AHFE 2023：第 *XX* 卷，
*2023*doi: 10.54941/ahfeXXXX

面向视障用户的移动设备文档创建与编辑全新对话式交互概念

Alireza Darvishy[1], Hans-Peter Hutter[1], Edin Beljulji[1], Zeno Heeb[1]

[1]苏黎世应用科学大学，8400 温特图尔，瑞士

摘要

本文介绍了一种对话式交互概念的持续开发过程，该概念可让视障用户主要通过语音输入，在移动设备上轻松创建和编辑文本文档。为验证这一概念，我们基于自然语言理解（NLU）平台 Google Dialogflow，开发并测试了适用于 iOS 和 Android 系统的原型应用。 视障用户和非视障用户多次测试了该应用程序和交互概念。根据用户反馈，我们不断完善、调整和改进这一概念，使其适用于两个移动平台。在迭代式、以用户为中心的设计过程中，我们探讨了以下研究问题：视障用户能否主要依靠语音命令，高效地在移动设备上创建和编辑文档？用户测试发现，基于对话式语音命令的交互概念对视障用户而言简单直观。然而，研究也发现，仅依靠语音命令会带来一些障碍，而结合手势和语音交互则更加可靠。未来应对视障用户开展更多研究和更广泛的可用性测试，以优化这一交互概念。

关键词： 视觉障碍、移动设备、非视觉交互、自然语言处理、语音输入、语音输出、文档创建

© 2022。由 AHFE 开放获取出版。保留所有权利。

引言与相关研究

据世界卫生组织统计，全球有超过2.5亿人受到视力障碍的影响（世界卫生组织，2012）。许多视觉障碍人士会使用屏幕阅读器等辅助技术朗读数字内容。屏幕阅读器是一种软件应用程序，可帮助盲人或低视力人士访问和使用计算机或智能手机的各项功能。屏幕阅读器会将屏幕上显示的文本和其他信息转换为合成语音输出，并允许用户通过键盘命令或手势与设备进行交互。移动设备上有多种屏幕阅读器，例如 iOS 上的“VoiceOver”和 Android 上的“TalkBack”。

除了语音输出之外，语音技术的最新进展也使用户能够利用语音输入与移动设备交互。例如，用户可以仅通过语音命令控制移动设备，例如在 iOS 设备上使用“语音控制”，或在 Android 设备上使用“语音访问”。一种应用场景是，用户可以通过语音命令编辑移动设备上的文本。与屏幕键盘、基于手势的文本输入（用户用手指或触控笔在屏幕上画出所需的字母或字符）或基于盲文的数字键盘等其他文本输入方式相比，这些工具对许多视觉障碍人士很有帮助（例如 Mattheiss 等人，2015年；Oliveria 等人，2011年）。然而，尽管这类新软件很有用，目前主流的语音输入程序并非专为视障用户设计。因此，仍存在许多无障碍问题。例如，在 iOS 中，删除某一段文本所需的语音控制命令是“删除那段”——这一命令默认用户能够看到当前高亮显示的是哪一段文本。

研究发现，对于视觉障碍用户而言，语音输入是最高效的文本输入方式之一。Azenkot 和 Lee 于 2013 年开展的一项用户研究发现，视觉障碍用户使用语音输入的速度几乎是使用标准触摸键盘的五倍。然而，该研究也指出，尽管输入文本相对容易，语音识别错误仍然是一个主要问题：参与者在撰写文本时，平均有 80% 的时间都花在查找和纠正错误上（Azenkot & Lee，2013）。

文本编辑和错误纠正会给视障用户带来重大障碍，尤其是在处理结构不清晰且无障碍性不足的文档格式时。标题、表格和列表等结构元素不仅更难创建（需要使用简单听写之外的额外命令），而且也是文档无障碍的重要组成部分，是借助屏幕阅读器高效导航所必需的（例如参见 Rajkumar 等人，2020；Darvishy 等人，2011；Munteanu 等人，1995）。

2015年末，谷歌推出了一项名为“语音输入”的功能。这是一种语音转文本听写功能，可与 Chrome 浏览器中的 Google 文档配合使用（Moynihan，2016）。这项技术有望为视障用户带来极大便利，因为与许多其他听写程序和应用程序不同，它支持插入结构元素（标题、表格、列表等）、跳转到文档中的指定词语或句子，以及通过语音命令调整格式。不过，目前只能在桌面版 Chrome 中使用语音输入创建 Google 文档。此外，这项功能的设计并未考虑视障用户的需求——例如，用户必须点击麦克风图标才能开始听写。

概念开发

这一创新交互概念的主要目标，是让用户能够通过语音输入，以对话方式轻松创建文档，包括标题、副标题、缩进、项目符号列表、编号列表等常用格式设置功能。该交互概念还支持用户以对话方式编辑文档，例如直接指明文本内容来添加、删除或修改文本。因此，我们也实现了语音输出，帮助用户在文档中导航并确定自身位置，而不必依赖光标或其他视觉线索。

为了以用户为中心开发相应的对话式交互模型，我们首先开展了用户研究并分析用户需求。随后，根据所识别的用户需求，完成了首轮设计与原型迭代，并使用该原型开展了大量用户测试。

为了明确用户需求，我们首先对视障潜在用户进行了非正式访谈。这些讨论表明，移动文档创建所需的主要任务，例如文本听写、编辑和基本结构化，对视障用户而言尤其繁琐且耗时。根据这些反馈，我们通过定义情境场景并为名为“Spectra”的原型应用编写示例对话，制定了初步概念。

情境场景

我们定义了一系列情境场景，用于展示原型应用中设想的交互方式。在每个情境场景中，都会描述虚构的用户角色使用应用程序执行任务。例如，以下是其中一个名为“校对文档”的情境场景：

Angelica 在一所大学担任技术助理。现在是周一下午，她受托在下周提交前校对几份文档，确保其技术准确性。由于 Angelica 患有视神经疾病，无法通过视觉来完成这项工作。于是，她在智能手机上启动 Spectra 应用，打开今天要处理的第一份文档。她可以坐在椅子上放松休息，同时让应用程序为她简要概述这份文档中的标题。Angelica 只关心技术上重要的章节，因此她指示应用程序跳转

到特定章节并朗读出来。她使用停止命令打断朗读，说：“请重复上一句。”Spectra 再次为她朗读上一句。由于句子内容有误，她使用语音命令在文档中插入评论：“插入评论：这句话应该重新表述。”Spectra 回复：“已插入评论：这句话应该重新表述。”Angelica 回答“继续”后，Spectra 接着往下读。

考虑了以下情境场景：

- 创建购物清单 • 撰写信件
- 撰写报告 • 审阅和编辑文档

示例对话

我们创建了若干示例对话，以构想用户与系统之间的具体交互，展示用户的语音输入和系统的语音输出。默认情况下，应用程序会确认并复述每条命令或听写的句子，以便用户发现并纠正任何错误或误解。表1展示了其中一段示例对话。

表1。 用户撰写和编辑一份介绍新产品（防盗系统）的文档的示例对话片段

Turn	话语
user	标题 «防盗系统»
系统	文档标题“防盗系统”
user	一级标题 «引言»
系统	标题 1 «说明»
user	将 «说明» 替换为 «说明»
系统	标题 1 «引言»
user	听写模式
系统	听写模式已启动
user	这套新系统应能实现防入室盗窃， 无论住户是否在家。
系统	这套新系统应能实现防入室盗窃，无论是在 居民不在场和在场的情况。
user	在“system”前插入“control”
system	这个新的控制系统应能防止入室盗窃， 无论居民不在场还是在场。

交互概念

该交互概念设想了两种模式：听写模式和命令模式。在听写模式下，用户通过语音口述文档的正文文本。因此，该模式主要提供语音转文本（STT）功能，将语音输入识别为文本。文本按话语逐句输入。听写模式还支持在口述文本中输入标点信息，例如逗号、分号等（见表1）。系统会自动回读识别出的话语作为确认，其中包括格式和标点信息

（句末的普通句号和问号除外，这些标点通过韵律手段表示）

朗读识别的话语后，系统会自动进入命令模式，用户可以立即替换、移动、插入或删除词语，编辑刚刚输入的文本。这样，用户可以直接引用之前输入的词语，例如：“在表1中，将‘控制’插入到‘系统’之前。”如果用户继续听写，系统会自动切换回听写模式。用户也可以使用相应命令，在听写模式和命令模式之间主动切换。

通过命令模式，用户可以检查、编辑、设置格式并组织听写的文本（如标题、段落、列表等）。输入错误或命令错误时，可以使用相应命令轻松撤销。用户还可以使用各种跳转命令，在文档中快速导航。为了更好地了解文档内容，用户还可以输入各种朗读命令，要求朗读特定词语、句子、段落或整篇文档。涉及相对位置的命令，例如“删除上一个词”，会根据当前焦点（虚拟光标）进行解释。焦点位于最后朗读的词语末尾，或最后一条编辑命令所指的位置。

除了纯语音交互外，该应用程序还支持屏幕阅读器，用户可通过点击手势让屏幕阅读器读出触摸屏上的元素。此外，应用程序还提供导出功能，可将口述的文档分享给他人。

原型实现

为了高效实现该交互概念的原型，我们采用了分布式语音识别方案，利用智能手机（iOS 或 Android）上的本地语音转文本组件。随后，识别出的文本会发送至远程的、基于意图的 Dialogflow 自然语言理解服务（Sabharwal & Agrawal，2020），以识别和解释命令；在此之前，系统会为基于意图的命令解释设置正确的上下文。最后，识别出的命令会发回移动设备，由应用程序执行；或者将识别出的句子读回给用户。紧接着，系统便开始下一轮识别循环。

用户测试

为测试初始原型的可用性，我们组建了一支测试用户团队，其中包括视障用户和非视障用户。我们请每位测试用户下载 Spectra 应用，随后向他们发送一段样本文本。测试任务是仅使用该应用（如有需要，也可使用设备内置的屏幕阅读器）重新创建这段样本文本。之后，测试用户将生成的文本文档提交给作者。所有测试用户还提交了非结构化的书面反馈，描述他们使用该应用的整体体验，并指出遇到的障碍。当前版本的 Spectra 应用根据这些反馈进行了修改和改进。

结果与讨论

这项持续进行的工作目前得出了两个重要发现：

> 1) 如下所述，在许多情况下，使用基于内容的命令对视障用户而言效率最高，而且通常比基于光标的命令更实用。2) 仅依赖语音命令和语音输出功能并不足以满足视障用户的需求。理想的方式是结合手势与语音交互。

在规划阶段和第一个原型实现中，该应用程序主要采用“基于光标”的命令，例如“选择单词”或“删除句子”，这些命令会根据当前光标位置进行解析。在这种情况下，如果要删除某个单词，就必须像使用普通文本编辑器那样，先将光标移至文本中的正确位置。然而，根据用户反馈，我们了解到，这类命令并不适用于非视觉文本编辑。为了找到想要的单词或句子，用户要么必须听读大段内容并在合适的位置停下来，要么必须使用导航命令四处移动。基于光标的命令的替代方案是更“基于内容的命令”，这类命令更接近人与人之间的沟通方式。例如，“将 x 替换为 y”这条命令，让人更容易联想到“抱歉，我想说的是 x，不是 y”这样的对话短语。这些命令无需指定位置，使用起来也比先导航找到“x”，再发出“删除单词”和“听写模式：y”等命令更简便。因此，在最新的原型中，我们实现了基于内容的命令，用户可直接在命令中指定旧内容和新内容，无需先执行任何导航命令。不过，基于内容的命令并非总是更容易使用，也可能被系统误解，因此原型中仍保留了标准的基于光标的命令，让用户可以选择使用任一类型。

当前版本的应用程序包含用于插入文本或结构化内容的命令，例如段落、列表或标题，也包含用于撤销或修改错误输入的命令。此外，还提供了朗读文档部分内容或整篇文档的命令。用户可以通过轻点一次来中断朗读，以便在所需位置进行修改。还可以使用“段落开头”“段落结尾”等导航命令来更改文档中的当前位置。

用户测试还发现，屏幕阅读器支持，尤其是 iOS VoiceOver 支持，是一项理想功能。这项支持意味着，只要 Spectra 和屏幕阅读器同时运行，用户也可以使用设备内置的屏幕阅读器功能来修改文本。尽管 iOS 和 Android 都内置了屏幕阅读器，但 iOS 的整体无障碍支持显著更完善，因此我们的目标群体越来越多地使用 iOS 操作系统；因此，当前版本的 Spectra 将重点支持 VoiceOver。

将自然语言理解交由 Dialogflow 处理带来了多项优势，因为分析过程独立于原生移动系统进行，且无需本地代理。因此，Dialogflow 代理可用的资源远多于单台移动设备所能提供的资源。不过，这种做法的缺点是，数据需要先发送至 Dialogflow 并由其评估，之后才能收到结果，这一过程会产生延迟。虽然在 Android 上这一过程并不会明显变慢，但在 iOS 操作系统上耗时更长。因此，与使用本地代理相比，使用速度会更慢。

一个重要的研究问题仍未得到解答：语音命令、手势和屏幕阅读器输出的最佳组合是什么，才能让视障用户在移动设备上高效地创建和编辑文本文档?

这一问题将在 Spectra 的后续开发中得到探讨，并通过更广泛的用户测试加以研究。

参考文献

1. Azenkot, S. 和 Lee, N.（2013）。探索盲人在移动设备上使用语音输入的情况。第15届 ACM SIGACCESS 计算机与无障碍国际会议论文集，ASSETS 2013。2. Azenkot, S. Wobbrock, J.O. Prasain, S. Ladner, R.E.（2012）。Perkinput 中用于非视觉触摸屏文本输入的手指输入检测。2012 年图形界面会议论文集，第 121-129 页。3. Darvishy, A., Hutter, HP., Mannhart, O.（2011）。用于分析、处理和生成无障碍 PDF 文档的 Web 应用程序。载于：Stephanidis, C.（编）《人机交互中的普适访问：应用与服务》。UAHCI 2011。《计算机科学讲义》，第 6768 卷。施普林格，柏林，海德堡。4. Mattheiss, E. Regal, G. Schra mmel, J. Garschall, M. Tscheligi, M.（2015）。EdgeBraille：面向触控设备的基于盲文的文本输入。《辅助技术期刊》。5. Darvishy, A., Leemann, T., Hutter, HP.（2012）。基于灵活软件架构创建完全无障碍 PDF 文档的两个软件插件。载于：Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W.（编）《计算机帮助有特殊需求的人》。ICCHP 2012。《计算机科学讲义》，第 7382 卷。施普林格，柏林，海德堡。6. Darvishy, A., Hutter, HP.（2013）。基于重要无障碍标准对不同无障碍插件的有效性进行比较。载于：Stephanidis, C., Antona, M.（编）《人机交互中的普适访问：生活质量应用与服务》。UAHCI 2013。《计算机科学讲义》，第 8011 卷。施普林格，柏林，海德堡。7. Moynihan, Tim.（2016）“现在可以通过语音编辑 Google 文档了。”《连线》，2月29日。https://www.wired.com/2016/02/now-can-type-google-docs-speaking/ 8. Munteanu, E., Guggiana, V., Darvishy, A., Schauer, H., Rauterberg, G. W. M., & Motavalli, M.（1995）。环境声音的物理建模。载于 F. Pedrielli（编），第2届国际声学与音乐研究会议论文集，CIARM '95（第 107-112 页）。费拉拉大学。9. Oliveira, J. Guerreiro, T. Nicolau, H. Jorge, J. Gonçalves, D.（2011）。BrailleType：在触摸屏手机上释放盲文的潜力。IFIP 人机交互会议，施普林格，柏林，海德堡，第 100-107 页。

10. Rajkumar, A., J. Lazar, J. B. Jordan, A. Darvishy, 和 H. Hutter.（2020）。研究论文的 PDF 无障碍性：评估与修复需要哪些工具？载于 HICSS。11. Sabharwal, N., Agrawal, A.（2020）。Google Dialogflow 简介。载于：《使用 Google Dialogflow 的认知型虚拟助手》。Apress，伯克利，加利福尼亚州。12. 世界卫生组织（WHO）。（2012）。2010年全球视力障碍数据。日内瓦，世界卫生组织。