之前，添加aed平滑系数设置接口
V1.5.0	赖陈潇	2023-10-31	更新哭声检测模型，添加confirm_prob设置接口
V1.5.1	赖陈潇	2023-11-03	修改log打印的格式，完善说明文档的内容

目录

Rockchip Sound Event Detection开发文档

1. 概述
2. 功能描述
 - 2.1 Abnormal Event Detection(AED)
 - 2.2 AI声音事件检测Sound Event Detection(SED)
 - 2.2.1 Baby Cry Detection(BCD)
 - 2.2.2 Buzz Detection(BUZ)
 - 2.2.3 Glass broken Detection(GBS)
3. 相关API介绍
 - 3.1 rkaudio_sed_init
 - 3.2 rkaudio_sed_destroy
 - 3.3 rkaudio_sed_init_res
 - 3.4 rkaudio_sed_process
 - 3.5 rkaudio_sed_param_init
 - 3.6 rkaudio_sed_param_destroy
 - 3.7 rkaudio_sed_param_aed
 - 3.8 rkaudio_sed_param
 - 3.9 rkaudio_agc_param_init
 - 3.10 rkaudio_sed_lsd_db
4. 参数介绍
 - 4.1 JUMP_FRAME
 - 4.2 RKAudioSedParam
 - 4.3 SedAedParam
 - 4.4 SedParam
 - 4.5 RKAudioSedRes
 - 4.6 RKAudioSedEnable
 - 4.7 RKSEDAGCParam
 - 4.8 RKFIRParam

1. 概述

本文档主要描述声音事件检测(Sound Event Detection)功能。当前包含婴儿哭声检测(Baby Cry Detection)、异常声检测(Abnormal Event Detection)和蜂鸣器检测(Buzz Detection)。三个模块通过统一接口调用，但功能相互独立，可通过开关使能关闭其中任何模块。

2. 功能描述

2.1 Abnormal Event Detection(AED)

AED实现实时异常声检测功能，包括超大声检测和信噪比检测，支持8k和16k的音频输入。超大声检测实现对dB的检测，超过设定的dB值输出1，否则输出0。信噪比检测主要实现对噪声和信号进行检测，这里说的噪声主要是环境中的平稳噪声和录音的底噪，当信噪比大于设定阈值后输出1，否则为0。

分贝（decibel），是量度两个相同单位之数量比例的计量单位，幅值为我们声音数据大小的绝对值，分贝(dB)和幅值X的关系为：

$$dB = 20 * \log_{10}(X)$$

所以幅值为倍数关系，dB为加减关系。幅值每上升一倍，dB值上升6dB。16bit的音频数据满幅值32767，此值设为0dB，所以我们这里说的dB值都在0dB以下。

信噪比(SNR)可理解为信号与噪声的比值，噪声设为0dB，信号比噪声高6dB，那么此时的信噪比为6dB。

2.2 AI声音事件检测Sound Event Detection(SED)

SED模块实现对婴儿哭声，蜂鸣器报警声及玻璃破碎声的实时检测，仅支持16k的音频输入。采用带有多头注意力机制的RCNN模型对约1.5s时间内的声音信息进行分析，从而实现对上述声音事件的检测，模块在信息信噪比高于6dB时有较好的效果。

2.2.1 Baby Cry Detection(BCD)

BCD实现实时检测婴儿哭声的功能。通过深度学习的方式进行婴儿哭声检测，信噪比高的时候效果好，从婴儿哭声出现开始计算，检测延迟约2s。

2.2.2 Buzz Detection(BUZ)

BUZ实现实时检测蜂鸣器报警声的功能。主要检测常见的警报声，包括烟雾报警、防空报警、防盗报警等。通过深度学习的方式进行警报声检测，信噪比高的时候效果较好，从蜂鸣器报警声出现开始计算，检测延迟约2s。

2.2.3 Glass broken Detection(GBS)

GBS实现实时检测玻璃破碎声的功能。通过深度学习的方式进行玻璃破碎声检测，信噪比高的时候效果较好，从玻璃破碎声出现开始计算，检测延迟约0.6s。

3. 相关API介绍

该功能模块为用户提供以下API:

- [rkaudio_sed_init](#): SED初始化。
- [rkaudio_sed_destroy](#): SED销毁。
- [rkaudio_sed_process](#): SED执行。
- [rkaudio_sed_init_res](#): 初始化结果。

3.1 rkaudio_sed_init

【描述】

初始化并返回SED的操作句柄，此句柄用于[rkaudio_sed_process](#)。使用结束后，执行[rkaudio_sed_destroy](#)销毁。

【语法】

void *rkaudio_sed_init(int fs, int bit, int chan, [RKAudioSedParam](#) *param)

【参数】

参数名	描述	输入/输出
fs	采样率，AED支持8k和16k，BCD、BUZ及GBS只支持16k。	输入
bit	每个数据的bit数，一般使用的都是16bit数据。	输入
chan	通道数，如果输入多通道数据，使用的是第一个通道的数据。	输入
param	SED参数，相关定义见 RKAudioSedParam 。可通过函数 rkaudio_sed_param_init 构建，也可自行构建相关函数和初始化系数。	输入

【返回值】

返回值	描述
NULL	失败。
非NULL	成功。

3.2 rkaudio_sed_destroy

【描述】

销毁SED句柄。

【语法】

void rkaudio_sed_destroy(void *st_)

【参数】

参数名	描述	输入/输出
st_	rkaudio_sed_init返回的句柄。	输入

【返回值】

无

3.3 rkaudio_sed_init_res

【描述】

输出SED句柄的初始化结果。

【语法】

char rkaudio_sed_init_res(void* st_)

【参数】

参数名	描述	输入/输出
st_	rkaudio_sed_init返回的句柄。	输入

【返回值】

返回8bit的char类型，从低位开始，第一字节表示agc的初始化结果，1表示初始化成功，0表示初始化失败；第二字节表示aed的初始化结果，1表示初始化成功，0表示初始化失败；第三字节表示sed的初始化结果，1表示初始化成功，0表示初始化失败；第四字节表示fir的初始化结果，1表示初始化成功，0表示初始化失败

3.4 rkaudio_sed_process

【描述】

进行声音事件检测，返回结果存于res中。

【语法】

int rkaudio_sed_process(void *st_, short *in, int in_size, [RKAudioSedRes](#) *res)

【参数】

参数名	描述	输入/输出
st_	句柄。	输入
in	输入数据的指针。	输入
in_size	输入数据的长度，8k数据size应为128的倍数，16k数据size应为256的倍数。	输入
res	检测结果结构体指针，此结构体需在外部申请，定义参见 RKAudioSedRes 。	输出

【返回值】

返回值	描述
大于等于0	执行成功，此返回值为执行数据的长度。
小于0	执行失败。

3.5 rkaudio_sed_param_init

【描述】

初始化SED模块参数，进行子模块使能，并调用各个子模块参数初始化函数。此函数源码对外开放，并且各参数默认值已设置，也可根据实际数据进行适当调整。使用完后，调用[rkaudio_sed_param_destroy](#)销毁。如果在SED调用过程中要对参数或者模块使能进行调整，需要将SED模块销毁后，重新初始化才能生效。

【语法】

```
RKAudioSedParam *rkaudio_sed_param_init()
```

【返回值】

SED参数指针，定义参见[RKAudioSedParam](#)。

3.6 rkaudio_sed_param_destroy

【描述】

销毁SED模块参数。

【语法】

```
void rkaudio_sed_param_destroy(RKAudioSedParam *param)
```

【参数】

SED参数指针。

3.7 rkaudio_sed_param_aed

【描述】

初始化AED模块参数，在rkaudio_sed_param_init函数中调用，在rkaudio_sed_param_destroy函数中销毁。

【语法】

[SedAedParam](#) *rkaudio_sed_param_aed()

【返回值】

AED模块参数指针。

3.8 rkaudio_sed_param

【描述】

初始化BCD/BUZ/GBS模块参数，在rkaudio_sed_param_init函数中调用，在rkaudio_sed_param_destroy函数中销毁。

【语法】

[SedBuzParam](#) *rkaudio_sed_param()

【返回值】

BUZ模块参数指针。

3.9 rkaudio_agc_param_init

【描述】

初始化AGC参数，在rkaudio_sed_param_init函数中调用，在rkaudio_sed_param_destroy函数中销毁。

【语法】

RKAGCParam* rkaudio_agc_param_init()

【返回值】

AGC模块参数指针。

3.10 rkaudio_sed_lsd_db

【描述】

返回aed超大声检测的音量的db值。

【语法】

float rkaudio_sed_lsd_db(void *st_)

【返回值】

dbfs值

4. 参数介绍

4.1 JUMP_FRAME

【说明】

SED算法相关参数，检测间隔帧数，数值越高则每秒检测频率越低，同时计算量越低。以数值20为例，则检测间隔为 $0.016 \times 20 = 0.32s$ ，即差不多平均一秒检测3次，建议数值：15-25。被跳过帧的检测结果会与上一帧保持一致。

【定义】

```
#define JUMP_FRAME 20
```

4.2 RKAudioSedParam

【说明】

SED算法相关参数。

【定义】

```
typedef struct RKAudioSedParam_
{
    int model_en;
    RKSEDAGCParam *agc_param;
    SedAedParam *aed_param;
    SedParam *sed_param;
    RKFIRParam *fir_param;
} RKAudioSedParam;
```

【成员】

成员名称	描述
model_en	通过设置bit位开启子模块，各bit定义参见 RKSedEnable ，如要开启AED和BCD，则应设为EN_AED EN_BCD。
SedAedParam	AED模块参数，定义参见 SedAedParam 。
SedParam	SED模块参数，定义参见 SedParam 。
RKSEDAGCParam	AGC模块参数，定义参见 RKSEDAGCParam
RKFIRParam	FIR模块参数，定义参见 RKFIRParam

4.3 SedAedParam

【说明】

AED算法相关参数。

【定义】

```
typedef struct SedAedParam_  
{  
    float    snr_db;  
    float    lsd_db;  
    int      policy;  
    float    smooth_param;  
} SedAedParam;
```

【成员】

成员名称	描述
snr_db	语音信噪比阈值，大于则输出1。
lsd_db	超大声阈值，大于则输出1。最大为0dB。
policy	信噪比检测算法灵敏度，取值范围为[0, 2]，值越大越灵敏，越容易满足检测阈值。默认取1。
smooth_param	超大声音量计算平滑系数，取值范围为(0, 1)，数值越大，平滑窗越大

4.4 SedParam

【说明】

BCD/BUZ/GBS算法相关参数。

【定义】

```
typedef struct SedParam_  
{  
    int      frm_len;  
    int      nclass;  
    int      babycry_decision_len;  
    int      buzzer_decision_len;  
    int      glassbreaking_decision_len;  
    float    babycry_confirm_prob;  
    float    buzzer_confirm_prob;  
    float    glassbreaking_confirm_prob;  
} SedParam;
```

【成员】

成员名称	描述
frm_len	统计的总帧数，建议数值：110-150，越长检测延迟越高，越低越容易漏检测或误检测。
nclass	关注的分类的总类别数目，固定值为3，不可修改。
babycry_decision_len	哭声检测确认帧长，数值应小于frm_len，建议长度100，越长检测延迟越高，越容易漏检测；越短越容易误检测。
buzzer_decision_len	蜂鸣器报警声检测确认帧长，数值应小于frm_len，建议长度100，越长检测延迟越高，越容易漏检测；越短越容易误检测。
glassbreaking_decision_len	玻璃破碎声检测确认帧长，数值应小于frm_len，大于JUMP_FRAME，建议长度25-50，越长检测延迟越高，越容易漏检测；越短越容易误检测。
babycry_confirm_prob	哭声检测确认概率，取值范围为(0，1)，建议取值0.5，取值越大越不容易误检，但容易漏检
buzzer_confirm_prob	蜂鸣器报警声检测确认概率，取值范围为(0，1)，建议取值0.5，取值越大越不容易误检，但容易漏检
glassbreaking_confirm_prob	玻璃破碎声检测确认概率，取值范围为(0，1)，建议取值0.8，取值越大越不容易误检，但容易漏检

4.5 RKAudioSedRes

【说明】

SED模块返回结果。

【定义】

```
typedef struct RKAudioSedRes_
{
    int snr_res;
    int lsd_res;
    int bcd_res;
    int buz_res;
    int gbs_res;
} RKAudioSedRes;
```

【成员】

成员名称	描述
snr_res	SNR返回结果，1为满足信噪比阈值，0不满足。
lsd_res	LSD返回结果，1为满足超大声阈值，0不满足。
bcd_res	BCD返回结果，1为检测到哭声，0没检测到。
buz_res	BUZ返回结果，1为检测到警报声，0没检测到。
gbs_res	GBS返回结果，1为检测到玻璃破碎声，0没检测到。

4.6 RKAudioSedEnable

【说明】

使能各模块，将此值赋给model_en，则可使能对应模块。如要使能多个模块，则使用EN_AED | EN_BCD方式。

【定义】

```
typedef enum RKAudioSedEnable_  
{  
    EN_AGC = 1 << 0,  
    EN_AED = 1 << 1,  
    EN_SED = 1 << 2,  
    EN_FIR = 1 << 3,  
} RKAudioSedEnable;
```

【成员】

成员名称	描述
EN_AGC	使能AGC模块，建议在接收音量较小时开启。
EN_AED	使能AED模块。
EN_SED	使能SED模块。
EN_FIR	使能FIR模块。

4.7 RKSEDAGCParam

【说明】

agc相关参数，不建议改动

定义

```
typedef struct RKSEDAGCParam_  
{  
    float    attack_time;  
    float    release_time;  
    float    max_gain;
```

```

float    max_peak;
float    fRth0;
float    fRk0;
float    fRth1;
int      fs;
int      frmlen;
float    attenuate_time;
float    fRth2;
float    fRk1;
float    fRk2;
float    fLineGainDb;
int      swSmL0;
int      swSmL1;
int      swSmL2;
} RKSEDAGCParam;

```

【成员】

成员名称	描述
attack_time	触发时间，即AGC增益下降所需要的时间
release_time	释放时间，即AGC增益上升所需要的时间
max_gain	最大增益，同时也是线性段增益，单位：dB
max_peak	经AGC处理后，输出语音的最大能量，范围：单位：dB
fRth0	扩张段结束能量dB阈值，同时也是线性段开始阈值
fRk0	扩张段斜率（未启用）
fRth1	压缩段起始能量dB阈值，同时也是线性段结束阈值
fs	数据采样率（未启用）
frmlen	处理帧长（未启用）
attenuate_time	噪声衰减时间，即噪声段增益衰减到1所需的时间（未启用）
fRth2	压缩段起始能量dB阈值
fRk1	扩张段斜率（未启用）
fRk2	扩张段斜率（未启用）
fLineGainDb	线性段提升dB数（未启用）
swSmL0	扩张段时域平滑点数（未启用）
swSmL1	线性段时域平滑点数（未启用）
swSmL2	压缩段时域平滑点数（未启用）

4.8 RKFIRParam

【说明】

fir相关参数，不建议改动

【定义】

```
typedef struct RKFIRParam_  
{  
    int fir_len;  
    float* fir_coeffs;  
} RKFIRParam;
```

【成员】

成员名称	描述
fir_len	fir长度
fir_coeffs	fir系数