



**UNIVERSIDADE FEDERAL DO PARÁ**  
**INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**TIAGO DAVI OLIVEIRA DE ARAÚJO**

**Interação por Voz em Ambientes de Realidade Aumentada Móvel:  
Avaliação e Sugestões de Boas Práticas**

Belém  
2016

Tiago Davi Oliveira de Araújo

**Interação por Voz em Ambientes de Realidade Aumentada Móvel:  
Avaliação e Sugestões de Boas Práticas**

Dissertação de Mestrado apresentada como requisito para obtenção do grau de Mestre em Ciência da Computação. Programa de Pós-Graduação em Ciência da Computação. Instituto de Ciências Exatas e Naturais. Universidade Federal do Pará.

Orientador Prof. Dr. Bianchi Serique Meiguins

Belém

2016

Tiago Davi Oliveira de Araújo

**Interação por Voz em Ambientes de Realidade Aumentada Móvel:  
Avaliação e Sugestões de Boas Práticas**

Dissertação de Mestrado apresentada como requisito para obtenção do grau de Mestre em Ciência da Computação. Programa de Pós-Graduação em Ciência da Computação. Instituto de Ciências Exatas e Naturais. Universidade Federal do Pará.

Orientador Prof. Dr. Bianchi Serique Meiguins

Belém

2016

Dados Internacionais de Catalogação-na-Publicação (CIP)  
Sistema de Bibliotecas da UFPA

---

Araújo, Tiago Davi Oliveira de, 1992-  
Interação por voz em ambientes de realidade  
aumentada móvel: avaliação e sugestões de boas práticas  
/ Tiago Davi Oliveira de Araújo. - 2016.

Orientador: Bianchi Serique Meiguins.  
Dissertação (Mestrado) - Universidade  
Federal do Pará, Instituto de Ciências Exatas e  
Naturais, Programa de Pós-Graduação em Ciência  
da Computação, Belém, 2016.

1. Interface de usuários (Sistemas de  
computador). 2. Realidade aumentada móvel  
(RAM)-Interação por voz. 3. Android (Recurso  
eletrônico). 4. Interação por voz. 5.  
Dispositivos móveis. I. Título.

CDD 22. ed. 005.269

---

Tiago Davi Oliveira de Araújo

**Interação por Voz em Ambientes de Realidade Aumentada Móvel:  
Avaliação e Sugestões de Boas Práticas**

Dissertação de Mestrado apresentada como requisito para obtenção do grau de Mestre em Ciência da Computação. Programa de Pós-Graduação em Ciência da Computação. Instituto de Ciências Exatas e Naturais. Universidade Federal do Pará.

Data da defesa: Belém, 26 de agosto de 2016

Conceito: \_\_\_\_\_

---

**Prof. Dr. Bianchi Serique Meiguins** – Orientador  
Programa de Pós-Graduação em Ciência da Computação – UFPA

---

**Prof. Dr. Nelson Cruz Sampaio Neto** – Membro Interno  
Programa de Pós-Graduação em Ciência da Computação – UFPA

---

**Prof.<sup>a</sup> Dr.<sup>a</sup> Marcelle Pereira Mota** – Membro Externo  
Faculdade de Computação – UFPA

*Aos meus pais, irmão, e amigos que sempre acreditaram que era possível.*

## **AGRADECIMENTOS**

Agradeço a Deus, pelo presente da vida e as oportunidades que me permitiu vivenciar.

Agradeço a minha família, por entender os vários momentos dedicados para conclusão desse trabalho e me apoiar por causa disso.

Ao meu orientador, Prof. Dr. Bianchi Serique Meiguins, que sempre me instigou a ir além, quebrar limites e sair da minha zona de conforto.

Ao LABVIS, especialmente aos que fui mais próximo nessa jornada. Msc. Gustavo Resque por muitas vezes me orientar em um caminho já percorrido por ele. Nikolas Carneiro, por vários momentos de apoio e conselhos. Brunelli Miranda, pela amizade de tempos antes desse trabalho começar e pelos momentos divertidos que várias vezes ele proporcionou. Ao Paulo Júnior, que foi um bom amigo, compartilhando boas lutas e seguindo o caminho do guerreiro.

A Priscila Ferreira, que foi como uma irmã durante esse período, sendo o suporte em vários passos difíceis dessa caminhada, e celebrando nos bons momentos.

A Janyne Palheta, uma amiga que sempre acreditou, e não cansou de expressar isso em nenhum momento.

Ao Arthur Masahiro e ao Alexandre Freitas, amigos que compartilharam bons momentos esses anos, suas companhias foram um alento durante o prosseguimento desse trabalho.

Ao PPGCC, na pessoa do Prof. Dr. Jefferson Morais, atual coordenador, que a boa gestão impediu empecilhos e dificuldades a realização deste trabalho.

Ao Prof. Dr. Nelson Neto e a Prof.<sup>a</sup> Marcelle Mota, por terem aceitado avaliar esse trabalho, participando assim desse momento especial.

## PUBLICAÇÕES

### Artigos Completos:

ARAÚJO, T. D. O.; SANTOS, C. G. R. ; MIRANDA, B. P. ; CARNEIRO, N. J. S. ; SOARES, A. G. M. ; MOTA, M. P. ; NETO, N. C. S. ; MEIGUINS, B. S. . Aspects of Voice Interaction on a Mobile Augmented Reality Application. In: HCI International 2016, 2016, Toronto. Proceedings of HCI International 2016, 2016.

ARAÚJO, T. D. O.; MIRANDA, B. P. ; SANTOS, C. G. R. ; CARNEIRO, N. J. S. ; MEIGUINS, B. S. . Aplicações Android de Realidade Aumentada em Arquitetura Extensível, Flexível e Adaptável. In: XI Simpósio Brasileiro de Sistemas de Informação, 2015, Goiânia. Anais do XI Simpósio Brasileiro de Sistemas de Informação. Goiânia, 2015. p. 63-70.

ARAÚJO, T. D. O.; MIRANDA, B. P. ; SANTOS, C. G. R. ; CARNEIRO, N. J. S. ; MOTA, M. P. ; NETO, N. C. S. ; MEIGUINS, B. S. . Usability Considerations for Voice Interaction in a Mobile Augmented Reality Application. In: WRVA 2015, 2015, Presidente Prudente. XVI Workshop de Realidade Virtual e Aumentada, 2015. p. 36-41.

### Resumo Extendido:

ARAÚJO, T. D. O.; MIRANDA, B. P. ; SANTOS, C. G. R. ; CARNEIRO, N. J. S. ; CARVALHO, M. B. ; SOUSA, G. I. S. ; MEIGUINS, B. S. ; NETO, N. C. S. . Implementation Aspects of a Voice Service in Mobile Augmented Reality Applications. In: XIV Workshop de Realidade Virtual e Aumentada, 2015, Presidente Prudente. Anais do XVI Workshop de Realidade Virtual e Aumentada, 2015. p. 191-195.



*“Esse é o caminho do meu destino” (Ryu, da série Street Fighter)*

## RESUMO

O contexto tecnológico atual permite que haja interações por comandos de voz em dispositivos móveis, o que pode facilitar a interação do usuário, principalmente se uma de suas mãos não estiver livre, já que retira a necessidade de tocar na tela; contudo, o uso dessa interação ainda apresenta desafios, tais como: tratamento de ruídos no ambiente, interfaces mais intuitivas, reconhecimento com maior precisão, síntese de voz mais próxima da fala natural, entre outros desafios. A Realidade Aumentada em dispositivos móveis (RAM – Realidade Aumentada Móvel) tem proporcionado aos usuários uma nova experiência de interação, já que inserem objetos virtuais em ambientes reais de interesse do usuário, tornando-os mais dinâmicos, tais como: pontos turísticos de cidade, navegação em museus, manutenção de equipamentos, não limitando a tecnologia a um computador de mesa. Alguns desafios para uma boa integração do uso da voz em aplicações de RAM são a necessidade de padrões de desenvolvimento, orientações na interface das aplicações para escolha de palavras pelo usuário, projeto padrão para interface das aplicações de RAM, considerando ícones, elementos gráficos da interface, etc. Apresentados alguns desafios atuais da área de interação por voz e aplicações RAM, e principalmente sua integração; este trabalho propõe boas práticas de desenvolvimento, projeto de interface gráfica e de interface de voz para inclusão de um serviço de reconhecimento de voz considerando o contexto aplicação RAM, e para tal um serviço de voz ser integrado na aplicação, e serão realizadas avaliações de usabilidade e comunicabilidade.

**Palavras-chave:** Realidade Aumentada Móvel, Interação por Voz, Android.

## **ABSTRACT**

The current technological context allows interaction by speech commands in mobile devices, easing user interaction, mainly if one hand is not free, as withdraws the need to touch in the screen; however, it still has challenges, such as: environment noise removal, intuitive interfaces, more accurate recognition, speech synthesis nearest natural voice, among others. The Augmented Reality in mobile devices (MAR – Mobile Augmented Reality) has provided to the users a new interaction experience, as the virtual objects insertion in real environments of user interest, making them more dynamic, such as: touristic spots, museum navigation, not limiting to a desktop computer. Some challenges for good integration of speech in MAR applications are development patterns, orientation in applications interface for words selection by the user, a default project for MAR applications interface, covering icons, graphical elements of interface, etc. Some challenges of speech interaction and MAR applications were presented, focusing in their integration; this work proposes good practices of software development, graphical user interface and voice interface to include a speech recognition service considering MAR applications context, performs usability and communicability evaluations.

**Keywords:** Mobile Augmented Reality, Speech Interaction, Android.

## LISTA DE ILUSTRAÇÕES

Figura 1. Fluxo de Atividades .....	19
Figura 2. Informações sobre artigos encontrados.....	27
Figura 3. Visão do mapa no ARGuideVoz .....	31
Figura 4. Visão do navegador RA no ARGuideVoz.....	32
Figura 5. Visão de QRCode no ARGuideVoz (a) Marcador de um POI (b) Leitura do marcador (c) reconhecimento do marcador (d) Exposição do conteúdo ligado ao marcador .....	33
Figura 6. Visão de Lista no ARGuideVoz .....	33
Figura 7. Visão de POI no ARGuideVoz.....	34
Figura 8. Funcionamento padrão de um modelo de reconhecimento de voz.....	36
Figura 9. Ilustração do fluxo de troca de mensagens na implementação do serviço .....	37
Figura 10. Aspectos gerais de implementação, em cinza os aspectos ligados a aplicação, em branco os aspectos relacionados a tecnologia. ....	39
Figura 11. Arquitetura genérica de uma aplicação de RA com um Serviço de Voz.....	42
Figura 12. Diagrama de Pacotes de um Serviço de Voz para aplicações RAM .....	42
Figura 13. Mudanças realizadas na arquitetura para acomodar o serviço de voz .....	44
Figura 14. Exemplo de alterações realizadas após a primeira reunião.....	47
Figura 15. Alterações realizadas após segunda reunião .....	48
Figura 16. Alterações realizadas após gerar definições gerais.....	50
Figura 17. Participante durante o cenário de testes .....	57
Figura 18. Experiência dos usuários com RA e Voz.....	58
Figura 19. Resultados dos testes com Usuários .....	59
Figura 20. Interface do Navegador RA - (a) Controlador de raio (b) Radar com POIs .....	62
Figura 21. Resultados do MAC baseado em falhas encontradas .....	63
Figura 22. Visão durante a atividade “Traçar rota a um determinado POI” .....	65

Figura 23. Visão durante a atividade “Visualizar a lista de Pontos de Interesse”.....	66
Figura 24. Visão do Navegador de Realidade Aumentada .....	67

## LISTA DE TABELAS

Tabela 1. Artigos encontrados e descrições .....	29
Tabela 2. Comandos disponíveis na aplicação.....	45
Tabela 3. Considerações realizadas para interface e voz .....	48
Tabela 4. Lista de tarefas para os testes, divididos em Unidades de Tarefa e Complexidade.....	53
Tabela 5. Tarefas propostas para avaliação.....	56
Tabela 6. Problemas encontrados e Melhorias sugeridas.....	61
Tabela 7. Etiquetagem total do MAC.....	64
Tabela 8. Conjunto de boas práticas em Interface Gráfica .....	67
Tabela 9. Conjunto de boas práticas em Interface de Voz .....	68

## LISTA DE SIGLAS

ADK	<i>Android Development Kit</i>
API	<i>Application Programming Interface</i>
ARGuide	<i>Augmented Reality guide</i>
GPS	<i>Global Positioning System</i>
GUI	<i>Graphical User Interface</i>
IHC	<i>Interação Humano Computador</i>
MAC	<i>Método de Avaliação da Comunicabilidade</i>
POI	<i>Points of Interest</i>
QR Code	<i>Quick Response Code</i>
RA	Realidade Aumentada
RAM	Realidade Aumentada Móvel
SDK	<i>Software Development Kit</i>
UI	<i>User Interface</i>
UML	<i>Unified Modeling Language</i>

# SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>16</b>
<b>1.1 Justificativa .....</b>	<b>17</b>
<b>1.2 Objetivos.....</b>	<b>18</b>
<b>1.3 Metodologia.....</b>	<b>19</b>
<b>1.4 Organização do trabalho .....</b>	<b>20</b>
<b>2. REFERENCIAL BIBLIOGRÁFICO .....</b>	<b>22</b>
<b>2.1 Realidade Aumentada Móvel .....</b>	<b>23</b>
<b>2.2 Interação por Voz .....</b>	<b>25</b>
<b>2.3 Trabalhos Relacionados.....</b>	<b>27</b>
<b>3. PROTÓTIPO DA APLICAÇÃO ARGUIDEVOZ .....</b>	<b>30</b>
<b>3.1 Mapa .....</b>	<b>30</b>
<b>3.2 Navegador de Realidade Aumentada .....</b>	<b>31</b>
<b>3.3 Leitor de QR Code.....</b>	<b>32</b>
<b>3.4 Lista .....</b>	<b>33</b>
<b>3.5 Visão de POI .....</b>	<b>33</b>
<b>4. PROPOSTA DE INTERAÇÃO POR VOZ E ASPECTOS DE PROJETO .....</b>	<b>35</b>
<b>4.1 Interação por Voz em dispositivos Móveis.....</b>	<b>35</b>
<b>4.2 Aspectos gerais de implementação.....</b>	<b>38</b>
<b>4.3 Arquitetura de um serviço de Voz em RAM.....</b>	<b>41</b>



4.3.1	Diagrama de Classes.....	42
<b>4.4</b>	<b>Serviço de Voz no ArguideVoz.....</b>	<b>44</b>
4.4.1	Definição de interface.....	46
4.4.2	Versão do protótipo para Avaliação.....	48
<b>5.</b>	<b>AVALIAÇÃO DE INTERAÇÃO DO ARGUIDEVOZ.....</b>	<b>51</b>
<b>5.1</b>	<b>Avaliação de Usabilidade.....</b>	<b>51</b>
5.1.1	Preparação.....	52
<b>5.2</b>	<b>Método de Avaliação de Comunicabilidade.....</b>	<b>54</b>
<b>5.3</b>	<b>Resultados.....</b>	<b>57</b>
5.3.1	Avaliação de Usabilidade.....	57
5.3.2	Método de Avaliação de Comunicabilidade.....	62
<b>5.4</b>	<b>Boas Práticas.....</b>	<b>67</b>
<b>6.</b>	<b>CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.....</b>	<b>69</b>
<b>6.1</b>	<b>Considerações Finais.....</b>	<b>69</b>
<b>6.2</b>	<b>Trabalhos Futuros.....</b>	<b>70</b>
	<b>REFERÊNCIAS.....</b>	<b>72</b>

## 1. INTRODUÇÃO

A popularização e evolução dos dispositivos móveis assim como o aumento da velocidade de internet móvel permitiram o surgimento de novas aplicações que antes não podiam estar presentes nos dispositivos móveis em função de limitações tecnológicas. O aumento de 21% para 35% da taxa de posse de *smartphones* em populações de países subdesenvolvidos (POUSHTER, 2016) mostra que o acesso a essa tecnologia está aumentando. Na pesquisa de Park. Associates (2016) é feita uma comparação de serviços de *stream* que os usuários dos EUA utilizam, e vídeos e voz são os serviços em que esses usuários utilizam mais de 40% de largura de banda. O uso de voz está se tornando mais comum e frequente entre os usuários de dispositivos móveis.

A voz é um dos meios de comunicação mais utilizados pelos seres humanos, estando entre as três formas de comunicação (voz, gestos e expressões faciais) mais utilizadas no dia-a-dia de uma pessoa (TEIXEIRA, 2013). A facilidade de interagir por voz pode proporcionar ao usuário uma nova experiência de uso em uma aplicação. A popularização do uso da voz também abre novos desafios para aplicações móveis, como padrões de desenvolvimento, em escolha de comandos, na disposição de elementos da Interface Gráfica (*Graphical User Interface – GUI*), entre outros. Integrar esse recurso e oferecer como uma alternativa de interação para uma aplicação ainda é um campo inicial em várias linhas de pesquisa. Uma área que vem se expandindo é a Realidade Aumentada Móvel (RAM), ou seja, aplicações de Realidade Aumentada para dispositivo móveis, com aplicações em vários setores da sociedade, tais como turismo, educação, divulgação de informação, entretenimento, entre outros, e que pode se beneficiar com o uso de interação por comandos de voz.

Com a popularização de dispositivos móveis, com preços mais acessíveis, uma gama de sensores, e o fator inovador e lúdico da RAM, é possível usar essas aplicações em novos cenários e com novas interações. Do mesmo modo que em interações por comandos de voz, existem desafi-

os a serem superados em aplicações de RAM, como estudos de usabilidade das ferramentas RAM (MARTÍNEZ, SKOURNETOU, *et al.*, 2014), a escassez de padrões de desenvolvimento e o pouco espaço para a apresentação de informações. Uma alternativa para apresentação e organização de conteúdo é o uso de interação por comandos de voz em aplicações de RAM.

Neste trabalho, uma aplicação RAM, denominada ARGUIDE (ARAUJO, CARNEIRO, *et al.*, 2015), foi adaptada para receber um serviço de reconhecimento por voz para ajudar na navegação e visualização de conteúdo aumentado. Essa nova versão da aplicação será denominada ARGuideVoz. Algumas orientações para implementação desse tipo de serviço são sugeridas, e com eles o projeto de desenvolvimento desse serviço. Para avaliação da interação por voz em na aplicação RAM foram executadas duas avaliações, uma avaliação de usabilidade foi realizada para validar a necessidade desse serviço, e uma avaliação de comunicabilidade foi feita para traçar o perfil semiótico da aplicação, ou seja, a percepção do usuário sobre as funções da aplicação.

## **1.1 Justificativa**

O uso de linguagem natural para aplicações de RAM não é difundido, logo é uma forma de interação que tem escassez de estudos e avaliações. A interação por voz apresenta uso prático em diversos cenários, porém ainda há poucos estudos que a utilizam juntamente com a RAM, portanto, apresenta necessidade de mais pesquisas nessa área.

A possibilidade de usar a aplicação segurando o dispositivo com uma mão abre possibilidades para vários tipos de perfis de usuários, como por exemplo, usuários com dificuldades motoras para segurar o dispositivo, que só podem utilizar a navegação por voz.

Além disso, também não se tem um padrão de arquitetura definido para integração de interações por voz. Mesmo para o uso somente de RAM existem poucos trabalhos relacionados mostrando a arquitetura desses sistemas (MARTÍNEZ, SKOURNETOU, *et al.*, 2014). O uso de voz possui alguns padrões de interação integrados com interfaces de toque em dispositivos móveis, como mostram Vidakis, Syntychakis (2012) e Xia, Kai (2010), mas com pouca aplicação a RAM, uma vez que sua interface difere das aplicações tradicionais, no sentido de que a área principal da aplicação é um vídeo (ou conjunto de imagens) capturado pela câmera do dispositivo, e o conteúdo de interação é dinâmico e misturado com as imagens do ambiente real. O estudo

do uso de interação por voz em um ambiente de RAM é um esforço para descobrir se e como essa interação pode ser adaptada para facilitar a execução de tarefas para o usuário.

A voz como interação foi escolhida, por ser uma das três formas mais comuns de comunicação no dia a dia de uma pessoa (TEIXEIRA, 2013), e o uso de comandos de voz em smartphones tem se intensificado com aplicações como o Siri no iOS (APPLE INC, 2016), Google Now no Android (GOOGLE INC, 2016) e Cortana no Windows Phone e Windows 10 (MICROSOFT INC, 2016). Isso mostra que os usuários estão se acostumando com interação por voz e o uso de RAM tem crescido durante os anos.

## 1.2 Objetivos

O objetivo principal deste trabalho é avaliar a interação por voz em aplicações de Realidade Aumentada Móvel (RAM), sugerindo boas práticas para integração desse tipo interação e projeto de interface para aplicação RAM.

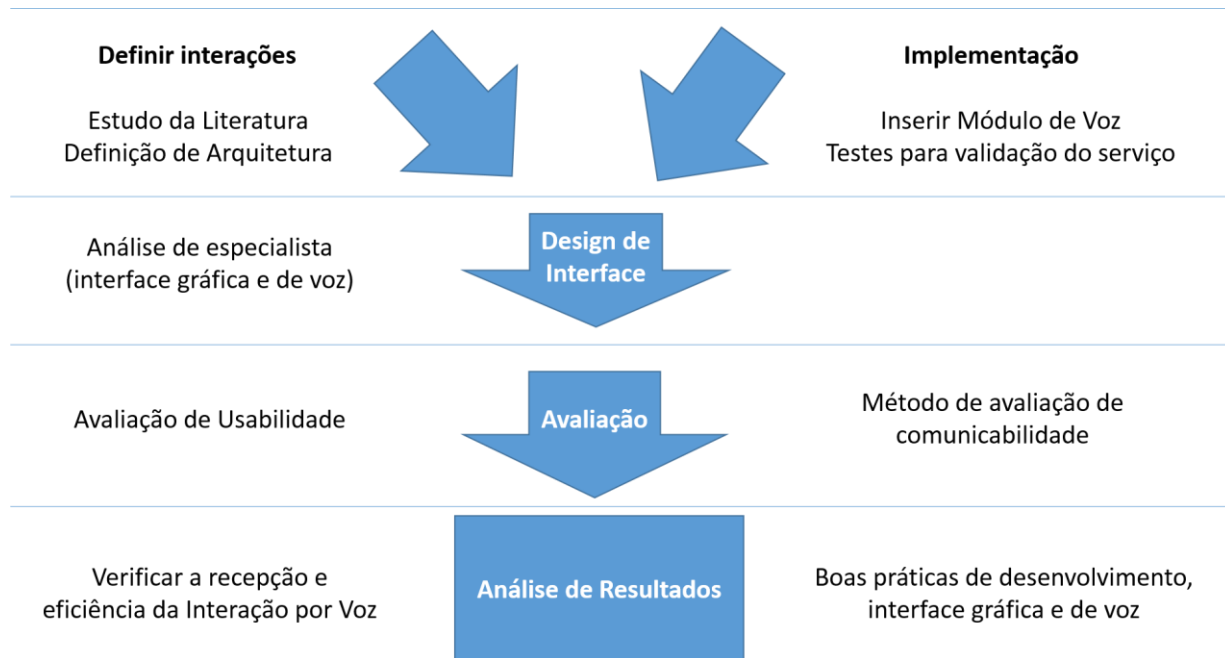
Os objetivos específicos para realização desse trabalho são os seguintes:

- **Integração de um serviço de voz em Android:** o processamento do reconhecimento dos comandos de voz será realizado por um serviço em nuvem, sendo necessário enviar o áudio captado para processamento em um servidor. Além disso, compreender as mensagens trocadas entre a aplicação, o serviço de voz e o servidor de processamento são importantes para definir as interações da aplicação.
- **Escolha de comandos de voz em Android:** A escolha do dicionário correto, com palavras que possam ser utilizadas facilmente e tenham significado para aplicação é importante para que as interações estejam implementadas da melhor maneira possível.
- **Realização da avaliação de usabilidade no ARGuideVoz:** A primeira avaliação será feita com testes em laboratório e questionários, esperando que se encontrem pontos positivos e negativos baseados na usabilidade da aplicação.
- **Realização de Avaliação de Comunicabilidade no ARGuideVoz:** A segunda avaliação será feita com testes de usuário para descobrir rupturas na comunicação entre designer-usuário.
- **Análise de Resultados:** A análise dos resultados avaliará a eficiência da interação por voz na navegação em RAM, e os pontos que precisam ser melhorados, apresentando

como resultado sugestões de boas práticas para integração da interação por voz e projeto de interface para aplicações RAM.

### 1.3 Metodologia

O desenvolvimento do serviço de voz foi realizado utilizando o ambiente Android Studio (GOOGLE INC., 2015), com o ADK (*Android Development Kit*). O serviço foi criado separadamente primeiro, para depois ser integrado na aplicação. A Figura 1 mostra o fluxo de atividade resumido para alcançar esse objetivo.



**Figura 1. Fluxo de Atividades**

No primeiro passo, de integração do serviço, será realizado a identificação de erros de lógica e integração de módulos no serviço de voz e toques, com implementação de tarefas de interação. O segundo passo diz respeito ao design de interface, com reuniões com especialista para definir as interações. O terceiro inclui testes de interações com usuários, e possíveis comparações com usuários que utilizaram somente a interação por voz, e ambas interações.

A partir da definição dos objetivos, a metodologia proposta para o desenvolvimento do projeto, como um todo, inclui as seguintes atividades:

- Interação com professores e estudantes pesquisadores de temas relacionados, inclusive de outras instituições e de outros países;

- Pesquisa individual, com acesso a bibliotecas e Internet;
- Participação em seminários, congressos e workshops;

Como atividades específicas, destacam-se:

- Levantamento bibliográfico das principais taxonomias de interação por voz, inclusive voltadas a área de realidade aumentada móvel.
- Definição do conjunto de interações a serem mapeadas em voz, e o dicionário para executar essas tarefas. E definir quais interações por voz selecionadas serão melhores utilizadas.
- Concepção e desenvolvimento de uma arquitetura.
- A interação por voz será realizada por um serviço de Voz do Google, utilizando a Google Speech API.
- Reuniões com especialistas, para confirmar interações e mudanças na interface. Serão geradas as diretrizes para interface e interação. As diretrizes serão implementadas para aplicação de avaliações com usuários.
- De maneira geral, para o desenvolvimento do serviço de voz serão utilizados padrões de projeto, modelagem baseada na notação UML, e um processo iterativo e incremental, contemplando as fases de concepção, elaboração, construção e implantação. Para o teste de usuário, serão utilizados um método de teste, em que serão gravados em vídeo para análise e encontrar as informações para definir as interações por voz e um método de avaliação de comunicabilidade, para traçar o perfil semiótico da aplicação.

## **1.4 Organização do trabalho**

Este trabalho está dividido em cinco seções e a considerações finais e trabalhos futuros. A presente seção é a primeira, a qual introduz o trabalho, com objetivos e a metodologia, mostrando a ideia central do trabalho como um todo. A seção 2 apresenta o referencial teórico utilizado para esse trabalho. A seção 3 apresenta o ARGuideVoz, aplicação adaptada para receber o serviço de voz para interação. A seção 4 mostra as decisões de projeto, e arquitetura para acomodar o novo serviço de voz na aplicação. A seção 5 tem os resultados das avaliações realizadas levando

em consideração as mudanças feitas na aplicação. Por fim, são apresentadas as considerações finais, a análise dos resultados e trabalhos futuros.

## 2. REFERENCIAL BIBLIOGRÁFICO

Esse trabalho tem como foco integrar e dar diretrizes sobre interação por voz em uma aplicação de RAM, que são basicamente os campos de conhecimento necessários para sua realização. O desafio de integrar essas duas áreas pede um estudo de ambas e trabalhos que já utilizaram RAM e interação por voz juntos são vistos nessa seção do trabalho.

A interação por toque, com gestos pré-definidos em tela, como tocar e deslizar, são interações que já possuem um certo padrão em dispositivos móveis, por exemplo, o uso de dois dedos para zoom em uma foto, e um toque em lista para seleção de um item. Logo não serão abordadas profundamente, mas em cada visão da aplicação serão dados detalhes de cada interação possível.

O estudo de RAM é necessário para criar e melhorar as ferramentas que a aplicação possui, fazer modificações na arquitetura e realizar testes com novos componentes de modo eficaz. Aplicações de RAM foram estudadas para buscar elementos comuns de interface de usuário (*User Interface* – UI), interação com gestos, colocação de objetos na cena e como foi realizado a distribuição de conteúdo. Foi levado em consideração também formas de interação que foram utilizadas nesses trabalhos, voz, gestos e etc.

Por outro lado, interações por voz são um componente novo na aplicação em que será implementado o protótipo. Entender interações por voz em dispositivos móveis, usos comuns, padrões de desenvolvimento e arquiteturas utilizadas é fundamental para a implementação e avaliação desse serviço na aplicação. O uso de voz vai acrescentar uma variedade de mudanças na aplicação, em arquitetura e interface de usuário, e para a base dessas mudanças essa seção é necessária.



## 2.1 Realidade Aumentada Móvel

A Realidade Aumentada (RA) trata da inserção de elementos virtuais no cenário real, visando aumentar a percepção do usuário sobre os elementos que o cercam (CAWOOD e FIALA, 2008). Com a inovação de dispositivos móveis, que agora possuem processadores e componentes mais avançados, capazes de suportar melhor a renderização de objetos 3D e o processamento de informações, surgiu uma boa oportunidade para o desenvolvimento da RAM. A RAM tem despertado interesse tanto no desenvolvimento de aplicações na área, quanto para a utilização dessas aplicações. Como mostrado em Olsson e Salo (2011), já existem algumas bibliotecas e Kits de Desenvolvimento de Software (SDK), que oferecem funcionalidades apropriadas à RA e facilitam o desenvolvimento de aplicações RAM.

Na pesquisa realizada em Olsson e Salo (2011), usuários foram convidados a utilizar várias aplicações RAM em diversos contextos, para citar alguns, ambientes *indoor*, *outdoor*, com vários pontos de interesse e em grandes áreas urbanas. Resultados mostraram que há interesse por parte dos usuários em aplicações RAM, mostrando uma aceitação positiva, porém, ainda são necessários estudos em como incluir essa tecnologia em serviços relevantes para o dia a dia do usuário.

Aplicações de RAM tem várias abordagens para posicionar conteúdo virtual em uma cena real entre elas as mais populares são a utilização de marcadores e uso de localização por GPS (Global Positioning System). Exemplos de ambas as abordagens podem ser vistos nas aplicações descritas em Awila (2014), Blippar (2015), Layar (2015), Metaio (2015) e Wikitude (2015) A abordagem de marcadores usa reconhecimento de imagem para reconhecer certos padrões, e com ele selecionar conteúdo. A abordagem de localização usa o GPS do dispositivo e a orientação da câmera para definir o conteúdo para o usuário baseado em sua localização. As principais características da RAM, segundo Srinivasa *et al.* (2014) são:

- **Apresentação de ambiente aumentado em dispositivos móveis:** Combinar o ambiente real, usando sensores, câmeras a/ou a posição do usuário, para prover o conteúdo correto.
- **Tecnologias de rastreamento:** Câmera, GPS, Acelerômetro, Compasso, Giroscópio e outros sensores podem ser usados para fazer o rastreamento do usuário, e ajudar a registrar o objeto virtual na tela do dispositivo.

- **O sistema gráfico:** é o responsável pela renderização dos objetos virtuais na imagem real, tornando ela aumentada.
- **A mistura de ambientes reais e virtuais:** O objetivo principal de uma aplicação de RAM, aumentar a percepção do usuário em um ambiente real através de conteúdo virtual

Os desafios no desenvolvimento de aplicações de RAM são, segundo Markov-Vetter (2013) e Doswell (2006):

- **Integração dos sensores do dispositivo:** Cada sensor do dispositivo responde para a aplicação uma ação ou estado diferente. Coordenar os sensores e o que está sendo mostrado em tela é essencial para a experiência de uso.
- **Baixa precisão em tecnologias de rastreamento:** Realizar o rastreamento exige o uso da câmera, posição do dispositivo, direção em que o dispositivo está apontando, entre outros. A precisão de cada um desses sensores pode prejudicar o rastreamento, já que o registro na tela depende deles para manter a sincronia na tela.
- **Dispositivos com características divergentes e limitantes:** A falta de um padrão de sensores, câmeras com diferentes resoluções, dispositivos de uma mesma família apresentando características diferentes, entre outros, são fatores não triviais que devem ser levados em conta no desenvolvimento de uma aplicação de RAM.
- **Variação na interface do usuário:** Botões, gestos, comandos, e uso dos sensores de um dispositivo não tem um padrão definido para esse tipo de aplicação, tornando esse um importante fator no planejamento desse tipo de aplicação.
- **Falta de padrão de desenvolvimento para aplicações de RAM:** Aplicações de RAM tem que usar vários recursos do dispositivo, sincronizar informações, e por vezes, utilizar a rede do dispositivo para manter e buscar dados em um servidor. A falta de um padrão arquitetural específico para aplicações de RAM gera uma variedade de problemas, já que a complexidade leva a problemas arquiteturais e dificuldades de adicionar novas características.
- **Consumo de energia:** O uso constante de sensores, o uso do processamento gráfico e câmera são grandes consumidores de energia do dispositivo. Usar somente o necessário dependendo da interação é essencial para a aceitação do usuário da aplicação.

Alguns princípios podem ajudar o desenvolvimento dessas aplicações, levando em conta os trabalhos de Nielsen (1994) e Pysylalo *et al.* (2000), pode-se dizer que são aplicáveis a RA:

- **Considerar:** o perfil dos usuários, se o uso será *outdoor*, com uma ou duas mãos no dispositivo e o tempo médio de uso da aplicação.
- **Seguir boas práticas de usabilidade em RA e dispositivos móveis:** Deixar a cena aumentada limpa, com ícones grane, camadas, interação com objetos 3D e sempre considerar o mundo real.
- **Sobre o dispositivo:** levar em consideração o peso, brilho, reflexo na tela, o uso dos sensores de baixa precisão para rastreamento.
- **Informação virtual:** Considerar o conteúdo que será mostrado em tela ao mesmo tempo, como ele será mostrado, os lugares em que serão mostrados e organizar múltiplas visões.
- **Avaliação:** Adaptar *guidelines* de avaliações de usabilidade de RA e dispositivos móveis e realizar avaliações de campo com usuários reais.

## 2.2 Interação por Voz

O reconhecimento por voz por si só não oferece valor ao usuário de uma aplicação. Os comandos utilizados devem ter significado para o usuário, e para isso uma série de medidas devem ser tomadas, como a escolha de palavras e organização da estrutura lógica da aplicação. Em interação humano computador, classicamente as atividades de interação estão agrupadas em quatro categorias principais: instruções; conversação; manipulação e navegação; e exploração (LARSEN, JENSEN, *et al.*, 2007), e depender somente da GUI para essas interações pode levar o usuário sempre a escolher uma interação em favor de outras. Como por exemplo, se um botão indica algo e pode ser clicado ou tocado, então o usuário pode entender esse *widget* somente desse modo, sem a necessidade de usar comandos de voz.

A computação ubíqua é um exemplo de aplicação de uso de interação por voz. Meios tradicionais de interação por teclado e mouse são ineficientes nesse cenário de computação ubíqua (JIANG, LU, *et al.*, 2006), sendo o uso de voz e gestos, principalmente voz, uma alternativa ideal para uso dessas aplicações. O uso de comandos de voz para jogos também tem ganhado popu-

laridade nos últimos anos, com uso de dispositivos como o Kinect. O uso de voz entre os próprios jogadores é um tópico antigo de pesquisa, e muito popular. O estudo de Carter *et al.* (2015) mostra que no ambiente de um jogo, a voz permite ligar a identidade do jogador com o imaginário do jogo. Quando a interação não está relacionada com o ambiente ou o avatar, uma dissonância é criada e a experiência de jogo é prejudicada. O trabalho de Blinov *et al.* (2011) utiliza uma mistura de gestos e voz para interações móveis e multimídia. A ferramenta desenvolvida utiliza várias interações que utilizam o polegar do usuário em uma tela de celular em conjunto com comandos de voz, para realizar operações em conteúdo multimídia.

O uso de botões em toque de tela para controle não é a melhor opção para uso em carros, que cada vez mais contém funções de assistência, de conforto e entretenimento. O uso de vários menus esconde funções do usuário, que não tem acesso simplificado a elas. O trabalho de Pflöging *et al.* (2012) encontrou uma solução no uso misto de voz e gestos. Um teste foi realizado em um protótipo, em dois cenários, um utiliza menus em uma tela de toque, e o outro somente os comandos de voz e uso de gestos. O resultado mostrou que apesar de interações mais lentas que com o menu, o uso de voz e gestos não atrapalhou o desempenho na direção e a carga visual sentida era menor.

Trabalhos mais antigos mostram que usos de comandos de voz sempre foram vistos com alguma utilidade prática. Na abordagem de Muller *et al.* (1998), os usuários podem interagir com um sistema de visualização de dados anatômicos em três dimensões derivadas de exames feitos. O usuário pode escolher o modo de interação, se por voz, gestos ou mouse 3D, e então fazer as interações características da sua escolha. Segundo Sweetser (1998), as pessoas mantêm uma certa relação com o que falam e os gestos que acompanham a conversa. Esse trabalho cerca vários modelos linguísticos, que relacionam os gestos utilizados em uma conversa com o que os participantes estavam conversando no momento. Problemas encontrados antes também podem ser vistos hoje como desafios, como o trabalho de Munteanu, Whittaker e Rajput (2013) que destaca alguns problemas com o uso da voz em aplicações, como a dificuldade de reconhecimento em reconhecedores e dificuldades de avaliação dessas interações.

Assistentes pessoais têm ganhado muito destaque como aplicações de dispositivos móveis. Com a agregação de recursos provenientes de sensores do dispositivo, assistentes pessoais podem usar essas informações e criar uma interação personalizada com o usuário. O uso de coman-

dos de voz se tornou importante em assistentes pessoais, como o Google Now e o Siri, que estão presentes em configurações padrão em vários dispositivos móveis.

### 2.3 Trabalhos Relacionados

A utilização de comandos de voz em aplicações de RAM ainda é rara e há uma notável carência de estudos da utilização de voz em aplicações de RAM em dispositivos como *smartphones* e *tablets*. Em uma pesquisa realizada nos principais indexadores de trabalhos acadêmicos da computação (IEEE Xplore, Science Direct e ACM Digital Library) foi constatada uma carência de estudos que contém RAM e comandos de voz.

Ao pesquisar com as seguintes palavras chaves: “Mobile”, “Augmented”, “Reality”, “Voice” e “Speech”, (sendo que *voice* e *speech* foram pesquisados separadamente por serem utilizadas como sinônimos) foram encontrados um total de 42 artigos, sendo 29 na IEEE, 13 na ACM e nenhum na Science Direct. Após a leitura dos resumos, remoção de artigos repetidos e remoção de artigos que não estão relacionados com assunto proposto, resultou um total de 12 artigos, sendo 9 na IEEE e 3 na ACM. A Figura 2 (a) mostra o resultado desta busca.

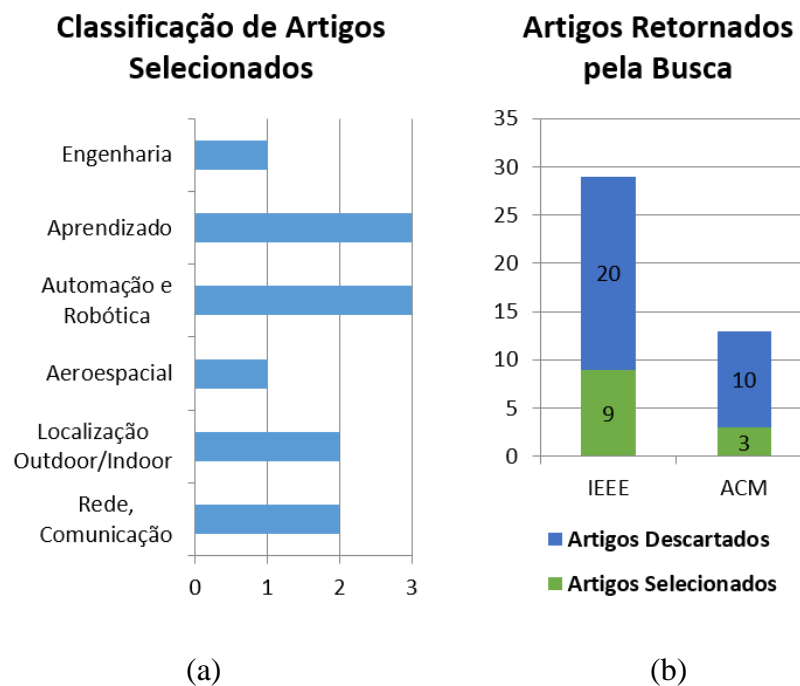


Figura 2. Informações sobre artigos encontrados

Após uma leitura mais aprofundada dos artigos selecionados, verificou-se que os artigos são sobre aplicações que utilizam comandos de voz, mas não há um estudo mais detalhado sobre os

impactos de sua utilização. A Figura 2 (b) mostra um gráfico dos artigos selecionados por área em que a voz e a RAM foram utilizados. Embora o número de aplicações com voz seja restrito existe uma ampla variedade de aplicações que vai do ensino e aprendizagem até as engenharias e robótica. A Tabela 1 apresenta as citações dos artigos selecionados na pesquisa por área. Ao realizar a mesma pesquisa, porém sem as palavras *voice* e *speech* foram retornados um total de 1927 resultados, que se comparado aos 42 resultados que incluíram a voz representam bem a ausência de estudos de RAM em conjunto com a voz.

Alguns trabalhos encontrados possuem interação por voz em uma aplicação de RAM e têm funcionalidades e métodos muito parecidos com os usados nesse trabalho, e serão destacados a seguir. O ARCity (DE LA NUBE AGUIRRE BRITO, 2015) é uma aplicação de RAM que auxilia o usuário a encontrar pontos turísticos em uma cidade. A aplicação tem um mapa integrado em um navegador de realidade aumentada. Usuário não sai da aplicação para usar o mapa, ao mesmo tempo que usava o navegador de RA foi um dos pontos positivos vistos na avaliação realizada.

O Immersive Tour Post (PARK, NAM e SHI, 2006) é um sistema que usa imersão com áudio e vídeo em um lugar importante de determinada cidade. O dispositivo deve ser colocado em frente ao POI e assim os turistas podem saber a história sobre atrações turísticas com um vídeo imersivo. A ideia de imersão em vídeo é a vantagem dessa tecnologia, então essa aplicação usa o conceito de navegação aumentada, mesmo que o usuário não esteja imerso totalmente, a menos que rotacione o dispositivo em torno do POI para ter a sensação aumentada.

O Time Machine (FENG, MENG, *et al.*, 2013) é uma aplicação de RAM que implementa o conceito de imagem panorâmica para fazer uma lista de atrações turísticas no seu estado atual e seu passado. O usuário pode seguir através de uma linha do tempo dessas visões. A aplicação mostra essa linha do tempo com imagens históricas de POIs. As imagens descrevem a história desses lugares conforme está descrito na imagem.

**Tabela 1. Artigos encontrados e descrições**

Área do conhecimento	Nº de Artigos	Citações	Descrição
Rede, Comunicação	2	(PYSSYSALO, REPO, <i>et al.</i> , 2000) (PULLI, PYSSYSALO, <i>et al.</i> , 1998)	Aplicação de RAM que usa várias formas de interação, como voz, mas não detalha delas
Localização Outdoor/Indoor	2	(DELAÏL, WERUAGA e ZEMERLY, 2012) (SRINIVASA, JAGANNATH, <i>et al.</i> , 2014)	Aplicações de contexto local, como universidades, que citam voz como forma de interação, mas somente implementação
Aeroespacial	1	(MARKOV-VETTER e STAADT, 2013)	Aplicação para apoiar o uso de RAM em concertos de alto risco, com o uso de voz somente em interações de menu
Automação e Robótica	3	(GIESLER, SALB, <i>et al.</i> , 2004) (TANG, ANG, <i>et al.</i> , 2015) (IGARASHI e INAMI, 2015)	Uso de RAM e voz em sistemas de automação em robôs, casas, mas sem explicar aspectos de interação
Aprendizado	3	(DOSWELL, 2006) (DOSWELL, BLAKE e BUTCHER-GREEN, 2006) (DOSWELL, 2006)	Uso da RAM em sistemas de ensino, com a voz sendo usada para interação indireta, com pessoas, não com sistema
Engenharia	1	(GOOSE, SUDARSKY, <i>et al.</i> , 2002)	RAM e uso de voz como uma conversa, com foco na performance

### 3. PROTÓTIPO DA APLICAÇÃO ARGUIDEVOZ

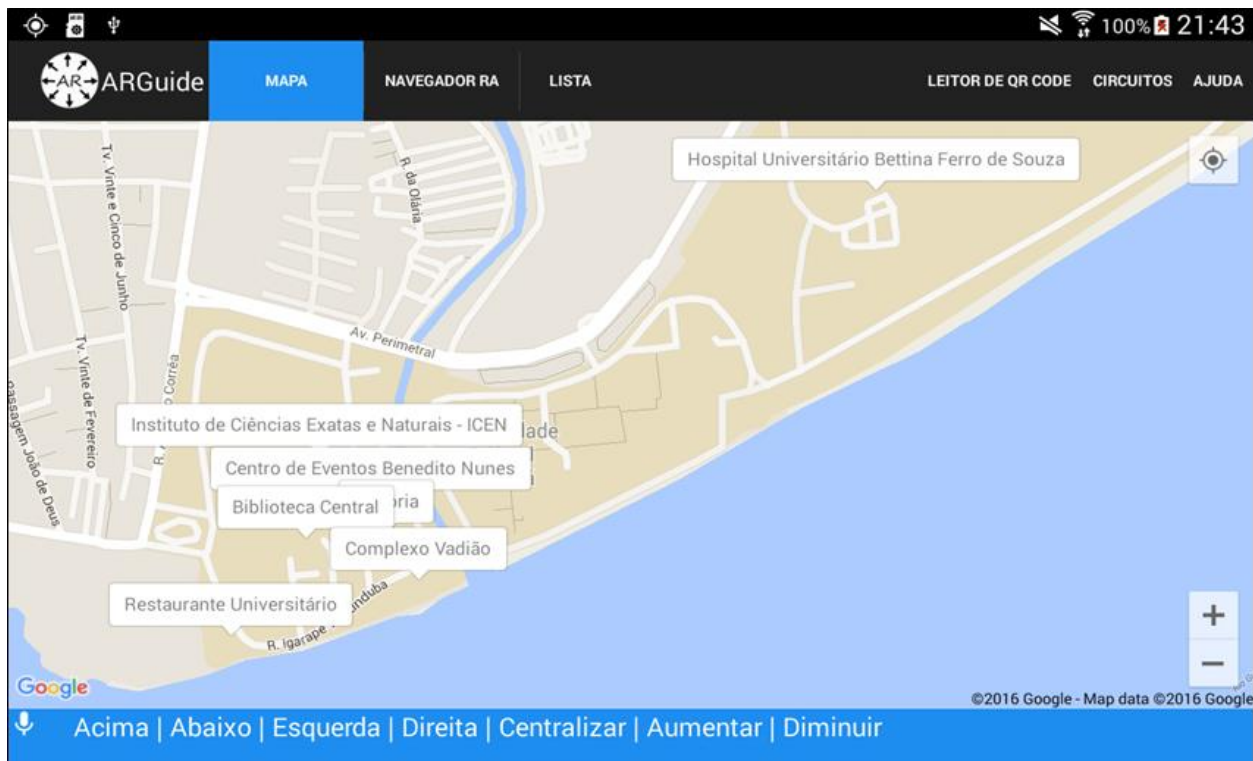
O ARGuideVoz é uma extensão do ARGuide com navegação por voz. O ARGuide é uma aplicação de RAM que utiliza conteúdos georeferenciados como base para a navegação do usuário (SANTOS, CARNEIRO, *et al.*, 2014), (SANTOS, CARNEIRO, *et al.*, 2014), (ARAUJO, CARNEIRO, *et al.*, 2015). A aplicação possui quatro visões principais, e permite que o usuário possa explorar pontos de interesse alternando entre várias visões coordenadas para acessar o conteúdo disponível. As visões da aplicação são coordenadas, ou seja, as alterações realizadas em uma visão refletem em todas as outras, essa sincronia permite que a mesma informação seja exposta por diferentes perspectivas. Todas as funções disponíveis com interação por toque em tela também podem ser realizadas com a voz no ARGuideVoz.

O conteúdo utilizado no ARGuideVoz é dividido por cada POI (Point Of Interest - Ponto de Interesse). Esse conteúdo pode ser em texto, vídeo ou fotos, que podem ser acessados através das visões da aplicação. Obrigatoriamente, cada POI deve ter um texto descritivo, uma foto que sirva como ícone e uma posição de GPS. Neste trabalho, os POIs utilizados na aplicação são da UFPA, com lugares que podem ser considerados relevantes para o dia-a-dia do estudante universitário, como a biblioteca e o restaurante. As visões disponíveis na aplicação são mostradas a seguir.

#### 3.1 Mapa

A visão de Mapa mostra os POIs que fazem parte do conteúdo dispostos em uma visão de mapa. O mapa utilizado para a aplicação é da API do Google Maps v2 (GOOGLE INC., 2015). Cada POI é representado por um marcador no mapa, e o usuário toca em um deles para acessar o conteúdo do POI. A Figura 3 mostra a visão do mapa na aplicação, onde os marcadores são indicados pelo nome do POI.





**Figura 3. Visão do mapa no ARGuideVoz**

As funções de navegação no mapa são as funções de zoom, para afastar e aproximar a visão do usuário, e Pan, para movimentar o mapa. Também é possível fazer uma rota do lugar que o usuário está até o POI selecionado.

### **3.2 Navegador de Realidade Aumentada**

A visão do navegador de realidade aumentada utiliza a câmera e os sensores do dispositivo para localizar os POIs no espaço e direções reais. A câmera, quando apontada corretamente para um POI, mostra um ícone com uma foto, nome e distância em metros até esse ponto, da posição do usuário. A Figura 4 mostra essa visão. Em 4(a), o *slider* é usado para controlar o alcance da visão dos POIs, e em 4(b) o radar mostra a posição de cada POI dentro do raio escolhido. Essa visão é apropriada para exploração de conteúdo em um ambiente aberto, em que a referência virtual do marcador ajude na localização do POI.

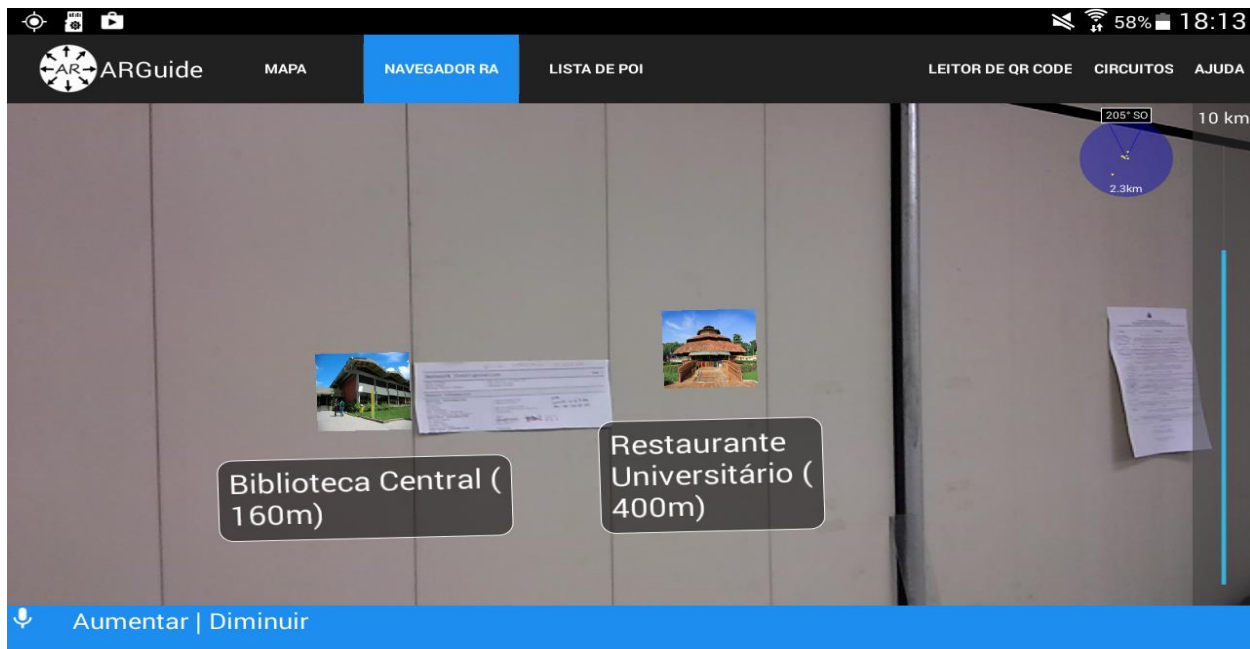


Figura 4. Visão do navegador RA no ARGuideVoz

### 3.3 Leitor de QR Code

O leitor de QR Code utiliza a câmera do dispositivo para ler um QR Code cadastrado e mostrar o conteúdo para o usuário. Esse tipo de interação é amplamente utilizada por vários serviços hoje, como na nota fiscal de alguns estabelecimentos e alguns pontos turísticos. No ARGuideVoz, o QR Code cadastrado chama a visão do POI correspondente para o usuário. Essa visão pode ser usada em ambientes com concentração de POIs, em que a proximidade entre eles atrapalha o funcionamento do mapa e do navegador de RA. Um exemplo de execução do leitor de QR Code pode ser visto abaixo, na Figura 5.

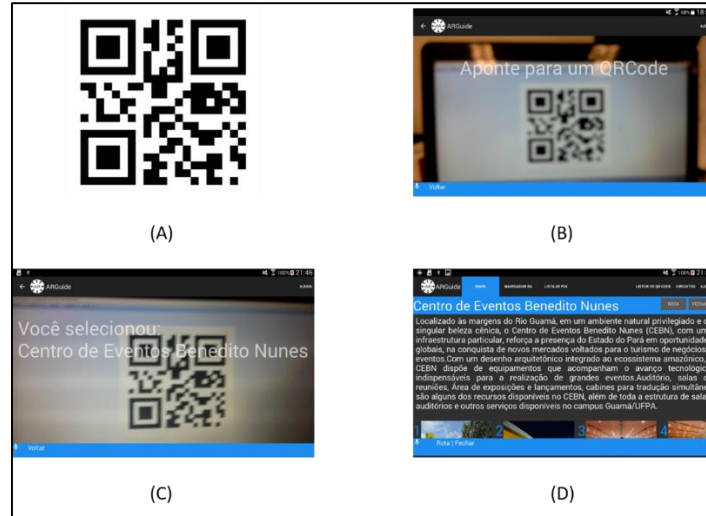


Figura 5. Visão de QRCode no ARGuideVoz (a) Marcador de um POI (b) Leitura do marcador (c) reconhecimento do marcador (d) Exposição do conteúdo ligado ao marcador

### 3.4 Lista

A lista de POIs mostra os POIs ordenados em forma de lista para o usuário. Para cada item da lista, é mostrado uma foto do POI como ícone, e uma pequena descrição do POI. Quando selecionado, a visão do POI correspondente é chamada na aplicação. Abaixo, na Figura 6, um exemplo da lista.

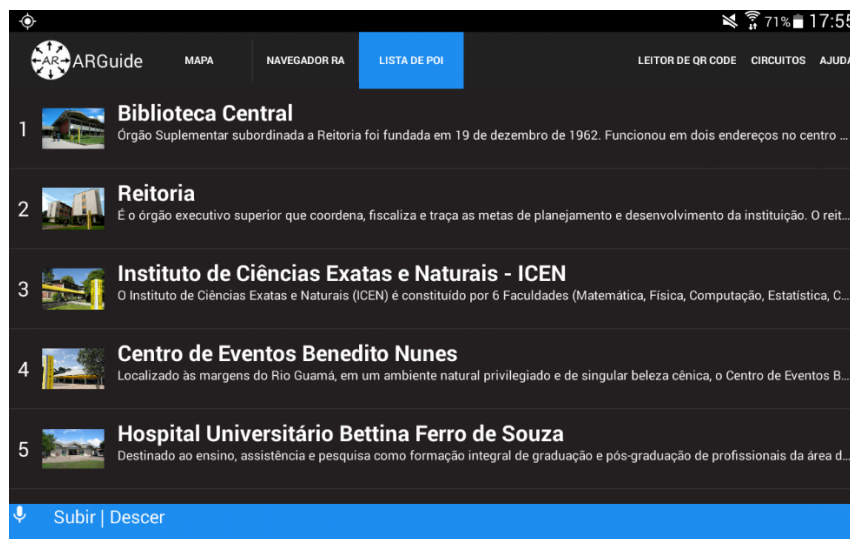


Figura 6. Visão de Lista no ARGuideVoz

### 3.5 Visão de POI

A visão de POI é a visão que mostra todo o conteúdo relacionado ao POI selecionado. Um texto descritivo é mostrado na parte central da tela, na barra em cima o botão de traçar rota faz

um caminho no mapa da aplicação do lugar que o usuário está para o POI. Na parte de baixo, o conteúdo multimídia pode ser acessado. Quando selecionado, uma tela com opções de interações é aberta para o controle do usuário sobre a mídia, conforme Figura 7.



**Figura 7. Visão de POI no ARGuideVoz**

## 4. PROPOSTA DE INTERAÇÃO POR VOZ E ASPECTOS DE PROJETO

Interações em dispositivos móveis atualmente estão comumente associadas com o uso de toque em tela. O avanço de pesquisas relacionadas a aprendizado de máquina e a velocidade de banda larga possível permitem que a RAM possa ser usada em conjunto com interações por voz. Se a tecnologia ajuda de um lado, dando as bases para o uso, cabe ao desenvolvedor desse tipo de aplicação, que deseja realizar essa integração, alterar a arquitetura da sua aplicação para acomodar esse tipo de interação.

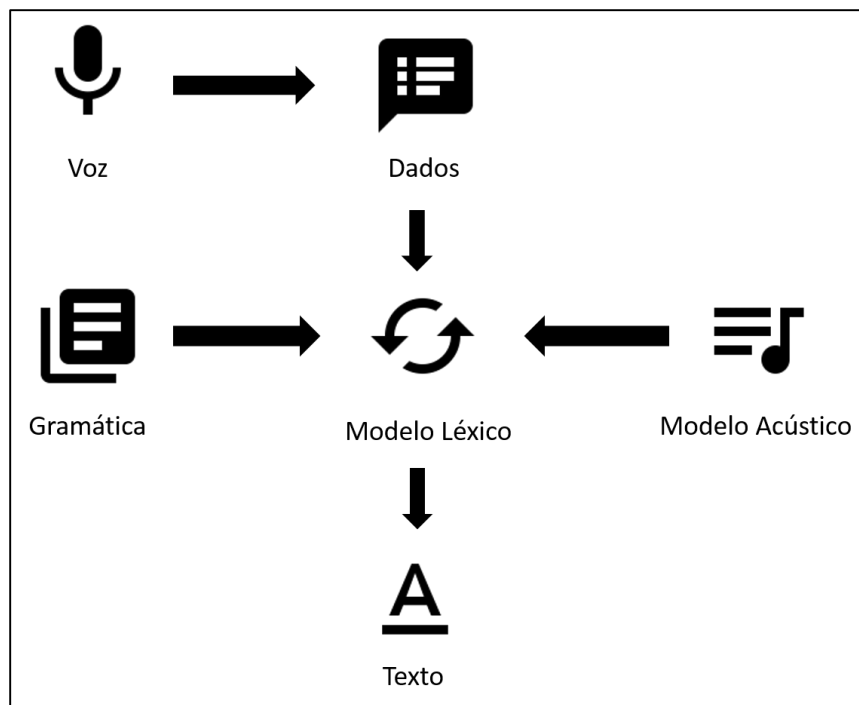
O uso de uma nova forma de interação expande as possibilidades de uso da uma aplicação de RAM. O uso por voz permite que, por exemplo, se use a aplicação enquanto está dirigindo, ou enquanto está com uma das mãos ocupadas, e até mesmo permite que um novo perfil de usuários; com dificuldades de tocar na tela possa utilizar a aplicação.

É fundamental que a interação por voz seja estável para que a experiência de uso não seja prejudicada por lentidão ou comportamentos inesperados. O desenvolvimento de um módulo novo para o reconhecimento de voz e a integração dele na aplicação de RAM escolhida será descrito nessa seção.

### 4.1 Interação por Voz em dispositivos Móveis

O uso de voz em *smartphones* e *tablets* apresenta vantagens para o usuário. O aumento de funcionalidades exige que as aplicações usem menus mais complexos, que podem acabar escondendo algumas funções do usuário ou demandar um tempo considerável até que a funcionalidade desejada seja encontrada. Nesse cenário, a interação por voz pode facilitar a busca do usuário por uma funcionalidade, ou até mesmo servir como atalho para a execução da funcionalidade em questão (XIA, KAI, *et al.*, 2010).

Apesar da utilização de comandos por voz poder variar de acordo com o tipo de aplicação, de maneira geral, o reconhecimento de voz nas aplicações segue um modelo padrão, que pode ser observado na Figura 8.



**Figura 8. Funcionamento padrão de um modelo de reconhecimento de voz**

No modelo apresentado, o usuário entra com a informação em forma de comando de voz, que é repassada para o modelo léxico, onde é analisada juntamente com a gramática e modelo acústico empregados na aplicação. O resultado desse processo é então retornado ao usuário em forma textual. Esse modelo geral exemplifica aplicações de interação por voz encontrados em vários sistemas de reconhecimento de voz (RABINER, 1993).

Para este trabalho, o sistema operacional Android foi utilizado e a Figura 9 mostra o fluxo do serviço de voz implementado utilizando o Google Speech API (GOOGLE INC, 2016), que é o serviço de reconhecimento de voz padrão do Android.

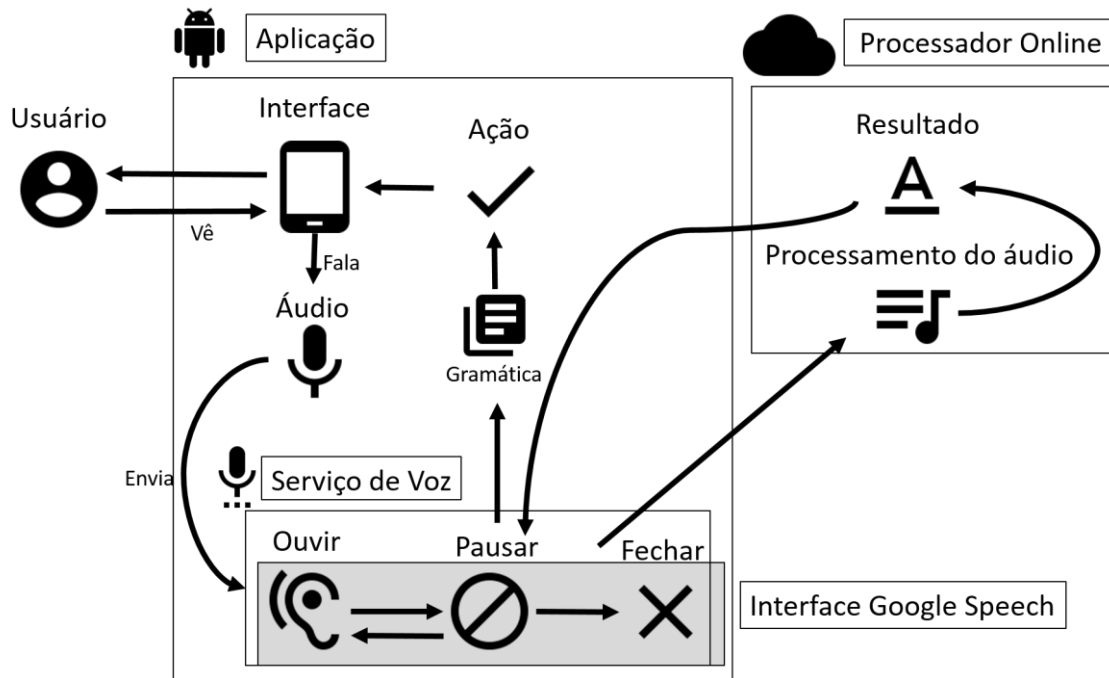


Figura 9. Ilustração do fluxo de troca de mensagens na implementação do serviço

Quanto ao funcionamento, uma interface reconhedora deve ser implementada para enviar os dados correspondentes a captura de áudio realizada, que vai ser processada em um conjunto de palavras correspondente a síntese realizada. Esse resultado é retornado para a aplicação em forma de um conjunto de palavras.

O serviço de voz integrado possui alguns estados para garantir a eficiência, acurácia e correteude do reconhedor. Esses estados e suas funções são listados a seguir:

- **Ouvir:** Esse estado liga o microfone do dispositivo e espera a entrada de áudio do usuário. Nesse estado o áudio puro é convertido em um formato de dados compatível com o reconhedor pelo processador. Assim que ele detecta silêncio, o reconhedor muda de estado para o estado de Parar.
- **Pausar:** Nesse estado é necessário parar o reconhecimento da entrada do usuário, e enviar os dados para o processador. Abre-se uma conexão e se espera o sinal positivo de entrega. De acordo com a ação realizada pelo serviço, de falha ou acerto de reconhecimento, pode-se passar para o estado de encerrar, ou voltar para o estado de Ouvir.

- **Encerrar:** Esse estado finaliza o serviço de reconhecimento. Faz as verificações necessárias para liberar o microfone para o sistema, e encerra qualquer conexão ainda aberta com a internet usada em função do reconhecedor.

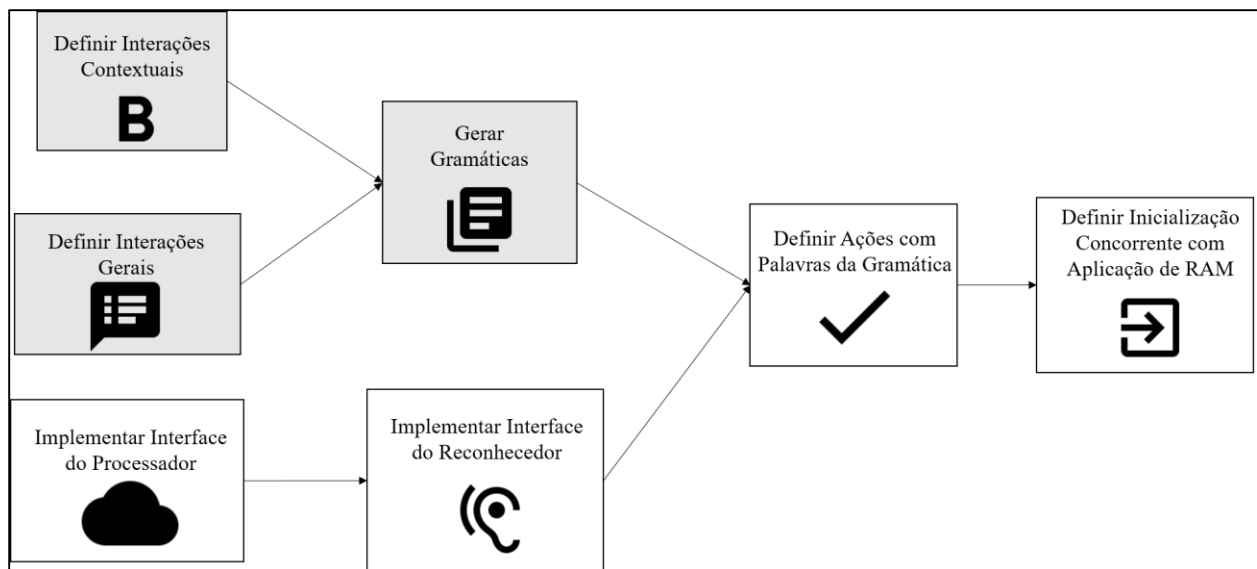
Esse modelo acompanha o desenvolvimento da aplicação, juntamente com a arquitetura, definida nas próximas subseções, para gerar uma forma geral de desenvolvimento de aplicações de RAM com interação por voz.

## 4.2 Aspectos gerais de implementação

A implementação desse serviço levou a uma série de passos que podem ser descritos na forma de aspectos para o desenvolvimento de serviços de voz para aplicações de RAM. Esses aspectos levam em consideração a arquitetura gerada e os pontos relevantes na utilização do Google Speech API para reconhecimento de voz.

Os aspectos podem ser divididos em duas partes principais: relacionadas a aplicação e relacionadas à tecnologia. Aspectos de aplicação são as medidas que devem ser tomadas para que o serviço de voz seja integrado a aplicação de RA. Esses passos são gerais a aplicação, como a escolha de palavras, mudanças de interface ou criação de novos componentes. Aspectos de tecnologia são as medidas que devem ser tomadas diante da tecnologia escolhida, como a implementação de uma interface para o processador e mudança no código da aplicação de RA para iniciar o serviço de voz. A Figura 10 mostra o processo de aplicação das diretrizes na aplicação de RAM.





**Figura 10. Aspectos gerais de implementação, em cinza os aspectos ligados a aplicação, em branco os aspectos relacionados a tecnologia.**

As diretrizes que devem ser seguidas para implementação do serviço, e estão relacionadas a aplicação estão listadas abaixo:

1) *Definir Interações Contextuais*: Definir comandos contextuais é necessário para que os comandos que são dependentes de certo contexto sejam executados somente nesse contexto. Exemplos de contexto podem ser a visualização de certo dado, a navegação em um mapa, a ação diante de um objeto virtual. As interações escolhidas devem ser mapeadas em um conjunto de palavras e essas palavras devem ser associadas a uma condição de contexto, como uma visão, presente na aplicação.

2) *Definir Interações Gerais*: Comandos gerais são necessários para ações que são gerais em mais de um contexto da aplicação. São comandos gerais as ações de menu que podem ser acessadas em qualquer momento, comandos padrões da própria tecnologia utilizada, como voltar, página principal (home) e configurações. Essas interações devem ser mapeadas para que o contexto não seja aplicado a elas, o conjunto de palavras não pode ser concorrente com nenhuma palavra de nenhum contexto.

3) *Gerar Gramáticas*: Definidas as interações que serão possíveis em contextos e de forma geral, o conjunto dessas palavras deve ser usado para geração de uma gramática que será usada para as ações da aplicação. Essa gramática deve ser organizada de modo que o reconhecedor

possa utilizar essas palavras para implementar essas ações, que depende do formato que o processador e o reconhecedor podem processar.

As diretrizes que devem ser seguidas para tecnologia estão listadas abaixo:

1) *Implementar Interface do Processador*: O processador pode estar integrado ou não ao telefone, mas é necessário ter uma interface de acesso a ele, ao microfone, ou ambos. Essa interface deve ser criada para processamento ou direcionamento da voz do usuário. Ela deve ter métodos para retorno do resultado do que foi captado para a interface do reconhecedor.

2) *Implementar Interface do Reconhecedor*: Essa interface deve ser implementada para direcionar o resultado do processador em alguma ação da aplicação. Deve implementar métodos para esperar os resultados do processador e métodos para perceber os estados possíveis do reconhecimento, como ‘ouvindo’, ‘começando a ouvir’, ‘parando de ouvir’ e ‘iniciando processamento’.

3) *Definir Ações com Palavras da Gramática*: As ações da aplicação devem ser associadas ao reconhecedor, com as palavras que foram definidas na gramática. Essas ações podem ser as gerais ou de contexto. Mais de uma palavra pode acionar a mesma ação, desde que esteja devidamente mapeada na gramática.

4) *Definir Inicialização Concorrente com Aplicação de RAM*: O passo final para implementação desse serviço é definir como o serviço vai ser iniciado na aplicação. Um botão que aciona o serviço é usado em algumas aplicações de interação por voz (Buti, 2012), mas pode ser definido do modo que for necessário para a aplicação. Essa implementação é altamente dependente da tecnologia, já que envolve paralelismo de execução e interação, e sistemas operacionais móveis implementam de forma diferenciada acesso a ferramentas como microfone e execução do modo de ouvir do mesmo.

Esses aspectos são relevantes no desenvolvimento de um serviço de voz em uma aplicação de RAM. O exemplo de reconhecedor de voz utilizado é padrão de sistemas Android, sendo assim esse modelo pode ser replicado com fidelidade nesse sistema. Outros sistemas móveis podem precisar de adaptação no sentido tecnológico, no modo como organiza-se os componentes base para essas aplicações e na manipulação.

### 4.3 Arquitetura de um serviço de Voz em RAM

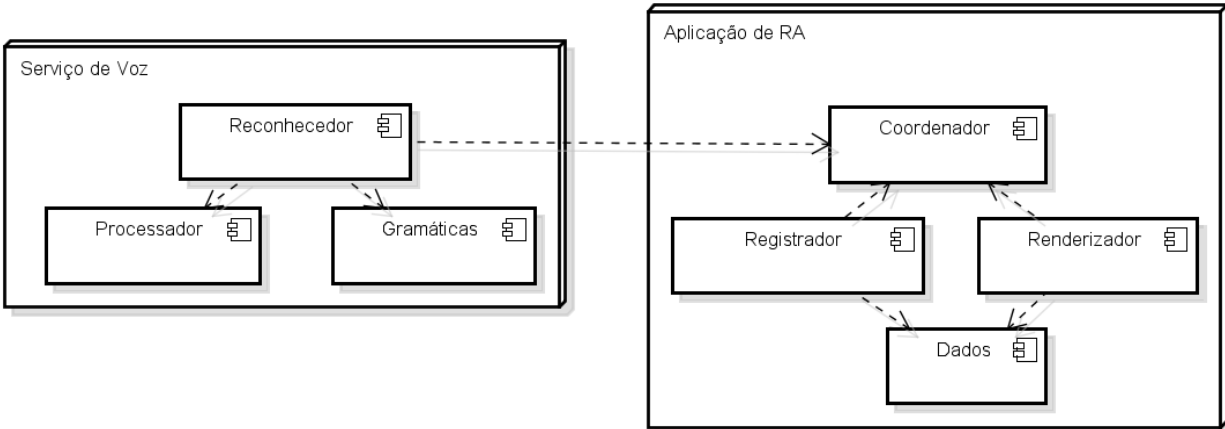
Um serviço de voz precisa de alguns elementos chaves para poder realizar a síntese e reconhecimento de voz. O processo visto na Figura 8 mostra o modelo geral independente da aplicação utilizada. Para aplicações de realidade aumentada móvel esse modelo se encaixa com algumas mudanças, relacionadas aos componentes de realidade aumentada.

Uma arquitetura para aplicações de RAM que utilizam serviços de voz pode ser utilizada para definir ações de implementação para os principais componentes desse tipo de aplicação. Os componentes utilizados são baseados em componentes de um dispositivo Android, que já contém alguns elementos prontos para reconhecimento e síntese de voz.

Estes componentes foram divididos em dois nós principais, relacionados aos dois principais módulos para esse modelo, serviço de voz e a Aplicação de RAM. Os componentes do serviço de voz são um reconhecedor, um sintetizador e as gramáticas. Os componentes de uma aplicação RA foram divididos em quatro, onde os trabalhos de (DELAILE, WERUAGA e ZEMERLY, 2012), (SANTOS, CARNEIRO, *et al.*, 2014), (TANG, ANG, *et al.*, 2015), (VULLAMPARTHI, NELATURU, *et al.*, 2013) e (D'ANTONI, DUNN, *et al.*, 2013) apresentam componentes semelhantes quando se foca em desenvolvimento. Foi definido um padrão dentro dessas aplicações, definidos pela sua função:

- **Dados:** Dados que a aplicação utiliza para gerar conteúdo para o usuário.
- **Renderizador:** Motor gráfico para gerar modelos de realidade aumentada na tela da aplicação.
- **Registrador:** Conjunto de algoritmos ou ferramentas utilizadas pela aplicação para gerar interação com usuário, como reconhecimento de imagens, posição georeferenciada e utilização de sensores.
- **Coordenador:** O componente que liga o Renderizador e o Reconhecimento para gerar o ambiente de realidade aumentada para o usuário.

A Figura 11 mostra um diagrama desses componentes para a obtenção deste serviço de voz para aplicações de RAM.

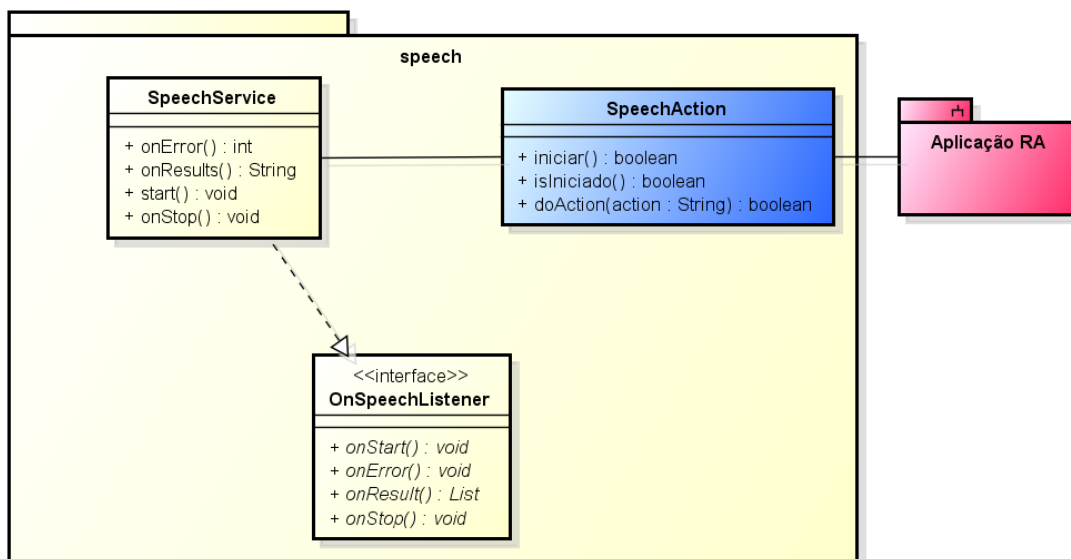


**Figura 11. Arquitetura genérica de uma aplicação de RA com um Serviço de Voz**

A ligação entre o reconhecedor do serviço de voz e do coordenador da aplicação de RAM é importante para manter a sincronização das interações. Um comando que tem sua ação muito atrasada ou uma ação da aplicação que seja inapropriada para um contexto podem dificultar a aceitação do serviço pelo usuário (OLSSON e SALO, 2011).

### 4.3.1 Diagrama de Classes

O diagrama da Figura 11 acima mostra o conceito geral da aplicação, Detalhes mais específicos podem ser vistos na Figura 12 abaixo, que mostra um diagrama de classes para o serviço de voz em uma aplicação de RAM.



**Figura 12. Diagrama de Pacotes de um Serviço de Voz para aplicações RAM**

A especificação de cada classe do diagrama e suas relações pode ser vista abaixo:

1) *SpeechService*: Essa classe é responsável pela comunicação realizada com o processador online. Deve perceber os eventos para enviar e receber os resultados do sintetizador. Os métodos dessa classe são explicados abaixo:

- **start**: Esse método inicia o reconhecimento do serviço de voz ligando o microfone. É encerrado assim que percebe silêncio por parte da entrada do usuário.
- **onStop**: Método que é chamado quando é percebido silêncio, para que a entrada de áudio pura seja encerrada, permitindo manipulação por parte da classe. Esse é o momento de enviar os resultados para o processador.
- **onResults(results : List)**: Nesse método deve-se encaminhar o conjunto de palavras para o reconhecedor poder realizar as ações necessárias fazendo comparação com a gramática.
- **onError(error : int)**: Esse método é chamado em caso de qualquer erro no processo de envio e recebimento do processo de reconhecimento de voz. Erros que podem acontecer estão relacionados a erros de rede, áudio não ser compreendido como voz, não entendimento do que foi dito e falta de resposta do sintetizador.

2) *OnSpeechListener*: É uma interface para ligar os eventos dependentes do sintetizador que ocorrem no serviço de voz. Como o resultado da voz só vem no final de todas as etapas da recuperação e sintetização do áudio, esses métodos são importantes para que o reconhecedor possa saber em que fase da síntese o serviço está e dar a resposta apropriada ao sistema. É implementada por *SpeechAction* e usada por *SpeechService*. Os métodos são chamados em *SpeechService* e suas chamadas manipuladas em *SpeechAction*.

3) *SpeechAction*: É a classe que implementa os serviços do reconhecedor. Ela recebe por meio da interface que implementa os resultados da síntese de voz e aplica as ações necessárias relacionadas a gramática. O método *doAction* recebe uma palavra da gramática utilizada pela aplicação e realiza uma ação ou cancela a mesma mediante definição já feita.

Esse diagrama mostra um conjunto básico de classes que devem ser implementados e como devem ser relacionados com componentes de uma aplicação de RAM. Serviços já implementados no sistema Android podem ser utilizados, como o acesso ao microfone e acesso ao que foi gravado.

A arquitetura do ARGuide é uma variação do modelo MVP (POTEL, 1996), que possui uma *View* com várias visões coordenadas, coordenação essa feita por um *Coordinator* que intermedia as informações e mídias de determinado POI no *Model* em uma visão. A Figura 13 mostra a mudança na arquitetura do ARGuide com a adição de comandos de voz, a interface em azul. A inclusão de um serviço de voz é uma interface de coordenação, já que além de acessar os POIs, também comanda outras visões. Esses comandos são paralelos aos comandos do toque do usuário, então a interação de voz passa a ser coordenada juntamente com o toque em tela. O componente *SpeechAction*, já visto anteriormente, é um conjunto de classes, interfaces e serviços que é responsável pelo conhecimento de voz, ele é o componente novo na arquitetura do ARGuideVoz, e coordena variadas ações do aplicativo.

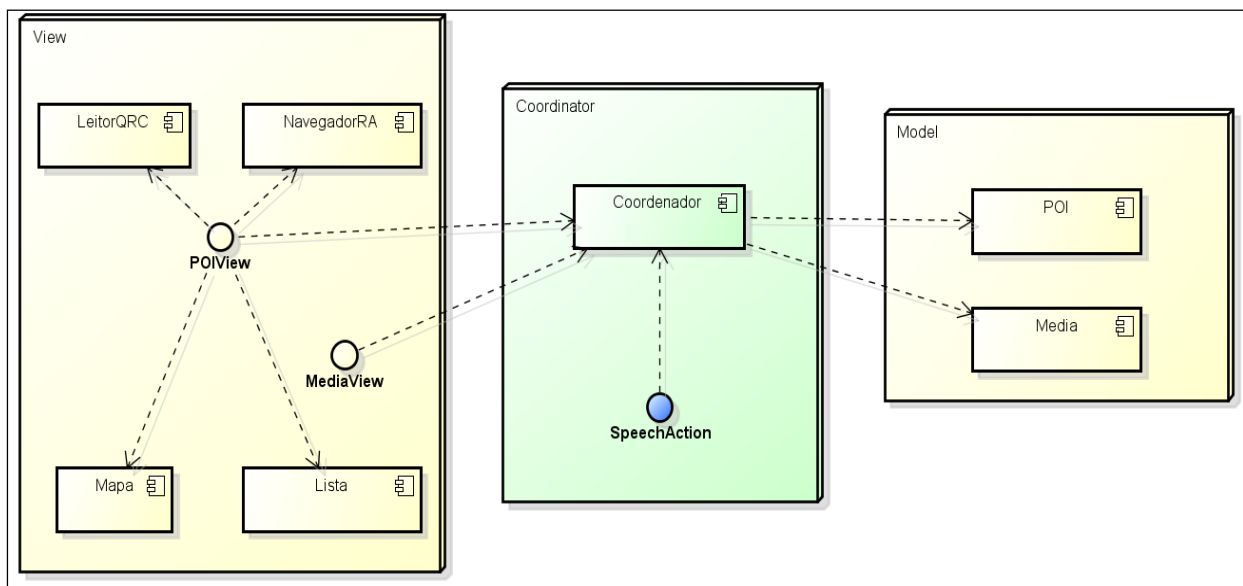


Figura 13. Mudanças realizadas na arquitetura para acomodar o serviço de voz

#### 4.4 Serviço de Voz no ArguideVoz

A navegação no ARGuide deve ser totalmente possível somente com o reconhecimento de voz, do mesmo modo que é possível somente com toques na tela, ou ambos. O reconhecedor fica sempre esperando ações do usuário e inicia junto com a aplicação. Para essas alterações no comportamento da aplicação, foi necessário fazer uma mudança na arquitetura do ARGuide. Um serviço de voz é ligado a um módulo da aplicação, como na Figura 13. A aplicação deve configurar ações de acordo com as palavras que serão reconhecidas.

As palavras da gramática são definidas de acordo com Xia *et al.* (2010), e a interação segue as orientações de Nielsen (1994) e Pyssylalo *et al.* (2000) para aplicações de RAM. As palavras selecionadas devem aceitar variações simples, como plural, sinônimos e variações de gênero. A Tabela 2 define algumas interações da ferramenta.

**Tabela 2. Comandos disponíveis na aplicação**

Comandos de Voz	Ação	GUI
esquerda, direita, cima, baixo	Pan	Mapa/Imagem
aumentar, diminuir	Zoom	Mapa / Imagem / Navegador RA
subir, descer	Pan	Lista
qrcode, circuitos, ajuda, mapa, navegador ra, lista	Menus & Visões	Leitor de QR Code, Circuitos, Ajuda e Visões
<Nome do POI>	Selecionar POI	Marcador do Mapa   Marcador de RA   Item da Lista

A escolha dessas palavras para a gramática visou tornar possível que toda a navegação seja realizada através de voz. Para navegação do mapa e da manipulação de imagem as mesmas palavras foram escolhidas por se tratar de ações análogas. Para o navegador RA, as palavras são as mesmas associadas ao zoom para as imagens e mapa, já que a única ação possível pelo usuário na visão do Navegador RA. A lista tem os comandos de sobe e desce, como o movimento da barra de rolagem.

Comandos de voz comuns para todas as visões são os de seleção de menu, uma vez que os menus ficam estáticos na parte superior da GUI principal; para troca de circuitos que servem como filtro para todas as visões; a ajuda que é contextualizada para cada visão; a troca de visões está disponível também para os usuários em qualquer momento; a seleção de POIs está disponível em todas as visões, com exceção para o leitor de QR Code, que só mostra o POIView se for feito o scan do QRCode correspondente. A seleção de POIs é feita através do comando de voz com o nome do POI.

Incluir uma nova forma de interação em uma aplicação de RAM pode acarretar em mudanças no modo em que essa aplicação exibe o conteúdo para o usuário. O trabalho de Lee e Grice (2006) identificaram os problemas relacionados à integração de comando de voz em aplicações móveis, e sugerem que esses problemas devem ser tratados desde o início da integração da aplicação com um serviço de reconhecimento de voz. Para definir quais mudanças devem ser feitas para inserir um serviço de reconhecimento de voz no ARGuide, foi definido um processo para a obtenção das mudanças que devem ser feitas.

Juntamente com o estudo da literatura, para dar apoio na inserção de interações por voz nessa aplicação, a ajuda de uma especialista em IHC foi necessária. Reuniões foram realizadas para gerar mudanças no desenvolvimento de forma em que as funções de interações por voz estivessem sendo adicionadas. Foram realizadas duas reuniões, e entre cada reunião uma nova versão do ARGuide foi gerada para passar por nova avaliação.

#### **4.4.1 Definição de interface**

O objetivo das reuniões realizadas era definir a interface a ser utilizada para o reconhecedor de voz, as mudanças a serem feitas, e a escolha inicial de palavras. Essas reuniões foram realizadas com um especialista em IHC, no intervalo de uma semana entre uma e outra. O resultado das reuniões pode ser visto a seguir.

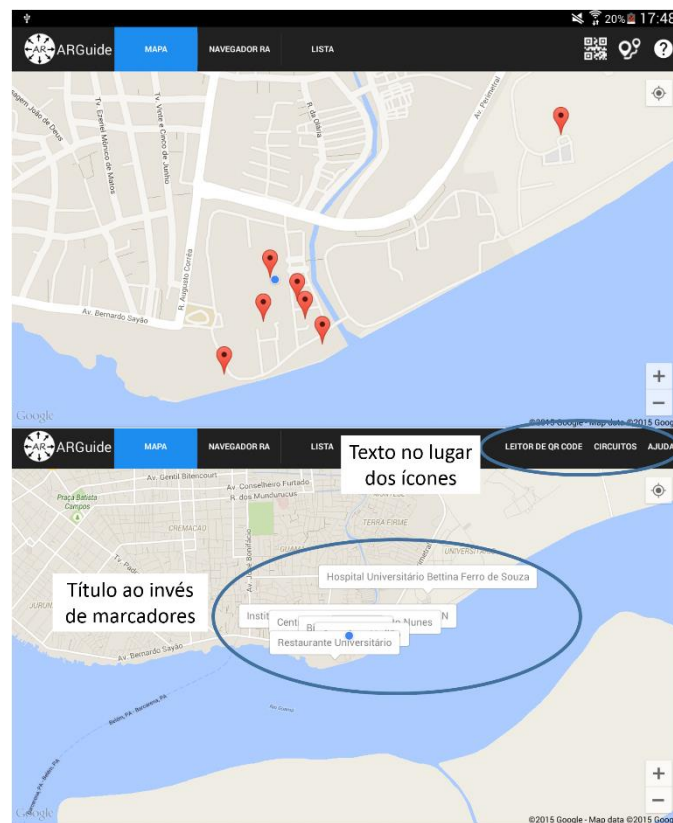
##### *1ª Reunião*

Sem componentes funcionais de reconhecimento de voz, essa versão recebeu avaliação baseada em mudanças de interface simples que facilitariam o uso do reconhecedor de voz. A primeira reunião foi conduzida com explicação prévia da ferramenta e observação do especialista enquanto utilizava a aplicação, mostrando as funcionalidades no decorrer da reunião. As considerações feitas na primeira reunião foram com elementos de interface. O especialista notou que vários componentes de interface estavam sem indicação de que poderiam ser acionados por um comando de voz. Elementos como os marcadores do mapa ou ícones de navegação tiveram que ser transformados para elementos textuais.

Um exemplo de retorno obtido através da reunião está na Figura 14. Essa figura mostra dois aspectos na interface que sofreram alteração em função da primeira reunião. A primeira observação é a troca dos ícones para texto. Os ícones que representavam ações da aplicação foram subs-



tituídos por palavras que podem ser usados como comandos de voz. Os marcadores utilizados para identificar os POI foram trocados rótulos de texto.



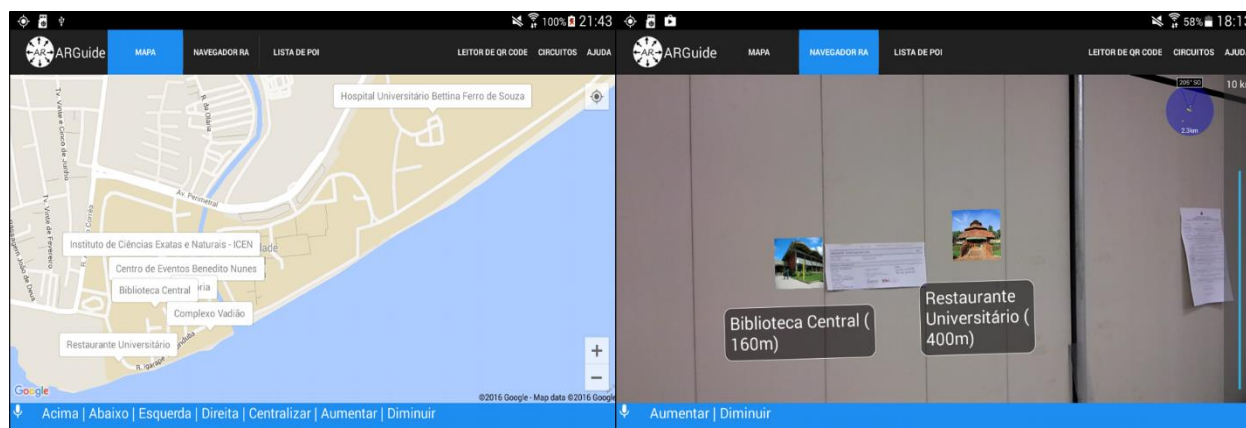
**Figura 14. Exemplo de alterações realizadas após a primeira reunião**

#### *2ª Reunião*

Para a segunda reunião, foi implementado o serviço de reconhecimento de voz, já utilizando o reconhecedor de voz padrão do Android. É um reconhecedor de voz que funciona somente online, a recepção e transmissão dos dados é feita no aparelho e enviada para o reconhecedor, que retorna o texto correspondente ao que foi falado. Com o conhecimento da ferramenta, a especialista focou a avaliação no reconhecedor de voz e uso da gramática escolhida.

Novas melhorias puderam ser realizadas com o reconhecedor funcionando. As considerações feitas nessa reunião foram relacionadas a interpretação e reconhecimento de voz. Comandos base tiveram que ser modificados para acomodar facilidades da linguagem natural, como relaxar comandos que eram somente verbos no infinitivo para substantivos e conjugações mais comuns. Como “sair”, que pode ser pronunciado “sai” ou “saia” que o reconhecedor entende como o mesmo comando. Outra mudança relevante é a presença de legendas constante, para que o usuá-

rio saiba que comandos pode fazer com voz em determinada visão. A Figura 15 exemplifica uma transição com relação à legenda das ações mapeadas para certa visão.



**Figura 15. Alterações realizadas após segunda reunião**

O usuário sempre tem disponível uma caixa de texto que e abre com os comandos de voz para que ele possa visualizar as ações naquela visão. Nesse exemplo, aumentar e diminuir o alcance de pontos de interesse no navegador RA e as ações do mapa. Em ações muito comuns, como mover uma lista, ou escolher um POI, os próprios marcadores visuais padrão podem indicar esses comandos.

#### 4.4.2 Versão do protótipo para Avaliação

As reuniões realizadas para gerar a implementação das interações de voz com o ARGuideVoz mostraram um conjunto de considerações a serem feitas. A Tabela 3 define as principais considerações que foram geradas através desse método. As colunas mostram em que sentido, se pela Interface Gráfica, ou pela Voz esse procedimento deve ser atendido.

**Tabela 3. Considerações realizadas para interface e voz**

Nº	Considerações	Interface Gráfica	Voz
1	Legendas sempre presentes	X	
2	Reconhecedor identificando variações		X
3	Legendas por visão	X	X
4	Ações de configuração sempre acessíveis		X
5	Ações comuns claras para o usuário	X	X
6	Elementos de conteúdo identificados através de texto	X	
7	Ícones devem ser trocados por palavras	X	
8	Ações gerais da aplicação devem ser identificados por palavras já conhecidas pelo público alvo		X

Cada consideração pode ser definida como:

- **Legenda presente:** A legenda para ações gerais sempre deve estar presente para o usuário.
- **Reconhecedor:** O reconhecedor deve ter uma gramática que aceita variações, não somente uma palavra em cada comando.
- **Legendas por visão:** O usuário deve ver as legendas correspondentes a cada visão corretamente.
- **Ações de configuração:** Para comandos de voz, ações de configuração devem ser sempre acessíveis.
- **Ações comuns:** Ações comuns da aplicação devem ser facilmente identificados por elementos de interface gráfica.
- **Identificar conteúdo:** Todo conteúdo disponível por toque deve ter identificação para acessar com voz.
- **Ícones por palavras:** Substituir ícones por palavras que serão usados como comandos.
- **Identificar ações gerais:** Ações gerais da aplicação devem ter comandos padrões ou fáceis de memorizar.

A Figura 16 mostra o efeito das considerações em outra parte da aplicação, de maneira mais geral.



**Figura 16.** Alterações realizadas após gerar definições gerais.

A troca de ícones realizada para as opções foi estendida para a visão de conteúdo do POI. A barra de rolagem sempre presente define a ação de subir ou descer o conteúdo que está sendo mostrado, que serve como exemplo, que mesmo sem legenda é uma ação comum para o usuário mover uma lista para cima ou para baixo, identificada esta ação pela barra de rolagem.

## 5. AVALIAÇÃO DE INTERAÇÃO DO ARGUIDEVOZ

Desenvolver um serviço de voz não é o bastante para interação de voz em uma aplicação, passa também pela escolha de palavras e ícones de interface para a aplicação. Em se tratando de RAM, o cuidado com a interação deve ser maior, por fatores como a pequena tela e a novidade na apresentação do conteúdo. Para este trabalho, duas avaliações foram feitas. Uma avaliação de usabilidade, que vai extrair informações sobre o serviço de voz na aplicação e uma avaliação de comunicabilidade, para verificar o perfil semiótico da aplicação.

### 5.1 Avaliação de Usabilidade

Para avaliação da interface em conjunto com os comandos de voz na aplicação, foram utilizados dois métodos de avaliação. Um dos métodos é um teste baseado em tarefas, que foi gravado para cada usuário participante. O segundo método utilizado é o de questionário, onde o usuário responde perguntas sobre a utilização da aplicação e sobre seu próprio perfil. Os dois métodos foram utilizados em duas das quatro avaliações realizadas em Schaeffer (2014), e nesse trabalho foi possível identificar vários pontos em que as aplicações podem ser melhoradas a partir do resultado da análise desses métodos.

O trabalho de Martínez e Bandyophyay (2014) mostra quatro avaliações de usabilidade e como podem ser utilizadas em aplicações de RAM, e baseado nele um método foi escolhido para ser utilizado nesse trabalho. O método de testes baseado em tarefas pode ser classificado como uma avaliação de laboratório neste trabalho, já que foram feitos em um ambiente com controle de ruído. Nessa avaliação, o avaliador deve conduzir os usuários a realização de tarefas pré-definidas. Essas tarefas podem ser estruturadas, se o avaliador cria as tarefas em forma de passo-a-passo para a execução, ou não estruturadas, onde o usuário escolhe o melhor jeito para realizar uma tarefa maior. Na avaliação realizada neste trabalho, foram escolhidas tarefas estruturadas, por permitir escolher interações por comandos de voz em tarefas específicas.

O uso de questionários é um método de avaliação que pode ser usado antes ou depois do teste. Neste trabalho, os questionários foram utilizados depois dos testes, para traçar o perfil dos usuários e perguntar sobre as funções e tarefas na aplicação. Esse método não pode ser utilizado sozinho em uma avaliação, já que só capta as informações que o usuário tem sobre a aplicação,

sem uma análise de especialista para confirmar. Este trabalho vai comparar os resultados dessa avaliação com os resultados do teste em laboratório.

### **5.1.1 Preparação**

Os testes de laboratório com foram realizados com 10 usuários, em um ambiente controlado, sem ruídos que atrapalhem o reconhecimento de voz. As tarefas utilizadas para o teste estão descritas na Tabela 4 abaixo. A coluna tarefa define o conjunto de tarefas realizada pelo usuário. Cada tarefa é dividida em unidades de tarefa, que são os passos mínimos necessários para realização da mesma. A complexidade de cada tarefa é definida pelo número de comandos de voz mínimos necessários para realiza-la. Para um comando de voz a tarefa foi classificada como fácil, para dois comandos foi classificada como médio e para três ou mais comandos foi classificada como difícil. As tarefas foram escolhidas com base nas ações que podem ser feitas na aplicação, em relação as visões e interações.

**Tabela 4. Lista de tarefas para os testes, divididos em Unidades de Tarefa e Complexidade.**

Número da Tarefa	Tarefa	Unidade de Tarefa	Complexidade
1.	Selecione um circuito	a. Selecione o menu circuitos	2-Médio
		b. Selecione um circuito	
2.	Encontre um POI utilizando o Navegador RA	a. Selecione o Navegador RA	3-Difícil
		b. Ajuste a visão do RA i) Aplique zoom	
		c. Selecione um POI	
3.	Encontre um POI utilizando o Mapa	a. Selecione o Mapa	2-Médio
		b. (Opcional) Ajuste a visão do mapa i) Use movimentar/zoom	
		c. Selecione um POI	
4.	Encontre um POI	a. Selecione a Lista	2-Médio
		b. (Opcional) Movimente a lista	
		c. Selecione um POI	
5.	Use o leitor de QR Code Reader para encontrar um POI	a. Selecione o leitor de QR Code	1-Fácil
		b. Leia um QRCode*	
6.	Navegue em uma imagem	a. Selecione uma imagem	2-Médio
		b. Use movimentar/zoom	

Essas tarefas cobrem as visões principais da aplicação, testando as mudanças na interface, e se os comandos de voz escolhidos são adequados para a aplicação. As perguntas feitas entre as tarefas foram as seguintes:

- Qual o nível de dificuldade desta tarefa?
- A interface ajudou você a realizar essa tarefa?
- Você achou o reconhecimento de voz satisfatório para essa tarefa?

Essas perguntas permitem avaliar três pontos importantes para a aplicação: o conforto do usuário na navegação por voz, o quanto a interface ajudou ele a realizar uma tarefa e se a navegação por voz ajudou na realização da tarefa.

Para definição de perfil as perguntas foram:

- Qual sua idade? Resposta livre.
- Você usa smartphone com frequência? Frequência de uso por dia, semana.
- Você já utilizou interação por comandos de voz em seu smartphone antes? Sim ou não.
- Você já usou alguma aplicação de Realidade Aumentada Móvel? Sim ou não.

Sobre a opinião geral do usuário:

- Se a função de interação por voz estivesse disponível para download, você usaria? Sim ou não
- O que você acha que são os pontos fortes e fracos do aplicativo? Resposta livre.
- De maneira geral, como você avalia o aplicativo? Resposta livre.

## **5.2 Método de Avaliação de Comunicabilidade**

O Método de Avaliação de Comunicabilidade (MAC) é uma técnica capaz de verificar rupturas na comunicação entre o usuário e o designer. Para tal, estabelece uma classificação quanto à qualidade da metacomunicação entre o usuário e o designer, procurando identificar como o usuário recebeu a informação. Enquanto a maioria das abordagens de avaliação de design e usabilidade se concentram em usuários e soluções oferecidas por designers, o MAC propõe uma mudança para se concentrar na conversa sobre soluções, fornecendo informações sobre problemas de usabilidade (SIECKENIUS DE SOUZA, PRATES e BARBOSA, 1999).

A comunicação deve ser realizada de forma que o usuário compreenda a informação que o designer deseja passar. O MAC avalia essa comunicação por meio de observação do usuário em um ambiente controlado, identificando falhas. Para a avaliação são registrados vídeos e anotações do avaliador, os quais fornecem informações para identificar as interpretações do usuário, as intenções de comunicação e rupturas de comunicação durante a interação.

As rupturas da comunicação observadas são etiquetadas por um conjunto de enunciados, descrevendo falhas completas, parciais e temporárias.



Para realizar o MAC são definidas cinco atividades a serem executadas (REIS e PRATES, 2012):

- **Preparação:** Consiste em inspecionar os signos (estáticos, dinâmicos e metalinguísticos), definir tarefas, preparar material, definir perfil dos participantes e executar um teste piloto;
- **Coleta de dados:** observação e registro, em vídeo e áudio da interação de cada participante;
- **Interpretação:** Os vídeos são etiquetados individualmente;
- **Consolidação dos resultados:** As etiquetas são interpretadas e é formado um perfil semiótico, no qual é feito o significado da metacomunicação designer-usuário;
- **Relato dos resultados:** É relatada a solução de IHC.

A partir das rupturas de comunicabilidade os problemas de interação podem ser identificados e explicados, bem como oferece suporte para realizar o re-design (SIECKENIUS DE SOUZA, LEITÃO, *et al.*, 2006). Em sistemas interativos, como o sistema avaliado neste trabalho, é fundamental que a comunicação entre o usuário e o sistema seja efetivo. A utilização do MAC apresenta ao designer o nível de interpretação que os usuários estão tendo do sistema.

Antes de tudo, foi necessário estabelecer o perfil dos usuários que iriam realizar os testes de interação com o aplicativo. Para isso foram escolhidas 13 pessoas que seguem o perfil abaixo:

- Faixa etária: 18 a 30 anos;
- Escolaridade: Estudantes em geral;
- Experiência: Faz uso de smartphone, tablet ou outra tecnologia semelhante;
- Status: Não está no ponto de interesse desejado;
- Atitude: Deseja encontrar um determinado ponto de interesse;
- Conhecimento: Não sabe como chegar em todos os locais;

Com a determinação do método a ser utilizado, e do perfil dos participantes, foi necessário determinar as tarefas que seriam realizadas por estes usuários do ARGuide, e foram definidas cinco tarefas a serem realizadas, descritas na Tabela 5.

**Tabela 5. Tarefas propostas para avaliação**

	<b>Tarefas</b>
1	Navegar e interagir com o mapa.
2	Visualizar e interagir com a lista de Pontos de Interesse cadastrados (POI) com seus nomes e fotos.
3	Utilizar a funcionalidade de Realidade Aumentada para descobrir POIs próximos.
4	Traçar uma rota no aplicativo a um determinado POI.
5	Utilizar leitor de QR Code.

Essas tarefas foram escolhidas visando fazer com que o usuário testasse as principais funcionalidades que já estão prontas do aplicativo. Para a realização das tarefas levantadas, foi construído um cenário a ser seguido pelo usuário durante os testes do método.

O cenário elaborado consistia em um possível usuário do aplicativo chamado João Maria que possui 21 anos, recém aprovado no curso de matemática da Universidade Federal do Pará (UFPA). No cenário, o João toma conhecimento do aplicativo ARGuideVoz e que serve para a localização de pontos de interesse da UFPA. Portanto ele decide baixar e testar o aplicativo. Quando já possui o aplicativo, João começa a usar a ferramenta, testando suas funcionalidades. Realizando as tarefas determinadas na tabela 5, João vai descobrindo como funciona o ARGuideVoz e percebendo todo o potencial da aplicação para as pessoas que frequentam a universidade, porém que não conhecem tudo que ela tem a lhes oferecer.

Para o desenvolvimento dos testes dos usuários e posteriormente para a avaliação dos resultados, foi necessário um ambiente de teste adequado ao método. Foi utilizada uma sala fechada e refrigerada onde ficava apenas o usuário que iria testar o aplicativo, e o guia da avaliação, que explicava o cenário a ser seguido e realizava as filmagens.

Também foram utilizados um tablet de 7 polegadas com o aplicativo ARGuideVoz instalado, um celular posicionado em um tripé a frente do usuário, para realizar a gravação das expressões, enquanto o guia, após a explicação do cenário, realizava uma filmagem lateral do participante através de um smartphone. O ambiente pode ser visto na Figura 17 abaixo.

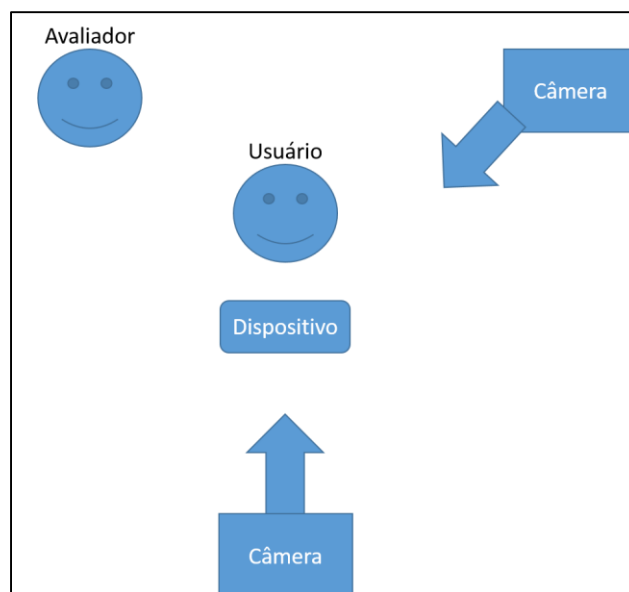


Figura 17. Participante durante o cenário de testes

## 5.3 Resultados

### 5.3.1 Avaliação de Usabilidade

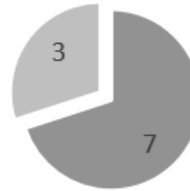
Os resultados da avaliação foram feitos com base nas respostas dos questionários e análise dos vídeos dos testes. Para a análise dos vídeos foram extraídos o número de erros de reconhecimento de voz em cada unidade de tarefa e número de erros em que a GUI não ajudou o usuário a realizar a tarefa. Os erros de reconhecimento são os erros em que o usuário falou algo e o reconhecedor não respondeu com uma ação, erros que o usuário fala algo e o reconhecedor não responde em tempo suficiente e o usuário entra com outro comando, ou erro em que a palavra é interpretada erroneamente. Para o perfil do usuário, a Figura 18 traz dois gráficos com as respostas tiradas dos questionários.

Você já usou alguma aplicação de Realidade Aumentada Móvel?



■ Sim ■ Não

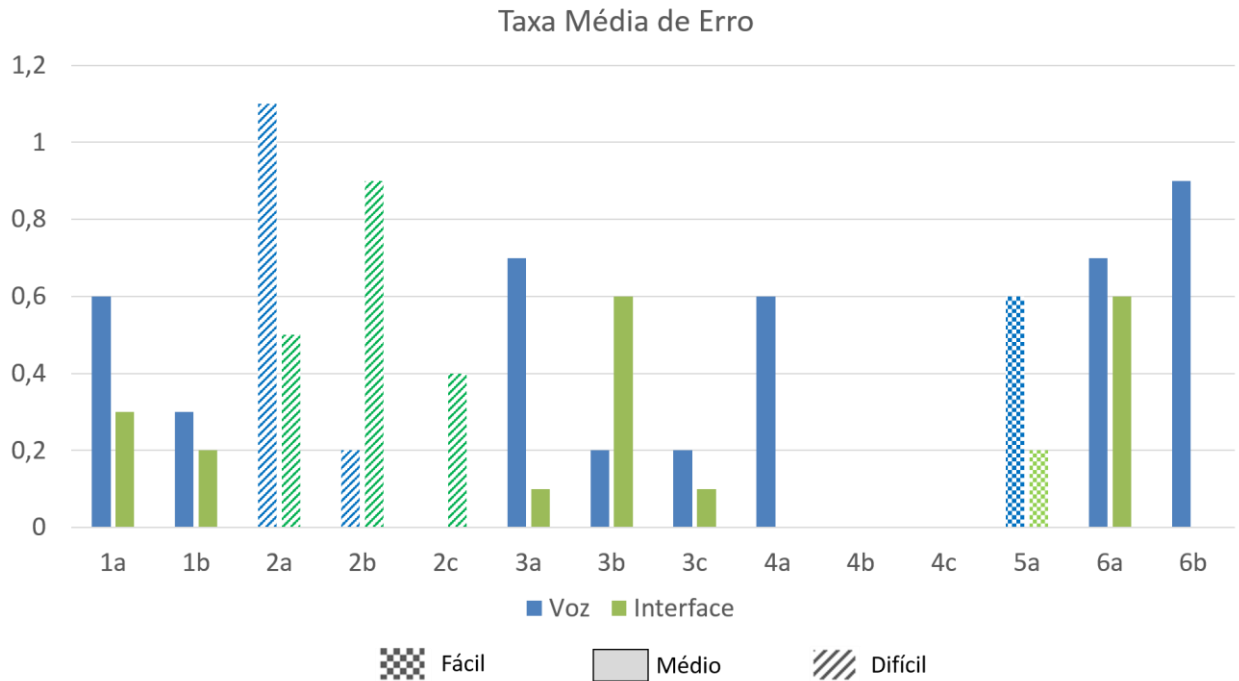
Você já utilizou interação por comandos de voz em seu smartphone antes?



■ Sim ■ Não

**Figura 18. Experiência dos usuários com RA e Voz**

A maioria dos usuários já tem experiência com interfaces de voz e metade deles já utilizou pelo menos uma vez uma aplicação de RA. A idade dos usuários está na faixa entre 21 e 42, e todos usam smartphone diariamente. A familiaridade que os usuários já têm com a tecnologia sendo testada pode ter sido um fator importante para a realização de todas as tarefas por todos os usuários. Não houve desistência de nenhum usuário. A Figura 19 mostra a média dos resultados para as tarefas realizadas pelos usuários, divididos em suas respectivas unidades de tarefas.



**Figura 19. Resultados dos testes com Usuários**

A tarefa que teve maior erro de reconhecimento de voz foi a tarefa 2a, para trocar da visão do mapa para a visão do navegador RA. A composição da palavra ("Navegador RA"), que poderia ser falada como a composição de uma palavra mais duas letras seguidas. O reconhecedor teve problemas em associar essa composição e grande parte dos usuários acabou usando apenas "navegador" como palavra de seleção. A mesma dificuldade pode ser vista na tarefa 5a, onde a tarefa de chamar o leitor de QR Code pede somente um comando de voz, mas tem uma média alta de erros de reconhecimento em comparação com erros relacionados à interface.

A tarefa em que se percebeu maior dificuldade na interface foi a tarefa 2b, para aplicar zoom no navegador RA. Usuários deviam aplicar o zoom e aumentar o raio do alcance do navegador RA para visualizar o POI. Usuários tiveram dificuldades para identificar o *widget* para aumentar o zoom, e quando identificaram não conseguiram aumentar o raio com facilidade, e alguns usuários tiraram todo o alcance da aplicação para poder entender que deviam aumentar o raio para ver os POIs mais distantes.

Erros de programação foram encontrados pelos usuários, que influenciaram o resultado da Tarefa 3, de navegação do mapa. Uma palavra para dar pan no mapa não estava sendo reconhecido, "acima", e os usuários tiveram que usar um sinônimo ("cima") para fazer a navegação.

Um erro de troca de contexto entre o POIView e o leitor de QR Code foi encontrado também, que gerou alguns problemas para os usuários, como encerrar a execução do programa inesperadamente. E dois usuários perceberam que ao escolher uma mídia no POIView depois de vir do contexto da Lista, o número para seleção da mídia selecionava um item da lista.

Problemas relatados pelos usuários no questionário, após o teste e percebidos durante a análise dos vídeos e as melhorias relacionadas estão na Tabela 6.

Dois usuários expressaram ao final dos questionários que a aplicação apesar de ter várias visões, mantém alguns comandos iguais, como trocar as visões e selecionar os POIs, o que fez que eles aprendessem com a aplicação, e não decorassem o que cada visão fazia separadamente. Nenhum usuário usou a ajuda, apesar de ser apresentada como uma função no treinamento e ser incentivado a usá-la durante o teste.

**Tabela 6. Problemas encontrados e Melhorias sugeridas.**

Problemas encontrados	Melhorias sugeridas
Algumas palavras no plural não têm reconhecimento equivalente no singular, e vice e versa.	Acrescentar mais palavras que os usuários possam falar para determinados comandos.
<ul style="list-style-type: none"> <li>- O ícone de microfone na legenda parece um botão para iniciar a navegação por voz.</li> <li>- A troca de circuitos não é clara para o usuário.</li> <li>- Um POI fora do circuito não foi selecionado apesar de um usuário já saber o nome dele.</li> </ul>	<ul style="list-style-type: none"> <li>- Dar mensagens de erro e acerto de algumas tarefas, como a seleção de circuitos e escolha de uma visão.</li> <li>- Uma opção de voltar para ações com comandos parecidos.</li> </ul>
As abas não indicam claramente que a visão pode ser trocada através da voz.	Fazer o separador de abas mais visível e as abas mais separadas do logo do ARGuideVoz.
O navegador RA utiliza uma câmera do mesmo modo que o leitor de QR Code, alguns usuários confundiam suas funções.	<ul style="list-style-type: none"> <li>- Indicar mais claramente na seleção de visão entre o QR Code e o Navegador RA qual visão o usuário está usando.</li> <li>- Remodelar a interface do navegador RA. A Figura 20(a) mostra o aumento do raio de visão do navegador e Figura 20 (b) o radar de navegação, onde cada ponto amarelo é um POI do circuito selecionado. Esses pontos precisam de mais enfoque, já que a maioria dos erros encontrados em razão da interface poderiam ser corrigidos se esses dois pontos tivessem suas funções mais claras para o usuário.</li> </ul>
Para o mapa, os comandos de voz estavam simulando a função de Pan como o toque realizado no mapa, onde se puxa ele para direção contrária para ele ir para direção desejada. Essa interação com a voz confundiu alguns usuários e foi bem recebida por outros.	Permitir customização de algumas ações por voz, como direção, zoom, e acesso a atalhos.

As unidades de tarefa 2c, 3c e 4c eram a mesma (Selecionar um POI) e foram as melhores recebidas pelo usuário. A seleção de um POI é uma interação simples, basta saber o nome do POI, e pode usar ele como atalho de navegação. Dois usuários destacaram a velocidade de reação e o feedback visual dados pela ferramenta na seleção de um POI.

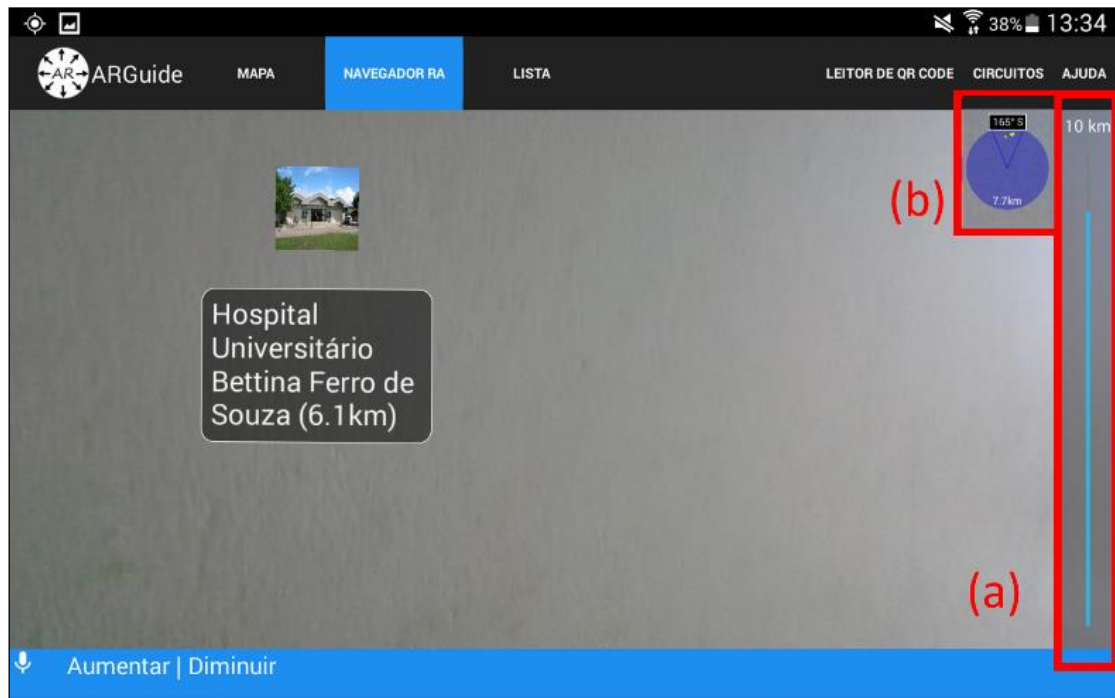


Figura 20. Interface do Navegador RA - (a) Controlador de raio (b) Radar com POIs

### 5.3.2 Método de Avaliação de Comunicabilidade

Após a etiquetagem dos vídeos, foi realizada uma interpretação do conjunto de etiquetas, na qual foram atribuídos significados.

- Quem o designer pensa ser o usuário do sistema por ele projetado?

Os usuários têm faixa etária entre 18 e 30 anos de idade, são estudantes em geral, gostam de tecnologia e fazem uso de smartphones, tablets ou similares várias vezes ao dia.

- Quais são os desejos e as necessidades dos usuários, segundo o designer?

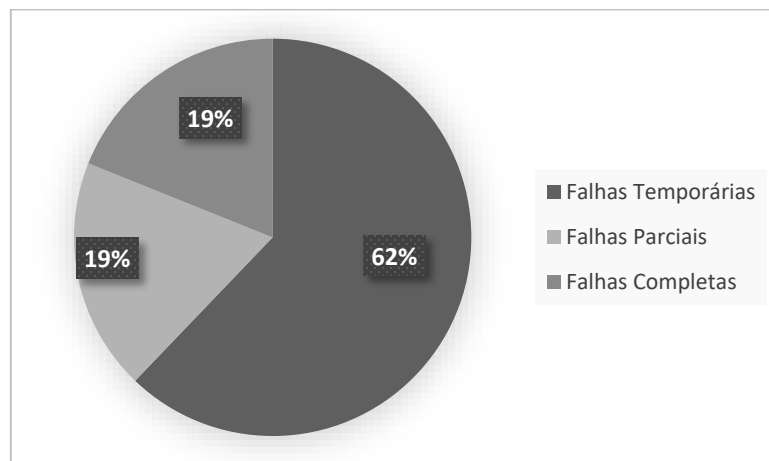
Os usuários não se encontram no POI desejado e não sabem como chegar até ele, e querem descobrir onde está ou como podem chegar até o POI. O usuário também deseja usar uma aplicação com sua interação feita através da voz.



- Na visão do designer, quais as preferências dos usuários de acordo com seus desejos e necessidades?

Os usuários demandam ilustrações em palavras ou ilustrações de fácil percepção e compreensão, do passo-a-passo a ser seguido desde compreender onde ele ou ela está, seguindo por descobrir como localizar o POI desejado até identificar a forma de chegar até ele.

O designer projetou para os usuários um aplicativo de RAM para o sistema operacional Android, com o intuito de localizar POI's através de Interação por Voz ou por gestos de toque. De acordo com os testes realizados utilizando-se o MAC com 13 voluntários, observaram-se 37 rupturas de comunicabilidade, totalizando uma média de 2,84 rupturas por voluntário. Dessas rupturas, 23 classificam-se como Falhas Temporárias (FT's), 7 como Falhas Parciais (FP's) e 7 como Falhas Completas (FC's), representando 62%, 19% e 19% do total, respectivamente, conforme Figura 21. As etiquetas correspondentes a essas falhas são mostradas na Tabela 7.



**Figura 21. Resultados do MAC baseado em falhas encontradas**

**Tabela 7. Etiquetagem total do MAC**

Falha	Aspecto semiótico	Característica	Etiqueta	Quant.
Completa	O usuário realiza uma semiose malsucedida e não inicia outra para obter o resultado esperado.	Não percebe que foi malsucedido.	Para mim, está bom.	4
		Percebe que foi malsucedido, mas não continua.	Desisto.	3
Parcial	O usuário abandona uma semiose e inicia outra para obter o resultado esperado.	Prefere outra solução ao invés da solução de IHC proposta.	Não, obrigado.	3
		Não entende solução de IHC proposta.	Vai de outro jeito.	4
Temporária	O usuário interrompe a semiose.	Não encontra expressão apropriada para sua comunicação.	Cadê?	12
		Não entende expressão do sistema.	Ué, o que houve?	1
		Não consegue realizar sua comunicação.	E agora?	1
	O usuário percebe que a comunicação não foi bem-sucedida.	Percebeu que havia “falado” no contexto errado.	Onde estou?	4
		Percebeu que havia “falado” errado.	Epa!	0
		Não obtém o resultado depois de conversar com o sistema, alternando outras falas.	Assim não dá.	1
	O usuário tenta compreender a comunicação.	Através da meta comunicação implícita.	O que é isto?	3
		Através da metacomunicação explícita.	Socorro!	1
		Realiza vários testes sobre o que o sistema comunicou.	Por que não funciona?	0

Após a das etiquetas de Comunicabilidade, foi possível avaliar a existência de tipos de falhas mais frequentes em determinadas tarefas:

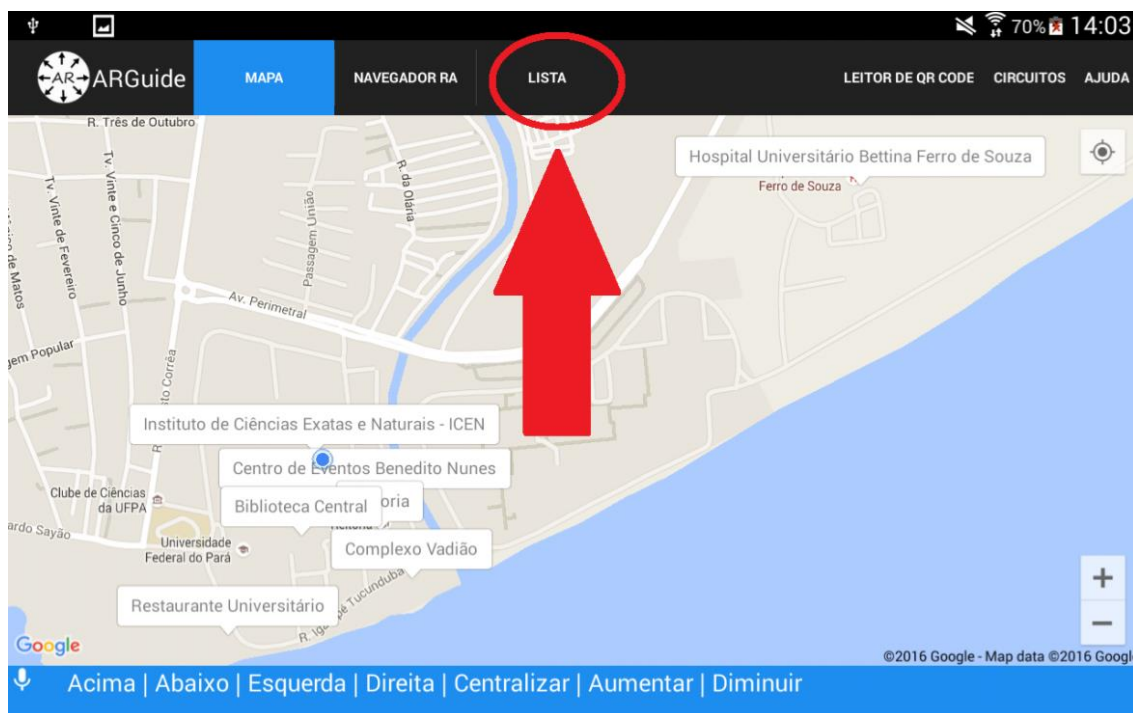
**Falhas Completas (FC’s):** Foram o tipo de falhas menos frequentes, onde existe uma interrupção total e malsucedida da semiose pelo usuário. Dentre as 7 FC’s, 4 delas, sendo 3 etiquetas “Desisto” e 1 das “Para mim, está bom” estavam relacionadas à tarefa “Traçar uma rota no aplicativo a um determinado POI”. Na visão do aplicativo onde os usuários se encontravam durante

a tarefa, a interface possui a opção para traçar a rota a partir do POI, mas alguns usuários não a identificaram, o que demonstra que essa interface possui um Problema Operacional, como mostra a Figura 22. Um maior destaque visual para a opção “ROTA” ou a troca da legenda para “TRAÇAR ROTA” seriam sugestões de solução para o problema.



**Figura 22. Visão durante a atividade “Traçar rota a um determinado POI”**

**Falhas Parciais (FP’s):** Ocorreram na mesma frequência que as FC’s com 7 ocorrências, o usuário abandona uma semiose mas inicia outra com o mesmo propósito. Dentre as 7 FP’s, 3 delas com as etiquetas “Vai de outro jeito” estavam relacionadas à tarefa “Visualizar a lista de Pontos de Interesse cadastrados (POI) com seus nomes e fotos”. Na visão do aplicativo onde os usuários se encontravam durante a tarefa, a interface possui a opção para acessar a visão da lista, mas alguns usuários não a identificaram e optaram por outras formas de visualizar um POI que não fosse pela lista, o que demonstra que essa interface possui um Problema Operacional, como mostra a Figura 23. Um maior destaque visual para a opção “LISTA” ou a troca da legenda para “LISTA DE PONTOS” seriam sugestões de solução para o problema.



**Figura 23. Visão durante a atividade “Visualizar a lista de Pontos de Interesse”.**

**Falhas Temporárias (FT's):** Foram o tipo de falhas mais frequentes, com 23 ocorrências. Na FT, o usuário pode interromper a sua semiose, perceber que o seu ato comunicativo foi malsucedido ou procurar compreender o ato comunicativo do sistema. Neste tipo de falhas, não houve uma atividade que se destacou das demais em relação à frequência de falhas, mas todas as falhas evidenciaram Problemas Operacionais semelhantes, sendo que os mais evidentes foram a falta de visibilidade ou de compreensão dos elementos gráficos referentes ao “radar” que mostra a existência de POI's a uma certa distância (Figura 24.A), à barra de regulagem da distância máxima dos POI's (Figura 24.B) e da visualização de POI's que não se encontram na área visível da câmera (Figura 24.C). Uma resposta visual aos itens A e B e uma referência visual ao item C seriam sugestões de solução para os problemas citados.

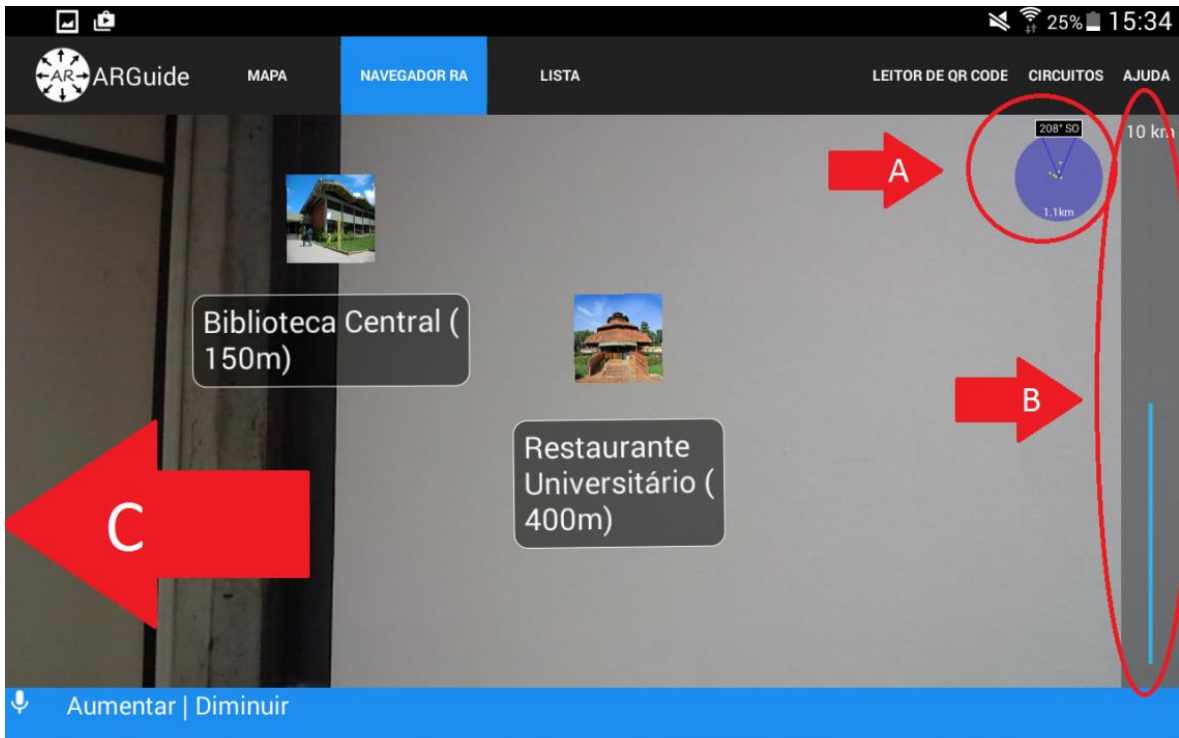


Figura 24. Visão do Navegador de Realidade Aumentada

## 5.4 Boas Práticas

O conjunto de boas práticas pode ser encontrados nas Seções 4 e 5, boas práticas de desenvolvimento e boas práticas de IHC, respectivamente. A Seção 4 possui modelos de arquitetura de RAM, modelos de reconhecimento de voz, e podem ser vistos na Seção 4 sem perdas. Os resultados de IHC estão espalhados entre as Seções 4 e 5, e estão organizados na Tabela 8 e na Tabela 9, resumindo as boas práticas de IHC, relacionadas a Interface Gráfica e Interface de Voz utilizadas na aplicação.

Tabela 8. Conjunto de boas práticas em Interface Gráfica

Interface Gráfica
Mensagens de erro e acerto
Deixar o separador de visões bem visível
Deixar claro as interfaces das visões que usam câmeras
Interface da visão do navegador RA sem Radar e <i>Slider</i> de alcance
Legendas sempre presentes e contextualizadas

Ícones devem ser sempre rotulados com palavras significativas
Feedback visual na aplicação
Evitar interação com elementos que somem
Indique visualmente a interação por voz
Não usar a área da câmera na visão aumentada para legendas

**Tabela 9. Conjunto de boas práticas em Interface de Voz**

<b>Interface de Voz</b>
Atender as variações de palavras selecionadas
Opção mais clara para voltar ações
Permitir customização de palavras pelo usuário
Acesso a ajuda da aplicação em qualquer contexto
Ações comuns com palavras já conhecidas pelo usuário
Evitar formalidade
Evitar palavras compostas para uma ação
Não usar palavras muito parecidas
Discretizar funções contínuas, como as de movimentação e zoom

## 6. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

### 6.1 Considerações Finais

O uso de linguagem natural para interação com conteúdo é um caminho que está se tornando mais simples com a popularização de aplicações e serviços que utilizam esses tipos de interação. As contribuições mais importantes geradas por esse trabalho são as considerações extraídas da implementação do serviço de reconhecimento de voz em uma aplicação de RAM, as de desenvolvimento e IHC. Essas considerações podem ser estendidas em futuras iterações do ARGuide-Voz, com mais visões ou funções que poderão integrar a aplicação.

A disposição de interface utilizada foi baseada em uma aplicação anterior, e foi alterada para suportar comandos de voz. As funcionalidades não foram alteradas, somente o modo que são acessadas, pois foi preciso mudar a arquitetura da aplicação e a interface gráfica para suportar o serviço de voz de forma que não atrapalhasse a navegação do usuário. Contribuições foram feitas no sentido de desenvolvimento e interface gráfica e voz, e um conjunto de boas práticas foi extraído.

Um modelo de arquitetura foi gerado por esse trabalho, os aspectos de implementação apresentados procuram facilitar o desenvolvimento de aplicações RAM que tem como objetivo integrar um serviço de voz. Uma arquitetura e diagramas foram apresentados para suportar esse modelo. A aderência a esses aspectos deve facilitar que aplicações de RAM possam melhorar suas interações e que os usuários não se sintam confusos para escolher e lembrar de variadas interações. A falta de padrões definidos para desenvolvimento em RAM é um desafio a ser superado e ter um padrão definido de interface de comandos de voz pode ajudar com essa dificuldade.

Quanto às interações, várias observações foram realizadas, em relação ao uso de visões, disposição de interface e escolha de comandos de voz. Como um resultado proeminente em sentido

de interação, a seleção de POI foi a melhor interação encontrada com comandos de voz. Usar o nome do POI como atalho para a interação facilitou o uso da aplicação nos testes realizados. O uso de atalhos, como já usado em alguns assistentes pessoais também se mostrou extremamente positivo para o usuário.

O navegador RA mostrou-se problemático em ambos os testes, e precisa ser remodelado. Vários problemas foram encontrados, os *widgets* não estão claros para os usuários, e mesmo os usuários experientes com aplicações de RAM tiveram problemas para interagir com o navegador. Novos modos de expressar a orientação geoespacial do usuário, o uso da câmera e a disposição de interface precisam ser estudados, para que os usuários possam utilizar melhor o componente mais inovador da ferramenta.

A continuidade desse trabalho será aplicar as mudanças extraídas das avaliações. Estudar novas formas de disposição para ao conteúdo aumentado, dar opções de customização, e mudar elementos de interface conforme vistos nos resultados dos testes realizados. Após isso, é preciso reavaliar as mudanças através de testes. Outros métodos de avaliação podem ser utilizados também, para comparar e validar com os testes realizados para essa versão da ferramenta. O teste com outros reconhecedores de voz pode ser feito para avaliar a acurácia dos reconhecedores na aplicação e como podem afetar a experiência do usuário. Espera-se poder usar um reconhecedor *off-line*, que funcione somente no dispositivo, sem precisar de conexão com Internet, e ver se a diferença de latência pode ajudar o usuário.

## **6.2 Trabalhos Futuros**

Como um trabalho futuro ainda em linguagem natural, o acréscimo de um sintetizador de voz, que possa atender a várias situações, falando para o usuário o contexto que ele se encontra, narrar o conteúdo que pode ser encontrado em determinado POI, usar o mapa para auxiliar na navegação, entre outras ações. Aplicações que usam comandos de voz costumam ter uma resposta sonora também, o que ajuda o usuário a “conversar” com a aplicação, e também abre espaço para novos usuários possuem alguma dificuldade visual. Uma resposta diferente de uma resposta visual exige novos testes, tanto para a escolha do sintetizador, quanto ao seu desempenho na aplicação, configurando um trabalho futuro ainda levando em conta linguagem natural.

No sentido de visões, usar visão computacional para detectar marcadores além do QR Code é uma extensão possível para esse trabalho. O uso de marcadores para posicionar mídias pode dilu-



ir ainda mais a necessidade de uso das mãos, e tornar o conteúdo mais dinâmico para usar com interações não convencionais. Outra visão possível é uma visão que utiliza o mapa juntamente com o navegador RA, para a referência geográfica do usuário ficar completa, em uma tela só.

Como uma adição as interações de linguagem natural, o uso de gestos pode ser utilizado além do toque em tela. Juntamente com visão computacional, gestos podem ser reconhecidos para realizar ações na aplicação, para acessar menus, utilizar atalhos ou simplesmente interagir com o conteúdo aumentado. Um objeto em 3D misturado com uma cena real pode ter interação em tela, e também uma interação direta utilizando a câmera e as mãos do usuário, com a voz para mudar os modos de interação, por exemplo. Técnicas para uso dessa interação teriam que ser feitas, assim como testes e avaliações.

As sugestões de melhorias apresentadas, os modelos criados e as análises apontam que o uso da voz pode ser interessante para o usuário, mas tem um caminho para ser percorrido para que a interação seja mais popular e fácil de utilizar. Este trabalho pretendeu ser um pequeno passo rumo à aceitação dessa interação em um campo tão inovador quanto a Realidade Aumentada Móvel.

## REFERÊNCIAS

Apple Inc. (2016). *iOS - Siri - Apple (BR)*. Retrieved Julho 24, 2016, from Apple: <http://www.apple.com/br/ios/siri/>

ARAUJO, T., CARNEIRO, N., MIRANDA, B., SANTOS, C. G., & MEIGUINS, B. S. (2015). Aplicações Android de Realidade Aumentada em Arquitetura Extensível, Flexível e Adaptável. *XI Simpósio Brasileiro de Sistemas de Informação* (pp. 63-70). Goiás: Instituto de Informática da Universidade Federal de Goiás.

AWILA. (2014). Retrieved junho 2015, from AWILA Augmented Technologies: <http://www.aug-tech.co.uk/augmented-reality/>

Blinov, M., Deru, M., & Sonntag, D. (2011). Mobile thumb interaction and speech. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM.

Blippar. (2015). Retrieved Junho 2015, from Blippar: <https://blippar.com/en/>

Carter, M., Allison, F., Downs, J., & Gibbs, M. (2015). Player Identity Dissonance and Voice Interaction in Games. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. ACM.

Cawood, S., & Fiala, M. (2008). *Augmented Reality: A Practical Guide* (1ª ed.). Dallas, Texas: Pragmatic Bookshelf.

D'Antoni, L., Dunn, A., Jana, S., Kohno, T., Livshits, B., Molnar, D., et al. (2013). Operating system support for augmented reality applications. *Proceedings of the 14th USENIX conference on Hot Topics in Operating Systems (HotOS'13)*.

de la Nube Aguirre Brito, C. (2015). Augmented reality applied in tourism mobile applications. *Second International Conference on eDemocracy & eGovernment (ICEDEG)*, (pp. 120-125).

- Delail, B. A., Weruaga, L., & Zemerly, M. J. (2012). CAViAR: Context Aware Visual Indoor Augmented Reality for a University Campus. *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 03* (pp. 286-290). Washington: IEEE.
- Doswell, J. (2006). Augmented Learning: Context-Aware Mobile Augmented Reality Architecture for Learning. *Sixth International Conference on Advanced Learning Technologies*, (pp. 1182-1183).
- Doswell, J. (2006). Context-Aware Mobile Augmented Reality Architecture for Lifelong Learning. *Sixth International Conference on Advanced Learning Technologies*, (pp. 372-374).
- Doswell, J., Blake, M., & Butcher-Green, J. (2006). Mobile Augmented Reality System Architecture for Ubiquitous e-Learning. *Fourth IEEE International Workshop on Wireless, Mobile and Ubiquitous Technology in Education*, (pp. 121-123).
- Feng, D., Meng, D., Zhang, Y., & Weng, D. (2013). Time Machine: A Mobile Augmented Reality System for Tourism Based on Coded-Aperture Camera. *2013 IEEE 10th International Conference and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC) Ubiquitous Intelligence and Computing*, (pp. 502-506).
- Giesler, B., Salb, T., Steinhaus, P., & Dillmann, R. (2004). Using augmented reality to interact with an autonomous mobile platform. *Proceedings of IEEE International Conference on Robotics and Automation*, (pp. 1009-1014).
- Google Inc. (2016). *The Google App - Voice Search, Answers and Assistance*. Retrieved Julho 24, 2016, from Google: <https://www.google.com/search/about/>
- Google Inc. (2015). *Android Developers*. Retrieved junho 1, 2015, from Fragments API Guide: <http://developer.android.com/guide/components/fragments.html>
- Google Inc. (2015). *Google Developers*. Retrieved Junho 1, 2015, from Google Maps Android API: <https://developers.google.com/maps/documentation/android/?hl=pt-br>
- Goose, S., Sudarsky, S., Zhang, X., & Navab, N. (2002). SEAR: towards a mobile and context-sensitive speech-enabled augmented reality. *2002 IEEE International Conference on Multimedia and Expo*, (pp. 849-852).

- Igarashi, T., & Inami, M. (2015). Exploration of Alternative Interaction Techniques for Robotic Systems. *Computer Graphics and Applications*, 35, 33-41.
- JIANG, L.-F., LU, G.-Z., XIN, Y.-W., & LI, Z.-L. (2006). Design and implementation of speech interaction system in pervasive computing environment. *International Conference on Machine Learning and Cybernetics*. IEEE.
- Larsen, B. L., Jensen, K. L., Larsen, S., & Rasmussen, M. H. (2007). Affordance in mobile speech-based user interaction. *Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*. ACM.
- Layar. (2015). *Layar*. Retrieved Junho 2015, from <https://www.layar.com/>
- Lee, K. B., & Grice, R. A. (2006). The Design and Development of User Interfaces for Voice Application in Mobile Devices. *IEEE International Professional Communication Conference*, (pp. 308-320).
- Markov-Vetter, D., & Stadt, O. (2013). A pilot study for Augmented Reality supported procedure guidance to operate payload racks on-board the International Space Station. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, (pp. 1-6).
- Martínez, H., & Bandyopadhyay, P. (2014). *Analysis of Four Usability Evaluation Methods Applied to Augmented Reality Applications*. Tampere University of Technology e University of Helsinki, Finlândia.
- Martínez, H., Skournetou, D., Hyppölä, J., Laukkanen, S., & Heikkilä, A. (2014). Drivers and Bottlenecks in the Adoption of Augmented Reality Applications. *Journal on Multimedia Theory and Applications*, pp. 20-26.
- Metaio. (2015). *Junaio*. Retrieved Junho 2015, from <http://www.junaio.com/>
- Microsoft Inc. (2016). *Cortana - Meet your personal Assistant - Microsoft USA*. Retrieved Julho 24, 2016, from Microsoft: <https://www.microsoft.com/en-us/mobile/experiences/cortana/>
- Muller, J., Krapichler, C., Nguyen, L. S., Englmeier, K.-H., & Lung, M. (1998). Speech interaction in virtual reality. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE.

- Munteanu, C., Whittaker, S., & Rajput, N. (2013). We need to talk: HCI and the delicate topic of spoken language interaction. *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM.
- Nielsen, J. (1994). Heuristic Evaluation. In J. Nielsen, R. Mack, & D. Cerra (Ed.), *Usability Inspection Methods*. New York, NY: Katherine Schowalter.
- Olsson, T., & Salo, M. (2011). Online User Survey on Current Mobile Augmented Reality Applications. *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality* (pp. 75-84). Washington: IEEE Computer Society.
- Park Associates. (2016). *Almost 40% of U.S. smartphone owners use voice recognition software*. Retrieved Julho 4, 2016, from Park Associates: <http://www.parksassociates.com/blog/article/pr-01272016>
- Park, D., Nam, T.-J., & Shi, C.-K. (2006). Designing an immersive tour experience system for cultural tour sites. *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)* (pp. 1193-1198). New York: ACM.
- Pfleging, B., Schneegass, S., & Schmidt, A. (2012). Multimodal interaction in the car: combining speech and gestures on the steering wheel. *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM.
- Potel, M. (1996). *MVP: Model-View-Presenter The Taligent Programming Model for C++ and Java*. Taligent Inc.
- Poushter, J. (2016). *Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies*. Retrieved Julho 4, 2016, from PewResearchCenter: <http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/>
- Pulli, P., Pyssysalo, T., Metsavainio, J.-P., & Komulainen, O. (1998). CyPhone-experimenting mobile real-time telepresence. *Proceedings of 10th Euromicro Workshop on Real-Time Systems*, (pp. 10-19).
- Pyssysalo, T., Repo, T., Turunen, T., Lankila, T., & Röning, J. (2000). Cy-Phone—bringing augmented reality to next generation mobile phones. *Proceedings of DARE 2000 on Designing augmented reality environments (DARE '00)*, (pp. 11-21). New York.

- Reis, S., & Prates, R. (2012). An initial analysis of communicability evaluation methods through a case study. *CHI '12 Extended Abstracts on Human Factors in Computing Systems(CHI EA '12)* (pp. 2615-2620). New York: ACM.
- SANTOS, C. G., CARNEIRO, N., LIMA, B., MEIGUINS, A., & MEIGUINS, B. S. (2014). Considerações de Projeto e Implementação de uma Aplicação Android de realidade aumentada móvel. *Conferência Íbero-Americana de Computação Aplicada*, (pp. 119-126). Porto.
- SANTOS, C. G., CARNEIRO, N., MIRANDA, B., & MEIGUINS, B. S. (2014). Uma Aplicação de Realidade Aumentada Móvel para Ambientes Indoor e Outdoor. *Livro dos Anais do XV Workshop de Realidade Virtual e Aumentada* (pp. 120-126). Marília: Caroline Kraus Luvizotto.
- Schaeffer, S. E. (2014). *Usability Evaluation for Augmented Reality*.
- Sieckenius de Souza, C., Leitão, C. F., Prates, R. O., & J., d. S. (2006). The semiotic inspection method. *Proceedings of VII Brazilian symposium on Human factors in computing systems*. ACM.
- Sieckenius de Souza, C., Prates, R. O., & Barbosa, S. D. (1999). A method for evaluating software communicability. *Monografias em Ciência da Computação*, (pp. 11-99).
- Srinivasa, K. G., Jagannath, S., Nidhi, P. S., Tejesh, S., & Santhosh., K. (2014). Augmented Reality Application: Cloud based Augmented Reality Android application to "know your world better". *In Proceedings of the 6th IBM Collaborative Academia Research Exchange Conference (I-CARE) on I-CARE 2014 (I-CARE 2014)* (p. Artigo 15). New York: ACM.
- Sweetser, E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. *Actes du 16e Congrès International des Linguistes*. Elsevier.
- Tang, L., Ang, K. S., Amirul, M., Bin Mohamed Yusoff, M., Tng, C. K., Bin Mohamed Alyas, M., et al. (2015). Augmented reality control home (ARCH) for disabled and elderlies. *IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, (pp. 1-2).
- Teixeira, A. (2013). Speech-Centric Multimodal Interaction for Easy-To-Access Online Services – A Personal Life Assistant for the Elderly. *5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion – DSAI*.

Vidakis, N., Syntychakis, M., Triantafyllidis, G., & Akoumianakis, D. (2012). Multimodal Natural User Interaction for Multiple Applications: The Gesture – Voice Example. *International Conference on Telecommunications and Multimedia – TEMU*.

Vullamparthi, A., Nelaturu, S., Mallaya, D., & Chandrasekhar, S. (2013). Assistive Learning for Children with Autism Using Augmented Reality. *IEEE Fifth International Conference on Technology for Education (T4E)* (pp. 43-46). IEEE.

Wikitude Inc. (2015). *wikitude See more*. Retrieved Junho 1, 2015, from Wikitude SDK: <http://www.wikitude.com/products/wikitude-sdk/>

Xia, L., Kai, K., W., X., & Dan, W. (2010). Research and Design of the “Voice-Touch-Vision” Multimodal Integrated Voice Interaction in the Mobile Phone.