

AI数字人语音交互盒子（AIBOX）手册

产品型号：MV-AIBOX

版本记录

版本	更新内容	更新时间
V1.0	初始文档	2024.11.05
V1.1	新增升级功能	2025.03.27

目录

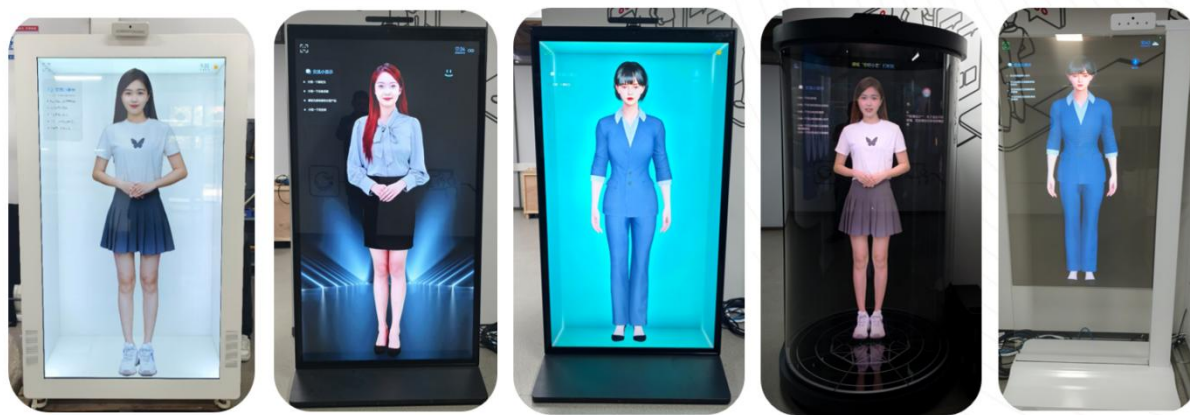
1. 产品概述	1
2. 产品实物图	2
2.1 麦克风摄像头部分	2
2.2 AI数字人计算盒子实图	2
3. 适用场景	3
3.1 摄像头+麦克风安装说明	3
3.2 交互距离说明	3
3.3 摄像头倾斜角度说明	4
4. 产品特点	5
5. 产品功能	6
6. 效果说明	7
7. 接口说明	7
7.1接口参数表	8
7.2 配件明细表	8
8. AI数字人多模态语音交互盒子规格参数	9
9. 硬件尺寸信息	10

1. 产品概述

AI数字人多模态语音交互盒子是一款用于公共环境的高噪场景下数字人语音交互产品，该产品基于摄像头人脸唇动算法和AI麦克风阵列进行多模态降噪，有效消除环境噪音和周围人声的干扰，搭配自主研发的数字人交互系统，实现复杂场景下人与设备使用语音进行自然交互。

AI数字人多模态交互语音盒子采用 500W 动态摄像头和AI4阵列降噪麦克风，外观采用全铝设计，表面氧化细沙纹工艺，设计简约，散热稳定等功能。集成神经网络处理器 NPU，算力高达 6.0Tops，兼容多种 AI 框架，该设备基于摄像头麦克风阵列进行多模态降噪，有效消除环境噪音和周围人声的干扰，灵活应用结合全息舱、政务一体机、透明屏等显示设备在公共环境的高噪场景下数字人交互方案。

使用场景示意图



(以上形象已授权)

2. 产品实物图

2.1 麦克风摄像头部分



2.2 AI数字人计算盒子实图



3. 适用场景

本套件产品适用于最佳交互距离在 1.5 米-2 米的数字人交互一体机、全息舱、透明广告屏，政务一体机等大屏在公共环境的高噪场景下数字人交互方案。

3.1 摄像头+麦克风安装说明

麦克风及摄像头水平方向安装，平行于大屏设备顶部。

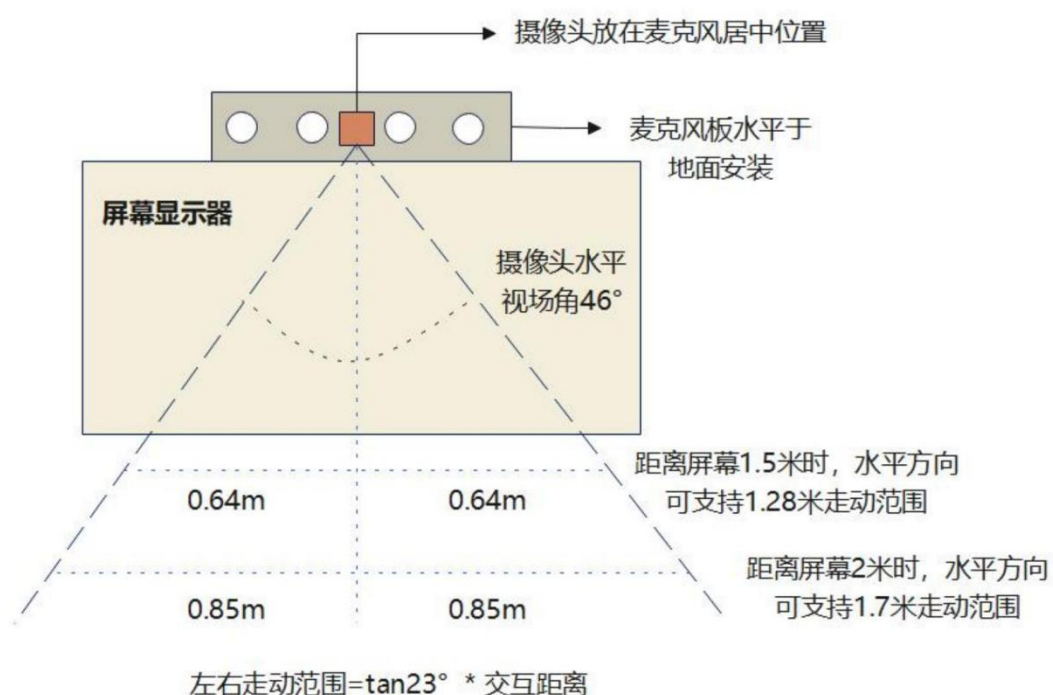


图 1 麦克风+摄像头安装方式说明

3.2 交互距离说明

当大屏设备高度在 1.6 米~2.1 米，建议在距离大屏 1.5 米~2米的位置进行语音交互。

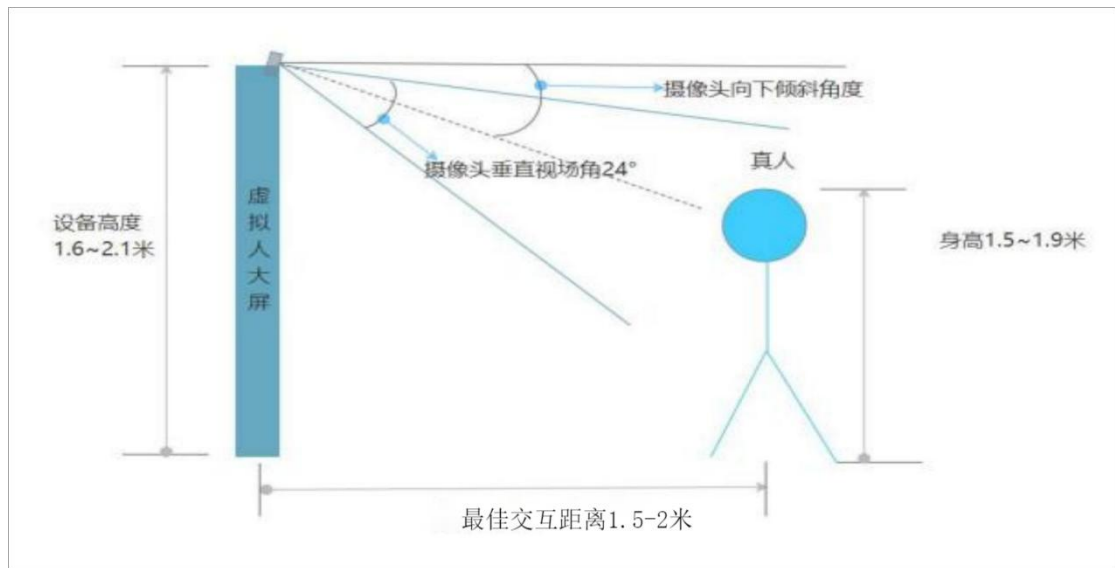


图 2 适用场景示意图

3.3 摄像头倾斜角度说明

需要根据设备实际高度调整摄像头倾斜角度。

基于目前摄像头垂直视场角 24 度计算	
设备高度	摄像头倾斜角度
1.6 米	0°
1.7 米	向下倾斜 5 度
1.8 米	向下倾斜 5 度
1.9 米	向下倾斜 10 度
2 米	向下倾斜 15 度
2.1 米	向下倾斜 15 度

表 1 摄像头倾斜角度参考表

4. 产品特点

- 本产品为摄像头人脸唤醒，不需要设置唤醒词，随问随答。
- 本产品使用人脸识别+唇形识别技术，锁定主交互人，不受其他人声音干扰。
- 本产品使用四麦窄波束技术，屏蔽波束外各类噪音。
- 本产品已集成自主研发的AI数字人语音交互系统，专用数字人后台管理系统，有分级账户管理权限，可管理多台数字人设备，自定义设置欢迎词、唤醒词，人物背景，提示词，城市天气。
- 有前端编辑界面，具备密码登录，可设置设备自启，待机广告，关闭系统任务栏，任意移动位置和大小，设置数字人服饰不同颜色功能。
- 本产品支持3D数字人\2D数字人\视频数字人，让用户获得更佳的人机交互体验。
- 本产品可支持私有化部署服务，保障客户数据隐私安全。

5. 产品功能

功能	特性	备注
回声消除	回声消除-25dB	标配
多模态降噪	噪声抑制-10dB	标配
唤醒方式	人脸唤醒、检测唇形，免语音唤醒，人脸检测准确率达99.99%	标配
开关机设置	支持远程自定义开关机设置	标配
串口通信	支持串口发送指令给上位机	标配
多模态降噪音频输出	通过 HPOUT 口输出多模态降噪后音频给上位机使用	标配
视频输出	通过千兆网口输出视频内容给上位机使用	标配
数字人后台管理系统	可远程管理数字人系统，修改数字人形象选择，设置UI背景、欢迎词、提示词，音色、升级知识库等功能，实时检测程序在线离线状态，远程管理多台数字人设备，开关机时间。	软件标配
大模型交互	自带体验版大模型交互，让人机交互更自然	软件标配
城市天气	支持自定义城市天气更新	软件标配
关闭系统任务栏	防止误触，程序界面快捷关闭系统任务栏	软件标配
前端编辑界面	具备密码登录，可设置设备自启，待机广告，关闭系统任务栏，任意移动位置和大小，设置数字人服饰不同颜色功能	软件标配
数字人角色身份设定	通过描述设定数字人的身份信息，设置完成后数字人会根据身份信息进行交流回复，如：“你现在的身份是康熙皇帝，请用该身份的角度与我对话”	软件标配
在线识别	支持中英文双语识别	标配
语音合成	支持中英文双语合成	标配
讲解弹窗	支持答案附带照片弹窗讲解功能	软件标配
播报模式	具备数字人播报模式（可语音指令切换）	软件标配
待机广告	支持休眠播放广告视频模式，在无人交流时可播放预制广告视频，摄像头感应到人体触发进入数字人对答模式。	软件标配
网络检测	实时检测联网状态，具备断网提醒功能	软件标配
屏幕比例选择	支持竖屏和横屏方案	软件标配
智能硬件控制	语义理解控制指令可对接中控系统以及其他智能硬件（支持TCP、UDP协议）	选配功能
AI大模型适配	第三方大模型（deepseek、千问、豆包、扣子、文心一言、智谱）	选配功能
支持私有化部署	保障企业信息安全	选配功能

表 2 AI数字人多模态语音盒子产品功能表

6. 效果说明

在摄像头高度 1.5-2.1 米，人身高 1.5-1.9 米，人声和噪音信噪比为-5db，光照度 300Lx 场景下，不同距离的识别准确率如下：

交互距离	1.5 米~2 米	2 米~2.5 米
字准确率	95%	93%

表 3 不同距离下多模态识别准确率

7. 接口说明



图3 AI数字人语音交互盒子产品接口

7.1接口参数表

编号	接口名称	说明
DC-12V	电源输入接口	DC12V3.5A 、DC5.5-2.5
USB-A	USB 接口	USB3.0-A*1 USB2.0-A*1
HDMI	HDMI 接口	HDMI 2.1 接口*1, 可最高输出 8K 60Hz, HDCP2.3
Audio in/out	音频输入/出	采用四段3.5国际标准OMTP定义 支持音频输出及输入
TF卡座	TF卡插槽	支持最大128GTF存储卡扩展
LAN	RJ45以太网	10/100/1000M自适应以太网
WIFI天线	SMA接口	2.4G/5.8G双频外置天线
DB9	RS232接口	RS232*1, 串口通讯
Mic/Camera	DB25接口	匹配定制麦克风和摄像头用

表 4 产品接口参数表

7.2 配件明细表

配件名称	数量	单位	备注
摄像头/麦克风	1	个	标配线长度约2.3米
AI主机盒子	1	套	标配rk3588\8G\64G
电源适配器	1	1.2m	12V3.5A
天线	1	20cm	外置棒状天线
盒子固定挂件	1	对	盒子固定挡板

表 5 产品配件详单表

8. AI数字人多模态语音交互盒子规格参数

产品参考图/ Product Reference Drawing		型号/model: MV-AIBOX
主机配置 /Host Configuration	菜单语言/OSD Language	中文，英文/Chinese, English
	系统/System	安卓/Android 12
	CPU/CPU	RK3588
	内存/RAM	8G
	硬盘/HD	64G (标配)
接口 / Interface	网口/Ethernet	RJ45 10M/100M/1000M
	USB	USB2.0*1, USB3.0*1
	WIFI	WIFI 天线*1
	其他接口/Other interfaces	HDMI*1, DC in*1,DB9-232*1,DB25*1, WIFI*1, audio in/out*1
语音交互模块 / Voice interaction module	摄像头/Camera	500W 宽动态
	麦克风/MIC	4 列阵AI降噪麦（支持噪声抑制、VAD 检测、 信噪比66db、灵敏度-38）
	音频系统/Audio Sytstem	Audio In/out*1
	多模态语音识别	唇动兼测多模态语音识别
	对话最佳距离	1.5米
	唤醒方式	人脸
	最佳对话区域	30度角扇形
	对话交互人数	单人
	识别精准率	在最佳距离1.5M位置，人声和噪音信噪比为- 5db，光照度300Lx场景下，识别精准率为（95 %）
	电源/Power	12V3.5A

其他配置 /Other configuration s	工作电流	裸板典型值 0.20A 最大值 0.25A
	最大功率/Maximum Power	< 40W
	工作环境	-10~75° 相对湿度≤80%
	外形尺寸	175*167*38mm
	联网要求	在线\500M

表 6 产品规格参数表

9. 硬件尺寸信息

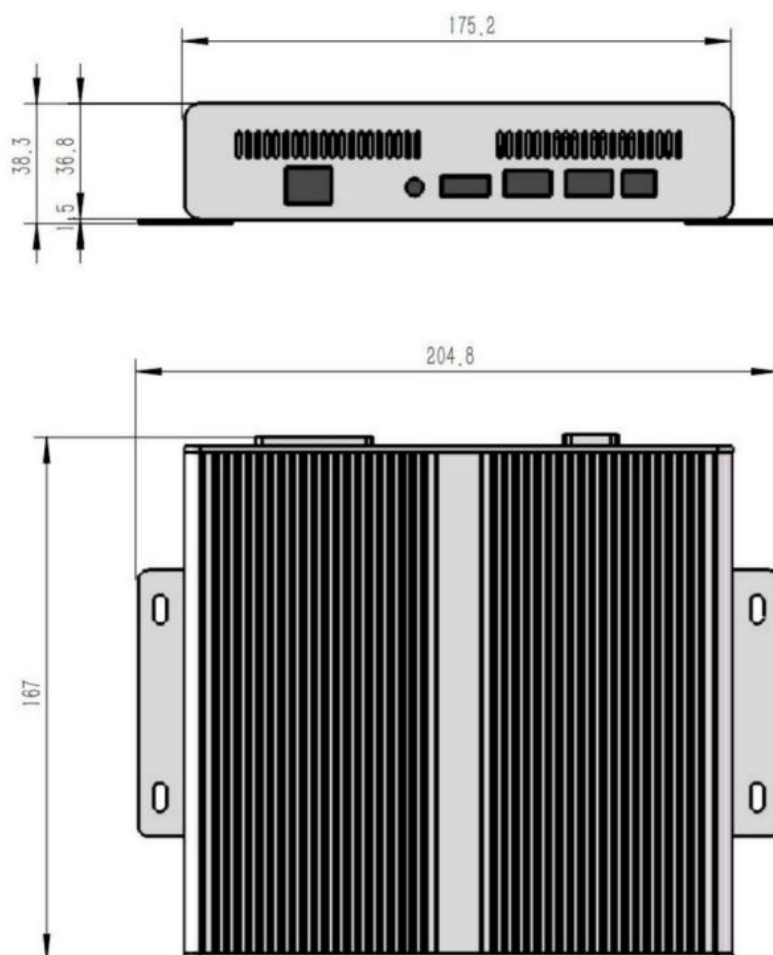


图4 AI语音盒子尺寸

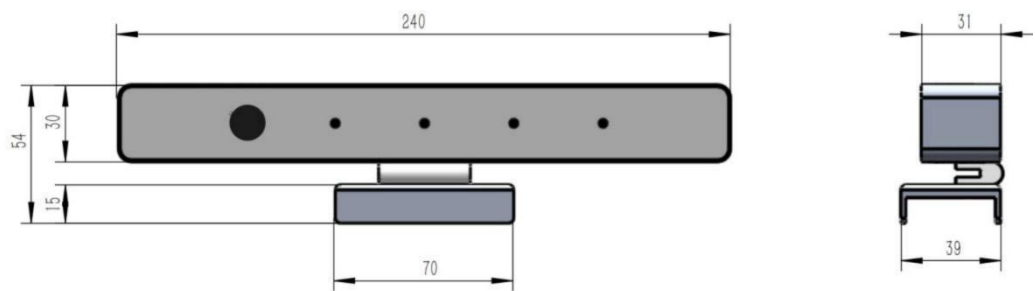


图5 麦克风摄像头尺寸