arXiv:2602.01201v1 [cs.HC] 1 Feb 2026

看着我，而不是幻灯片：帮助演讲者改善演讲中眼神交流的实时可穿戴助手

杜凌宇 代尔夫特理工大学 代尔夫特，荷兰
lingyu.du@tudelft.nl

Xucong Zhang
代尔夫特理工大学
代尔夫特, 荷兰
xucong.zhang@tudelft.nl

Guohao Lan 代尔夫特理工大学 荷兰代尔夫特
g.lan@tudelft.nl

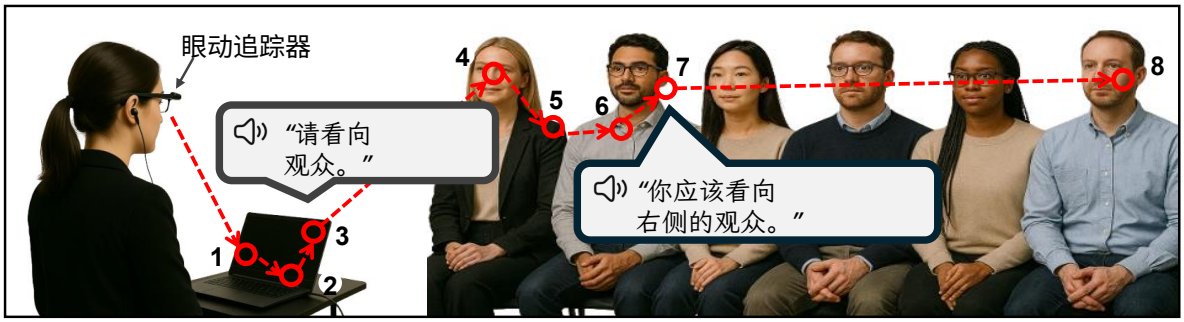


图 1：示意图展示了 SpeakAssis 如何帮助演讲者在演讲中保持有效的眼神交流。SpeakAssis 持续分析演讲者的视觉注意力，并提供实时反馈，以促进演讲者与观众之间平衡的互动。红色虚线表示演讲者的注视轨迹。起初，演讲者注视笔记本电脑（注视点 1–3）。SpeakAssis 检测到这一情况后，发出眼神接触提示：“请看向观众。”随后，演讲者将注意力转向观众，但只注视左侧（注视点 4–7）。对此，SpeakAssis 提示：“您也应该看看右侧的观众。”随后，演讲者将视线转向右侧（注视点 8）。

摘要

有效的眼神交流是成功进行公开演讲的基石。它能增强演讲者的可信度，并促进观众参与。然而，掌握有效的眼神交流需要大量培训和练习，这往往给新手演讲者带来巨大挑战。本文介绍 SpeakAssis，这是首个旨在现场演讲期间主动帮助演讲者保持有效眼神交流的实时情境可穿戴系统。SpeakAssis 利用头戴式眼动追踪器采集注视信息和场景视图，持续监测并分析演讲者在观众区域和非观众区域的注视分布。当检测到无效的眼神交流模式（例如眼神接触不足，或忽视某些观众区域）时，SpeakAssis 会通过耳机及时提供结合情境的音频提示，引导演讲者的注视行为。我们通过一项涉及 8 名演讲者和 24名观众的用户研究对 SpeakAssis 进行了评估。定量结果显示，SpeakAssis 平均使演讲者的眼神接触时长增加了 62.5%，并促进视觉注意力分布更加均衡。此外，基于观众调查的统计分析表明，演讲者眼神交流行为的改善，显著提升了观众在演示中的感知参与度和互动性。

关键词

眼神交流、实时辅助、公开演讲、可穿戴注意力追踪

1 引言

眼神交流是公开演讲中最有力、最重要的非言语线索之一 [21]。有效的眼神交流能让演讲者吸引并保持听众的注意力，建立联结感，并提升整体参与度 [10]。研究表明，能够进行有意义眼神交流的演讲者会被认为更可信，展现出更强的能力、诚实和自信 [12]。相反，眼神交流不足或不得当，往往会被解读为焦虑 [13, 30, 50] 或缺乏自信 [34, 35], ，从而削弱演讲者的权威性，降低演讲的有效性 [49]。鉴于眼神交流的重要性，有效的眼神交流被广泛视为成功公开演讲的核心要素 [10] ，专业公开演讲组织（如国际演讲会）也经常在表现评估中强调这一点 [17]。

建立有效的眼神交流，不只是看着观众；熟练的演讲者会长时间保持眼神交流，避免目光过度集中在某一个人身上，并主动将视觉注意力分散到房间的各个区域，确保每个人都感受到自己受到关注 [18, 21, 27, 46]。然而，这项技能并非与生俱来 [23]。在公开演讲所带来的认知与情绪负荷下学会管理注视，通常需要有意识的训练和反复练习 [49]。许多新手演讲者难以将这项技能内化，尤其是在现场演示这种动态环境中。

目前已有多种注意力追踪系统，可针对演讲者的注视行为提供反馈 [16, 17, 19, 50]。例如，基于虚拟环境的培训平台 [16, 17, 50]会模拟虚拟观众，并提供简单、即时的

反馈，例如演讲者进行眼神交流时，虚拟化身会点头或微笑。此外，可穿戴眼动追踪解决方案 [19, 33] 可让演讲者录下演讲过程中的注视行为，并在事后回看，以反思注意力模式，从而在未来加以改进。尽管这些解决方案适用于离线训练，但并非为现场演讲期间支持演讲者而设计。它们缺少实时现场反馈，无法帮助演讲者及时调整注视行为。因此，演讲者只能依靠先前的训练和内化的策略，而这些对于承受压力的新手演讲者来说并不足够。

为填补这一空白，我们推出 SpeakAssis，这是首款旨在帮助演讲者在现场演讲中改善眼神交流的实时、现场可穿戴助手。SpeakAssis 利用头戴式眼动追踪器采集的场景视图和注视信号，持续监测并分析演讲者对所有观众及非观众区域（如笔记本电脑、天花板和背景）的注视分布。当检测到无效的眼神交流模式时，例如长时间注视非观众区域、过度盯着某一位观众，或忽视坐在边缘位置的听众成员，SpeakAssis 会通过无线耳机向演讲者生成并传递及时且结合语境的眼神交流提示。图 1 展示了 SpeakAssis 如何在现场演讲中帮助演讲者。当系统检测到演讲者一直盯着笔记本电脑、没有与观众进行眼神交流时，便会发出提示 “看向观众”；当演讲者总是忽视房间右侧的观众时，系统便会生成诸如 “看向右侧的观众” 之类的提示。这种实时指导能让演讲者立即调整，并在整个演讲过程中保持有效的眼神交流。

提供恰当且及时的指导，需要系统实时准确地跟踪演讲者的注视目标。具体而言，系统必须识别演讲者正在注视哪位观众成员，即目标人脸，并监测演讲者的注视如何随时间分布。然而，在现场演讲中，这尤其具有挑战性，因为演讲者的头部持续移动、场景视图不断变化，摄像机角度变化和遮挡也会导致观众面部外观发生变化。此外，观众成员可能会移动，或间歇性地出现在场景视图中，进一步增加了跨帧可靠跟踪的难度。

为应对这一挑战，我们开发了一种新颖的识别方法，利用观众成员之间的相对空间位置来识别目标人脸。具体而言，我们将人脸识别与人脸跟踪相结合，以便在场景视图中高效识别一位参考观众，并将其指定为锚点人脸。确定锚点人脸 后，即可根据锚点人脸的相对位置推断目标人脸 的身份。即使摄像机频繁移动且观众的面部外观不断变化，这种方法也能让 SpeakAssis 稳健而高效地跟踪演讲者对不同观众成员及非观众区域的注意力。

我们使用市售硬件实现 SpeakAssis：使用现成的可穿戴眼动追踪器捕捉场景和注视，使用笔记本电脑（通常用于显示幻灯片）处理数据，并使用一只耳机传递隐蔽的音频提示。

我们提供了一个简单的界面，以确保易用性和无障碍使用。总而言之，我们的主要贡献有三点：

- 我们提出了 SpeakAssis，这是首个可在真实场景中实时提供可执行反馈的可穿戴系统，旨在帮助演讲者在公开演讲中改善眼神交流。

- 我们提出了一种稳健的识别技术，通过利用观众布局中的相对空间关系来识别演讲者的注视目标。该技术能够持续追踪场景视图中演讲者对所有观众及非观众区域的注视分布。这些能力对于系统在频繁的头部运动和动态场景变化下提供可靠的眼神接触提示至关重要。

- 我们邀请 28 名参与者（四名演讲者和 24 名观众）开展了一项全面的用户研究，以评估 SpeakAssis 在真实演讲场景中的表现。定量分析表明，SpeakAssis 平均可将演讲者的眼神接触时长提高 62.5%，并促进视觉注意力分布更加均衡。此外，基于研究结束后开展的观众调查所进行的统计分析证实，演讲者眼神交流行为的改善显著提升了观众的感知参与度和互动性。

2 相关工作

2.1 眼动追踪及其应用

眼动追踪系统大致分为两类：远程眼动追踪和近眼眼动追踪。远程眼动追踪使用放置在距离用户一定距离处的摄像头，可在多种设备上实现，包括手机 [25], 笔记本电脑 [53], 网络摄像头 [52], 和平板电脑 [24]。然而，远程系统通常存在追踪精度有限的问题，并且对环境光照条件高度敏感。相比之下，近眼眼动追踪采用放置在眼睛附近的摄像头，通常集成于智能眼镜等可穿戴头戴式设备 [1, 3, 26]中。这种设置能够显著提高追踪精度，并增强对头部运动和用户移动性的鲁棒性，因此非常适合用于动态和非受控环境。在本研究中，我们采用头戴式眼动追踪器，从演讲者处获取准确的注视数据，并专注于设计实时辅助功能，以支持公开演讲中的有效眼神交流。

眼动追踪技术的进步推动了其在各种领域中的广泛应用。特别是，由于眼球运动与视觉感知相关认知过程之间存在很强的相关性，眼动追踪已被广泛用于认知感知系统。典型应用包括早期检测自闭症谱系障碍 [22, 39],、心理工作负荷估计 [37, 51], 和驾驶员注意力监测 [2, 8]。此外，注视也是一种快速、直观的互动方式 [32], ，因此眼动追踪成为人机交互（HCI）中的重要组成部分。例如，Gaze-Guided Narratives [28]通过引导用户的注视，帮助他们找到参照物，例如城市全景中的建筑物。G-VOILA [48]利用大语言模型，支持用户在日常情境中基于注视的信息检索。CASES [38] 分析用户阅读时的注视模式，以提供情境辅助。眼动追踪也被用于研究公开演讲中演讲者的注视行为

，从而揭示注视模式如何影响观众参与。基于这些研究成果，我们探索了基于注视的可穿戴辅助，旨在帮助演讲者改善现场演讲中的眼神交流。

2.2 演讲辅助

我们的工作与先前旨在评估并改进演讲表现的研究相关 [9, 11, 15 17, 29, 31, 36, 40 42, 44, 45, 49, – –。54] [41] 例如，Strangert 等人分析了声学韵律特征以评估演讲者表现，表明优秀的演讲者往往展现出流畅表达、适当的语速以及动态的音高变化。为支持声音表达，Tanveer 等人 [45] 开发了一套使用谷歌眼镜的反馈系统，该系统监测演讲者的音量和语速，并生成诸如“更大声”和“更快”等实时语音提示，以在演示过程中引导演讲者。

其他研究探索了利用虚拟环境训练演讲者保持有效眼神交流 [16, 17, 49, 50]。这些系统通过响应式的虚拟观众行为提供反馈，例如当演讲者进行眼神交流时虚拟化身会点头或微笑 [17], ，或通过视觉指示，例如进度条变为绿色以表示眼神交流已足够 [16]。然而，这些系统是为在虚拟环境中练习而设计的，并不在真实世界的现场演示中提供反馈。此外，它们主要根据眼神交流相对于讲话时间的总体时长来评估眼神交流。相比之下，我们的工作聚焦于现场演讲期间的实时眼神交流辅助。我们不仅向演讲者提供现场反馈，还通过评估注视在观众中的空间分布，超越基于时长的指标，从而促进更平衡、更有效的观众参与。

3 系统设计

本节首先概述 SpeakAssis，然后详细介绍 SpeakAssis 的设计。

3.1 设计概述

3.1.1 设计目标。SpeakAssis 旨在帮助演讲者在现场演讲中保持专注且均衡的眼神交流行为。为此，我们将 SpeakAssis 设计为一种可穿戴系统，通过持续监测和分析演讲者的注视行为，提供实时现场反馈。该系统以两个主要目标为指导：（1）确保
演讲者始终将注意力放在观众身上，以及 (2) 促使注意力在各位听众成员之间均匀分布。因此，SpeakAssis 旨在检测并响应以下无效的眼神交流模式：

- 眼神接触不足。当演讲者很少看向观众时，例如演讲者注视观众的次数少于 20%，SpeakAssis 会提示演讲者与观众进行眼神交流。

- 注意力不均衡。 当某些观众区域受到的注意力显著较少时，SpeakAssis 会建议演讲者将注视转向关注不足的观众，以实现所有观众成员之间均衡的注意力分布。



图 2：SpeakAssis 的硬件平台。演讲者佩戴眼动追踪器和耳机，两者均连接到笔记本电脑。眼动追踪器配有一个前向场景摄像头和两个近眼摄像头。

3.1.2 硬件平台。 我们使用市售硬件实现 SpeakAssis：一台可穿戴眼动追踪器、一台笔记本电脑和一只耳机。如图 2 (a) 所示，演讲者佩戴眼动追踪器和耳机，二者均连接到笔记本电脑。各组件的作用如下：

- 眼动追踪器。我们使用 Pupil Core 眼动追踪器 [26] 捕捉演讲者的注视和场景视图，以追踪演讲者的注意力。如图 2 所示，眼动追踪器配有一个前向场景摄像头和两个近眼摄像头。场景摄像头以 1280×720 的分辨率捕捉场景视图，而近眼摄像头以 120 Hz 拍摄演讲者的眼部图像，用于注视估计。借助开源平台 Pupil Capture，我们将演讲者的注视位置获取为场景图像中的像素坐标。
- 笔记本电脑： SpeakAssis 软件运行在演讲者的笔记本电脑上，该设备通常用于显示演示文稿幻灯片。在本研究中，我们使用一台搭载 M1 Pro 芯片的 2021 年款 MacBook Pro，代表标准的通用型笔记本电脑，而非高端设备。SpeakAssis 软件会分析眼动追踪器捕获的注视信号和场景视频，并根据演讲者的注视分布生成实时眼神交流建议。
- 耳机： 演讲过程中，单只通过蓝牙连接的耳机用于向演讲者提供音频反馈。虽然我们的设置简单且有效，但也可以使用其他反馈渠道（例如智能眼镜或增强现实设备中的近眼显示器 [6, 7] ）来替代耳机。

3.1.3 SpeakAssis 工作流程。 图 3 展示了 SpeakAssis 工作流程。该系统分为两个主要阶段：演讲前的观众登记，以及演讲期间的注视监测与反馈。

注册阶段。在演讲开始前，SpeakAssis 会进行听众登记，为每位听众成员分配一个唯一标识符和一个人脸模板，以便之后进行身份识别。为启动此过程，演讲者会请观众面向自己，然后扫描房间，录制一段包含所有观众的短视频。由于观众通常分散在房间各处，难以在单个画面中捕捉到所有人，因此这一步是必要的。SpeakAssis 会处理场景视频，以检测每张人脸，并根据其在场景视图中从左侧到右侧的空间位置，为每位听众成员分配一个从 S_1 到 S_N的唯一标识符，其中 N 是观众的总人数

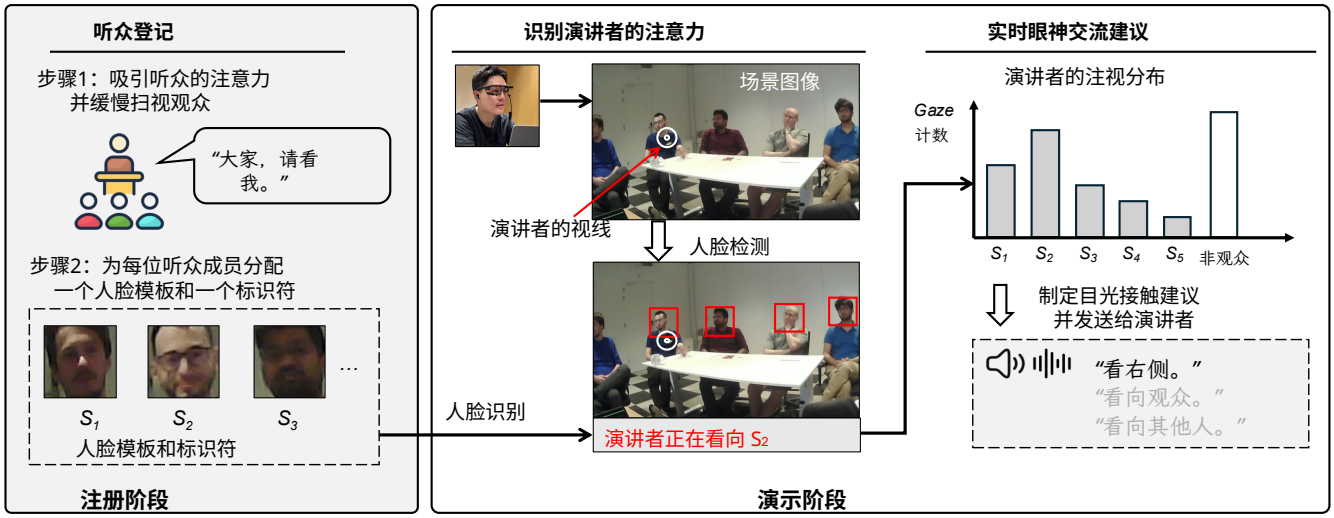


图 3：SpeakAssis 工作流程。在演讲前，SpeakAssis 执行听众登记。具体而言，演讲者首先吸引听众的注意力，然后使用眼动追踪器的场景摄像头扫描并记录所有观众。之后，SpeakAssis 在视频中检测人脸，并为每位听众成员分配一个人脸模板和唯一标识符。在演示阶段，SpeakAssis 持续跟踪演讲者的注视分布，涵盖观众和非观众区域。通过实时分析该注视分布，SpeakAssis 检测无效的眼神交流模式，生成适当的眼神交流反馈，并悄悄地将其传递给演讲者。

。每位观众还会关联一张从视频中提取的人脸图像，用作演讲期间进行人脸识别的模板。

演示阶段。在演讲期间，SpeakAssis 持续跟踪演讲者的注视分布，该分布反映了演讲者如何在不同观众和非观众区域之间分配视觉注意力。对于眼动追踪器传入的每张场景图像和相应的注视点，SpeakAssis 首先执行人脸检测，并通过测量注视点与检测到的人脸中心之间的距离来确定演讲者是否正在看某位听众成员。如果演讲者确实在看观众，SpeakAssis 会调用专为现场演讲场景定制的人脸识别模块（第 3.2 节），以确定正在被注视的是哪位听众成员。系统实时更新演讲者的注视分布。基于此分布，SpeakAssis 生成眼神交流提示，以帮助演讲者保持对观众的注意力并促进均衡的注视分配（第 3.3 节）。这些眼神交流提示通过耳机悄悄地传递给演讲者。

3.2 通过锚点跟踪在场景视频中实现稳健高效的人脸识别

为生成准确的实时眼神交流建议，SpeakAssis 需要及时更新的注视分布，以反映演讲者如何在观众和非观众区域之间分配视觉注意力。这又取决于能否快速、可靠地识别每张场景图像中的观众成员。具体而言，这项任务包括检测场景视图中的所有人脸，并识别演讲者当前正在注视的人脸，即目标人脸。形式上，如果检测到的人脸与演讲者的注视点之间的距离小于预设阈值 L，则该人脸被视为目标人脸。这样设计是为了弥补眼动追踪器注视估计中的微小误差。如果场景图像中未检测到目标人脸，则认为演讲者正在注视非观众区域（例如笔记本电脑）。下文将首先概述在真实演示场景动态中可靠识别目标人脸所面临的挑战，随后介绍我们提出的解决方案的具体细节。

3.2.1 挑战。 在标准笔记本电脑上以低延迟准确识别场景视图中的目标人脸，是 SpeakAssis 的一项关键要求。常见方法是先使用预训练的人脸检测模型（如 YOLOv8n-face [47]），再使用 Inception-ResNet-V1 [43] 等人脸识别模型提取检测到的人脸的人脸表征。随后，将该表征与已注册的人脸模板进行比对，以确定听众成员的身份。然而，由于演讲期间观众的面部外观与注册阶段采集的相应模板之间存在显著差异，这种传统流程往往导致身份识别准确率较低。

具体而言，这些差异源于现实世界中的若干因素。首先，演讲者在不同观众、笔记本电脑和显示屏之间转移视觉注意力时，会频繁改变头部姿态，从而导致场景图像中观众面部的视角发生变化。其次，观众自身在演讲过程中也常常移动或转动头部。此外，观众的面部有时会被手部部分遮挡。例如，当他们用手托住下巴、调整眼镜、挠头，或将手放在额头上时。这些姿态变化与遮挡共同造成了显著的

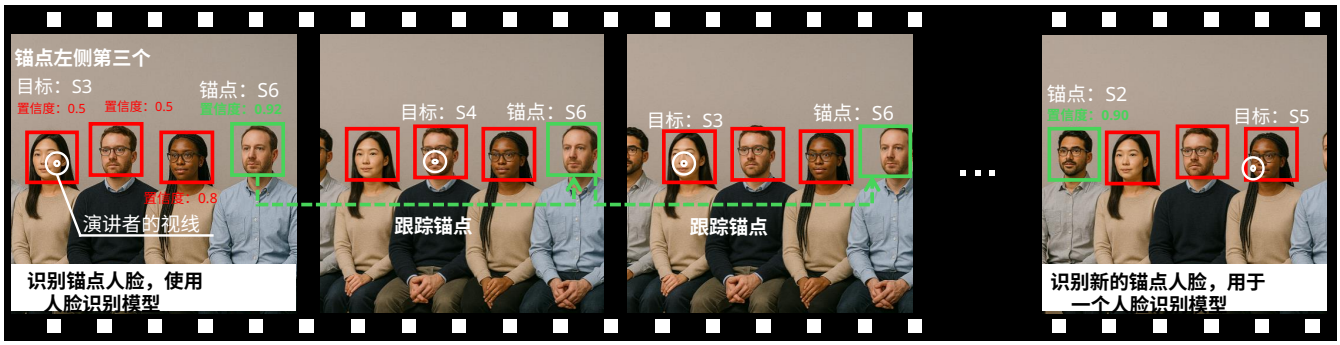


图4：所提出的人脸身份识别方法在场景视频中的示意图。我们引入了锚点人脸**的概念，即可以通过人脸识别模型或人脸跟踪算法被可靠识别的人脸。例如，在第一帧中，我们将人脸识别模型应用于所有检测到的人脸。最右侧的人脸获得了最高置信度分数** 0.92 **（量表范围为**0 **到**1**），被选为锚点人脸（以绿色边界框高亮显示）。一旦完成身份识别，锚点人脸会在后续帧中被跟踪（例如第二帧和第三帧），直到跟踪失败——例如，当锚点人脸在最后一帧中从场景视图中消失时。只要锚点人脸能够被成功跟踪或识别，我们就可以根据目标人脸相对于锚点人脸的相对空间位置来推断其身份。例如，在第一帧中，锚点人脸被识别为** S_6**。由于目标人脸位于锚点人脸左侧三个位置，因此推断其为** S_3**。如果锚点人脸跟踪失败或不再可见，我们会在当前场景帧上重新运行人脸识别模型以选择新的锚点人脸，并继续该过程。**

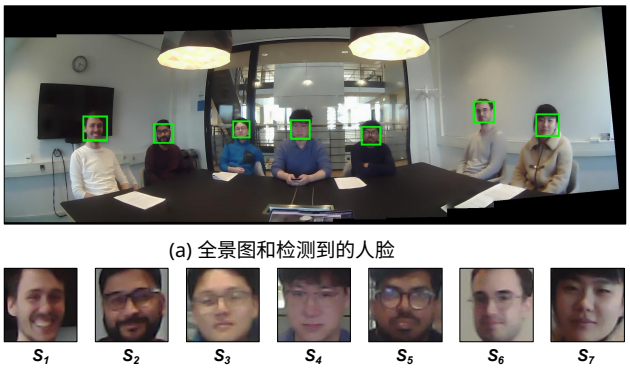
外观差异，使得直接从场景图像进行人脸识别变得尤为困难。

3.2.2 我们的解决方案。 为应对这一挑战，我们引入了锚点人脸这一概念，它指场景中能够由人脸识别模型或人脸跟踪算法可靠识别的人脸。

如图 4 所示，在场景视频的第一帧中，我们对所有检测到的人脸应用人脸识别模型。最右侧的人脸 S_6以绿色边界框标记，其识别置信度分数较高，为 0。92（范围为 0 到 1），因此被选为锚点人脸。随后，我们使用基于位置的人脸跟踪算法，在后续帧（例如第二帧和第三帧）中跟踪 锚点人脸，该算法依赖检测到的人脸的空间连续性，而非面部外观。 因此，即使面部外观发生变化或出现部分遮挡，也能实现稳健的人脸跟踪。锚点人脸会持续被跟踪，直到其从场景视图中消失，如图 4 的最后一帧所示。此时，我们会重新运行人脸识别模型，以选出新的锚点人脸。

在给定的场景图像中确定锚点人脸后，我们可以根据目标人脸相对于锚点人脸的位置推断其身份。这利用了演讲过程中观众的位置布局保持稳定这一观察结果。例如，在图4的第一帧中，目标人脸位于锚点人脸左侧第三位，因此我们可以推断其身份为 S_3。这些相对位置是在注册阶段确定的（见第3.2.3节）。

此外，我们避免在每张场景图像上运行计算开销巨大的face identification模型，转而依赖锚定人脸跟踪和基于位置的推断，从而显著降低系统延迟，并实现更高的帧处理速率。这会导致连续场景图像中的人脸位置出现轻微偏移，进而增强



(a) 全景图和检测到的人脸

(b) 从左到右排列的每位个体的已注册人脸模板

图5：由眼动追踪器拍摄的场景图像构建的全景图，以及按从左到右的相对位置为所有观众注册的、带有唯一标识符的人脸模板。

锚定人脸跟踪的鲁棒性。正如我们将在后续章节中展示的那样，我们的方法通过锚定人脸跟踪和基于位置的身份推断，有效减轻外观变化带来的影响，从而实现鲁棒的人脸识别。

3.2.3 详细设计。现在，我们将详细介绍所提方法，该方法用于在演讲者的场景视频中识别目标人脸。

推断观众的相对位置。在注册阶段，演讲者缓慢地从左向右扫视观众，SpeakAssis 同时录制场景视频。随后，SpeakAssis 从视频中采样 M 帧来构建全景图，如图5(a)所示。之后，它将人脸检测模型（预训练的 YOLOv8n-face） [47]，应用于全景图，并根据检测到的人脸的水平位置确定观众的相对位置。如图5

(b)所示，每张检测到的人脸都会被分配一个唯一标识符：最左侧的人脸标记为 S_1，最右侧的人脸标记为 S_N，其中 N 为检测到的人脸总数。带有标识符的已注册人脸模板如图5(b)所示，它将在演讲阶段由人脸识别模型用于确定锚点人脸。请注意，为确保演讲期间识别可靠，SpeakAssis 不会使用拼接全景图中的失真人脸作为模板。相反，它会从原始采样帧（用于构建全景图）中提取对应的人脸作为人脸模板，如图5(b)所示。

选择锚点人脸。 锚点人脸是当前场景图像中的一张人脸，可以通过人脸识别模型或人脸跟踪算法进行高置信度识别。如果前一帧中没有可用的锚点人脸，例如在演讲的第一帧中，SpeakAssis会将人脸识别模型（例如，在VGGFace2数据集 [14]上预训练的 Inception-ResNet-V1 [43],）应用于当前帧中的所有检测到的人脸。它会选择识别置信度最高的人脸作为候选锚点人脸。如果置信度分数超过阈值 p，则接受该候选人脸作为锚点人脸；否则，SpeakAssis会将选择过程推迟到下一帧。

如果前一帧中存在锚点人脸，如图4的第二帧和第三帧所示，SpeakAssis会尝试在当前帧中跟踪该人脸。具体而言，它会选择与前一帧中的锚点人脸在空间位置上距离最近的人脸作为候选人脸。如果两者之间的距离小于阈值 D，则将该候选人脸指定为锚点人脸；否则，例如当锚点人脸从当前帧中消失时（如图4的最后一帧所示），SpeakAssis会认为锚点人脸丢失，并使用人脸识别模型重新启动选择过程。

识别目标人脸。选定锚点人脸后，SpeakAssis 会通过计算目标人脸相对于锚点人脸的位置，推断同一帧中目标人脸的身份。例如，在图4的第一帧中，锚点人脸被识别为 S_6，目标人脸位于其左侧三个位置；系统由此推断目标人脸的身份为 S_3。

3.2.4 有效性。下面，我们将在真实环境中评估人脸身份识别方法的有效性。

设置。 我们招募了八名参与者进行评估，其中包括一名演讲者和七名观众。演讲者佩戴眼动追踪器，并与观众进行为时一分钟的交谈，同时录制场景视图。演讲者和观众都可以自然地活动身体。作为比较基线，我们采用第 3.2.1 节所述的直接方法。具体而言，我们使用与本方法相同的人脸识别模型，即 Inception-ResNet-V1 [43],，提取目标人脸的特征表示，并将其与所有已注册的人脸模板进行比较。如果目标人脸的特征表示与任一模板的相似度超过阈值 D，则将该目标人脸判定为与此人脸模板对应的身份。我们将人脸识别准确率 和处理延迟 作为评估指标。录制的视频包含 855 张场景图像。我们使用人工标注作为真实值，评估这些场景图像中所有检测到的人脸 的人脸识别准确率。我们对每张场景图像中检测到的人脸的身份进行人工标注

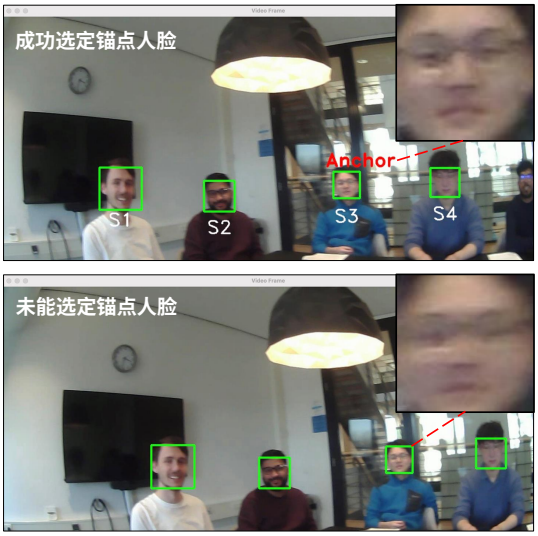


图6：真实场景中的人脸识别示例。上图展示了 SpeakAssis 成功选定锚点人脸的一帧。在下方的连续帧中，由于人脸位置发生显著变化，SpeakAssis 未能继续跟踪锚点人脸；同时，由于运动模糊，系统也无法选定新的锚点人脸，右上区域尤为明显。

。对于处理延迟，我们测量从接收到场景图像到生成识别结果所需的时间。

结果。表1报告了所提方法和基线方法的人脸识别准确率及每帧平均延迟。我们的方法准确率显著更高，延迟也更低。展示所提方法人脸识别性能的演示视频可通过此 Google 云端硬盘链接观看[1]。所提方法的大多数误分类都发生在演讲者快速转头时，导致 SpeakAssis 跟踪丢失锚点人脸。图6展示了一个示例，其中下图是上图的连续帧。在上方帧中，SpeakAssis 成功选定了锚点人脸。然而，在后续帧中，由于人脸位置发生显著变化，系统未能继续跟踪锚点人脸。此外，下方帧右上区域尤为明显的运动模糊降低了人脸识别模型的置信度，导致 SpeakAssis 无法选定新的锚点人脸。在这种情况下，SpeakAssis 会将这些帧视为不可靠帧，不为其分配任何身份，这也是观察到身份识别错误的原因之一。值得注意的是，运动模糊也会导致基线方法无法识别观众。

在效率方面，SpeakAssis 的处理延迟显著低于基线方法，因为它仅对 6% 的场景图像应用人脸识别模型。具体而言，对于一张场景图像，人脸识别模型在评估设备（2021 款 MacBook Pro，搭载 M1 Pro 芯片）上的处理时间为 30 毫秒，而人脸跟踪算法仅需 0。1 毫秒，因此处理速度更快，也更具可扩展性。

[1]人脸识别结果的可视化：http://bit.ly/4lzBums

表1：在录制的场景视频中，所有检测到的人脸的人脸识别准确率（百分比），以及所提出的方法和基线方法的平均每帧延迟（毫秒）。与基线方法相比，所提出的方法在真实世界中的动态演讲者与听众互动场景下，准确率显著更高，同时延迟也显著更低。

方法	准确率（%）	延迟（毫秒）
所提方法	93.3	28
基线方法	30.8	57

3.3 生成眼神交流建议

基于所提出的人脸身份识别方法，SpeakAssis 计算演讲者在观众和非观众区域中的注视分布。该分布构成了检测演讲者眼神交流行为无效的特定情况并生成适当反馈的基础。

3.3.1 计算演讲者的注意力分布。考虑一个包含 X 个连续场景图像和注视点、时长为 P 秒的时间窗口。令 $\bar{X}$ 表示选定目标人脸的场景图像数量，令 X_i 表示目标人脸被识别为 S_i 的场景图像数量。我们将此时间段内的**眼神交流比例**（EP）定义为：

$$EP = \bar{X}/X \cdot 100\%. \tag{1}$$

该指标反映了演讲者将注意力转向观众的帧所占的百分比。我们还将此时间段内特定观众成员 S_i 的**眼神交流分布**（ED）定义为：

$$ED(S_i) = X_i/\bar{X} \cdot 100\%. \tag{2}$$

ED 表示演讲者指向 S_i 的注意力占其指向所有观众成员的总注意力的比例。

基于这些指标，SpeakAssis 检测两种特定的无效眼神交流情况，并生成相应的反馈以支持演讲者。

3.3.2 眼神交流情境与建议。在当前设计中，SpeakAssis 会检测并响应演讲期间的两种特定眼神交流模式：

• 眼神接触不足：指演讲者很少看向观众的情况。SpeakAssis 每隔 n 秒检查一次这种情况。每个检测周期结束时，如果 EP 低于阈值r_P，系统便会判定演讲者与观众互动不足且缺乏眼神交流。SpeakAssis 会生成建议：“看向观众”。

• 注意力不均衡：指演讲者对某些观众的关注明显较少的情况。SpeakAssis 每隔 k 秒检查一次这种情况，并找出演讲者与其眼神交流分布最少的听众成员。为了向演讲者提供简单且可执行的反馈，SpeakAssis 会生成一条简单的建议，鼓励演讲者看向参与度较低的观众。具体而言，如果这位听众位于观众席左侧，SpeakAssis 会建议：“多看左侧”；否则，会建议：“多看右侧”。

我们通过实证将 $r_p = 20\%$、$n = 30$ 秒和 $k = 75$ 秒设定为相应值。生成的建议通过耳机传递给演讲者

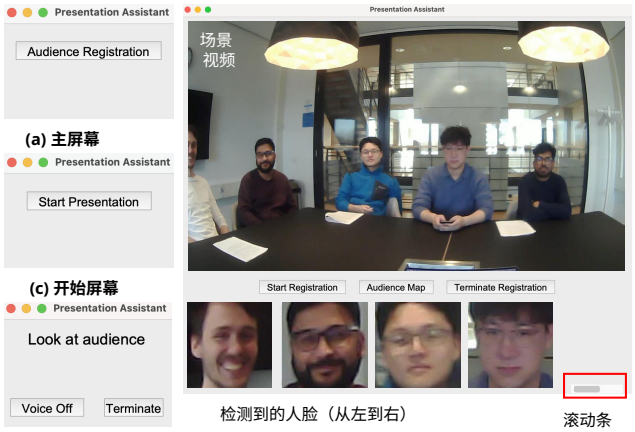


图 7：SpeakAssis 的用户界面（UI）。该用户界面包含四个屏幕：(a) 主屏幕；(b) 观众注册屏幕；(c) 开始屏幕；以及 (d) 演示屏幕。

耳机通过蓝牙连接到笔记本电脑。我们使用 Python 中的 pyttsx3 [4] 文本转语音转换库，将文本建议转换为音频提示。

3.4 用户界面设计

我们使用 Tkinter [5] 为 SpeakAssis 设计用户界面（UI）。如图 7 所示，该用户界面包含四个用户界面屏幕。下文将逐一详述。

• 主屏幕。 启动系统后，演讲者将看到主屏幕（图 7(a)）。点击“听众登记”按钮将进入观众注册屏幕。

• 观众注册屏幕。此屏幕用于观众注册阶段。如图 7(b) 所示，场景视频显示在顶部，中间有三个控制按钮。演讲者首先点击“开始登记”，该按钮随后切换为“停止登记”。录制进行期间，演讲者缓慢扫视观众，以捕捉所有成员。完成后，演讲者点击“停止登记”以结束录制。接下来，演讲者点击“观众地图”按钮，为每位听众成员登记模板人脸图像和唯一识别号码。这些模板人脸图像显示在屏幕底部，并根据观众的相对位置从左到右排列。当观众数量超出可见区域时，可使用滚动条浏览列表。最后，演讲者点击“终止登记”以进入“开始屏幕”。

• 开始屏幕。此过渡屏幕出现在注册之后、演讲开始之前。它采用紧凑布局，使演讲者能够打开演示幻灯片，并根据需要重新调整界面位置。准备就绪后，演讲者点击“开始演讲”即可开始。

• 演示屏幕。 演讲期间，此屏幕处于活动状态。生成的目光接触建议显示在屏幕顶部。屏幕底部有两个按钮：“VoiceOff”按钮用于静音语音建议，“Terminate”按钮用于退出 SpeakAssis。

表2：28名参与者的人口统计信息。

总人数	28（4名演讲者；24名观众）
性别	女性（4人），男性（22人）
Age	最小值：20，最大值：32
种族	东亚人（18）、欧洲人（7）、南亚人（3）



图 8：用户研究所用的会议室。演讲者可站在或坐在显示屏前进行演讲，观众则坐在按从最左侧到最右侧（以演讲者视角为准）排列的椅子上。显示屏用于展示演示文稿幻灯片。

4 评估

在本节中，我们通过在真实演示场景中对演讲者和观众进行用户研究，评估 SpeakAssis 的性能。

4.1 评估设置

4.1.1 参与者。我们招募了28名参与者参与该研究。他们的人口统计信息如表2所示。我们将参与者分为四组，每组由一名演讲者和六名观众组成。我们要求每位演讲者使用自己选择且熟悉的幻灯片进行时长约八分钟的演讲。该研究已获我们机构伦理委员会批准。

4.1.2 用户研究设计。用户研究在会议室内进行，如图8所示。在演讲过程中，演讲者可以自由选择站在或坐在显示屏前方，而观众则坐在从演讲者视角看由左侧至右侧排列的黑色椅子上。笔记本电脑连接至显示屏，用于展示演示文稿幻灯片。每组参与两个环节，演讲分为两个四分钟的部分，每个环节进行一部分。在两个环节中，演讲者均佩戴眼动追踪器，同时采集注视数据和场景视频。在每个环节开始前，使用眼动追踪供应商提供的校准软件为演讲者校准眼动追踪器。

在第一场中，演讲者不借助 SpeakAssis 完成演讲的第一部分。在第二场中，演讲者在 SpeakAssis 的辅助下完成演讲的第二部分。通过分析和比较两个场次记录的客观数据，以及所有参与者的主观反馈，我们可以考察 SpeakAssis 是否能改善演讲者与观众之间的眼神交流。为确保主观反馈的有效性，我们要求观众

表 3：面向演讲者的问卷。每个问题均采用 5点李克特量表作答。

问题	描述
1. **可学习性**	这个系统易于使用。
2. **注意力**	演讲助手帮助我 与观众进行更多眼神交流。
3. **平衡性**	演讲助手帮助我平衡 我对观众的注意力。
4. **精确度**	演讲助手能准确反映我的 眼神交流情况，例如，在我盯着 笔记本电脑时提醒我看向观众。
5. **有用性**	这些建议总体上有助于改善 与观众的眼神交流。
6. **舒适度**	硬件眼动追踪器不会影响 我的演讲过程。
7. **侵扰性**	演讲过程中，建议不会造成侵扰 。
8. **意愿**	我愿意在未来再次使用该系统来 改善眼神交流。

密切关注演讲者在两个场次之间眼神交流的变化。

为了避免影响演讲者自然的眼神交流行为，我们不会在第一场之前向演讲者透露研究的具体目的。相反，我们会请演讲者佩戴眼动追踪器，像往常一样进行演讲，并告知他们将监测和分析其注视行为。在第二场之前，我们会向演讲者说明研究的完整目的，并提供 SpeakAssis 的使用说明。随后，演讲者会进行演讲的第二部分。

第二场结束后，我们会请演讲者和观众填写一份问卷（详见第 4.1.3 节），并邀请他们提供开放式反馈。每组研究约持续 30 分钟。

4.1.3 评估方法。我们希望探讨以下研究问题（RQs）：

- **RQ1：** *SpeakAssis* 是否能增加演讲者在演讲过程中与观众的眼神接触时长？
- **RQ2：** *SpeakAssis* 是否能帮助演讲者将注意力更均匀地分配给各位听众成员？
- **RQ3：** 演讲者是否认为 *SpeakAssis* 能有效改善他们在演示过程中的眼神交流？
- **RQ4：** 使用 *SpeakAssis* 是否能提升观众参与度，并使演讲更具互动性？

我们引入了不同的性能指标进行评估。具体而言，为回答 **RQ1**，我们使用 眼神接触比例（ *EP* ）作为定量指标。一个环节中的 EP 定义为演讲者注视观众的视频帧（由场景摄像头捕获）所占的百分比。该指标近似表示演讲者在整个环节中与观众进行眼神交流的时间比例。EP 越大，表明演讲者投入给观众的视觉注意力越多，反之亦然。

表4：面向观众的问卷。

问题	描述
1. **重要性**	您认为演讲者与 观众在演讲过程中的眼神交流是否重要?
2. **差异**	你认为演讲者在眼神交流 模式上，两场次之间是否有所不同?
3. **参与度**	在哪一场次中，演讲者通过眼神交流 更能吸引你的注意力?
4. **互动**	在哪个环节，演讲者通过眼神交流让你 感觉互动性更强?

对于 **RQ2**，我们引入了一项称为 注视分布熵（GDE）的指标，用于量化演讲者在一次演讲环节中跨听众成员的注视分布的分散程度。令 X_i 表示演讲者注视听众成员 S_i 时的视频帧数，并令 N 表示观众成员的总人数。我们定义 $X_{sum} = \sum_{i=1}^{N} X_i$。然后，注视分布熵计算如下：

$$GDE = -\sum_{i=1}^{N} \frac{X_i}{X_{sum}} \log \frac{X_i}{X_{sum}}, \tag{3}$$

其中 X_i/X_{sum} 表示演讲者注视听众成员 S_i的比例。较高的 GDE 表明注意力在所有观众之间分布得更加均匀。对于 $N = 6$, GDE 的范围为 0 至 1。79。 GDE 为 0 表示演讲者只关注某一位听众，而 GDE 为 1.79 则表示演讲者将相同数量的注视分配给观众中的每位听众。

对于**RQ3**和**RQ4**，我们通过研究后问卷收集主观反馈，问卷由各用户研究组的参与者填写。演讲者填写表 3 中所用的问卷。第一题评估 SpeakAssis 的可学习性，接下来的四题评估其在改善眼神交流方面的有用性。随后两题询问演讲过程中可能产生的负面影响，最后一题询问演讲者未来使用 SpeakAssis 的意愿。每题均采用 5点李克特量表进行回答和评分（1：非常不同意；2：不同意；3：中立；4：同意；5：非常同意）。我们对收集到的 5点李克特分数进行单侧 Wilcoxon符号秩检验。

观众填写一份单独的问卷，如表4所示。听众问卷首先询问观众对演讲中眼神交流**重要性** 的看法，随后询问他们对演讲者在两个环节中的眼神交流模式**差异** 的观察。这两个问题的回答均为

是 或否。接下来的两个问题询问他们觉得哪一个环节更**投入** 且**互动性更强**。研究人员请观众在两个环节之间二选一，也可以选择相同。他们还可以提供开放式评论。

4.2 结果与分析

下面我们将讨论各项研究问题的评估结果与发现。

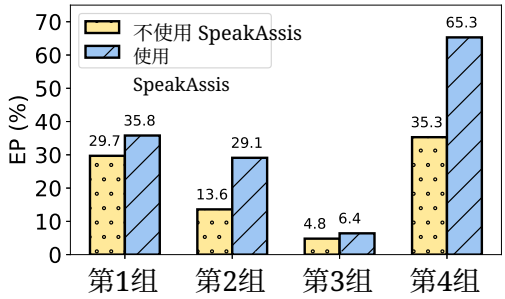


图9：演讲者使用和未使用 SpeakAssis 时的眼神接触比例（EP，以百分比表示）。总体而言，SpeakAssis 提高了四组演讲者的眼神接触比例，平均提升了 62.5%。

RQ1 的研究发现：SpeakAssis 显著增加了演讲者的眼神接触时长。

图9展示了四个组在两场演讲中的眼神接触比例。如前所述，EP 值越大，说明演讲者越关注观众。第一场中，演讲者在不使用 SpeakAssis 的情况下进行演示；第二场中，他们则会收到 SpeakAssis 提供的实时眼神接触指导。

总体而言，四组在使用 SpeakAssis 后的 EP 均有所提高，这表明 SpeakAssis 能有效帮助演讲者与观众进行更多眼神交流。平均而言，眼神接触比例提升了 62.5%。值得注意的是，第2组和第4组的提升比第1组和第3组更为显著。我们推测，这些差异可能源于演讲内容或演讲者特征的不同。例如，收到 SpeakAssis 的反馈后，第 2 组的演讲者比第 3 组的演讲者更愿意也更能够持续关注观众。此外，在演讲过程中，第3组的演讲者经常需要查看并讲解幻灯片上的内容，这减少了他们与观众保持眼神交流的机会。

此外，第2组的演讲者表示：“我在演示中的一个个人习惯是，有时我会面向观众，却并没有真正看着他们。*SpeakAssis*能够准确检测到这一点，并提醒我看向观众。”这一反馈进一步支持了SpeakAssis在演示中增加演讲者眼神接触时长的能力。

RQ2的研究发现：SpeakAssis帮助演讲者更均匀地将注意力分配给观众。

两个场次和四个组别的注视分布熵如图10所示。更高的 GDE 表示注视在观众中分布得更均匀，反之亦然。总体而言，SpeakAssis提高了全部四个组别的GDE，这表明所有演讲者在使用SpeakAssis时都拥有更平衡的注意力分布。平均而言，GDE提高了17.4%。

为了深入了解，图11展示了第1组和第4组中，演讲者对不同座位位置观众的注视分布。如图11(a)所示，在未使用 SpeakAssis 的情况下，第1组的演讲者往往会忽略坐在右侧的观众成员，尤其是位于“右侧”和“最右侧”的观众

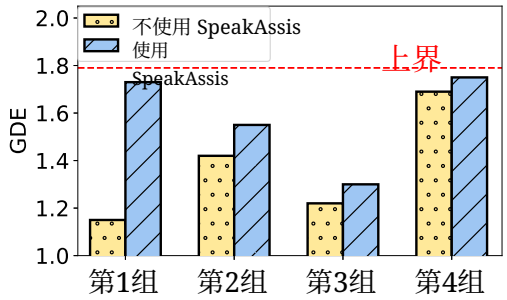


图10：演讲者使用和不使用 SpeakAssis 时的注视分布熵（GDE）。总体而言，SpeakAssis 鼓励演讲者更均匀地分配注视。平均而言，GDE 提高了 17.4%。

位置。使用 SpeakAssis 并收到眼神交流指导“多看右侧”后，演讲者将更多注意力转向这些观众成员，从而使注意力在观众中的分布更加平衡。

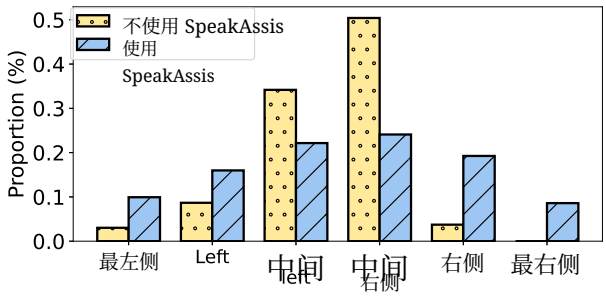
图11(b)显示，在第4组中，演讲者在第一场（不使用 SpeakAssis）期间没有忽略任何观众，因此其注视分布熵高于其他演讲者（见图10）。然而，与第1组的演讲者类似，这位演讲者在不使用 SpeakAssis 时，也较少关注坐在右侧的观众。相比之下，在第二场使用 SpeakAssis 后，演讲者将注视更均匀地分配给所有观众，使注视分布熵接近最大值。一位观众的评论进一步印证了这一量化观察：“作为坐在角落的一员，我在第二场获得了更多关注”，凸显了 SpeakAssis 在平衡演讲者视觉注意力方面的有效性。

RQ3 的研究发现：演讲者认为 SpeakAssis 能有效改善演示中的眼神交流。

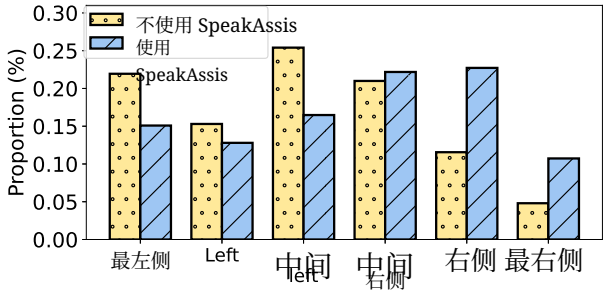
我们在图12中呈现了演讲者对问卷的回答。总体而言，结果表明，演讲者普遍认同 SpeakAssis 具有以下优点：(1) 系统易于使用（**可学习性**，$p = 0.0625$）；(2) SpeakAssis 有助于改善眼神交流（**有用性**，$p = 0.0625$），尤其能够延长眼神交流的持续时间（**注意力**，$p = 0.0625$），并促进在观众之间更加平衡地分配注意力（**平衡性**，$p = 0.0625$）。这项主观反馈与定量结果一致，后者显示所有演讲者的眼神交流比例（图9）和注视分布熵（图10）均有所改善。然而，演讲者也指出了当前实现的一项局限性：在演示过程中，语音建议会被认为具有侵扰性（**侵扰性**，$p = 0.7181$）。我们承认这是当前原型的一项局限性，并将在第4.3节中进一步讨论。尽管如此，仍有三位演讲者表示未来有兴趣使用 SpeakAssis（

意愿，$p = 0.1875$）。第4组的演讲者强调了 SpeakAssis 作为培训工具的巨大潜力，可帮助新手演讲者提升眼神交流技巧。

RQ4 的发现：SpeakAssis 提升了观众的参与度和演讲的互动性。



(a) 第1组演讲者在观众中的注视分布。



(b) 第4组演讲者在观众中的注视分布。

图11：演讲者的视线注视不同位置观众的比例（百分比），分别见于（a）第1组和（b）第4组。在两组中，演讲者最初（不使用 SpeakAssis）较少注意坐在右侧的听众成员，即位于“右侧”和“最右侧”的听众成员。在 SpeakAssis 的辅助下，他们的注视分布变得更加平衡，注意力更均匀地分布在观众之间。

表5总结了观众对表4中问卷的回答。具体而言，对于有关**重要性** 和**差异**的问题，答案可以是 是 或否。对于有关**参与度**和**互动**的问题，答案可以是第一场、第二场或相同。表5汇总了四个小组共24名观众的回答。总体而言，所有观众都认可在演示中保持眼神交流的重要性。

此外，如表5所示，在19名表达了偏好的观众中，有18人更喜欢使用SpeakAssis的演讲，仅有一人更喜欢未使用它的演讲。为评估这一偏好是否具有统计显著性，我们进行了单尾二项检验 [20]。零假设假定观众对两种条件没有偏好，而备择假设则认为观众更有可能偏好使用SpeakAssis的演讲。在显著性水平为 $\alpha = 0.05$时，检验得到的 p值为0.000038，表明在零假设下出现该观察结果的可能性极低。因此，我们拒绝零假设，并得出结论：观众认为使用SpeakAssis的演讲显著更具参与度和互动性。

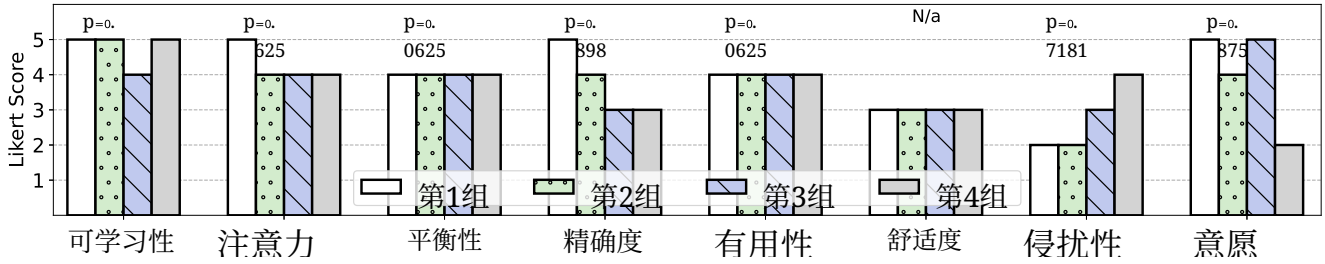


图12：sp的回应　　eakers在用户研究中对不同组别问卷的回应，以及单尾的 p值　　ed Wilcoxon符号秩检验fo　　r每个问题。

表5：观众对问卷回应的总结。75%的观众在第二场感觉更加投入，并认为演讲更具互动性；在该场中，演讲者从SpeakAssis获得实时眼神交流建议。

问题	回答		
	Yes	*No*	
重要性	24	0	
差异	19	5	
	第一场	第二场	*Same*
参与度	1	18	5
互动	1	18	5

为获得更深入的见解，我们对报告称第一场比第二场更投入的那位听众成员进行了后续访谈。访谈显示，该听众对演讲主题兴趣有限，并在第一场中很快失去了专注。这使其在两个场次中总体上都感到缺乏参与，尽管其认为第一场略更具参与感。值得注意的是，所有报告在两个场次之间的**参与度**或**互动**方面未感知到差异的观众均来自第4组。该组的演讲者比其他演讲者拥有更丰富的演讲经验。图9和图10所示的结果进一步支持了这一点：第4组的演讲者在两个场次中的EF和ADE均持续取得更高的数值，表明其演讲技巧更为精湛，眼神交流行为也更为有效。

4.3 局限性与未来研究方向

当前设计的一个局限性在于，现场演讲期间的语音建议可能会造成一定干扰。几位演讲者指出，音频提示虽有帮助，但偶尔会分散他们的注意力，尤其是在呈现复杂或技术性内容时。这凸显了探索其他反馈模态的必要性：它们既能有效引导演讲者行为，又能减少干扰。未来研究的一个有希望的方向是探索非侵入式反馈渠道，例如通过近眼显示器在演讲者的周边视觉范围内呈现不显眼的文本建议。Halliday 眼镜 [6] 和 Vuzix 智能眼镜 [7]便是这类设备的实例。这类显示器可以让演讲者接收反馈，同时不打断其口头表达或与观众的视觉互动。借助更先进的可穿戴原型，我们可以打造更无缝、自适应且易于使用的 SpeakAssis 系统，通过实时反馈提升公开演讲效果。

5 结论

本文介绍了 SpeakAssis，这是首个能够在现场演讲中实时提供情境化建议、帮助演讲者保持有效眼神交流的可穿戴系统。SpeakAssis 借助头戴式眼动追踪器，追踪并分析演讲者在不同观众成员及非观众区域之间的注视分布。基于此，SpeakAssis 能够检测无效的眼神交流，并生成恰当的建议，帮助演讲者将注意力集中、均衡地分配给观众。我们通过一项涉及四名演讲者和24名观众的用户研究对 SpeakAssis 进行了评估。定量结果表明，SpeakAssis 在眼神接触时长和注意力分布方面显著改善了演讲者的眼神交流。此外，对观众主观反馈的统计分析表明，SpeakAssis 显著提升了他们的参与度。

参考文献

[1] 2014。SMI 眼动追踪眼镜。https://imotions.com/hardware/smi-eye-tracking-glasses/。[2] 2023。Tobii 驾驶员监控系统。https://www.tobii.com/products/ automotive/tobii-dms。[3] 2023。Tobii Pro Glasses 3。https://www.tobiipro.com/product-listing/tobii-pro- glasses-3/。[4] 2024。适用于 Python 3 的文本转语音（TTS）库。https://pypi.org/project/pyttsx3/。[5] 2024。tkinter——Python 接口。https://docs.python.org/3/library/tkinter.html。[6] 2025。Halliday Proactive AI 眼镜。https://hallidayglobal.com/。[7] 2025。Vuzix 智能眼镜：面向医疗及更多领域的透视显示屏。https://www.vuzix.com/products/m4000-smart-glasses。[8] Sam Abuelsamid。2020。Seeing Machines 与宝马合作开展 CSE 驾驶员监控。https://www.forbes.com/sites/samabuelsamid/2020/01/07/seeing-machines-and-bmw-team-up-on-driver-monitoring-at-ces/。[9] Manuel Bachmann、Abimanju Subramaniam、Jonas Born 和 David Weibel。2023。虚拟现实公开演讲训练：有效性与用户技术接受度。*Frontiers in virtual reality* 4（2023），1242544。[10] Leila Barati。2015。教师与学生之间眼神交流对二语学习的影响。*Journal of Applied Linguistics and Language Research* 2，7（2015），222–227。[11] Patryk Bartyzel、Magdalena Igras-Cybulska、Daniela Hekiert、Magdalena Majdak、Grzegorz Łukawski、Thomas Bohné 和 Sławomir Tadeja。2025。探索用户对用于语音与公开演讲训练的语音控制虚拟现实环境的接受度。*Computers & Graphics* 126（2025），104160。[12] Steven A Beebe。1974。眼神交流：决定演讲者可信度的非言语因素。*Communication Education* 23，1（1974），21–25。[13] Scott Campbell 和 James Larson。2013。公开演讲焦虑：面对面演讲与网络演讲的比较。*Journal of Instructional Pedagogies* 10（2013），1–8。[14] Qiong Cao、Li Shen、Weidi Xie、Omkar M Parkhi 和 Andrew Zisserman。2018。VGGFace2：一个用于识别不同姿态和年龄人脸的数据集。在*Proceedings of*

第 13 届 IEEE 国际自动人脸与手势识别会议（FG 2018）。[15] Mathieu Chollet、Pranav Ghate、Catherine Neubauer 和 Stefan Scherer。2018。使用虚拟观众进行公开演讲训练时个体差异的影响。在第 18 届国际智能虚拟代理会议论文集。[16] Mathieu Chollet、Torsten Wörtwein、Louis-Philippe Morency 和 Stefan Scherer。2016。用于评估公开演讲能力与焦虑的多模态语料库。在第十届国际语言资源与评估会议论文集（LREC'16）。[17] Mathieu Chollet、Torsten Wörtwein、Louis-Philippe Morency、Ari Shapiro 和 Stefan Scherer。2015。探索改善公开演讲的反馈策略：交互式虚拟观众框架。在 2015 年美国计算机协会普适与泛在计算国际联合会议论文集。[18] Kai S Cortina、Kevin F Miller、Ryan McKenzie 和 Alanna Epstein。2015。低推断与高推断数据的交汇：使用移动眼动追踪与专家教师和新手教师验证 CLASS 评估中的数学教学。国际科学与数学教育杂志 13（2015），389–403。[19] Atakan Coskun 和 Kursat Cagiltay。2021。使用眼动追踪技术研究课堂管理技能。《教育与信息技术》26（2021），2501–2522。[20] David Roxbee Cox。2018。二元数据分析。劳特利奇。[21] Elisabeth Gareis。2006。公开演讲指南。巴鲁克学院/CUNY 传播学系，纽约（2006）。[22] Quentin Guillon、Nouchine Hadjikhani、Sophie Baduel 和 Bernadette Rogé。2014。自闭症谱系障碍中的视觉社会注意：来自眼动追踪研究的见解。《神经科学与生物行为评论》42（2014），279–297。[23] John Hart、Jonathan Gratch 和 Stacy Marsella。2017。虚拟现实训练如何赢得朋友并影响他人。在《国防训练与模拟的基本问题》中。CRC出版社，235–249。[24] Qiong Huang、Ashok Veeraraghavan 和 Ashutosh Sabharwal。2017。Tabletgaze：用于移动平板电脑中无约束基于外观的注视估计的数据集与分析。《机器视觉与应用》28（2017），445–461。[25] Sinh Huynh、Rajesh Krishna Balan 和 JeongGil Ko。2021。iMon：移动设备上的基于外观的视线追踪系统。《美国计算机协会交互、移动、可穿戴与泛在技术学报》（2021）。[26] Moritz Kassner、William Patera 和 Andreas Bulling。2014。Pupil：用于普适眼动追踪和移动端基于视线的交互的开源平台。在 2014 年美国计算机协会普适与泛在计算国际联合会议论文集：附录出版物。[27] Osias Kit T Kilag、Glendon M Quimada、Marlon B Contado、Hanelyn E Macapobre、Jason Isaac III A Rabi 和 Cyd C Peras。2023。肢体语言在公开演讲中的运用。《科学与教育》4，1（2023），393–406。[28] Tiffany C.K. Kwok、Peter Kiefer、Victor R. Schinazi、Benjamin Adams 和 Martin Raubal。2019。"Gaze-Guided Narratives"：使语音导览内容适应虚拟与现实环境中的注视。在 2019 年计算机系统人为因素 CHI 会议论文集（CHI '19）。美国计算机协会，纽约，NY，美国，1–12。[29] Matthew M Laske 和 Florence D DiGennaro Reed。2022。远程视频训练对公开演讲的有效性。《应用行为分析杂志》55，4（2022），1124–1143。[30] Heera Lee 和 Andrea Kleinsmith。2019。真实课堂中的公开演讲焦虑：迈向反思系统的开发。在 2019 年计算机系统人为因素 CHI 会议扩展摘要中。[31] Joanne Leong、Florian Perteneder、Muhender Raj Rajvee 和 Pattie Maes。2023。"想象一下观众……"：探索用于在线公开演讲的私密增强现实面部滤镜。在 2023 年计算机系统人为因素 CHI 会议论文集（CHI '23）。美国计算机协会，纽约，NY，美国，文章 512，13 页。[32] Päivi Majaranta 和 Andreas Bulling。2014。眼动追踪与基于眼睛的人机交互。伦敦施普林格，伦敦，39–65。[33] Alaeddine Mihoub 和 Grégoire Lefebvre。2019。可穿戴设备与社会信号处理：实现更智能的公开演讲。《美国计算机协会交互、智能系统学报》9，2，3，9（2019），24——文章（3 月），页码。[34] Nadiah Nadiah 等。2019。学生在公开演讲中的自信心。《精英期刊》1，1（2019），1–12。[35] Gregory Nash、Gail Crimmins 和 Florin Oprescu。2016。如果一年级学生害怕公开演讲评估，教师可以做些什么来缓解这种焦虑？《高等教育评估与评价》41, 4 (2016), 586–600.[36] Fabrizio Palmas, Ramona Reinelt, Jakub E Cichor, David A Plecher, and Gudrun Klinker. 2021. 虚拟现实公开演讲训练：直接反馈技术接受度的实验评估。载于 *2021 IEEE*虚拟现实与三维用户界面（*VR*）。IEEE。[37] Bastian Pfleging, Drea K Fekety, Albrecht Schmidt, and Andrew L Kun. 2016. 一种描述瞳孔直径与心理工作负荷及照明条件之间关系的模型。载于 美国计算机协会计算机系统中的人为因素年会论文集（*CHI*）。[38] Xiangyao Qi, Qi Lu, Wentao Pan, Yingying Zhao, Rui Zhu, Mingzhi Dong, Yuhu Chang, Qin Lv, Robert P. Dick, Fan Yang, Tun Lu, Ning Gu, and Li Shang. 2023. CASES：认知感知智能眼镜系统，用于理解人们的阅读方式。《*ACM*交互、移动、可穿戴与普适技术汇刊》7, 3, 文章 115（2023年9月），31页。[39] Emiko Shishido, Shiori Ogawa, Seiko Miyata, Maeri Yamamoto, Toshiya Inada, and Norio Ozaki. 2019. 眼动追踪器在理解精神障碍方面的应用：精神分裂症和自闭症谱系障碍案例。*Neuropsychopharmacology reports* 39, 2 (2019), 72–77.[40] Haania Siddiqui, Hafsa Irfan, Aliza Saleem Lakhani, Batool Ahmed, Sadaf Shaikh, Muhammad Mobeen Movania, and Muhammad Farhan. 2023. Manifest：利用虚拟现实进行公开演讲训练。载于 *2023 IEEE*混合与增强现实国际研讨会附属会议（*ISMAR-Adjunct*）。IEEE。[41] Eva Strangert and Joakim Gustafson. 2008. 什么样的演讲者才算优秀？受试者评分、声学测量与感知评估。载于 *Interspeech*，第8卷。[42] Robin E Sülter, Paul E Ketelaar, and Wolf-Gero Lange. 2022. SpeakApp-Kids！通过虚拟现实训练减少儿童对公开演讲的恐惧——概念验证。《计算机与教育》178 (2022), 104384.[43] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. 2017. Inception-v4、Inception-ResNet及残差连接对学习的影响。载于 *AAAI*人工智能会议论文集，第31卷。[44] Marcel Takac, James Collett, Kristopher J Blom, Russell Conduit, Imogen Rehm, and Alexander De Foe. 2019. 多次虚拟现实训练课程可降低公开演讲焦虑。《公共科学图书馆·综合》14, 5 (2019).[45] M Iftekhar Tanveer, Emy Lin, and Mohammed Hoque. 2015. Rhema：帮助人们进行公开演讲的实时现场智能界面。载于 第*20*届智能用户界面国际会议论文集。[46] Niek van den Bogert, Jan van Bruggen, Danny Kostons, and Wim Jochems. 2014. 初步探索教师对课堂事件的视觉感知。《教学与教师教育》37 (2014), 208–216.[47] Rejin Varghese and Sambath M. 2024. YOLOv8：一种性能和鲁棒性有所提升的新型目标检测算法。载于 *2024*年数据工程与智能计算系统进展国际会议（*ADICS*）。[48] Zeyu Wang, Yuanchun Shi, Yuntao Wang, Yuchen Yao, Kun Yan, Yuhan Wang, Lei Ji, Xuhai Xu, and Chun Yu. 2024. G-VOILA：日常场景中的视线辅助信息查询。《*ACM*交互、移动、可穿戴与普适技术汇刊》8, 2, 文章78（2024年5月），33页。[49] Torsten Wörtwein, Mathieu Chollet, Boris Schauerte, Louis-Philippe Morency, Rainer Stiefelhagen, and Stefan Scherer. 2015. 多模态公开演讲表现评估。载于 *2015*年美国计算机协会多模态互动国际会议论文集。[50] Torsten Wörtwein, Louis-Philippe Morency, and Stefan Scherer. 2015. 公开演讲焦虑的自动评估与分析：一项虚拟观众案例研究。载于 *2015*年情感计算与智能交互国际会议（*ACII*）。[51] Yasunori Yamada and Masatomo Kobayashi. 2018. 从观看视频时收集的眼动追踪数据中检测精神疲劳：对年轻人与老年人的评估。《医学人工智能》91 (2018), 39–48.[52] Xucong Zhang, Seonwook Park, Thabo Beeler, Derek Bradley, Siyu Tang, and Otmar Hilliges. 2020. ETH-XGaze：一个用于极端头部姿态和注视变化下注视估计的大规模数据集。载于 欧洲计算机视觉会议论文集（*ECCV*）。[53] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2015. 野外环境中的基于外观的视线估计。载于 *IEEE*计算机视觉与模式识别会议论文集（*CVPR*）。[54] Hangyu Zhou, Yuichiro Fujimoto, Masayuki Kanbara, and Hirokazu Kato. 2021. 将虚拟现实作为公开演讲训练的反思技术。《应用科学》11, 9 (2021), 3988。