

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE

Arikleyton de Oliveira Ferreira

*MHNSS: Um Middleware para o Desenvolvimento de Aplicações
Móveis com Interações Baseada na Fala*

São Luís
2014

Arikleyton de Oliveira Ferreira

*MHNSS: Um Middleware para o Desenvolvimento de Aplicações
Móveis com Interações Baseada na Fala*

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Eletricidade da Universidade Federal do Maranhão como requisito parcial para a obtenção do grau de MESTRE em Engenharia de Eletricidade com área de concentração em Ciência da Computação.

Orientador: Francisco José da Silva e Silva

Prof. Doutor em Ciência da Computação

Universidade Federal do Maranhão

São Luís

2014

Ferreira, Arikleyton de Oliveira

MHNSS: Um Middleware para o Desenvolvimento de Aplicações Móveis com Interações Baseada na Fala / Arikleyton de Oliveira Ferreira. – São Luís, 2014.

102 f.

Orientador: Francisco José da Silva e Silva.

Impresso por computador (fotocópia).

Dissertação (Mestrado) – Universidade Federal do Maranhão, Programa de Pós-Graduação em Engenharia de Eletricidade. São Luís, 2014.

1. Aplicações móveis - Acessibilidade. 2. Interações baseada na Fala. 3. Sistemas de diálogo falado I. Título.

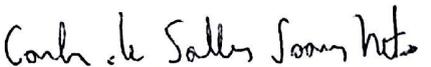
CDU 004.5

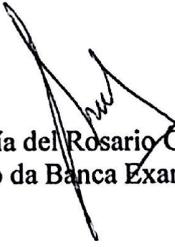
**MHNSS: UM MIDDLEWARE PARA O DESENVOLVIMENTO DE
APLICAÇÕES MÓVEIS COM INTERAÇÃO BASEADA NA FALA**

Arikleyton de Oliveira Ferreira

Dissertação aprovada em 04 de agosto de 2014.


Prof. Francisco José da Silva e Silva, Dr.
(Orientador)


Prof. Carlos de Salles Soares Neto, Dr.
(Membro da Banca Examinadora)


Profa. Maria del Rosario Girardi, Dra.
(Membro da Banca Examinadora)

*Dedico este trabalho aos
meus pais, minha irmã,
meus amigos e professores.*

Resumo

Aplicações para ambientes computacionais móveis usualmente apresentam diversas limitações de acessibilidade por dependerem da interação com o usuário através da tela dos dispositivos móveis, o que dificulta seu uso às pessoas que possuem limitações para ler, escrever (digitar) e que tenham pouca fluência no uso de tecnologias.

Neste trabalho de mestrado propomos um *middleware* que fornece suporte ao desenvolvimento de aplicações móveis com recurso de acessibilidade através do diálogo falado. Essa modalidade de acesso é capaz de manter uma conversa com o usuário, proporcionando uma interface de interação natural que não requer prévio aprendizado. Assim, aplicações móveis podem utilizar o *middleware* para proporcionar acessibilidade ao usuário que supera a necessidade do contato físico ou visual, pois eles podem apenas dialogar entre si.

O *middleware* proposto está inserido no contexto do projeto MobileHealthNet, onde auxiliará aplicações móveis focadas ao domínio da saúde a atingir usuários com diferentes perfis, com especial atenção a moradores de comunidades carentes e distantes. No processo de validação do *middleware* proposto foi utilizado um estudo de caso de uma aplicação dedicada a acompanhar o estado de saúde de pacientes portadores de fibrilação atrial, realizando-se uma avaliação com 10 sujeitos na qual obteve-se resultados bastante positivos.

Palavras-chave: Acessibilidade, Aplicações Móveis, Interações Baseada na Fala, Sistemas de Diálogo Falado.

Abstract

Applications for mobile computing environments usually have several accessibility limitations due to the dependency on the interaction with the user through the device display, which hinders its use to people who have limitations to read, write (type) and/or have little fluency in the use of technology.

In this master thesis we propose a middleware that provides support for developing mobile applications with accessibility features through spoken dialogue systems. These systems are able to hold a conversation with the user, providing a natural interaction interface that does not require prior learning. Thus, mobile applications can use the middleware to provide accessibility to the user that overcomes the physical or visual contact needs.

The proposed middleware was developed in the context of the MobileHealthNet project, where it will help mobile applications with focus in the health domain to reach users with different profiles, with particular attention to underserved and remote communities. To perform the middleware evaluation, we used a case study based on a mobile application for evaluating the health condition of patients with atrial fibrillation. The evaluation involved 10 individuals, and the results obtained were very positive.

Keywords: Accessibility, Mobile Applications, Speech-Based Interactions, Spoken Dialogue Systems.

Agradecimentos

Deixo aqui os meus agradecimentos às pessoas que tanto me ajudaram nesta caminhada.

A Deus, por sua infinita misericórdia e graça.

Aos meus pais, José Arimatea e Maria José, por terem me apoiado durante toda minha vida, em especial, nos últimos anos.

A minha irmã, familiares e amigos, pelo incentivo e estímulo.

Ao Prof. Dr. Francisco José da Silva e Silva, por sua paciência, confiança e sábia orientação durante toda minha pesquisa e produção deste trabalho.

Aos professores Carlos de Salles Soares Neto e María del Rosario Girardi Gutiérrez, por aceitarem fazer parte da banca avaliadora.

Aos colegas de laboratório, pela ajuda e apoio, em especial a Dejailson Pinheiro e Ariel Teles, pelo auxílio contínuo e consultoria, e a Berto de Tércio, pelos momentos de reflexão e conselhos.

À UFMA e ao Programa de Pós-Graduação em Engenharia de Eletricidade (PPGEE), pela oportunidade e estrutura oferecida para a conclusão deste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte durante esses anos.

"A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê."

Arthur Schopenhauer

Lista de Figuras

1.1	Arquitetura Básica de um Sistema de Diálogo	23
1.2	Grafo do Diálogo para um Sistema de Viagem	26
1.3	Conteúdo Abordado pelo OpenPhone	33
1.4	Exemplo de Interação do OpenPhone	34
1.5	Arquitetura do Mobile Companion	36
1.6	Fragmentos de um Plano de Diálogo e suas Interações	37
1.7	Arquitetura do sistema HWYD	39
1.8	Exemplo do Uso da Emoção no Diálogo do HWYD	40
1.9	Rede com as Transições de Estado do FalaBrasil	41
1.10	Regra Gramatical para Pesquisa no FalaBrasil	42
2.1	Arquitetura Geral do <i>Middleware</i> MobileHealthNet	49
2.2	Arquitetura da Infraestrutura de Comunicação do MobileHealthNet	51
2.3	Cenário de Usabilidade do PBB	53
3.1	Arquitetura do <i>Middleware</i> MobileHealthNet com o Speech Service	59
3.2	Modelo de Arquitetura Descentralizada do MHNSS	62
3.3	Arquitetura do MHNSS-Mobile	63
3.4	Métodos da Interface do MHNSS-Mobile	64
3.5	Arquitetura do MHNSS-Servidor	65
3.6	Arquitetura da Comunicação no MHNSS	68
3.7	Fluxograma do Gerenciador de Diálogo-Mobile	73
3.8	Processamento de Linguagem Natural no MHNSS	76

4.1	Questionário Preenchido com o MHNSS	82
4.2	Questionário Qualitativo	86
4.3	Resultado da Avaliação Qualitativa	89

Lista de Tabelas

1.1	Tabela Comparativa dos SDSs Analisados	44
3.1	Tópicos Utilizados na Comunicado do MHNSS	69
3.2	Comparação da Proposta com os Trabalhos Relacionados	79
4.1	Perfil dos Sujeitos Participantes da Avaliação	87
A.1	Resultados da Avaliação dos Sujeitos P1 a P5	101
A.2	Resultados da Avaliação dos Sujeitos P6 a P10	101
B.1	Resultados Qualitativos dos Sujeitos P1 a P5	102
B.2	Resultados Qualitativos dos Sujeitos P6 a P10	102

Lista de Siglas

API *Application Programming Interface.*

DCPS *Data-Centric Publish-Subscribe.*

DDS *Data Distribution Service.*

MHNSS *MobileHealthNet Speech Service.*

MOBHA *MobileHealthnet on Android.*

MobileHealthNet *Mobile Social Networks for Health Care in Offside Regions.*

MRUDP *Mobile Reliable UDP.*

PBB *Patient-Buddy-Build.*

PLN *Processamento de Linguagem Natural.*

RSMs *Redes Sociais Móveis.*

SDS *Spoken Dialogue System.*

Sumário

Lista de Figuras	vi
Lista de Tabelas	viii
Lista de Siglas	ix
1 Introdução	16
1.1 Objetivos	18
1.2 Estrutura da Dissertação	18
1.3 Definições e Características	20
1.4 Componentes de um SDS	21
1.4.1 Reconhecimento da Fala	23
1.4.2 Compreensão da Linguagem	24
1.4.3 Gerenciador do Diálogo	25
1.4.4 Gerador de Linguagem	28
1.4.5 Síntese Texto-para-Fala	29
1.5 Sistemas de Diálogo na Saúde	30
1.6 Trabalhos Relacionados	32
1.6.1 OpenPhone	32
1.6.2 Health and Fitness Companion	35
1.6.3 HWYD - How was your day?	37
1.6.4 FalaBrasil	40
1.6.5 Análise Comparativa dos Trabalhos Relacionados	42
1.7 Conclusão	44

2	Projeto MobileHealthNet	46
2.1	Redes Sociais Móveis na Saúde	46
2.2	O Projeto MobileHealthNet	47
2.3	O <i>middleware</i> MobileHealthNet	49
2.4	Comunicação no MobileHealthNet	50
2.5	Aplicações do MobileHealthNet	52
2.5.1	Patient-Buddy-Build	53
2.6	Conclusão	55
3	MobileHealthNet Speech Service	57
3.1	O MHNSS	57
3.2	Requisitos	60
3.3	Arquitetura da Solução	62
3.3.1	O MHNSS-Mobile	62
3.3.2	O MHNSS-Server	65
3.3.3	A Comunicação no MHNSS	67
3.4	Aspecto da Implementação	71
3.4.1	Serviço de Conversão de Áudio	71
3.4.2	Processamento no Gerenciador de Diálogo-Mobile	72
3.4.3	Processamento de Linguagem Natural no MHNSS	74
3.5	Análise Comparativa com os Trabalhos Relacionados	77
3.6	Conclusão	79
4	Avaliação do Trabalho	81
4.1	Estudo de Caso	81
4.2	Metodologia de Avaliação	83
4.3	Resultados e Discussão	86
4.4	Conclusão	89

5	Conclusões e Trabalhos Futuros	91
5.1	Contribuições	92
5.2	Trabalhos Futuros	93
	Referências Bibliográficas	95
A	Resultados da Avaliação dos Sujeitos	101
B	Resultados Qualitativos dos Sujeitos	102

1 Introdução

Muitas tecnologias são destinadas a encurtar a distância entre as pessoas, otimizar a realização de tarefas ou até mesmo automatizar processos repetitivos. Contudo, poucas são aquelas empenhadas em tornar as atividades mais acessíveis, considerando as diferentes necessidades e dificuldades dos usuários. Os dispositivos móveis são um exemplo de tecnologia popular que, apesar das várias inovações do setor, ainda apresentam diversas limitações de acessibilidade. Nas aplicações móveis, por exemplo, prevalece a dependência pelo enfoque visual, uma vez que seu conteúdo é baseado em textos e imagens, o que torna sua interação mais complexa por parte dos deficientes visuais e, portanto, outras formas de interação são necessárias [10].

Os assistentes pessoais virtuais para dispositivos móveis, como o Siri¹ da Apple e o Google Now², surgiram como uma proposta das respectivas empresas para fornecer um certo nível de acessibilidade a seus usuários. Apesar de concorrentes, esses assistentes são iguais no objetivo de auxiliar seus donos a efetuar atividades básicas através de comandos de voz, como redigir mensagens de texto, iniciar um aplicativo ou fazer buscas na Internet. Para fazer uma chamada telefônica, por exemplo, basta que o usuário inicie o assistente e fale o comando juntamente com o nome da pessoa que deseja ligar (por exemplo, “Ligar para Maria”), assim o assistente se encarregará de buscar o número do telefone dessa pessoa na agenda e efetuará a chamada. Este foi um grande passo para a acessibilidade nos dispositivos móveis, pois permitiu a realização de algumas tarefas de forma rápida, além de viabilizar o uso dos dispositivos móveis por pessoas com certos tipos de limitações físicas. Entretanto, são poucas as atividades que os assistentes virtuais móveis conseguem realizar e, entre os mais populares, apenas o Google Now possui suporte ao idioma português brasileiro.

As aplicações encontradas nos dispositivos móveis modernos, que atualmente são mais utilizadas que as funções básicas de telefonia, são pouco beneficiadas pelos assistentes pessoais virtuais. Eles conseguem apenas iniciar as aplicações através de comandos de voz (por exemplo, “Iniciar Calculadora”), mas não

¹<http://www.apple.com/ios/siri/>

²<http://www.google.com/landing/now/>

auxiliam no uso da mesma. Isso ocorre devido os assistentes serem tratados como aplicações independentes, não possuindo relação com as demais aplicações. Suas funcionalidades podem ser utilizadas apenas quando o assistente está em primeiro plano e, por isso, seu uso em paralelo com as aplicações é limitado. No sistema Android é possível que o usuário ative serviços que o auxiliam a ler e escrever textos através da fala³, mas é uma opção do usuário recorrer a esses serviços, pois as aplicações não são desenvolvidas considerando esse recurso ou qualquer outro tipo de acessibilidade como requisito. Desta forma, as limitações de acessibilidade ainda estão fortemente presentes nos dispositivos móveis, pois são poucas as aplicações que tentam amenizar esse problema.

O projeto *Mobile Social Networks for Health Care in Offside Regions* (MobileHealthNet), no qual o presente trabalho está inserido, viabiliza a criação de aplicações sociais móveis colaborativas dedicadas à atenção a saúde [38]. Este é um projeto destinado a rede pública de saúde e parte dos seus potenciais usuários possuem baixa escolaridade e pouco conhecimento tecnológico. Suas condições financeiras também são limitadas e, por isso, utilizam dispositivos móveis de baixo custo. Então, uma das preocupações para que os objetivos do projeto sejam atingidos é garantir que este público alvo possa utilizar os serviços oferecidos através dos seus telefones celulares de forma apropriada e sem grandes dificuldades. Por isso, opções de acessibilidade para suas aplicações são necessárias.

Nesse contexto, em que aplicações móveis necessitam prover acessibilidade aos usuários, propomos o *MobileHealthNet Speech Service* (MHNSS), um *middleware* que fornece suporte ao desenvolvimento de aplicações móveis com recursos de acessibilidade através da fala. Este *middleware* conta com um sistema de diálogo falado que proporciona às aplicações interagirem com o usuário utilizando a fala em diversas situações, como para coletar informações através de perguntas, ler textos ou receber comandos para executar atividades específicas. Assim, aplicações móveis de diversas finalidades podem garantir acessibilidade sem que os desenvolvedores se preocupem com elementos extras, pois o *middleware* fornece os recursos necessários para a interação via voz.

³<https://support.google.com/accessibility/android/>

1.1 Objetivos

A pesquisa à qual se refere esta dissertação de mestrado tem como objetivo geral desenvolver uma infraestrutura baseada em sistemas de diálogo falado para melhorar a interação homem-computador de aplicações móveis no contexto do projeto do MobileHealthNet, proporcionando aos usuários melhor usabilidade e facilidade de acesso.

Os objetivos específicos desta pesquisa são:

- Levantar o estado da arte em aplicações móveis no domínio da saúde;
- Levantar o estado da arte em sistemas de diálogos falado;
- Identificar ferramentas para desenvolver um sistema de diálogo em ambientes móveis e em português brasileiro;
- Projetar e implementar um sistema de diálogo que possa ser usado pelas aplicações do projeto MobileHealthNet e por outras aplicações móveis;
- Avaliar as melhorias de interação homem-computador proporcionadas pelo sistema de diálogo através de um estudo de caso.

1.2 Estrutura da Dissertação

Esta dissertação está organizada como segue:

- O Capítulo 1.2 apresenta uma breve fundamentação teórica sobre sistemas de diálogo falado, mostrando suas principais características, componentes e aplicação na área da saúde, além de citar alguns trabalhos relacionados.
- O Capítulo 2 descreve o projeto MobileHealthNet e o ambiente onde este trabalho dissertativo está inserido, mostrando o *middleware* contido no projeto e as aplicações desenvolvidas para ele.
- O Capítulo 3 aborda nossa proposta, o MHNSS, apresentando seus requisitos, arquitetura e comunicação, além de descrever os principais aspectos da sua implementação e fazer uma análise comparativa com os trabalhos relacionados.

-
- O Capítulo 4 apresenta a avaliação do trabalho, citando a metodologia utilizada, quais as métricas foram avaliadas e os resultados alcançados.
 - O Capítulo 5 apresenta as considerações finais, relatando as contribuições obtidas a partir desta pesquisa e citando alguns trabalhos futuros que podem ser desenvolvidos a partir deste esforço inicial.

Sistemas de diálogo falado são geralmente aplicados em situações onde os olhos e mãos do usuário estão ocupados, pois utilizam uma forma simples de comunicação e não dependem de contato visual ou físico para sua utilização. Esses sistemas proporcionam um alto nível de acessibilidade e melhoria na interação homem-computador, pois as pessoas podem utilizar algo que elas têm usado a vida toda - a habilidade de falar e entender conversas.

Neste capítulo apresentamos uma breve fundamentação teórica sobre sistema de diálogo falado, uma das bases da nossa pesquisa para levar acessibilidade a diversas aplicações móveis. Ele está organizado como segue. Na Seção 1.3 apresentamos definições e características dos sistemas de diálogo falado. Na Seção 1.4 descreveremos sua arquitetura básica e seus principais componentes. Na Seção 1.5 abordamos sua relação com a área da saúde, enquanto que na Seção 1.6 destacamos alguns trabalhos relacionados à nossa proposta.

1.3 Definições e Características

Os Sistemas de Diálogo Falado (em inglês, *Spoken Dialogue System (SDS)*) são aplicações, ou um conjunto de aplicações, onde o mecanismo primário de comunicação entre o usuário e o computador é a fala [5]. A fala é o meio mais natural de comunicação e seu uso torna a obtenção de informação e serviço mais conveniente ao usuário [25]. Por isso, através da fala os SDS podem guiar, perguntar, responder, ler e escrever para o usuário, proporcionando uma interface que não requer prévio aprendizado e permitindo sua utilização por uma variedade de público, desde que a linguagem de comunicação seja a mesma.

Também conhecido como aplicações baseadas na fala [22] [27] ou agentes conversacionais [16] [33] [40], esses sistemas aceitam linguagem natural como entrada, geram linguagem natural na saída e, como o próprio nome sugere, mantém uma conversa com o usuário [21]. Sua aplicação costuma estar relacionada a atividades que exijam um alto número de interações entre usuário e sistema, como os sistemas coletores de informações (por exemplo, obtenção de dados de voo em sistemas de viagem), controladores de dispositivos (por exemplo, ajuste de uma cama em um hospital, mesmo que o paciente esteja impossibilitado fisicamente) e assistentes

virtuais pessoais (por exemplo, marcar uma reunião na agenda eletrônica) [35]. São também muito utilizados em ambientes inteligentes e pervasivos como um dos elementos de interação que garante a comunicação efetiva, fácil e transparente entre os usuários e os componentes do ambiente [19]. Em casas inteligentes, por exemplo, ele pode intermediar a comunicação entre o usuário e o sistema para ajustar a iluminação e temperatura, alertar sobre compromissos e horários para tomar medicações, ajudar em receitas culinárias ou ainda fazer uma lista de compras [28].

Sistemas de diálogo falado podem ser do tipo liderados pelo sistema, liderados pelo usuário ou de iniciativa mista. Os liderados pelo sistema são aqueles em que a aplicação tem o controle sobre o diálogo e tenta obter informações do usuário para um fim específico, como preencher um formulário, por exemplo. Nos liderados pelo usuário as perguntas e instruções são feitas por ele e a aplicação fornece os dados ou executa algo. Estes, por exemplo, são utilizados em serviços de consultas a base de dados, como os de consulta a saldo bancário e a previsão do tempo. Já os de iniciativa mista são utilizados por aplicações que necessitam de uma interação parecida com uma conversa entre duas pessoas, pois o controle do diálogo é alternado e tanto o usuário quanto o sistema podem tomar a iniciativa de perguntar, iniciar tópicos, requisitar esclarecimento, entre outros. Os assistentes pessoais virtuais e os serviços de casa inteligentes são um exemplo de SDSs de iniciativa mista.

Várias áreas de pesquisa são combinadas para compor um SDS efetivo, como o reconhecimento automático da fala, a síntese da fala, processamento de linguagem natural e controle do diálogo. Cada uma dessas áreas está relacionada com um componente do SDS e é responsável por determinada parte do diálogo, conforme descrito a seguir.

1.4 Componentes de um SDS

Vários componentes precisam trabalhar em conjunto para que o SDS funcione corretamente, o que normalmente inclui:

1. Reconhecer as palavras que o usuário falou - Reconhecimento da Fala;
2. Atribuir um significado a essas palavras - Compreensão da Linguagem;

3. Processar o diálogo e decidir o que fazer a seguir - Gerenciador do Diálogo;
4. Utilizar bases de dados para auxiliar no processamento - Fontes Externas;
5. Transformar a instrução do Gerenciador do Diálogo em uma frase estruturada - Gerador de Linguagem;
6. Reproduzir a frase para o usuário - Síntese Texto-para-Fala;

Em geral, conforme ilustrado na Figura 1.1 baseada em [21], um Dispositivo de Acesso, que pode ser um telefone ou um equipamento com microfone embutido, captura a voz do usuário e a entrega para ser processada pelo componente de Reconhecimento da Fala. Este componente recebe o sinal de áudio da fala e o converte em uma sequência de palavras que representa textualmente o que o usuário falou. Depois de reconhecidas, as palavras são interpretadas pelo componente de Compreensão da Linguagem que dará um significado sintático e semântico a elas. O resultado da interpretação é então passado para o Gerenciador do Diálogo que decidirá qual ação a ser tomada pelo sistema, seja responder a uma pergunta, executar uma ação, solicitar esclarecimento em caso de dúvidas ou má compreensão das palavras, entre outros. Para auxiliar no processamento do diálogo, o gerenciador pode consultar a Fontes de Informações Externas, que são bases de dados e serviços que complementam o conhecimento do sistema de diálogo, como serviços de previsão do tempo, informações sobre o usuário, etc. Caso necessite se comunicar de volta com usuário, o gerenciador gera uma mensagem de comando, que pode significar uma confirmação de entendimento, uma pergunta adicional para esclarecer a requisição ou uma mensagem de erro. Essa mensagem é enviada para o Gerador de Linguagem que criará uma estrutura textual para ela de forma que ela seja compreendida pelo usuário. Esse texto é então convertido em áudio pela Síntese Texto-para-Fala e reproduzido pelo Dispositivo de Acesso através de alto-falantes.

Por ser um fluxo normalmente sequencial, a performance de cada componente é dependente dos antecessores. Por exemplo, caso o Reconhecimento da Fala falhe, o Gerenciar do Diálogo pode trabalhar com informações erradas e comprometer todo o diálogo. Contudo, nem todos os sistemas de diálogo seguem essa arquitetura completamente. As aplicações desenvolvidas comercialmente, por

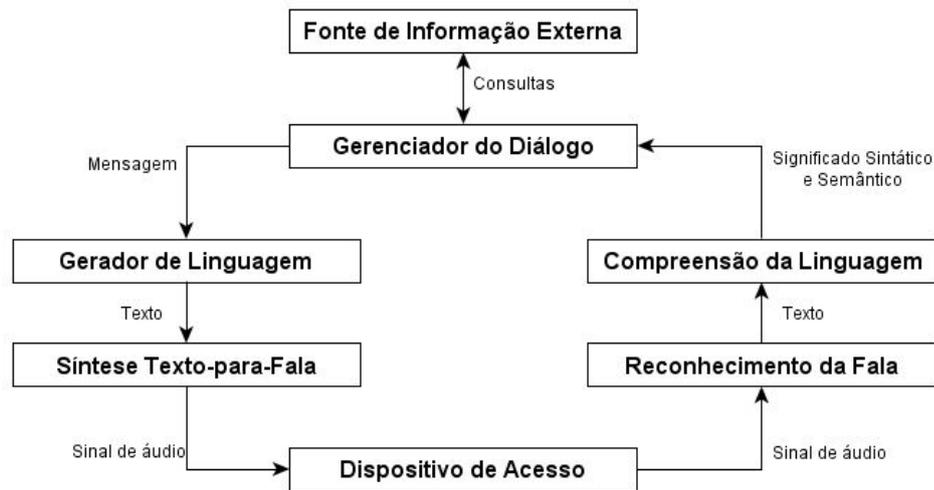


Figura 1.1: Arquitetura Básica de um Sistema de Diálogo
Baseada em [21]

exemplo, buscam simplificar cada componente da arquitetura pra reduzir custos e facilitar a produção. Alguns módulos podem ainda ser mesclados ou desconsiderados, dependendo do desenvolvedor e da finalidade da aplicação. Por outro lado, sistemas mais avançados podem envolver uma arquitetura mais elaborada, com interações diretas de componente a componente, com a adição de novos componentes e sem necessariamente seguir um fluxo sequencial [35].

1.4.1 Reconhecimento da Fala

A principal tarefa do componente de Reconhecimento da Fala (*Speech Recognition*, em inglês) é capturar o áudio de entrada e convertê-lo em uma sequência de palavras. Este componente está diretamente relacionado com a área de Reconhecimento Automático da Fala [3] e é uma tarefa complexa de ser realizada, pois não há garantias de uma interpretação correta dos dados de entrada uma vez que o som captado pode sofrer interferências diversas, dificultando a decodificação do áudio original. De acordo com [21], os principais problemas que dificultam o reconhecimento da fala são:

- **Variabilidade Inter-Orador:** Há diferenças entre oradores com relação ao jeito como eles falam e pronunciam as palavras. Algumas dessas diferenças são atribuídas a fatores físicos, como o timbre da voz. Outras diferenças são devidas a fatores como idade, sexo e regionalismo (sotaque).

- Variabilidade Intra-Orador: A mesma palavra falada em ocasiões diferentes pelo mesmo orador pode ter diferentes propriedades acústicas. Fatores como cansaço, resfriado e alterações no humor, têm influencia em como as palavras são pronunciadas.
- Variabilidade no Canal: Diferenças no canal de comunicação – microfone ou telefone, por exemplo – dão origem a diferenças no sinal acústico.
- Ruído Ambiente: O sinal acústico pode ser distorcido por barulho e ruídos, que podem ser constantes (por exemplo, barulho do ar-condicionado) ou transitórios (por exemplo, uma porta batendo).

Para converter a fala em texto, o componente de Reconhecimento da Fala recebe o áudio capturado por um microfone, o fragmenta e o classifica em uma sequência de fonemas usando modelos acústicos previamente definidos. Essas sequências de fonemas combinadas sugerem as palavras que o usuário falou. Quanto maior o número de palavras que o sistema for capaz de reconhecer, mais difícil será a identificação correta delas, pois a similaridade e ambiguidade fonética das palavras podem dificultar o reconhecimento [5]. Por exemplo, a análise fonética da frase "Ela encanta" pode ser confundida com "Elen canta".

Uma forma de melhorar o reconhecimento da fala é torná-lo dependente do orador, ou seja, treiná-lo para um padrão de voz de um orador específico. Isso vai garantir que não haja variabilidade inter-orador, reduzindo as chances de um reconhecimento errado. Contudo, este método limita o sistema a ser usado por apenas uma pessoa e, por isso, os componentes de reconhecimento da fala independente do orador são mais adotados.

1.4.2 Compreensão da Linguagem

O papel do componente de Compreensão da Linguagem (*Language Understanding*, em inglês) é analisar as palavras identificadas pelo Reconhecimento da Fala e associar um significado que possa ser usado pelo Gerenciador de Diálogo. Tradicionalmente, a compreensão das palavras envolve dois processos: análise sintática, que estuda as palavras de acordo com a função que elas desempenham em meio a um determinado contexto linguístico [1], e análise

semântica, que estuda o significado das palavras e do conjunto resultante delas [13]. Nestes processos, podemos enfrentar dois problemas muito comuns em sistemas de diálogo falado: ambiguidade na linguagem natural e má formação da entrada [21].

Ambiguidade na linguagem natural pode trazer mais de um sentido a palavras, frases ou textos. Ela pode ser classificada como ambiguidade lexical ou estrutural. A lexical geralmente ocorre quando a palavra possui mais de um sentido, como em "O garoto está sentado perto do banco"(este "banco"é um assento ou uma agência bancária?). Já a ambiguidade estrutural ocorre devido a maneira como a frase foi escrita, como no exemplo: "Preciso da chave do armário que estava na cozinha"(quem estava na cozinha, a chave ou o armário?). Um dos métodos usados para interpretar corretamente a frase em caso de ambiguidade é analisar o contexto da situação, principalmente históricos do diálogo.

A má formação dos dados de entrada pode ser um problema de reconhecimento ruim das palavras pelo módulo de Reconhecimento da Fala, ou peculiares do diálogo espontâneo como sentenças fragmentadas, auto correção do usuário, língua presa ou combinações não gramaticais [21]. Uma das formas de tratar esse problema é usar verificações frequentes, do tipo: "Você falou [isso]?".

Componentes de Compreensão da Linguagem com análise rígida de sintaxe e semântica são pouco comuns em sistemas de diálogo falado. A maioria dos sistemas usa técnicas simples de extração dos dados relevantes da expressão a partir de palavras-chave e regras que auxiliam na compreensão [20]. Outra alternativa é a utilização de técnicas de Processamento de Linguagem Natural, alternativa esta que é utilizada neste trabalho e apresentamos na Seção 3.4.3.

1.4.3 Gerenciador do Diálogo

Entre outras funcionalidades, o Gerenciador do Diálogo (*Dialogue Manager*, em inglês) é responsável por processar os dados de entrada, produzir as mensagens que serão comunicadas ao usuário, interagir com fontes de informações externas (banco de dados ou outros serviços) e, geralmente, controlar o fluxo do diálogo. O tipo de sistema de diálogo (por exemplo, se é liderado pelo sistema, liderado pelo usuário ou de iniciativa mista) está diretamente relacionado em como o

Gerenciador do Diálogo realiza essas atividades e, por isso, este é o componente central do sistema de diálogo [21].

Para processar um diálogo e tomar uma ação, este componente depende de como o diálogo está sendo controlado, que pode ser baseado no estado finito, baseado em *frames* ou baseado em agentes.

Controle Baseado no Estado Finito

A estrutura do diálogo nesse tipo de sistema é representada na forma de uma rede de transição de estado (ou grafo) em que os nós representam as perguntas do sistema e a transição entre os nós representa todos os possíveis caminhos que o diálogo pode seguir. Cada estado representa uma etapa em que alguma informação é dita ou confirmada pelo usuário, ou alguma ação é executada pelo sistema. Geralmente utilizam perguntas que exigem respostas curtas e diretas, do tipo Sim ou Não e, por isso, são também conhecidos como sistemas de diálogo direto. A figura 1.2 ilustra um exemplo de uma rede de transição entre estados de um sistema de viagens.



Figura 1.2: Grafo do Diálogo para um Sistema de Viagem

Neste exemplo, o sistema pergunta sobre o destino da viagem e depois verifica a resposta recebida. Se o usuário confirmar a verificação, o sistema segue para a próxima questão sobre o dia da viagem. Caso a confirmação não seja verdadeira, a

Condições	Perguntas
[Destino=?] + [Data=?] + [Hora=?]	Para onde você gostaria de viajar?
[Data=?] + [Hora=?]	Quando você deseja viajar?
[Hora=?]	A que horas você deseja viajar?

pergunta sobre o destino da viagem é refeita. Esse processo continua até que toda a rede tenha sido percorrida.

A maior vantagem dos sistemas com controle baseado em estados finito é a simplicidade, e são ideais para SDSs liderados pelo sistema que envolvem tarefas bem estruturadas ou informações que precisam ser transmitidas em uma sequência pré-determinada. Entretanto, este modelo é inflexível, pois os caminhos do diálogo são especificados antecipadamente e, por isso, não há meios de fazer desvios como, por exemplo, se o usuário quiser corrigir um item ou informar uma opção que não foi prevista no fluxo.

Controle Baseado em *Frames*

Sistemas de diálogos baseados em *frames* são mais apropriados para tarefas com diálogo aberto, sem restrição de fluxo, principalmente por parte do usuário. Seu processamento é baseado em encontrar valores para os *frames*, que são espaços de dados correspondentes a informações de interesse do sistema. Por exemplo, para efetuar a compra de uma passagem aérea o sistema precisa obter valores para os *frames* [Destino = ?], [Data = ?] e [Hora de saída = ?]. Inicialmente os valores são desconhecidos pelo sistema e, assim, temos:

Sistema: *Para onde você gostaria de viajar?*

Usuário: *Eu quero um voo para São Paulo no sábado.*

Baseado nessa resposta do usuário, dois valores de *frames* foram obtidos: [Destino = São Paulo] e [Data = Sábado]. Mas a [Hora de saída] ainda não foi identificada e, neste caso, um algoritmo de interpretação determina a próxima mensagem a ser transmitida ou ação a ser executada baseado em uma condição, que leva em conta o *frame* desconhecido. Exemplo:

Assim, para este caso onde apenas a [Hora de saída] é desconhecida, a pergunta "A que horas deseja viajar?" seria realizada.

A flexibilidade é a maior vantagem dos sistemas baseados em *frames*, pois podem processar várias informações em um único turno, o que pode reduzir o tempo dos diálogos e torná-los mais natural. Além disso, não necessita seguir uma sequência pré-determinada, pois a ordem em que os *frames* são obtidos não é relevante. Contudo, é muito dependente de uma compreensão precisa dos dados de entrada e o algoritmo de interpretação deve considerar todas as condições possíveis para cada *frame* especificado, sendo assim mais utilizado para uso em domínio específico.

Controle de Diálogo Baseado em Agentes

A abordagem baseada em agentes utiliza técnicas da Inteligência Artificial e foca na modelagem de diálogos colaborativos entre agentes inteligentes para processar tarefas mais complexas, como negociações e solução de problemas. Exemplo:

Usuário: *Estou procurando por emprego em São Luis. Há alguém contratando?*

Sistema: *Não, não há ninguém contratando em São Luis. Entretanto, estão contratando em Rosário e Chapadinha. Você está interessado em algum desses?*

Neste exemplo, a resposta do sistema foi negativa. Mas ao invés de simplesmente responder "não", o sistema tenta fornecer uma resposta que atendesse às necessidades do usuário. Para isso, o sistema usa expectativas para tentar prever e interpretar a próxima possível pergunta do usuário.

Sistemas baseados em agentes tendem a ser de iniciativa mista, onde o usuário pode tomar controle do diálogo, iniciar um novo tópico ou fazer contribuições que não foram pedidas pelo sistema. Assim, não há como prever os dados que o usuário vai fornecer ou o que ele vai solicitar e, por isso, um módulo robusto de Compreensão da Linguagem é necessário para processar essas informações.

1.4.4 Gerador de Linguagem

Uma vez que o sistema de diálogo necessita se comunicar com o usuário, o componente Gerador de Linguagem (*Language Generation*, em inglês) é responsável

por construir a mensagem em linguagem natural que será reproduzida a ele. Para tanto, o Gerenciador do Diálogo passa uma mensagem de comando e o Gerador de Linguagem aplica conceitos semânticos a essa mensagem e formula a expressão de saída. Exemplo:

Mensagem enviada pelo Gerenciador: *Confirmar: Sintoma = Dor de cabeça*

Expressão construída pelo Gerador de Linguagem: *Você está sentindo dor de cabeça?*

O processo de construção da oração geralmente é feito em três estágios: 1) determina qual informação será incluída na mensagem e como a mensagem deve ser estruturada (por exemplo, se a mensagem deve confirmar um sintoma, perguntar se o usuário "está sentindo"aquele sintoma); 2) combina o conteúdo em porções mais significativas usando conjunções, caso necessário (por exemplo, "está sentindo dor de cabeça [e] está cansado?"); e 3) converte o conteúdo em texto linguístico para ser reproduzido pelo componente de Síntese Texto-para-Fala caso ele necessite de um padrão específico (por exemplo, um esquema XML⁴) [21].

Tecnologias de geração de linguagem natural são utilizadas apenas em abordagens que buscam um diálogo próximo ao humano [34]. Sistemas mais simples, como os desenvolvidos comercialmente, utilizam frases pré-definidas como saída, pois o foco maior desses sistemas está na interpretação dos dados de entrada.

1.4.5 Síntese Texto-para-Fala

O componente Texto-para-Fala (*Text-to-Speech Synthesis*, em inglês) converte a mensagem de texto construída pelo componente Gerador de Linguagem em forma falada. Nos casos mais simples, conversas pré-gravadas podem ser usadas e, algumas vezes, possuem espaços que podem ser preenchidos com informações relevantes, como em: "Você recebeu uma ligação de <Ana Maria>". Várias aplicações utilizam esse método, pois é bastante eficiente desde que as mensagens sejam similares e constantes. Em casos onde o texto é variante e imprevisível, quando um grande número de informação tem que ser reproduzido ou quando é requisitado certa consistência na voz, uma análise sintática da fala é necessária [21].

⁴<http://www.w3schools.com/xml/default.asp>

A análise sintática envolve dois estágios: análise de texto e geração da fala. Análise de texto produz uma representação linguística do texto de entrada enquanto a geração da fala dá ritmo e entonação à mensagem. São também conhecidos como "conversão texto-para-fonema" e "conversão fonema-para-fala" respectivamente.

A Síntese Texto-para-Fala é um dos principais componentes do sistema de diálogo, pois a forma como algo é dito é tão importante quanto as palavras pronunciadas [36]. A voz que o sistema emite influencia diretamente na aceitação do usuário em utilizar qualquer sistema baseado na fala, pois uma pronúncia muito mecânica ou robótica tira a atenção do usuário da informação que está sendo passada. Por isso, este componente está em constante evolução para que o diálogo pareça cada vez mais natural.

1.5 Sistemas de Diálogo na Saúde

Sistemas de diálogo falado podem ser utilizados por diversos domínios e áreas de interesse. O termo *Health Dialog System* é utilizado para caracterizar sistemas de diálogo que estabelecem uma comunicação com pacientes sobre assuntos relacionados a saúde [4]. Sua aplicação pode ser destinada ao auxílio de auto-medicação, acompanhamento remoto e prevenção de doenças, educação de pacientes, entre outras finalidades.

Mesmo que não sejam tão eficientes quanto os encontros com os profissionais da saúde, os *Health Dialog Systems* podem atender as necessidades do paciente em diferentes formas. Uma dessas formas é o fator tempo, pois muitos dos encontros com os médicos e especialistas são em curto tempo para que possam atender o maior número de pessoas, o que pode intimidar o paciente a fazer perguntas ou pedir para que uma informação seja repetida [9]. Neste caso, um sistema de diálogo poderia tirar dúvidas e fornecer informações para o usuário sempre que necessário, sem que haja a necessidade de um profissional durante esse processo. Outro ponto favorável desses sistemas é o alcance, pois grande parte dos pacientes não possuem acesso a todos os profissionais que eles precisam, seja por questões financeiras ou pela demora em conseguir marcar uma consulta. Assim, os sistemas de diálogo podem fornecer atendimento especializado a um maior número de pessoas [41]. Além disso, sistemas

de diálogo podem fornecer assistência a pacientes em idiomas diferentes, aumentando ainda mais o alcance dos atendimentos [24].

Segundo [4], diversos fatores fazem o diálogo na saúde um domínio desafiador. Alguns desses fatores são:

- **Criticidade** - Sistemas de diálogos na saúde podem ser usados em situações de urgência, como os sistemas que auxiliam pacientes com doenças crônicas no auto-tratamento. Então, esses sistemas devem ficar disponível a maior parte do tempo, para auxiliar o paciente em momentos críticos, além de serem capazes de tomar medidas de emergência, como ligar para alguém, caso necessário;
- **Privacidade e Segurança** - O conteúdo do diálogo e as mídias de comunicação devem atender questões de privacidade e segurança dos dados. Informações de contexto podem ser usadas para auxiliar esse processo como, por exemplo, não fornecer dados confidenciais em locais públicos ou fornecer somente as informações relacionadas com a doença dos usuário.
- **Continuidade na Interação** - Algumas aplicações necessitam de várias interações com o usuário em um determinado período de tempo, seja várias vezes ao dia ou com intervalos de semanas. Em casos como o monitoramento de doenças crônicas onde a interação não é isolada, é necessário manter uma persistência das informações para otimizar o diálogo. Ou seja, através dos diálogos anteriores é possível gerar uma interação melhorada, principalmente em casos de mudanças no estado de saúde do paciente.
- **Mudanças na linguagem** - Quanto mais o usuário utiliza um aplicação, mais ele se familiariza com ela e, progressivamente, a aplicação obtêm mais informações sobre o usuário. Essa experiência adquirida por ambas as partes faz com que algumas perguntas e respostas sejam evitadas, pois já foram discutidas anteriormente. Essas mudanças na linguagem são importantes para manter o interesse do usuário no sistema, pois a repetitividade de interação desmotiva o usuário a continuar usando a aplicação ou a seguir suas recomendações.
- **Iniciativa do Diálogo e Negociação** - Durante o diálogo entre um profissional da saúde e um paciente, o profissional é quem mantém a iniciativa do diálogo na maior parte do tempo. Isso quer dizer que é ele quem faz as perguntas e tem o

controle do diálogo. Contudo, o melhor meio de motivar o paciente a aderir ao tratamento é através de um acordo igualitário de ambas as partes. A iniciativa deve ser então centrada no paciente e o sistema negocia com ele os processos do tratamento.

1.6 Trabalhos Relacionados

Na literatura podemos encontrar sistemas de diálogo falado com abordagem voltada para diversos cenários. Muitos destes são destinados a fornecer acesso a serviços que, por algum motivo, não estava ao alcance do usuário. Um exemplo é o apresentado em [32], onde o SDS viabiliza o acesso a informações de saúde que antes estavam disponibilizadas em idioma diferente do local. Outro exemplo é o apresentado em [31], onde o SDS é utilizado para promover acesso psicológico a veteranos de guerra através da Internet, disponibilizando atendimento especializado para este público.

Dependendo da área de aplicação, da finalidade, do tipo de diálogo que será construído e vários outros aspectos, os SDSs podem apresentar diferentes características entre si. A seguir, citamos alguns trabalhos com características relacionada com nossa proposta, apontando peculiaridades de cada um, para então compará-los ao final da seção.

1.6.1 OpenPhone

O *Open Phone* é um serviço de diálogo apresentado em [15] para que pessoas que cuidam de crianças com HIV possam obter informações sobre a doença e seus tratamentos. Ele foi desenvolvido com base no atendimento prestado pela Clínica Baylor, especializada em pediatria e em tratamento contra HIV/Aids, e através das necessidades e desafios apresentados por seus pacientes, dos quais o artigo cita:

- Cuidadores (pessoas que cuidam de crianças com HIV) têm muitas dúvidas sobre informações gerais de saúde, por exemplo, como lidar com infecções, necessidades nutricionais, requisitos de higiene e equívocos comuns sobre o HIV.

- A maioria dos cuidadores são pessoas analfabetas e semi-alfabetizadas. Portanto, o material escrito com essas informações são pouco utilizados.
- A clínica oferece treinamento e palestras sobre a doença, mas o deslocamento até ela nem sempre é possível para os cuidadores devido responsabilidades de trabalho/família, custos de transporte e limitações de tempo.
- As palestras da clínica e todas as interações com os cuidadores são em Setswana, idioma local, mas o material informativo ainda é em inglês, o que dificulta a acessibilidade às informações.
- Os cuidadores são incentivados a ligar para a clínica e tirar dúvidas sobre o tratamento, mais a maioria é relutante ou incapaz de fazê-la devido aos elevados custos de chamadas de telefonia móvel.

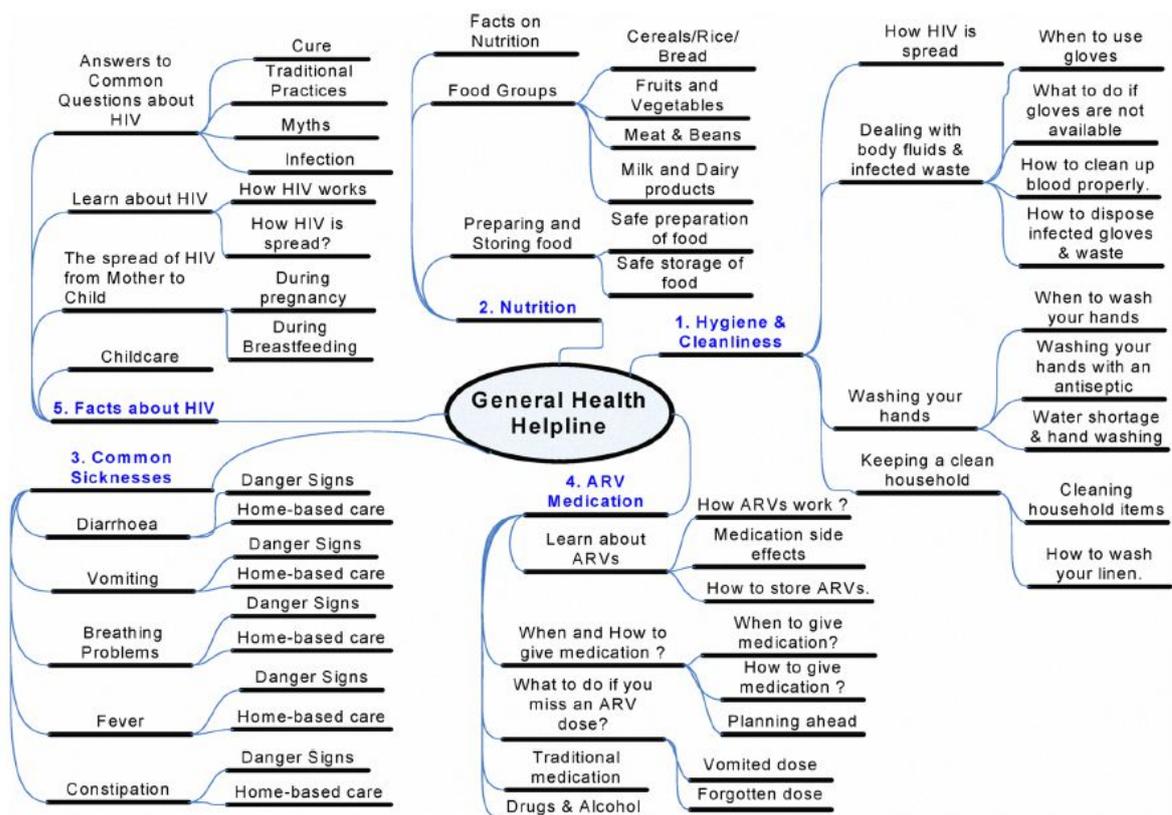


Figura 1.3: Conteúdo Abordado pelo OpenPhone

Fonte: [15]

Com base nessas necessidades e através de entrevistas com médicos, enfermeiros e nutricionistas, os conteúdos relevantes que seriam abordados pelo SDS foram identificados. Estes conteúdos foram agrupados e categorizados em menus e

sub-menus, onde cada um apresenta uma informação específica, como pode ser notado na Figura 1.3.

O *OpenPhone* SDS utiliza um plataforma telefônica como forma acesso a ele. Para obter informações sobre determinado assunto, o cuidador realiza uma ligação telefônica para esta plataforma, iniciando assim a interação por voz com o sistema de diálogo. Esta interação é controlada pelo usuário, ou seja, é ele quem decide sobre qual assunto será falado, o sistema apenas fornecer as informações. A Figura 1.4 é um exemplo dessa interação:

User: [Dials number...]

System (Introduction): *Hello and Welcome to the Health Helpline, I am Nurse Lerato and I know you probably have many questions about caring for someone with HIV.*

System (Overview): *I can tell you about Hygiene & Cleanliness, Nutrition, Common Sickneses, ARV Medication, and Facts about HIV. If at any time during your call you want to start our conversation again, you say Main Menu.*

System (Main Menu): *For Hygiene & Cleanliness, say Hygiene and Cleanliness, for Nutrition, say Nutrition, for Common Sickneses, say Common Sickneses, for ARV medication, say Medication or for Facts about HIV, say Facts.*

User: *Nutrition.*

System: *Eating a balanced diet of different foods helps people with HIV stay healthy and strong. A healthy diet does not have to be costly and contains food from all the different food groups. Healthy food is always prepared and stored in a clean environment...*

Figura 1.4: Exemplo de Interação do *OpenPhone*

O processamento do diálogo no *OpenPhone* é baseado em estado finito, pois limita-se a acompanhar o fluxo determinado pelos menus e sub-menus. As mensagens de *Introduction*, *Overview*, *Main Menu* e informações sobre cada assunto são oriundas de áudios pré-gravados, pois o sistema não possui componente de síntese texto-para-fala. O reconhecimento da fala do usuário não é robusto e reconhece apenas as opções contidas nos menus. E, por conta desses fatores, o *OpenPhone* não realiza nenhum tipo de processamento de diálogo, pois se propõe a ser apenas uma alternativa de interação simples para acessar informações sobre cuidados de HIV.

1.6.2 Health and Fitness Companion

No trabalho apresentado em [40], um assistente virtual chamado de *Companion* auxilia seus donos a manterem um estilo de vida saudável através de exercícios físicos e alimentação controlada. Diferente da maioria das abordagens pra assistentes pessoais que são baseados em tarefas únicas e bem definidas, o projeto *Companion* se considera orientado a tarefas, pois se propõe a fornecer suporte de longa duração e um certo tipo de companheirismo no acompanhamento de tarefas diárias do usuário. Assim, ele tenta estabelecer uma relação de parceria para apoiar o usuário a mudar hábitos diários e manter uma vida saudável, em vez de apenas dar conselhos sobre questões relacionadas a saúde.

Para atingir esse nível de relacionamento com o usuário, o *Health and Fitness Companion* utiliza um sistema de diálogo falado pra interagir com o usuário em várias situações do cotidiano. Este SDS está presente: na casa do usuário (*Home Companion interface*) para estabelecer uma conversa antes dele ir ao trabalho ou iniciar uma atividade física; no dispositivo móvel do usuário (*Mobile Companion interface*) para orientá-lo em atividades ao ar livre, como caminhadas e passeios de bicicleta; e em uma plataforma Web (*Cooking Companion interface*) para auxiliar o usuário em questões nutricionais.

No dispositivo móvel, conforme ilustrado na Figura 1.5, há dois componentes responsáveis pelo *Mobile Companion*: um aplicativo Java que lida com a interface gráfica e com o diálogo realizado com o usuário, e um servidor de fala que executa operações de reconhecimento da fala (ASR) e síntese texto-para-fala (TTS) a pedido do aplicativo. Seu processamento é baseado em planos de diálogo recebidos do *Home Companion* através da Internet, via WLAN ou 3G/GPRS.

Os planos de diálogo são modelos que o sistema considera como melhor rota para concluir determinada tarefa. Estes planos são criados no *Home Companion* a partir do conteúdo de itens de planejamento e novas informações relevantes (predicados) que são incluídas ao plano durante a interação, como o estado do usuário (exemplo: cansado), os compromissos do usuário para com o sistema (exemplo: vai a pé para o trabalho), as preferências do usuário (exemplo: não gosta de cafeterias) e relatórios sobre atividades passadas (exemplo: tomou um táxi para o trabalho).

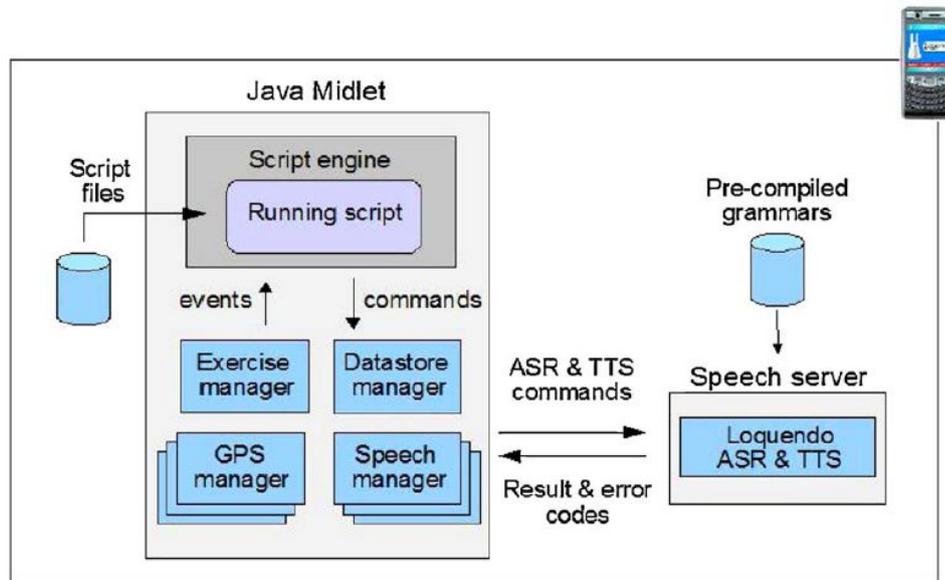


Figura 1.5: Arquitetura do Mobile Companion

Fonte: [40]

Na Figura 1.6 temos partes de um plano de diálogo que está sendo construído durante a interação com o usuário. Este exemplo começa com o item de planejamento `QUERY-PLANNED-ACTIVITY`, que foi interpretado pelo sistema em perguntar ao usuário se há alguma atividade planejada para aquele dia. Como o usuário pretendia fazer uma corrida leve, o sistema incluiu essas informações no plano e questionou se ele faria o exercício antes do jantar. O usuário respondeu que seria após o jantar e, assim, o plano foi criado contendo a atividade que será realizada e o horário da mesma.

Após criado, o plano de diálogo é disponibilizado para que outras interfaces, como a do dispositivo móvel, tenham acesso a qualquer momento. Em posse do plano, o *Mobile Companion* o utiliza como base do diálogo com o usuário durante o exercício. Além do plano, o *Mobile Companion* mantém dados persistentes no dispositivo sobre o usuário (nome, idade, sexo, peso, etc) e resultados passados. Salvar esses dados permite que o Companion compare o progresso de um exercício com exercícios anteriores do mesmo tipo. Por exemplo, se o Companion sabe que o usuário está indo de bicicleta para o trabalho, ele pode fornecer *feedback* sobre como o usuário está se saindo em relação às sessões anteriores (exemplo, ao dizer: "Você está 1 minuto e 23 segundos atrás do tempo de ontem"). À medida que o exercício é completado ou abortado, essas informações são repassadas ao *Home Companion* para que sirva de base para futuros planos de diálogo.

```
( <plan-item>
<action>QUERY-PLANNED-ACTIVITY</action>
</plan-item> )
```

C1: Good morning. Anything interesting organized for today?

U1: I'm going jogging.

```
( <pred>
<action>PLANNED-ACTIVITY</action>
<param>ACTIVITY-JOGGING</param>
<param>unknownTime</param>
</pred> )
```

C2: Is that jogging exercise before dinner?

U2: No, its after.

```
( <pred>
<action>PLANNED-ACTIVITY</action>
<param>ACTIVITY-JOGGING</param>
<param>AFTER-DINNER</param>
</pred> )
```

Figura 1.6: Fragmentos de um Plano de Diálogo e suas Interações

Como pode-se notar, o *Health and Fitness Companion* apresenta um proposta diferente de diálogo, baseado em planos, o que possibilita uma interação de longa duração e entre múltiplas plataformas. Ele é um dos poucos trabalhos que utilizam sistemas de diálogo em ambiente móvel, apesar de não ser o principal foco da proposta. O SDS apresentado é de iniciativa do sistema, onde ele tem o controle do diálogo ao fazer as perguntas e dar sugestões, e o usuário apenas fornece as informações. Seu processamento, apesar de não ser citado pelo autor, está associado com o gerenciamento baseado em *frames*, onde pode ser notado na criação dos planos de diálogo devido aos itens de planejamento e aos predicados. Para realizar a compreensão da linguagem, o H&F Companion utiliza uma interpretação semântica para o reconhecimento da fala, o que é suficiente para identificar os itens de planejamento. Assim, este é um dos trabalhos mais interessantes encontrado na pesquisa.

1.6.3 HWYD - How was your day?

O HWYD [29], sigla para '*How was your day*', é um assistente virtual que objetiva estabelecer uma relação confortante com o usuário através de conversação

sobre vários tópicos relacionados a seu trabalho. Esta conversação não está necessariamente conectada a uma tarefa específica, pois o HWYD propõe-se a atender a diálogos aberto referente a assuntos de trabalho, que vão desde os relativamente banais (exemplo, chegar atrasado devido ao trânsito e prazos de projetos), até assuntos mais importantes (exemplo, promoções, demissões e políticas do escritório).

Este SDS apresenta uma abordagem de gerenciamento de diálogo misto, tanto baseado em estado finito quanto baseado em *frames*. Inicialmente há subtópicos pré-definidos que categorizam os assuntos do diálogo, o que é característico dos sistemas baseado em estado finito. Mas, uma vez que a categoria foi definida, há *slots* que induzirão o gerenciador de diálogo a fazer perguntas para preenchê-los, estimulando um diálogo mais extenso, assim como acontece nos sistemas baseados em *frames*.

Outro diferencial do HWYD é que ele reconhece e responde apropriadamente a diferentes emoções do usuário. Isso acontece a partir de um conjunto de componentes que detectam a emoção do usuário através da análise da voz e fazem com que o sistema adote um tom de voz e um tipo de processamento baseado nessa análise. Para isso, a arquitetura do HWYD é composta por três grupos de componentes: o grupo dos componentes de entrada afetiva multimodal, que captura e analisa voz do usuário, composto pelo reconhecimento da fala (ASR), segmentação e marcação do diálogo (DAT), compreensão da linguagem (NLU), reconhecimento emocional da fala (AA), detector de turnos (ATT), analisador sentimental (SA) e modelador emocional (EM); o grupo do diálogo afetivo, responsável pelo processamento e pela entonação do diálogo de saída, composto por um gerenciador do diálogo (DM), modelagem do usuário e bases de conhecimento (KB/UM), módulo de estratégia afetiva (ASM) e gerador de linguagem (NLG); e o grupo dos componentes de saída afetiva multimodal, que reproduz a voz do sistema e gera expressões faciais e textuais no avatar, composto por um controle multimodal de avatar (MFM), um avatar (ECA) e uma síntese texto-para-fala (TTS). Esses componentes estão apresentados na Figura 1.7.

Esse sistema também inclui um gerenciador de interrupções (IM), que detecta interrupções e sobreposições de diálogo feitas pelo usuário, resultando na suspensão imediata da voz do sistema, caso ele esteja falando algo. Esta interrupção

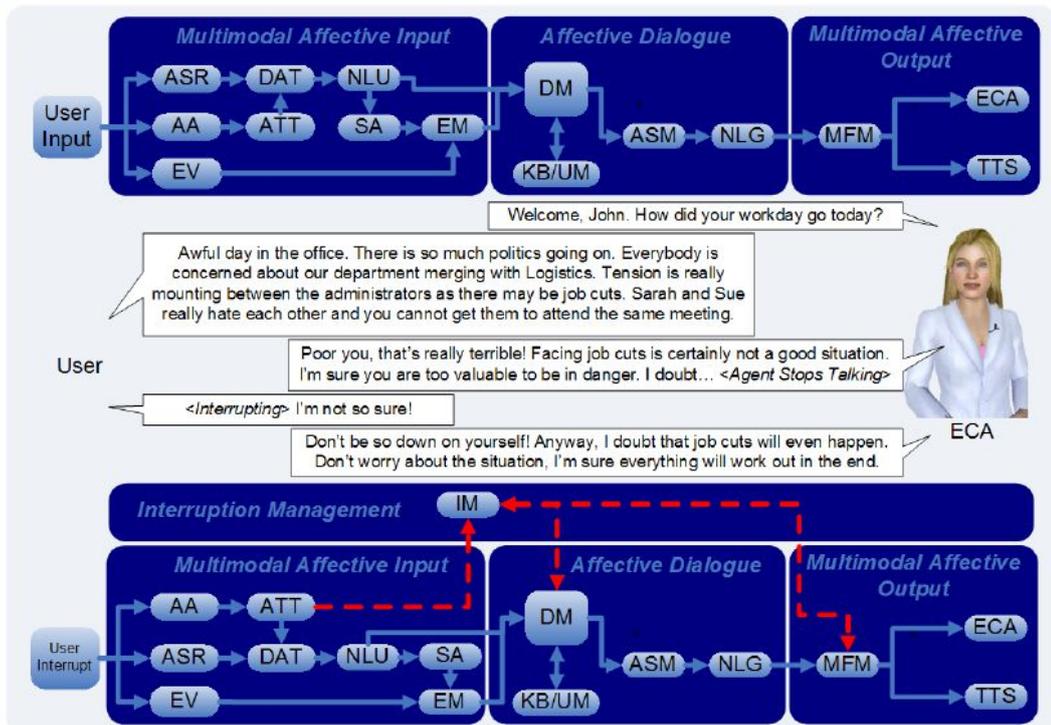


Figura 1.7: Arquitetura do sistema HWYD

Fonte: [29]

ativa um novo processamento no gerenciador do diálogo, que produzirá uma nova resposta apropriada ao usuário, assim como ilustrado na parte de baixo da Figura 1.7.

Para demonstrar como as emoções passadas pela voz do usuário afetam no processamento do HWYD, temos o exemplo da Figura 1.8. Nela o usuário interage com o sistema falando que seu escritório irá se juntar com outro de Barcelona. Ao receber essa informação, o sistema avaliou que a emoção contida na voz do usuário era positiva, o que representa que aquela era uma boa notícia. O mesmo acontece na segunda parte do diálogo, onde o usuário informa que uma cafeteira nova havia chegado no escritório, onde também foi detectado emoções positivas.

Para gerenciar um diálogo tão aberto, o HWYD possui um componente de compreensão da linguagem (NLU) adequado para essa necessidade. Ele não é tão robusto, pois atende apenas a um domínio específico, mas é capaz de compreender frases extensas. Para isso, ele utiliza algumas técnicas de processamento de linguagem natural, como classificar cada palavra gramaticalmente (exemplo, identificar o que substantivo e verbo) e identificar se é uma frase nominal ou um grupo verbal. Além disso, as palavras referentes a entidades nomeadas são identificadas, como nomes de pessoa, organização e hora. Esse tipo de processamento é importante na proposta

User: I have just learnt that my department is going to be merged with the Barcelona one
EMOTION: POSITIVE ACTIVE
System: You seem a bit too ecstatic! I'm doubtful that your workload could be maintained.
User: Yes but one good thing anyway we had a new coffee machine arrived today
EMOTION: POSITIVE ACTIVE
System: It's always reassuring to hear good news! It's right to feel good about such a good situation.

Figura 1.8: Exemplo do Uso da Emoção no Diálogo do HWYD

do sistema, para que as palavras referente ao ambiente de trabalho do usuário sejam identificadas e o processamento a respeito seja correto.

1.6.4 FalaBrasil

O trabalho apresentado em [23] pelo grupo de pesquisa Fala Brasil⁵, da UFPA, visa tornar a Web mais acessível para deficientes físicos e visuais. Para isso, é proposto um sistema de diálogo falado que permite a realização de pesquisas na Web utilizando apenas a fala e destaca-se, principalmente, por ser destinado ao português brasileiro.

Esse sistema utiliza a SAPI (*Speech Application Programming Interface*), uma interface de reconhecimento e síntese de voz para o desenvolvimento de aplicações baseadas nos sistemas operacionais Windows. Através do SAPI, o sistema estabelece um diálogo falado com o usuário questionando-o sobre o site e a palavra que deseja pesquisar na Internet. Após a busca, o conteúdo da página principal é lido para o usuário, permitindo assim uma navegação Web não-visual.

Por estar em estágio inicial de desenvolvimento, o sistema de diálogo limita-se a buscar informações apenas sobre países e a utilizar o Google e Wikipedia como opções de sites de busca. Para fazer a interação com o usuário, ele utiliza uma rede de transição de estado que é a base para a construção do diálogo. Essa rede de transição é mostrada na Figura 1.9.

⁵<http://www.laps.ufpa.br/falabrasil/>

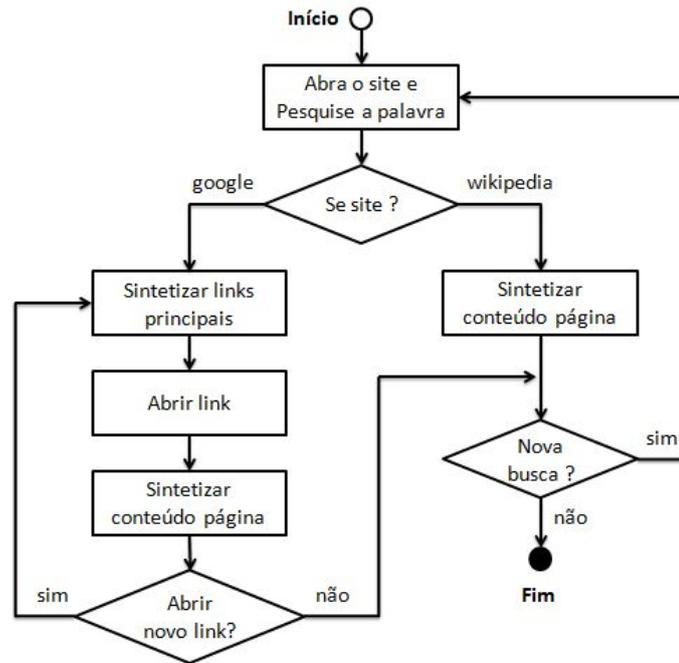


Figura 1.9: Rede com as Transições de Estado do FalaBrasil

Fonte: [23]

Primeiramente, o usuário deve dizer qual o site e a palavra que ele deseja consultar. Caso o site escolhido seja a Wikipedia, o conteúdo (texto) principal da página resultante da pesquisa é sintetizado (lido) para o usuário. Caso o site escolhido seja o Google, uma lista com âncoras resultantes da pesquisa é lida para o usuário e é solicitado que ele escolha um dos links disponíveis para, então, sintetizar o conteúdo da página escolhida. Por fim, o usuário tem a opção de acessar outro link ou realizar uma nova busca.

Para iniciar a pesquisa, uma regra gramatical foi construída para reconhecer o comando que corresponde ao primeiro estado da rede de transição. Uma amostra dessa regra pode ser conferida na Figura 1.10, onde as palavras entre as *tags* [P] são passíveis de reconhecimento e as de *tags* [O] são opcionais. Assim, é possível reconhecer frases como "pesquisa no google...".

Nesse sistema, é possível que o usuário interrompa a leitura de uma página sempre que necessário, seja para iniciar uma nova busca ou voltar atrás na decisão tomada. Além disso, para tentar prevenir erros de reconhecimento por parte do SAPI (dado as restrições de vocabulário e de domínio), o sistema de diálogo faz confirmações explícitas dos dados solicitados sempre que possível, permitindo que o usuário volte ao estado anterior e corrija sua solicitação.

```
<RULE NAME="pesquisa" TOPLEVEL="ACTIVE">
  <L>
    <P>pesquisa</P>
    <P>pesquisar</P>
  </L>
  <O><P>no</P></O>
  <L PROPNAME="SITE" PROPID="SITE">
    <P VALSTR="google">google</P>
    <P VALSTR="wikipedia">wikipedia</P>
  </L>
</RULE>
```

Figura 1.10: Regra Gramatical para Pesquisa no FalaBrasil

Esse trabalho, apesar de simples, consegue atender bem ao que se propõe. O próprio sistema coordena as interações, ou seja, a aplicação guia o usuário ao longo das transições de estado, o que diminui a possibilidade de solicitações fora do contexto. Isso viabiliza o uso de componentes menos robustos, mas ainda assim eficazes, o que muito se parece com nossa proposta. Além disso, é um dos raros sistemas de diálogo que utilizam o português brasileiro como idioma padrão.

1.6.5 Análise Comparativa dos Trabalhos Relacionados

Embora os trabalhos relacionados citados não apresentem o mesmo objetivo ou cenário de aplicação, todos utilizam sistemas de diálogo como base de suas propostas devido a sua facilidade de interação com o usuário. Cada trabalho apresentado possui um SDS com características específicas e comportamento diferente, o que torna difícil sua comparação. Mas, alguns aspectos em comum destacam-se entre eles.

O domínio de atuação do sistema de diálogo influencia em vários aspectos do seu desenvolvimento. Quando um SDS é voltado para um domínio específico, componentes como o de compreensão da linguagem e de reconhecimento da fala costumam ser mais simples, pois não haverá muita variação das palavras. É o que acontece no OpenPhone, onde o reconhecimento da voz é limitado a identificar apenas as opções do menu listado pelo sistema, o que induz o usuário a falar uma dessas opções e, por isso, não há necessidade de um reconhecimento além dessas palavras. No projeto FalaBrasil acontece algo semelhante, pois há um comando específico para

fazer as buscas na Web e apenas nomes de países podem ser utilizados como objetivo da pesquisa. Já em HWYD, o reconhecimento da fala é mais robusto, pois é capaz de identificar frases longas do usuário. Contudo, a compreensão da fala é limitado pelo domínio e atende apenas a assuntos relacionados ao ambiente de trabalho do usuário. E em H&F Companion, o domínio de aplicação é a área da saúde, mas não apresentou limitações no diálogo devido a isso.

Quanto a plataforma de acesso ao SDS, cada trabalho apresentou uma opção diferente. No OpenPhone, o usuário interage com o sistema de diálogo através de ligações telefônica. Em HWYD, a interação é realizada através do computador. Assim como no FalaBrasil, onde o navegador do Internet Explorer é utilizado nas buscas, o que exigia um computador. E em H&F Companion, o SDS pode ser utilizado no computador, no dispositivo móvel e a Web. Este aspecto comparativo, contudo, busca identificar os sistemas de diálogo que funcionam em ambiente móvel, que é um dos focos da nossa proposta. Assim, apenas o H&F Companion possui suporte a essa plataforma.

O gerenciamento do SDS está relacionado com o controle que o gerenciador adota para o diálogo, que pode ser baseado em estado finito ou baseado em *frames*. Entre os trabalhos relacionados, o OpenPhone e o projeto FalaBrasil adotam um gerenciamento baseado em estado finito, pois as ações são pré-determinadas e o diálogo com o usuário é direto. Já em H&F Companion, o diálogo é mais aberto e o gerenciamento é baseado na obtenção de informações para criar os planos de diálogos, seguindo a ideia do *frames*. E o HWYD adota ambas as abordagens um gerenciamento misto.

A iniciativa do diálogo corresponde a quem tem controle sobre o diálogo, o que geralmente é definido por quem faz as perguntas e quem responde. Em H&F Companion, o sistema é quem faz as perguntas e dá sugestões e, por isso, a iniciativa do diálogo é do sistema. Assim como no FalaBrasil, onde o sistema guia o usuário ao longo das transições de estado, o que é diferente no OpenPhone, onde o usuário que decide o assunto do diálogo ao escolher a opção no menu do sistema. Em HWYD, o controle do diálogo é do usuário, pois o sistema apenas reage às expressões deles.

O idioma é um dos aspectos utilizado nessa comparação, pois há poucos sistemas de diálogo diferentes do inglês. Mais precisamente, busca-se SDS no idioma português brasileiro, o que foi encontrado apenas no trabalho do FalaBrasil.

Por fim, compara-se se o SDS realiza alguma compreensão da fala mais detalhada, o que normalmente envolve *Processamento de Linguagem Natural* (PLN) no diálogo. O OpenPhone não necessita desse tipo de processamento, pois a participação do usuário no diálogo é limitada. Assim como no FalaBrasil, que apenas recebe instruções de buscas que não precisam de interpretação detalhada. Em H&F Companion e em HWYD, o diálogo aberto exige uma compreensão mais específica. Contudo, o H&F Companion utiliza um tipo de interpretação semântica para o reconhecimento da fala, o que não envolve técnicas de processamento de linguagem natural. Diferente do HWYD que utiliza várias dessas técnicas.

	OpenPhone	H&F Companion	HWYD	FalaBrasil
Domínio	Específico	Específico	Específico	Específico
Plataforma	Telefonia	Computador, Web e Móvel	Computador	Web
Gerenciamento	Estado Finito	Planos de Diálogo (Frames)	Misto	Estado Finito
Iniciativa	Usuário	Sistema	Usuário	Sistema
Idioma	Inglês	Inglês	Inglês	Português
Processamento	Não	Interpretação Semântica	PLN	Não

Tabela 1.1: Tabela Comparativa dos SDSs Analisados

A Tabela 1.1 mostra um breve resumo comparativo dos trabalhos utilizados como relacionados. Os critérios de avaliação foram baseados nos aspectos previamente analisados, e que também serão utilizados para caracterizar o SDS apresentado em nossa proposta. Assim, na seção 3.5 será retomada essa discussão, onde faremos uma nova comparação dos trabalhos apresentados com o proposto.

1.7 Conclusão

Neste capítulo são conhecidos alguns conceitos de sistemas de diálogo falado, que são aplicações que utilizam a fala como forma de interação com usuário e proporcionam um auto nível de acessibilidade por permitir sua utilização ainda que os olhos e mãos do usuário estejam ocupados. Esses sistemas também são conhecidos

como aplicações baseada na fala ou agentes conversacionais e costumam ser aplicados em atividades que exijam um alto número de interações com o usuário.

Vários componentes e várias áreas de pesquisa são combinadas para compor um SDS efetivo. Por isso, seu desenvolvimento não é trivial. Cada componente possui características individuais e aspectos que dificultam seu funcionamento, o que demanda atenção redobrada na utilização de cada um deles.

Sua aplicação na saúde é bastante promissora, pois é capaz de superar necessidades do usuários com relação ao tempo de consultas e no alcance dos atendimentos, além do acesso a conteúdo especializado.

Na área acadêmica os sistemas de diálogo falado ainda são pouco difundidos, apesar de ser um assunto muito comentado nos últimos anos devido a popularização de assistentes pessoais virtuais. Nos trabalhos relacionados citados é possível perceber que eles podem ser aplicados para diversas finalidades e podem ser desenvolvidos de diversas formas, mas seu uso ainda é propenso para ambientes não-móveis, pois requer muitos recursos computacionais para funcionar adequadamente. O idioma português brasileiro também é pouco adotado por esse tipo de sistema e não foi encontrado nenhum trabalho que pudéssemos relacionar.

2 Projeto MobileHealthNet

Neste capítulo apresenta-se o projeto MobileHealthNet, no qual o presente trabalho está inserido. O MobileHealthNet viabiliza, entre outras coisas, a construção de redes sociais móveis dedicadas aos cuidados à saúde e, por isso, iniciamos o capítulo com uma breve fundamentação teórica sobre este ambiente e os benefícios esperado pelo uso dele. Em seguida, descrevemos maiores detalhes sobre o projeto, abordando seus objetivos, características e aplicações previstas. Em especial, apresentamos a aplicação *Patient-Buddy-Build* (PBB), que será utilizada como estudo de caso no processo de avaliação do *middleware* proposto neste trabalho de mestrado.

2.1 Redes Sociais Móveis na Saúde

Com a evolução dos dispositivos móveis várias tarefas que antes só podiam ser feitas no computador, como acessar Internet, agora são realizadas no celular. Os *smartphones* estão cada vez mais baratos e com mais recursos, permitindo a execução de aplicativos cada vez mais sofisticados, dentre eles os que disponibilizam acesso às redes sociais.

De modo geral, a rede social é uma estrutura social cujos membros se relacionam em grupos e cuja interação é realizada através de tecnologias da informação e comunicação [42]. Neste ambiente, indivíduos de interesses semelhantes ou comuns podem interagir criando e visitando perfis, enviando mensagens, compartilhando arquivos multimídia ou trocando informações em comunidades. Estas interações permitem a quebra de barreiras geográficas e temporais devido ao acesso e interações simultâneas de diversas pessoas em qualquer lugar, possibilitando inclusive o intercâmbio de informações, colaboração e integração social para uma área específica de interesse, como a da saúde.

As *Redes Sociais Móveis* (RSMs) são uma extensão das redes sociais. Elas são caracterizadas pelo uso das redes sociais a partir de dispositivos móveis e tecnologias de comunicação sem fio (rede celular, rede local sem fio, bluetooth, etc), agregando

mobilidade ao usuário e ricas informações de contexto [42]. Seus conceitos podem ser aplicados a funções envolvidas na prática e entrega de cuidados médicos para estabelecer um rede social dedicada a atenção à saúde. Uma rede social na área da saúde pode ser definida como um grupo de pessoas (e a estrutura social que elas coletivamente constroem) que utiliza tecnologias da informação e comunicação com o propósito de conduzir coletivamente ações relacionadas à assistência médica e sua educação [8]. Estas ações podem incluir o próprio provimento de serviços de saúde, a educação em saúde envolvendo profissionais e pacientes, uma plataforma para o suporte e discussão sobre questões e problemas relacionadas a tratamentos, compartilhamento de documentos e consultorias com especialistas. Além disso, a manutenção do contato e relacionamento entre as pessoas envolvidas no processo de atendimento à saúde pode se prolongar além dos encontros presenciais. Entre os agentes que possam estar envolvidos nesse processos cita-se os profissionais da saúde (por exemplo, médicos, enfermeiros, fisioterapeutas, terapeutas ocupacionais), pesquisadores da saúde (professores e alunos de graduação e pós-graduação), pacientes e seus familiares, bem como membros da comunidade em geral [7].

Quando devidamente estabelecidas, RSMs podem contribuir com o processo de atendimento à saúde, tornando-o mais eficaz e eficiente. Entre os benefícios esperados, destacam-se um melhor fluxo de informação e maior colaboração entre profissionais dos diversos níveis de atendimento à saúde; melhorias nos níveis de comprometimento e informação do paciente, o que contribui com o processo terapêutico [17]; melhor gestão de pacientes portadores de doenças crônicas, levando a diminuição da ocorrência de complicações; melhoria na qualidade da tomada de decisão relativa ao tratamento por parte dos pacientes; melhor suporte emocional aos pacientes e redução de custos do sistema de atendimento à saúde, diminuindo-se a necessidade de deslocamentos e a ocorrência de complicações no tratamento de pacientes [11].

2.2 O Projeto MobileHealthNet

O projeto MobileHealthNet [37] [39] tem por objetivo desenvolver um *middleware* que permita a construção de RSMs e facilite o desenvolvimento de serviços colaborativos para o setor da saúde, a troca de experiências e a comunicação entre

pacientes e profissionais da saúde, além de uma melhor gestão dos recursos da saúde por órgãos governamentais.

Este projeto está sendo desenvolvido em parceria entre o Laboratório de Sistemas Distribuídos (LSD) da Universidade Federal do Maranhão (UFMA) e o *Laboratory for Advanced Collaboration (LAC)* da Pontifícia Universidade Católica do Rio de Janeiro (Puc-Rio) e conta com apoio institucional do Hospital Universitário da UFMA (HUUFMA). Em particular, duas unidades do HUUFMA estão diretamente envolvidas com o desenvolvimento do projeto: o Programa de Assistência a Pacientes Asmáticos (PAPA) e a Casa da Dor. O PAPA possui profissionais com especialidade em pneumologia e tem como objetivo principal o tratamento de pacientes portadores desta doença crônica. A Casa da Dor, por sua vez, possui profissionais especializados no tratamento de pacientes que sofrem com dores agudas, independentemente de sua etiologia.

O projeto foi concebido para ser aplicado em especial a comunidades carentes e distantes, em que o custo do deslocamento de pacientes para os postos de atendimento é alto devido as suas condições de acesso. Dessa forma, pode-se explorar o uso de redes sociais móveis na prática e entrega de cuidados médicos. Alguns aspectos adotados no desenvolvimento do MobileHealthNet e das aplicações, definidas a partir de interações com profissionais da computação e da saúde, foram:

- Componentes de software para dispositivos móveis devem apresentar baixo consumo de recursos, de maneira que possam ser executados em uma grande variedade de equipamentos, incluindo aqueles considerados de baixo custo, já que os resultados do projeto devem ser aplicáveis a comunidades carentes;
- Deve-se prover suporte a diversos tipos de comunicação sem fio (em especial sistemas celulares e redes locais sem fio), de forma a se atingir uma ampla área de cobertura, minimizando-se custos de comunicação sempre que possível para não gerar gastos ao usuário;
- A plataforma móvel visada é o Android, pois está presente em maior quantidade nos dispositivos móveis populares atuais, tornando os serviços acessíveis ao maior número de pessoas;

- As aplicações devem respeitar critérios de privacidade e segurança, para não expor informações dos pacientes e de seus tratamentos;
- Fornecer opções de acessibilidade que permitam ao usuário utilizar os serviços oferecidos sem dificuldades.

Este último é o principal aspecto que atacaremos neste trabalho dissertativo, mas sem ignorar os demais citados. Para isso, propõe-se interações baseadas na fala para proporcionar acessibilidade aos usuários no uso das aplicações móveis. No MobileHealthNet especificamente, estas aplicações móveis são desenvolvidas com base em serviços básicos obtidos através do *middleware* do projeto, conforme descrito a seguir.

2.3 O *middleware* MobileHealthNet

O *middleware* construído no projeto MobileHealthNet fornece recursos para o desenvolvimento de RSMs e a criação de novos serviços e aplicações voltadas para a área da saúde. Este *middleware* possui arquitetura organizada em camadas, que são ilustradas na Figura 3.1.

A camada de aplicações (*Applications*) refere-se às aplicações previstas no projeto para serem desenvolvidas na etapa inicial do MobileHealthNet. São elas: *Health Education*, *Professional Collaboration*, *HUPD Care* e *Patient-Buddy-Build*. Mais detalhes sobre elas serão apresentados na Seção 2.5.

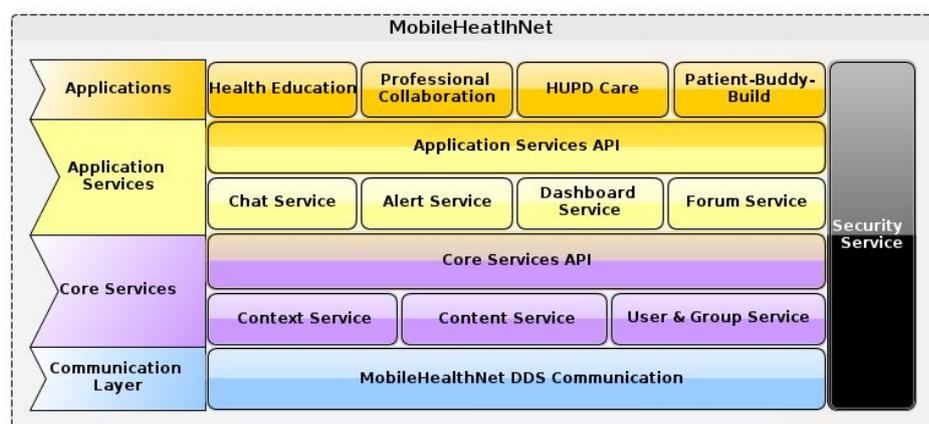


Figura 2.1: Arquitetura Geral do *Middleware* MobileHealthNet

A camada de Serviços de Aplicação (*Application Services*) disponibiliza serviços típicos de redes sociais para serem utilizados pelas aplicações da camada superior, como serviços de publicação de mensagem em murais, fórum e chat, além de um serviço de notificações (alertas) para os usuários.

A camada de serviços do núcleo (*Core Services*) disponibiliza serviços básicos que são compartilhados pelos demais serviços e aplicações do MobileHealthNet. O Serviço de Contexto (*Context Service*) é o responsável pelo armazenamento e disponibilização de informações de contexto. Por exemplo, as informações de localização de um determinado usuário podem ser compartilhadas com os outros usuários da rede social. O Serviço de Conteúdo (*Content Service*) tem como principal tarefa o compartilhamento de mídia (texto, fotos, áudio, filmes, etc) entre usuários da rede social. Este serviço permite rotular cada elemento de mídia com metainformações definidas pela aplicação como, por exemplo, a posição geográfica na qual a mídia foi obtida ou o modelo de uma bomba de asma ao qual um tutorial se refere. O Serviço de Usuários e Grupos (*User & Group Service*) é responsável pelo cadastro e gerenciamento de usuários e grupos na rede, bem como os relacionamentos entre os usuários da rede, o que é característico das redes sociais. Como exemplo, podemos citar o relacionamento de "amizade" entre os usuários, os quais podem enviar solicitações de amizade uns aos outros, podendo ser aceito ou não pelo usuário solicitado.

A camada de segurança *Security Services* é transversal a todo o código gerado no MobileHealthNet e disponibiliza os componentes responsáveis pelos mecanismos de privacidade e segurança implementados no *middleware* [12] [11].

Por fim, a camada de comunicação (*Communication Layer*) provê os mecanismos de comunicação necessários para o compartilhamento de dados no MobileHealthNet, cujos detalhes são descritos a seguir

2.4 Comunicação no MobileHealthNet

A infraestrutura de comunicação do MobileHealthNet proposta em [2] e apresentada em [7], provê uma interface de comunicação *publish/subscribe* voltada a colaboração em RSMs. Esta baseia-se em dois protocolos, o *Mobile Reliable UDP*

(MRUDP) [6] e o *Data Distribution Service* (DDS) [14], e é intermediada por *gateways* que fazem a transcodificação entre esses protocolos, conforme mostra a Figura 2.2.

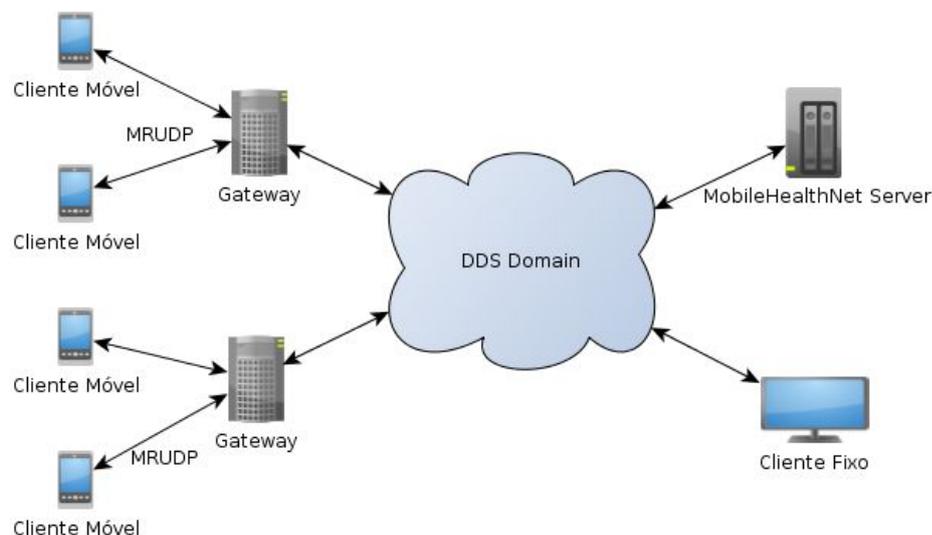


Figura 2.2: Arquitetura da Infraestrutura de Comunicação do MobileHealthNet

O MRUDP é a base para comunicação entre clientes móveis e o *gateway*. Ele é baseado no protocolo UDP, mas diferencia-se por proporcionar confiabilidade na entrega das mensagens ao incluir uma confirmação de recebimento (um *ack*) na comunicação entre cliente e *gateway* e faz várias tentativas de entrega da mensagem antes de considerar a perda da conexão de um deles [6]. Além disso, ele foi customizado para lidar com conectividade intermitente, travessia de *Firewalls/NAT*, suporte a mudança de endereços de IP e um reduzido consumo dos recursos computacionais dos dispositivos móveis. Para enviar mensagens MRUDP, os clientes móveis utilizam uma API do *MobileHealthnet on Android* (MOBHA), que fornece interfaces e abstrações necessárias para a comunicação com o *gateway*.

A partir do *gateway*, a comunicação no núcleo da infraestrutura do MobileHealthNet utiliza o protocolo DDS. O DDS é uma especificação da OMG [14] (*Object Management Group*) e é um padrão para comunicação *publish/subscribe* com alta qualidade de serviço que visa a distribuição crítica de informações em sistemas distribuídos em tempo real. Ele foi idealizado sobre um modelo *Data-Centric Publish-Subscribe* (DCPS) baseado em tópicos, onde a comunicação é gerenciada por um *middleware* que administra automaticamente a entrega das mensagens dentro de um determinado domínio. Assim, o desenvolvedor se preocupa somente com o conteúdo que será transmitido, pois a comunicação não requer qualquer intervenção das

aplicações no processo de comunicação. O CoreDx¹ é o *middleware* DCPS utilizado para gerenciamento da comunicação no MobileHealthNet.

O domínio DDS é um espaço lógico onde as aplicações enviam e recebem dados. Somente participantes do mesmo domínio podem se comunicar, o que ajuda a isolar e otimizar a comunicação dentro de uma comunidade que partilha interesses em comum. Esta comunicação é realizada através da publicação e subscrição em tópicos, que são uma representação das mensagens. Um tópico conecta um *data writer* (aquele que publica os dados) a um *data reader* (aquele que se inscreve e recebe os dados). A comunicação ocorre somente se o tópico publicado por um *data writer* equivaler ao tópico subscrito pelo *data reader*. Esta comunicação é anônima e transparente, pois os *publishers* e *subscribers* não precisam se preocupar como os tópicos são criados nem com quem está escrevendo ou lendo o tópico, uma vez que o *middleware* DCPS DDS gerencia estas questões.

O DDS disponibiliza recursos que podem ser utilizados como filtros no processo de subscrição. Os filtros são descritos através de expressões com uma sintaxe bem definida, as quais são baseadas nos atributos de cada tópico. Os filtros são essenciais, pois diminuem a carga com tratamento de mensagens que não são do interesse do usuário. Por exemplo, um usuário pode estar interessado em receber somente os tópicos publicados pelo usuário cujo atributo "user_name" seja igual a "João". Esta restrição é então expressa através de um filtro.

Na Seção 3.3.3 retomaremos essa discussão sobre comunicação no MobileHealthNet, pois apresentaremos como nossa proposta fará uso da mesma.

2.5 Aplicações do MobileHealthNet

Atualmente, três aplicações estão em desenvolvimento para o projeto MobileHealthNet. Essas aplicações foram priorizadas devido as necessidades das unidades do Hospital Universitário, parceiras do projeto, e abordam o uso do *middleware* em propósitos diferentes. São elas:

¹<http://www.twinoakscomputing.com/coredx>

- *Health Education* - possui ferramentas para o compartilhamento de arquivos multimídia e seu objetivo é aprimorar a educação de pacientes e profissionais da saúde através de mídias com conteúdo educacional.
- *Professional Collaboration* - visa diminuir a distância entre os profissionais da saúde através de recursos como *chat* multiusuário, fórum de discussão e serviço de notificação, permitindo assim informar sobre a urgência de se obter o resultado de um exame, discutir a respeito do tratamento e acompanhamento de um determinado paciente.
- *HUPD Care* - tem por objetivo explorar os conceitos das redes sociais móveis para promover a assistência remota de especialistas responsáveis pelo atendimento de alta complexidade a profissionais da atenção básica a saúde no atendimento a casos específicos.

Uma quarta aplicação já foi finalizada, a PBB, cujos detalhes descritos a seguir.

2.5.1 Patient-Buddy-Build

O PBB é uma ferramenta para geração de aplicações móveis contendo um questionário customizado para o monitoramento a distância de pacientes com doenças crônicas [26]. Na Figura 2.3 temos o cenário de usabilidade do PBB que ilustra como esse monitoramento acontece.

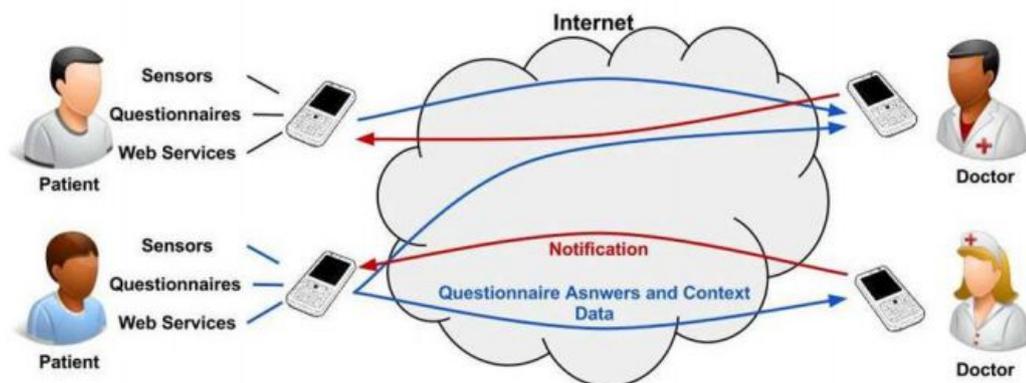


Figura 2.3: Cenário de Usabilidade do PBB

Fonte: [26]

Neste cenário, cada paciente possui um *smartphone* Android com uma aplicação móvel responsável por coletar seus dados. O principal objetivo dessa aplicação é obter informações utilizando sensores (adaptados no paciente ou carregados por ele), serviços Web (exemplo, serviço de previsão do tempo, caso o médico necessite saber das condições climáticas do local onde o paciente se encontra) e questionários (gerados automaticamente pela aplicação e respondidos pelo paciente).

Os questionários são customizados (variam de paciente a paciente) e sua criação é baseada em parâmetros e descrições formais, como as preferências do paciente, tipo de doença crônica, o processo de monitoramento exigidos pelo médico, medicação prescrita e informações sobre o contexto do paciente (exemplo, clima). Eles são gerados no servidor do PBB e são enviados para a aplicação através da Internet móvel. Entre as perguntas padrão que podem ser utilizadas pelo servidor na criação do questionário personalizado, temos:

- Você está com febre?
- Você está sentindo algum tipo de cansaço que não costumava sentir?
- Está com falta de ar?
- Você desmaiou recentemente?
- Você costuma sentir dores de cabeça?
- Está sentindo dores no peito?
- Você vomitou alguma vez hoje?
- Sentiu alguma palpitação na região do coração recentemente?
- Você possui plano de saúde?

Essas perguntas devem ser rápidas e práticas de se responder. Por isso, aceitam como resposta apenas Sim ou Não. Contudo, algumas delas exigem mais detalhes sobre a condição do usuário (exemplo, se o paciente está sentindo dores no peito, qual a intensidade dessas dores?) e, por isso, baseiam-se em respostas de múltipla escolha do tipo “Muito Intenso”, “Pouco” ou “Nenhum”. A definição da intensidade é importante para alertar ao médico sobre a criticidade da situação do

paciente que está sendo monitorado e, por isso, também podem ser utilizadas como resposta.

Uma vez que gerado, o questionário é preenchido pelo paciente com sua condição e possíveis sintomas. Essas informações serão enviadas pelo aplicativo móvel, juntamente com os dados dos sensores, para o médico responsável que avaliará suas condições de saúde. Contudo, ressalta-se que o PBB não apresenta diagnósticos, mas apenas fornece meios para os médicos obterem informações atualizadas sobre seus pacientes, de forma a permitir o monitoramento remoto dos mesmos.

O preenchimento correto do questionário é de grande importância para o monitoramento preciso do paciente. Por isso, meios de facilitar o uso do PBB pelo usuário são bem aceitos. Nessa perspectiva, utilizamos a aplicação PBB em um estudo de caso em que avalia-se a facilidade de interação que o MHNSS pode proporcionar, conforme descrito no capítulo 4 de avaliação dessa dissertação.

2.6 Conclusão

Neste capítulo abordamos conceitos de redes sociais móveis, que quando devidamente estabelecidas permitem o acesso e interação simultânea de diversas pessoas em qualquer lugar, possibilitando inclusive o intercâmbio de informações, colaboração e integração social para uma área específica de interesse, como a da saúde. Sua importância para este trabalho deve-se ao projeto MobileHealthNet, que faz parte desse ambiente.

O projeto MobileHealthNet viabiliza, entre outras coisas, a construção de aplicações dedicadas aos cuidados à saúde a partir de serviços básicos de RSMs. Ele é um projeto destinado a comunidades carentes e parte dos seus potenciais usuários possuem dificuldades para ler, escrever (digitar) e utilizar aplicações móveis de forma geral. Por isso, necessita-se de opções de acessibilidade que permitam aos usuários utilizar os serviços oferecidos pelo projeto MobileHealthNet sem dificuldades, que é o objetivo deste trabalho dissertativo.

Foi apresentado ainda o *middleware* MobileHealthNet, descrevendo as camadas de sua arquitetura e seus serviços oferecidos. Estes serviços são utilizados

pelas aplicações móveis criadas a partir do *middleware* para executar atividades típicas de redes sociais, como publicar mensagens em mural, forum e chat.

Por fim, o capítulo listou as aplicações previstas para desenvolvimento utilizando os recursos e serviços do *middleware* MobileHealthNet, das quais uma já foi finalizada e será utilizada no estudo de caso apresentado na seção de avaliação da proposta.

3 MobileHealthNet Speech Service

Este capítulo apresenta o *middleware* proposto para o desenvolvimento de aplicações móveis com recursos que possibilitam sua interação através da fala, denominado MHNSS. Para melhor compreensão do *middleware*, iniciamos o capítulo detalhando e caracterizando sua utilização, seguido por uma descrição dos requisitos elicitados que expõem o contexto para o qual o *middleware* foi concebido. Em seguida os componentes de sua arquitetura são apresentados, ressaltando ainda alguns aspectos de sua implementação. Ao final, fazemos uma análise comparativa do MHNSS com os trabalhos relacionados, onde demonstra-se o diferencial e as contribuições da presente proposta.

3.1 O MHNSS

A maior parte das aplicações móveis são desenvolvidas com base em textos e imagens de vários formatos, necessitando assim que sua interação com o usuário seja realizada a partir do enfoque visual. Essas aplicações não apresentam nenhuma outra opção de acesso ao usuário e muitas vezes não utilizam nem mesmo os efeitos sonoros para ajudar na navegação, o que restringe seu uso apenas às pessoas que conseguem enxergar bem a tela do celular e ler seus textos. Ainda assim, mesmo entre essas pessoas que conseguem ver e ler os textos, há aquelas que são incapazes de manusear o dispositivo móvel devido a problemas de articulação motora, deficiência física ou simplesmente estão com as mãos ocupadas. Por isso, melhorias de acessibilidade nessa área são tão necessárias.

O *MobileHealthNet Speech Service* (MHNSS) é um *middleware* que fornece suporte ao desenvolvimento de aplicações móveis com recursos de acessibilidade através da fala. Seu objetivo é proporcionar uma melhor usabilidade para as aplicações móveis e facilitar o acesso a elas. Para isso, o *middleware* conta com um sistema de diálogo falado que proporciona às aplicações interagirem com o usuário utilizando a

fala em diversas situações, como para coletar informações através de perguntas, ler textos ou receber comandos para executar atividades específicas.

Como exemplo do uso do MHNSS, podemos considerar um cenário em que um usuário chamado José necessita enviar uma mensagem de texto, mas ele não tem habilidade de escrever utilizando as teclas pequenas de seu *smartphone*. Ele nem mesmo sabe como utilizar a aplicação que envia mensagens, pois nunca precisou utilizá-la antes. Contudo, a aplicação foi desenvolvida considerando os recursos do MHNSS e José se surpreendeu ao ouvir a aplicação falar-lhe após iniciada: "Olá. Qual mensagem desejas enviar?". Sem acreditar muito no que estava acontecendo, José falou: "Querida, chegarei atrasado para o jantar". A aplicação então falou novamente: "E para quem desejas enviar essa mensagem?". Assim, José falou um número de celular e ficou ainda mais surpreso ao ouvir da aplicação por fim: "Mensagem enviada com sucesso!".

Neste cenário, José conseguiu enviar uma mensagem de texto sem precisar digitar nada. Ele apenas dialogou com a aplicação que utilizava os recursos do *middleware* como meio de interação. Isso proporcionou um nível de acessibilidade a José que superou sua falta de habilidade para digitar textos no *smartphone* e sua inexperiência com o uso de uma nova aplicação. E, além disso, é possível que esta operação tenha ocorrido em menor tempo se comparado com José digitando a mensagem nos meios tradicionais.

As aplicações que necessitam desse tipo de interação são aquelas que, geralmente, coletam informações do usuário para um determinado propósito como, por exemplo, para preencher um formulário, fazer uma busca na Internet ou escrever um e-mail. Com o uso do MHNSS elas deixam de ser passíveis em ficar aguardando uma informação do usuário a qualquer hora e tornam-se ativas na interação a partir do momento que são iniciadas. Isso acontece por que elas precisam acionar os comandos do MHNSS para realizar uma interação, mesmo que seja no turno do usuário, pois é a aplicação que chama o serviço de captura do áudio.

Para o usuário, além da acessibilidade, o MHNSS proporciona praticidade no uso da aplicação, pois ele não precisa nem mesmo ler as mensagens que são exibidas na tela, já que a aplicação pode reproduzi-las via voz. Sua participação na interação geralmente resume-se a responder perguntas ou conceder informação para as

aplicações, o que é suficiente para realizar inúmeras atividades que vão desde buscar um contato na agenda até mudar configurações do celular. Só depende de como a aplicação foi desenvolvida e em como ela utiliza o MHNSS.

O *middleware* MHNSS está inserido no contexto do MobileHealthNet e se propõe a facilitar a interação e uso das aplicações móveis desenvolvidas pelo projeto. Para isso, o MHNSS foi inserido no *middleware* MobileHealthNet como serviço de diálogo (*Speech Service*), conforme ilustrado na Figura 3.1, para que as interações das aplicações do MobileHealthNet possam ser realizadas através do diálogo falado.

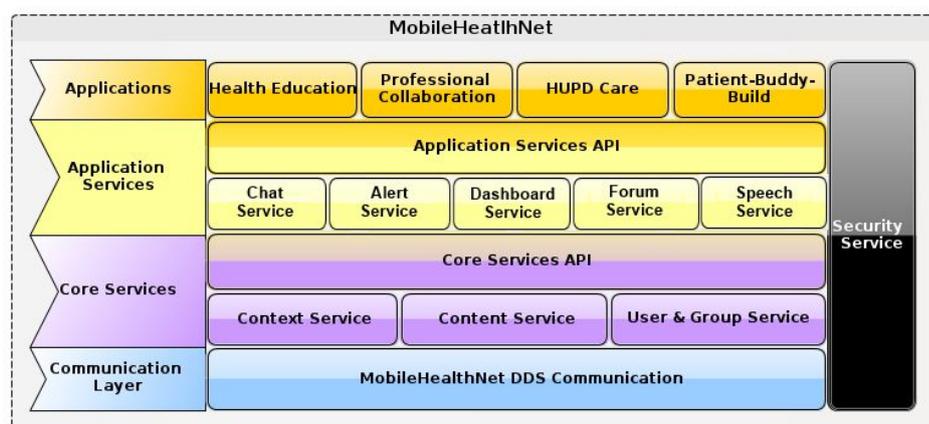


Figura 3.1: Arquitetura do *Middleware* MobileHealthNet com o Speech Service

Apesar de ser referenciado na arquitetura do MobileHealthNet como serviço, o MHNSS pode ser instanciado e utilizado de forma independente do MobileHealthNet, o que justifica chama-lo de *middleware* MHNSS em nossa proposta. Este conceito é reforçado pelo fato de qualquer aplicação móvel - mesmo que não faça parte do projeto MobileHealthNet - poder utilizar os recursos do MHNSS, desde que tenha, no mínimo, a versão Android 2.3.3 (Gingerbread) e que possua conexão com a Internet.

Entre as opções de uso do MHNSS pelas aplicações do MobileHealthNet, podemos citar:

- No *Health Education*, o sistema de diálogo pode auxiliar o usuário na busca por arquivos, leitura de documentos e no uso de algum medicamento, equipamento médico ou como atuar em determinadas situações;

- Em *Professional Collaboration*, o MHNSS pode auxiliar na leitura e escrita de mensagens no bate-papo e fórum, possibilitando que o profissional da saúde utilize a aplicação mesmo que esteja com as mãos ocupadas;
- No *HUPD Care*, o MHNSS poderá auxiliar os profissionais da saúde no uso da rede social em geral, através do diálogo falado;
- E no *Patient-Buddy-Build*, os pacientes poderão preencher o questionário customizado utilizando a fala;

Para o desenvolvimento do MHNSS, considerou-se alguns requisitos pertinentes às limitações e necessidades do ambiente em que será aplicado, conforme descrito a seguir.

3.2 Requisitos

Opções de acessibilidade devem estar presentes em todo tipo de aplicação, mas são poucas as ferramentas e práticas que auxiliam as aplicações nesse sentido. No ambiente móvel elas são ainda mais difíceis de serem adquiridas, pois a falta de recursos computacionais, que é uma característica desse domínio, dificulta sua implementação. Essa dificuldade estende-se aos sistemas de diálogo falado, pois sua demanda por elevados requisitos computacionais faz sua aplicação no ambiente móvel bastante desafiadora.

Sabendo disso, para o desenvolvimento do MHNSS definiu-se alguns requisitos importantes a serem atendidos, mediante às necessidades e limitações do ambiente em que será aplicado, dos quais destacam-se:

- *Permitir seu uso em ambientes móveis a partir de dispositivos de baixo custo e restringir o consumo da bateria dos dispositivos móveis:* A duração da bateria ainda é um fator crítico para os dispositivos móveis, o que limita o uso de aplicações móveis exigentes em recursos como o SDS. Assim, além de considerar a capacidade de processamento para viabilizar o uso de sistemas de diálogo, a bateria do dispositivo também exige demasiada atenção, pois nos dispositivos móveis considerados de baixo custo, os recursos de hardware são ainda mais restritos, principalmente em níveis de bateria.

- *Controlar o uso da conexão com a Internet e limitar gastos com comunicação, licenças e serviços proprietários:* Alguns sistemas de diálogo utilizam servidores robustos para processar algumas etapas do diálogo, superando assim o fraco poder de processamento encontrado em alguns ambientes de aplicação, como o móvel [25]. Neste caso, conexões com a Internet ou com a rede local são utilizadas para se comunicar com o servidor de forma que as atividades que demandam mais processamento sejam realizadas fora do equipamento em que o sistema de diálogo está executando. Essa comunicação, contudo, necessita ser controlada, para não causar falhas no diálogo devido a quedas na conexão. No ambiente do MobileHealthNet, o uso da Internet exige ainda mais atenção, pois o projeto é destinado a comunidades carentes e possíveis gastos financeiros com comunicação, licenças e serviços proprietários devem ser evitados ou minimizados.
- *Disponibilizar o serviço para o maior número de pessoas e atender a um perfil diversificado de usuários:* Para tornar as aplicações disponíveis ao maior número de pessoas, os serviços e aplicações do MobileHealthNet são desenvolvidas na plataforma móvel Android, pois está presente em maior quantidade nos dispositivos móveis populares atuais. O perfil dessas pessoas é de difícil previsão e tende a ser diversificado, o que influencia diretamente no desenvolvimento de um sistema de diálogo, pois ele precisa ser capaz de atender a usuários de diferentes sexos, idades e níveis de escolaridade.
- *Utilizar o português brasileiro como idioma padrão:* A maior parte das ferramentas de desenvolvimento e das técnicas de processamento para sistemas de diálogo são baseadas e utilizam o inglês. Contudo, as aplicações e serviços do projeto MobileHealthNet são destinadas à população brasileira, exigindo assim que o sistema de diálogo trabalhe em idioma português brasileiro. Portanto, é necessário utilizar ferramentas e serviços que atendam ao idioma local.

Considerando esses requisitos, o MHNSS foi projetado e desenvolvido. O modelo de arquitetura adotado é um dos principais fatores que nos permite atingir esses objetivos, conforme descrito a seguir.

3.3 Arquitetura da Solução

Tendo em vista que o MHNSS deve auxiliar aplicativos móveis encontrados em dispositivos considerados de baixo custo e de configuração de hardware modesta, a arquitetura adotada no desenvolvimento do MHNSS segue uma abordagem descentralizada como forma de minimizar o consumo de recursos do dispositivo móvel. Nessa abordagem, uma parte do serviço é executada no próprio dispositivo para auxiliar as aplicações móveis em atividades simples, fornecendo recursos de diálogo que não dependem de muito poder de processamento. Já as atividades que exigem um diálogo mais completo, em condições computacionais mais intensas, são executadas em um servidor. A essas partes deu-se o nome de MHNSS-Mobile e MHNSS-Server respectivamente, conforme representado na Figura 3.2.



Figura 3.2: Modelo de Arquitetura Descentralizada do MHNSS

O maior benefício deste modelo descentralizado de arquitetura é fazer com que o serviço MHNSS encontrado no dispositivo móvel execute o mínimo de processamento possível e, assim, não dependa de muitos recursos de hardware e de bateria. Isso o torna compatível com qualquer especificação de dispositivos móvel, desde que atenda às versões de software exigidas. Por outro lado, é necessário estabelecer uma conexão via Internet com o servidor para que os processamentos mais intensos sejam realizados. Entretanto, o processamento do diálogo não é totalmente dependente do serviço servidor, pois o MHNSS-Mobile tem autonomia para tratar diversos diálogos complementares. Assim, nem sempre uma conexão com o servidor será necessária.

3.3.1 O MHNSS-Mobile

A aplicação cliente do MHNSS, o *MHNSS-Mobile*, é um serviço Android instalado no dispositivo móvel que contém os recursos básicos do *middleware* e se

comunica com o servidor quando necessário. A partir desse serviço Android, que fica executando em segundo plano no dispositivo móvel, as aplicações podem requisitar recursos para executar atividades diversas via fala, como fazer uma pergunta ao usuário, ler um texto para ele, capturar sua voz para inserir como texto em um *chat*, entre outras.

Conforme ilustrado na Figura 3.3, o MHNSS-Mobile possui três componentes principais: o Reconhecimento da Voz, Síntese Texto-para-Fala e Gerenciador de Diálogo-Mobile.

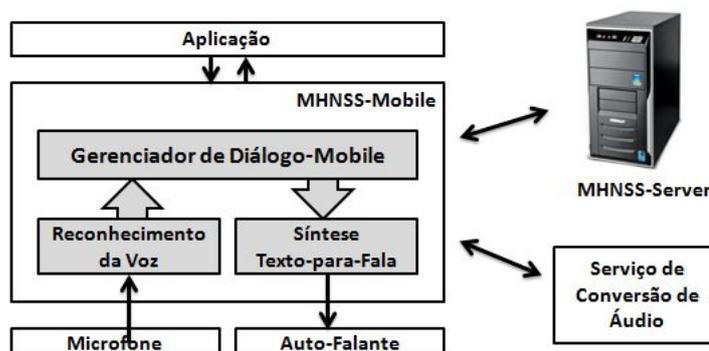


Figura 3.3: Arquitetura do MHNSS-Mobile

O componente de Reconhecimento da Voz é responsável por capturar o áudio de entrada e convertê-lo em texto para ser processado pelo gerenciador de diálogo. Este áudio é obtido através do microfone do dispositivo móvel e sua conversão em texto é realizada com o auxílio de um serviço de conversão de áudio, consultado através da Internet. A utilização desse componente é necessária sempre que a aplicação deseja obter alguma informação do usuário, como para saber qual mensagem ele deseja publicar no mural de uma rede social, por exemplo.

O componente de Síntese Texto-para-Fala é responsável por reproduzir o áudio de saída do sistema de diálogo. Para isso, ele recebe um texto do gerenciador de diálogo, transforma-o em sinal de áudio e o reproduz pelos alto-falantes do dispositivo móvel. Assim como o componente de Reconhecimento da Voz, ele também utiliza um serviço para a conversão do áudio, mas aqui a conversão é de texto para áudio. Este componente tem grande participação durante o diálogo, pois geralmente há mais turnos de interação por parte do sistema que do usuário.

O Serviço de Conversão de Áudio utilizado pelos componentes de Reconhecimento da Voz e Síntese Texto-para-Fala é um serviço de

terceiro. Na implementação atual, utiliza-se o *Dragon Mobile SDK* para realizar essas conversões (mais detalhes em 3.4.1). No entanto, o código do *middleware* foi estruturado de forma a poder ser estendido e contemplar outros serviços de conversão, como o nativo do Android. O uso de um Serviço de Conversão de Áudio é essencial para esta proposta descentralizada do MHNSS, pois os processos de conversões de áudio exigiriam muitos recursos de hardware caso fossem realizados no próprio dispositivo móvel. Além disso, ele é diretamente responsável pela precisão do reconhecimento e qualidade na reprodução da fala, o que influencia na avaliação geral do sistema.

O Gerenciador de Diálogo-Mobile é o componente central do MHNSS-Mobile. Ele faz o processamento do diálogo no dispositivo móvel, define qual a próxima ação a ser tomada pelo sistema e, caso precise, comunica-se com o servidor. Ele também gerencia os demais componentes do MHNSS-Mobile e os aciona sempre que necessário. Por exemplo, caso a aplicação queira ajuda para inserir uma mensagem do usuário em um *chat*, ela solicita ao Gerenciador de Diálogo-Mobile a captura da fala dele. Então, o Gerenciador de Diálogo-Mobile aciona o componente Reconhecimento da Voz, que captura o áudio da mensagem do usuário e o converte em formato de texto. Esse texto é entregue para o Gerenciador de Diálogo-Mobile que o repassará para a aplicação solicitante. Desta forma, ela poderá inserir a mensagem do usuário no *chat* e continuar suas atividades, assim como seria caso o usuário tivesse digitado sua própria resposta.

Para fazer uma solicitação ao Gerenciador de Diálogo-Mobile a aplicação móvel utiliza uma API que possui os métodos necessário para que ela utilize os recursos do MHNSS. Esta API foi incluída ao MOBHA e integra o conjunto de interfaces dos serviços do MobileHealthNet. A Figura 3.4 apresenta os métodos contido nessa interface.

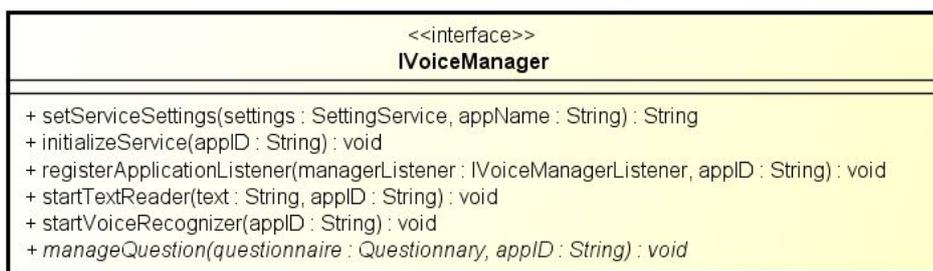


Figura 3.4: Métodos da Interface do MHNSS-Mobile

Na interface `IVoiceManager`, o `setServiceSettings` é um método utilizado pela aplicação para passar dados de configurações do serviço de conversão utilizado. Além disso, este método faz um registro da aplicação solicitante no serviço MHNSS e gera um identificador único (*appID*) que associa aquela aplicação à configuração setada. Este *appID* é utilizado pela aplicação para acessar os outros métodos do serviço, como o `initializeService`, que ativa o serviço MHNSS e checa se as configurações previamente definidas para aquele *appID* ainda são válidas. Caso o `initializeService` não retorne erros, a aplicação pode registrar um *listener* através do método `registerApplicationListener` para receber o status da atividade em execução, os logs de interação e as respostas obtidas, além das mensagens de erros apresentadas. Por fim, a aplicação pode solicitar: a reprodução de um texto em áudio, através do método `startTextReader`; a captura do áudio do usuário, através do `startVoiceRecognizer`; e fazer uma pergunta pro usuário ou deixar que o sistema gerencie um questionário com ele, através do método `manageQuestion`.

Mais detalhes sobre o processamento realizado no Gerenciador de Diálogo-Mobile e sobre os métodos `setServiceSettings` e `manageQuestion` são apresentados na Seção 3.4 de aspectos da implementação do MHNSS.

3.3.2 O MHNSS-Server

O *MHNSS-Server* é um serviço responsável por auxiliar o MHNSS-Mobile no processamento das etapas do diálogo que demandam maior poder computacional. Este serviço fica hospedado em um dos servidores da infraestrutura do MobileHealthNet e é composto por um gerenciador de diálogo mais robusto e um módulo de compreensão da linguagem, conforme ilustrado na Figura 3.5.

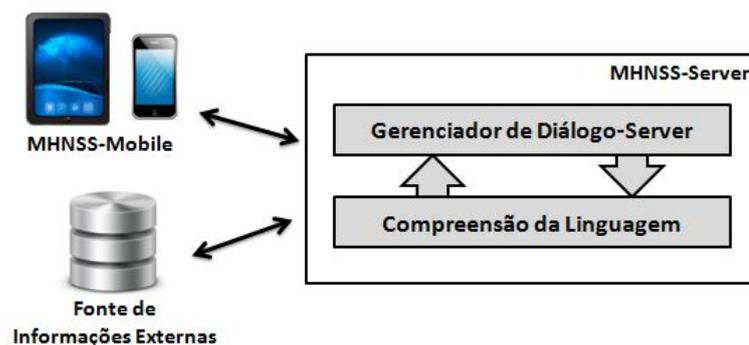


Figura 3.5: Arquitetura do MHNSS-Servidor

O componente Gerenciador de Diálogo-Server auxilia o Gerenciador de Diálogo-Mobile na tomada de decisões e no tratamento de erros através do processamento de textos do diálogo. Por estar situado em um servidor, sua capacidade de processamento é superior se comparado com os dispositivos móveis. Portanto, quando determinado processamento não puder ser realizado pelo Gerenciador de Diálogo-Mobile devido a sua simplicidade, o Gerenciador de Diálogo-Server é acionado.

As atividades que normalmente necessitam do auxílio do Gerenciador de Diálogo-Server são as que envolvem diálogo contínuo, como no preenchimento de um questionário. Os questionários costumam possuir múltiplas opções de respostas em que o usuário deve escolher uma delas. Assim, o sistema de diálogo interage com o ele quantas vezes forem necessárias até obter uma resposta correta, baseada nessas opções de respostas. Contudo, o usuário pode fornecer uma resposta diferente dessas opções esperadas ou responder utilizando palavras diferentes. Assim, o Gerenciador de Diálogo-Server é acionado para analisar essas mensagens adequadamente e tentar identificar se elas correspondem a uma opção válida de resposta ou não.

Para analisar as mensagens, o Gerenciador de Diálogo-Server conta com o auxílio do componente de Compreensão da Linguagem. Este componente tenta compreender determinado texto e atribuir-lhe um significado dentro do contexto do diálogo. Para isso, ele utiliza técnicas de PLN para fazer uma análise linguística do texto, além de utilizar Fontes de Informações Externas para otimizar a compreensão.

As Fontes de Informações Externas representam toda informação externa ao serviço que podem auxiliar no processamento do diálogo, como por exemplo, bancos de dados, dicionários e serviços de terceiros. Estas fontes de informações são fracamente acopladas ao MHNSS-Server, podendo assim serem utilizadas de acordo com a necessidade do desenvolvedor. Atualmente um dicionário de sinônimos é utilizado para auxilia no PLN realizado pelo componente de Compreensão da Linguagem, cujos detalhes são descritos na Subseção 3.4.3.

Dependendo do resultado do processamento realizado pelo componente de Compreensão da Linguagem, o Gerenciador de Diálogo-Server responde

ao MHNSS-Mobile sugerindo que ele tome a decisão mais apropriada para cada situação, que pode ser:

- Aceitar a resposta como válida: indica que a resposta do usuário condiz com uma resposta válida para a pergunta realizada. Isso ocorre quando o sistema espera um resposta específica (baseada nas respostas de referência enviadas pela aplicação), mas o usuário disse algo diferente. Então, após análise da resposta, constata-se que elas eram equivalentes e, por isso, é considerada como válida. Por exemplo, quando o sistema espera por uma resposta "Sim", mas o usuário respondeu "Positivo".
- Repetir a pergunta: indica que o sistema não foi capaz de compreender o texto, mesmo após analisado. A repetição é sugerida para coletar uma nova resposta do usuário, pois pode ter ocorrido erros no reconhecimento da fala que dificultaram a compreensão da resposta correta ou o usuário falou algo muito diferente do esperado. Contudo, a repetição só é indicada quando não houve outra melhor solução para sugerir.
- Confirmar uma resposta: indica que o sistema tem dúvidas sobre a resposta do usuário. Geralmente ocorre quando o texto analisado foi parcialmente compreendido ou quando a resposta obtida foi um pouco diferente da esperada. Por exemplo, quando espera-se "Sim", mas obteve-se "Também".
- Notificar erro: indica que, devido a um erro, o processamento não pôde ser iniciado/completado. Pode ocorrer devido a falta de informações suficientes para o processamento, como a ausências das fontes de informações externas.

Portanto, a partir de uma dessas sugestões enviadas pelo MHNSS-Server, o MHNSS-Mobile volta a interagir com o usuário, executando a ação mais adequada com a situação do diálogo.

3.3.3 A Comunicação no MHNSS

O MHNSS utilizada a infraestrutura de comunicação disponibilizada pelo projeto MobileHealthNet. Assim, a comunicação entre MHNSS-Mobile e MHNSS-Server é realizada através da publicação e subscrição de tópicos, onde quem quer

enviar uma mensagem publica o respectivo tópicos em um espaço global de dados e quem quer receber a mensagem se inscreve a este tópicos. A Figura 3.6 ilustra como é feita essa comunicação.

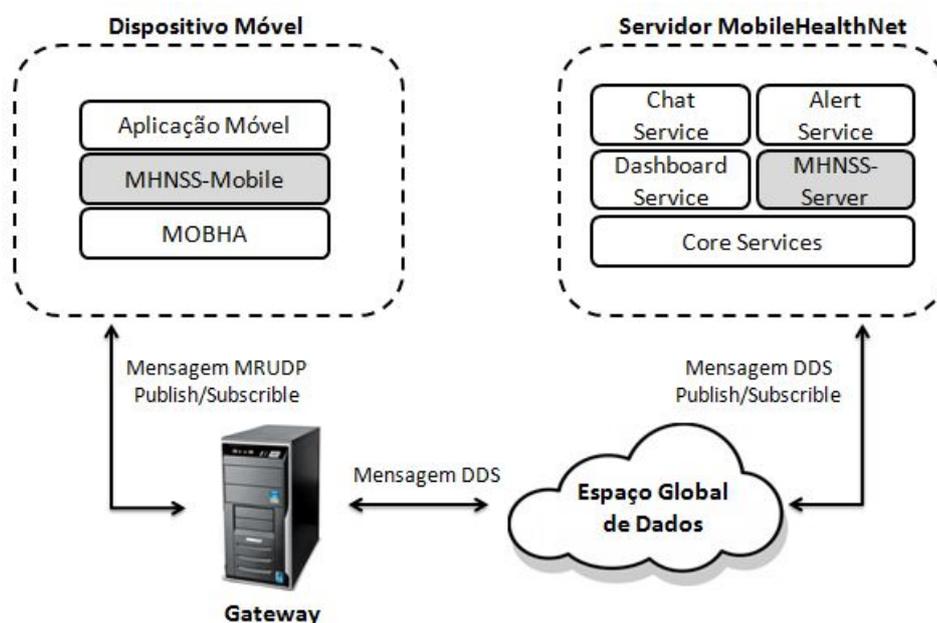


Figura 3.6: Arquitetura da Comunicação no MHNSS

No dispositivo móvel, os tópicos de publicação/subscrição do MHNSS-Mobile são criados/recebidos através do MOBHA, que auxilia dispositivos Android na comunicação com o MobileHealthNet e estrutura os tópicos em mensagens MRUDP, conforme foi descrito na Seção 2.4. Essas mensagens são a base da comunicação dos dispositivos móveis com os *gateways* e são transmitidas entre eles através da Internet. O *gateway* é responsável por codificar mensagens MRUDP em mensagem DDS e vice-versa, fazendo assim a ligação dos dispositivos móveis com o espaço global de dados onde encontram-se os tópicos. No espaço global de dados os tópicos são agrupados em domínios e tanto o MHNSS-Mobile quanto o MHNSS-Server pertencem ao domínio único do MobileHealthNet, permitindo assim a comunicação entre eles. A troca de mensagem com o espaço global de dados é realizada através de mensagens DDS e, pois isso, o *gateway* precisa fazer a codificação da mensagem MRUDP para esse padrão. Por fim, os serviços do MobileHealthNet, incluindo o MHNSS-Server, utilizam mensagens DDS para publicação/subscrição de tópicos diretamente no espaço global de dados e esta comunicação acontece via rede local, pois eles pertencem ao mesmo núcleo de rede na infraestrutura do MobileHealthNet.

Tópico	Conteúdo	Quando Utilizado
dialogueTextToProcess	appKey	Tópico utilizado pelo cliente para encaminhar uma resposta ao servidor para ser processada.
	question	
	userAnswer	
	expectedAnswers	
processedText	appKey	Tópico utilizado pelo servidor para informar ao cliente que a resposta processada é válida.
	question	
	userAnswer	
	secure	
dialogueNotification	appKey	Tópico utilizado pelo servidor para notificar algo ao cliente, como para que repita a pergunta.
	notificationMessage	
	secure	
dialogueConfirmation	appKey	Tópico utilizado pelo servidor para sugerir que o cliente confirme uma resposta com o usuário.
	question	
	confirmAnswer	
	secure	
dialogueError	appKey	Tópico utilizado para alertar o cliente sobre um erro no processamento.
	errorMessage	
	secure	

Tabela 3.1: Tópicos Utilizados na Comunicado do MHNSS

Quando o MHNSS-Mobile e o MHNSS-Server são ativados, eles se inscrevem a seus respectivos tópicos de interesse no domínio MobileHealthNet. Os tópicos utilizados por eles estão listados na Tabela 3.1. Assim, quando a comunicação for necessária, basta publicarem no tópico equivalente ao tipo de mensagem que desejam transmitir. Por exemplo, para que o MHNSS-Mobile envie uma mensagem para ser processada pelo servidor, ele publica no tópico `dialogueTextToProcess`. Neste tópico são passadas as informações necessárias para o processamento no MHNSS-Server, o qual está subscrito para receber essas informações sempre que forem publicadas. Entre os dados contido nos tópicos utilizados no MHNSS temos:

- *appKey* - Identificador único da aplicação, gerado ao se registrar no MHNSS-Mobile. Ele é utilizado como filtro para ajudar a selecionar um tópico específico durante o processo de comunicação;

- *question* - Indica a última pergunta realizada ao usuário e é referente à resposta que irá ser processada. Também é utilizada como filtro de identificação, pois apenas o *appKey* não é suficiente;
- *userAnswer* - Informa a resposta obtida do usuário e que necessita de uma compreensão mais detalhada;
- *expectedAnswers* - Indica a lista de respostas de referências informadas pela aplicação que auxiliarão no processamento.
- *notificationMessage* - Mensagem enviada pelo servidor para notificar ou sugerir algo à aplicação, como para que ela repita a última pergunta feita.
- *confirmAnswer* - Aponta uma das respostas de referências que o servidor considera equivalente à resposta do usuário, mas que precisa ser confirmada.
- *errorMessage* - Mensagem enviada pelo servidor para alertar o cliente sobre um erro.
- *secure* - Tag utilizada pela infraestrutura do MobileHealthNet para garantir a segurança e privacidade dos dados.

As mensagens recebidas pelo MHNSS-Server através do tópico `dialogueTextToProcess` são analisadas e processadas pelo componente Gerenciador de Diálogo-Server. Dependendo do resultado do processamento, ele responde ao MHNSS-Mobile utilizando o tópico referente a ação mais apropriada com o contexto do diálogo. Como a resposta do servidor pode variar, o MHNSS-Mobile se inscreve aos quatro tópicos utilizados pelo servidor no processo de resposta (`processedText`, `dialogueNotification`, `dialogueConfirmation`, `dialogueError`) e aguarda a atualização de um deles para então tomar uma decisão de como prosseguir no diálogo. Contudo, o servidor atende a requisições de vários MHNSS-Mobile diferentes e responde a eles enviando tópicos para o mesmo espaço global de dados. Para assegurar que apenas os respectivos destinatários recebam a resposta do servidor, utilizamos o atributo *appKey* no processo de subscrição para a filtragem dos tópicos. Ainda assim, há casos em que apenas o *appKey* não é suficiente, pois a mesma aplicação pode fazer várias requisições ao servidor e, por isso, também utilizamos o atributo *question* para alguns tópicos. Com isso é possível garantir

que apenas o destinatário recebera os respectivos tópicos, mesmo que haja vários assinantes para ele.

3.4 Aspecto da Implementação

Em consonância com a arquitetura apresentada na seção anterior, o *middleware* MHNSS foi implementado em linguagem de programação Java¹ e utilizando o Ambiente Integrado de Desenvolvimento (*Integrated Development Environment* (IDE)) Eclipse². Para a criação do MHNSS-Mobile, contou-se ainda com Android SDK³ que forneceu as bibliotecas e ferramentas necessárias para o desenvolvimento do serviço que será executado no dispositivo móvel do usuário. Vários componentes da arquitetura do MHNSS possuem aspectos de implementação que merecem atenção, conforme descritos nas próximas subseções.

3.4.1 Serviço de Conversão de Áudio

O serviço de conversão de áudio utilizado pelos componentes de Reconhecimento da Voz e Síntese Texto-para-Fala tem grande importância nessa proposta descentralizada de sistema de diálogo do MHNSS. Isso por que a conversão áudio-texto e texto-áudio necessita de processamento elevado para sua realização, o que dificultaria sua realização nos dispositivos móveis dos usuários do projeto MobileHealthNet. O uso de um serviço externo que faça essa conversão reduz o consumo de recursos no dispositivo móvel e possibilita a execução de sistemas de diálogo nesse ambiente.

Na versão atual do *middleware* MHNSS, o serviço de conversão utilizado é o *Dragon Mobile SDK*⁴, uma ferramenta proprietária de serviços de voz para aplicações móveis. Este serviço diferencia-se por ser específico para o ambiente móvel, fornecendo suporte a diversas plataformas de desenvolvimento e a um mecanismo de compressão de dados que possibilita seu uso em redes de baixa velocidade como a 3G. Este serviço de compressão foi um dos motivos que levaram a adesão do serviço de

¹<http://www.java.com>

²<http://www.eclipse.org>

³<http://developer.android.com/sdk/index.html>

⁴<http://www.nuance.com/for-developers/dragon-mobile-sdk/>

conversão a partir do MHNSS-Mobile, pois assim o tráfego de dados em formato de áudio na rede é reduzido, o que seria diferente se a conversão fosse realizada a partir do MHNSS-Server já que haveria tráfego de áudio até o servidor.

Outros serviços poderiam substituir o *Dragon Mobile SDK* como serviço de conversão de áudio, bastando apenas estender a classe de configuração do MHNSS (contendo endereço, porta e outras informações do servidor de conversão) e passá-la através do método `setServiceSettings` da *Application Programming Interface (API)* de diálogo do MOBHA. O serviço de conversão nativo do Android é outra opção para a conversão do áudio, mas que não foi utilizado nessa proposta por que não apresentava suporte ao idioma português no início do desenvolvimento.

A escolha do *Dragon Mobile SDK* como serviço de conversão deu-se pelo fato de requerer versão do Android 2.3 ou superior, sendo assim compatível com os requisitos mínimos das aplicações do MobileHealthNet e não limitando o serviço às versões do Android encontradas apenas nos dispositivos de alto custo. Além disso, ela fornece suporte ao idioma português brasileiro, os dados trafegados são criptografados, o áudio gerado é de fácil compreensão e a precisão de reconhecimento é satisfatória, atendendo assim aos requisitos do projeto.

3.4.2 Processamento no Gerenciador de Diálogo-Mobile

Gerenciador de Diálogo-Mobile é responsável por fazer o processamento do diálogo no dispositivo móvel e controlar os demais componentes do MHNSS-Mobile, entre outras coisas. Ele recebe requisições das aplicações móveis através dos métodos apresentados posteriormente na Figura 3.4 e se encarrega de gerenciar os eventos que intermeiam a atividade solicitada, retornando para a aplicação apenas o resultado, quando obtido. Entre esses métodos, o que merece maior atenção é o `manageQuestion`.

O `manageQuestion` é utilizado quando a aplicação necessitar de várias interações com o usuário (por exemplo, no preenchimento de um questionário) ou para realizar uma pergunta em que a resposta do usuário esteja de acordo com o esperado pela aplicação (por exemplo, em uma pergunta de múltipla escolha onde a resposta deve ser uma das opções sugeridas). Nestas ocasiões, um diálogo será estabelecido onde o usuário e sistema podem interagir por inúmeros turnos, até que

todas as respostas sejam identificadas. Para tal, a aplicação passa como argumento do método `manageQuestion` seu `appID` e uma instância da classe `Questionary`, onde é possível definir a pergunta (ou várias perguntas no caso de um questionário) e uma lista de respostas de referência para cada pergunta. Essas respostas de referência são utilizadas na etapa do processamento para auxiliar na identificação das respostas que a aplicação considera como satisfatórias. Por exemplo, em “Você está com febre?”, respostas de referências seriam “Sim” ou “Não” e qualquer resposta muito diferente disso não é aceita como válida. A partir dessas informações, o Gerenciador de Diálogo-Mobile irá seguir uma série de etapas para obter as respostas do usuário, o que normalmente envolve a realização da pergunta, o processamento das repostas e a reformulação da pergunta em caso de dúvidas do usuário. Essas etapas são baseadas no controle de diálogo por estado finito, onde há uma rede de transição de estados a serem seguidos, conforme ilustrado na Figura 3.7.

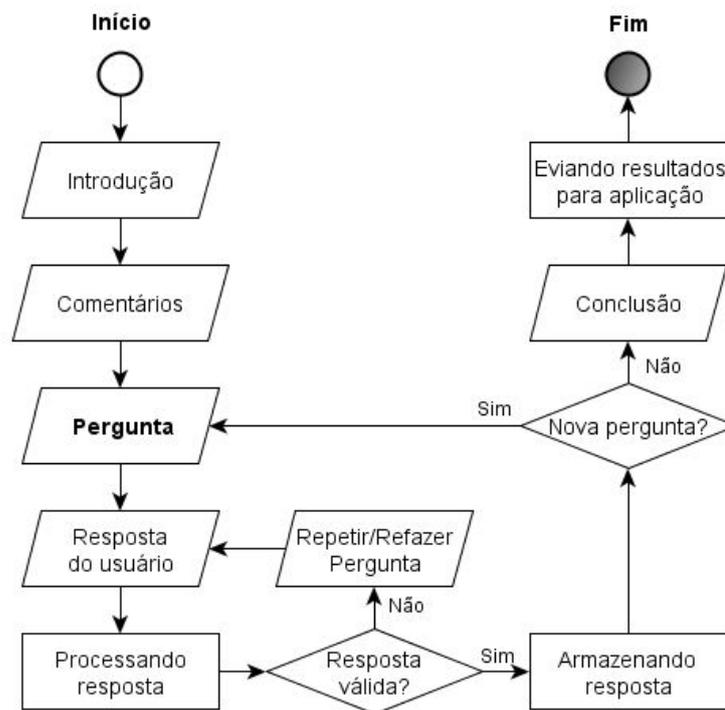


Figura 3.7: Fluxograma do Gerenciador de Diálogo-Mobile

Para fazer a pergunta, o Gerenciador de Diálogo-Mobile inicia a interação reproduzindo mensagens de Introdução e Comentários, que podem ser utilizadas opcionalmente como mensagens de apresentação, um resumo dos objetivos do questionário ou instruções de como o usuário deve responder corretamente. Em seguida, a primeira pergunta é reproduzida em forma de áudio, com o auxílio

do componente de Síntese Texto-para-Fala. Após a leitura da questão, o Gerenciador de Diálogo-Mobile aciona o componente de Reconhecimento da Voz para coletar a resposta do usuário e, uma vez que obtida, a resposta do usuário é processada com base nas respostas de referência. Caso esta seja uma resposta válida, ela é armazenada e a próxima pergunta (se houver) é realizada. Caso contrário, a pergunta é repetida ou refeita, dependendo no número de vezes que a mesma pergunta já foi realizada e do resultado do processamento da resposta. Ao final das perguntas e caso necessário, uma mensagem de Conclusão é reproduzida, seja com agradecimentos ou apenas para informar ao usuário o fim das questões. E, por fim, as respostas coletadas para cada pergunta são repassadas para a aplicação que solicitou o questionamento.

O processamento das respostas no Gerenciador de Diálogo-Mobile limita-se a verificar se a resposta obtida do usuário condiz com uma das respostas de referências. Caso elas sejam iguais, a resposta é aceita. Caso contrário, uma melhor compreensão da resposta será necessário e, então, o MHNSS-Server é acionado. Essa simplicidade no processamento é uma das opções tomadas no desenvolvimento do MHNSS para limitar a execução de atividades no dispositivos móvel, reduzindo assim a necessidade por recursos de hardware elevado e mantendo o consumo de bateria controlado.

3.4.3 Processamento de Linguagem Natural no MHNSS

O componente de Compreensão da Linguagem encontrado no MHNSS-Server utiliza técnicas de PLN para compreender determinado texto e auxiliar no processamento do diálogo. Atualmente quatro técnicas de PLN são utilizadas:

- O *End of Sentence Detection* é utilizada para identificar se o texto analisado é uma sentença composta, ou seja, se é formado por mais de uma frase. A separação das sentenças é necessária por que o *Tokenization* (próxima técnica) geralmente opera em sentenças individuais.
- A técnica *Tokenization* é utilizada para dividir o texto em *tokens* (unidades mais simples) como números, pontuações e palavras. Na maioria dos casos os espaços em branco e sinais de pontuação são usados como delimitadores na divisão dos

tokens. A partir dos *tokens* é possível fazer uma análise individual em cada palavra e conseguir uma melhor compreensão do texto.

- O **Stemming** consiste em reduzir uma palavra (*token*) a seu radical, ou seja, são retirados seus afixos (prefixos e sufixos). A partir dessa técnica é possível identificar variações linguísticas da mesma palavra (por exemplo, muit-o, muita, muit-íssimo) e, assim, tratá-las como um único termo em vez de termos diferentes. A *Stemming* auxilia ainda a identificar palavras sinônimas mais facilmente, uma vez que os dicionários geralmente não possuem variações linguísticas.
- O **Part-of-Speech Tagging** é responsável por fixar em cada *token* uma marcação (*tag*) de sua respectiva categoria gramatical como substantivo, adjetivo ou verbo. A partir dessas *tags* é possível diferenciar palavras ambíguas, ou seja, que possuem diversos sentidos ou múltiplos significados. Por exemplo, “paciente” pode ser um substantivo, e significar uma pessoa que está doente, ou um adjetivo, e significar uma pessoa que tem paciência.

Para processar determinado texto, nem sempre o MHNSS utiliza todas essas técnicas de PLN. Na maioria dos casos é possível compreender o texto com apenas algumas delas. Como exemplo, a Figura 3.8 ilustra uma sequência comum para processamento de um texto no MHNSS. Neste caso, o usuário foi questionado sobre “Você vomitou hoje?”. Como resposta, o usuário disse “Nenhuma vez hoje” e esse texto foi processado usando as técnicas de PLN.

A princípio, o texto de entrada passa por três técnicas de processamento de linguagem natural: *End of Sentence Detection*, *Tokenization* e *Stemming*. Em negrito, o exemplo mostra como fica o texto após a aplicação de cada técnica. Após a *Stemming*, o sistema faz uma verificação se alguma das palavras representa uma resposta satisfatória para a pergunta. Neste caso, o sistema busca por palavras afirmativas ou negativas. Caso essa verificação seja satisfeita, a resposta é dada como válida. Caso contrário, o processamento do texto continua através do uso da *Part-of-Speech Tagging* e da busca por palavras sinônimas em um dicionário. A *Part-of-Speech Tagging* auxilia a identificar sinônimos das palavras de acordo com a classificação gramatical delas naquela sentença. Através das palavras sinônimas, o sistema tenta identificar novamente palavras afirmativas ou negativas na resposta do

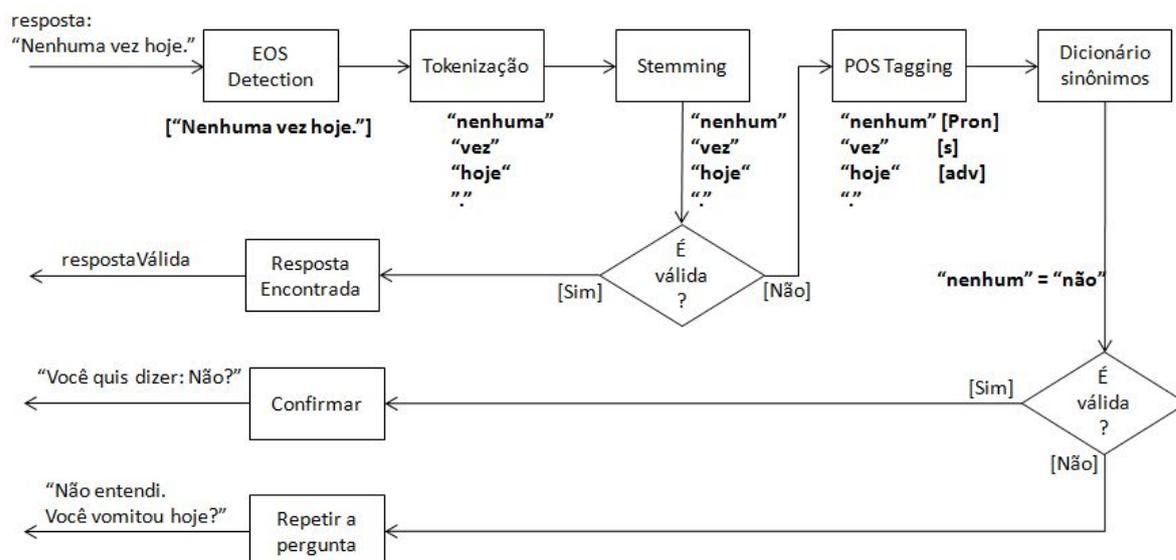


Figura 3.8: Processamento de Linguagem Natural no MHNSS

usuário. Contudo, caso encontre, a resposta não é dada como totalmente válida e, por isso, é necessário confirmar se o usuário realmente quis dizer aquilo. Esta é uma forma do MHNSS prevenir-se de tomar conclusões precipitadas, pois nem sempre a análise individual das palavras representa o sentido da frase. Por exemplo, as palavras da sentença “muito pouco” (uma possível resposta dada pelo usuário) se analisadas individualmente possuem sentidos opostos, o que pode confundir o sistema de diálogo e, por isso, é necessário confirmar com o usuário a sua real resposta. Por fim, caso nenhuma das técnicas tenha ajudado a compreender a resposta do usuário, a pergunta é refeita para coletar uma nova resposta.

Para o uso dessas técnicas, utilizamos a ferramenta *Natural Language Toolkit (NLTK)*⁵ que é uma biblioteca *Python open source* para processamento de linguagem natural. Esta ferramenta fornece suporte a PLN em vários idiomas, incluindo o português brasileiro. Contudo, para o uso adequado do *POS Tagging* em português, foi necessário treinar um *corpus* com textos analisados morfossintaticamente. Para tal, utilizamos um acervo disponibilizado pelo Projeto Floresta⁶, que possui mais de nove mil sentenças classificadas e revisadas por linguistas a partir de textos extraídos do jornal Folha de São Paulo. Com isso, foi possível treinar o *corpus* para fazer a classificação gramatical de diversas palavras e, assim, tornar o NLTK ideal para os propósitos do MHNSS.

⁵<http://www.nltk.org/>

⁶<http://www.linguateca.pt/Floresta/>

O dicionário de sinônimos aplicado para auxiliar no PLN foi construído em linguagem Java com base nas palavras mais utilizadas pelos usuários durante nosso testes. Essas palavras são as mais comuns para responde a perguntas diretas, como as realizadas pela aplicação *Patient-Buddy-Build*, e podem ser classificadas em:

- Respostas Afirmativas: sim, afirmativo, isso, positivo, exatamente, correto, certo;
- Respostas Negativas: não, jamais, negativo, nunca, nenhum;
- Respostas de Intensidade: intenso, forte, profundo, muito, pouco, médio.

A essas palavras inclui-se também suas variações, como fortemente, pouca, nenhuma, etc.

3.5 Análise Comparativa com os Trabalhos Relacionados

Na Seção 1.6 foram descritas alguns sistemas que utilizam diálogo falado para prover acessibilidade em diversos cenários. Nesta seção, confrontamos algumas características desses sistemas com o apresentado nessa proposta, fazendo uma análise comparativa entre eles.

O MHNSS foi desenvolvido para utilização em ambiente móvel e na área da saúde. Contudo, diferente dos sistemas apresentados nos trabalhos relacionados que são dependentes do domínio (por exemplo, no HWYD a compreensão do diálogo limita-se a assuntos de trabalho), o MHNSS utiliza componentes genéricos de diálogo de forma a torná-lo independente do domínio. Isso permite que seus recursos sejam reutilizados com mais facilidade por aplicações diversas e em diferentes situações.

A arquitetura adotada no desenvolvimento do MHNSS o torna mais apropriado para o ambiente móvel, pois foi projetada considerando as limitações dos dispositivos (exemplo, poder de processamento e nível de bateria) e da conexão móvel (exemplo, baixa velocidade e intermitência). A utilização de um protocolo de comunicação específico para a comunicação sem fio (o MRUDP, Seção 2.4) e a utilização de serviços externos para tarefas que exigem maior poder de processamento, como a conversão do áudio, são os principais responsáveis pela viabilização do diálogo em ambiente móvel. Assim, o MHNSS juntamente com o H&F Companion oferecem

sistemas de diálogo destinados a esse ambiente, apesar deste último não considerar as limitações do ambiente móvel como requisito de desenvolvimento.

Para gerenciar o diálogo em caso de questionários, o MHNSS realiza um processamento baseado em estado finito, onde ele segue uma rede de transição de estados para obter respostas do usuário de forma independente da aplicação. Isso quer dizer que ele é responsável por gerenciar os processos que intermeiam essa atividade (exemplo, fazer perguntas ou coletar e processar respostas) sem que a aplicação tenha participação direta, pois ela apenas ativa o serviço e aguarda o resultado com as respostas coletadas. Os trabalhos relacionados que também utilizam esse método de processamento são o OpenPhone e o FalaBrasil, pois também abordam diálogos simples e de objetivos bem definidos. Isso possibilita que o diálogo entre usuário e sistema seja mais direto, reduzindo assim as chances de erro e possibilitando que os componentes de reconhecimento e compreensão sejam menos robustos e complexos.

Quanto a iniciativa do diálogo, o MHNSS é controlado pelo sistema. Isso quer dizer que a participação do usuário acontece em momentos definidos pela aplicação, tornando-a previsível. Assim, o sistema não necessita ficar aguardando uma instrução do usuário para executar, o que otimiza o tempo que o serviço fica ativo e diminui o consumo de recursos como a bateria do dispositivo móvel. Nos trabalhos relacionados, o H&F Companion e o FalaBrasil também possuem iniciativa do sistema, mas por propósitos distintos, pois o primeiro faz aconselhamentos e necessita desse controle por parte do sistema, enquanto o segundo precisa controlar a forma como as informações são apresentada para o usuário, assim como acontece em nossa proposta.

Tornar o sistema de diálogo capaz de falar e compreender ao idioma português brasileiro é outro ponto favorável da proposta MHNSS, pois são poucos os sistemas que utilizam esse idioma. Entre os trabalhos relacionados, apenas o apresentado pelo projeto FalaBrasil faz uso do português, mas ele utiliza uma ferramenta que só funciona em sistemas operacionais Windows (SAPI), diferente do MHNSS que é baseado em um serviço de conversão de áudio específico para o ambiente móvel (Dragon Mobile SDK). Além disto, o MHNSS utiliza a ferramenta NLTK para processamento de linguagem natural com suporte ao português do Brasil, enquanto o FalaBrasil não realiza este tipo de processamento.

As técnicas de processamento de diálogo utilizadas pelo MHNSS também são diferentes da maioria apresentada nos trabalhos relacionados. Entre eles, apenas o HWYD utiliza PLN para auxiliar na compreensão do diálogo, o qual serviu como base para a definição das técnicas de PLN utilizadas no MHNSS.

	OpenPhone	H&F Companion	HWYD	FalaBrasil	MHNSS
Domínio	Específico	Específico	Específico	Específico	Independente
Plataforma	Telefonia	Computador, Web e Móvel	Computador	Web	Móvel
Gerenciamento	Estado Finito	Planos de Diálogo (Frames)	Mista	Estado Finito	Estado Finito
Iniciativa	Usuário	Sistema	Usuário	Sistema	Sistema
Idioma	Inglês	Inglês	Inglês	Português	Português
Processamento	Não	Interpretação Semântica	PLN	Não	PLN

Tabela 3.2: Comparação da Proposta com os Trabalhos Relacionados

A Tabela 3.2 mostra uma comparação dos sistemas de diálogo apresentados nos trabalhos relacionados com o proposto pelo MHNSS. Apesar de não demonstrar uma característica totalmente exclusiva do MHNSS, ela aponta alguns pontos que foram pouco enfatizado em outros trabalhos, mas que são importantes em nossa proposta, como abstrair o domínio para ser utilizado por diversos tipos de aplicações, adotar a plataforma móvel para acesso aos serviços, falar e compreender ao idioma português brasileiro, além de utilizar técnicas de processamento de linguagem natural para uma melhor compreensão dos textos.

3.6 Conclusão

Neste capítulo apresentamos o MHNSS, um *middleware* que fornece suporte ao desenvolvimento de aplicações móveis com recursos de acessibilidade através da fala. A partir do *middleware* as aplicações móveis podem construir um diálogo com o usuário, proporcionando facilidade de acesso e superando limitações de usabilidade dos dispositivos móveis.

Sua arquitetura foi projetada para permitir o uso do *middleware* a partir de dispositivos móveis de baixo custo e utilizando redes de baixa velocidade como o 3G. Além disso, seus componentes foram construídos de forma modular, possibilitando que melhorias em cada um deles possam ser realizadas sem impactar nos demais.

As características atribuídas ao sistema de diálogo contido no *middleware* não são exclusivas, mas atacam pontos pouco difundidos do setor, como utilização em dispositivos móveis, em idioma português e independente do domínio. Assim, o *middleware* MHNSS é um passo a mais para tornar as atividades mais acessíveis, considerando diferentes necessidades e dificuldades dos usuários.

4 Avaliação do Trabalho

Para avaliarmos o MHNSS, conduzimos um estudo de caso que demonstra a utilização dos recursos do *middleware* pela aplicação PBB. Através desse estudo é possível perceber como uma aplicação móvel dentro do domínio da saúde pode usufruir do sistema de diálogo proposto de forma a proporcionar uma maior acessibilidade a seus usuários. Descrevemos neste capítulo detalhes sobre o estudo de caso, a metodologia de avaliação utilizada e os resultados obtidos. Apresentamos também uma discussão dos resultados, onde analisamos e comparamos os dados com valores ideais.

4.1 Estudo de Caso

Como citado na seção 2.5, a aplicação PBB utiliza dados coletados de sensores, o histórico do paciente e a situação atual dele (obtida através de um questionário) para identificar mudanças no seu estado de saúde. Este questionário é composto de nove perguntas do tipo “Está com falta de ar?” ou “Sentiu palpitação na região do coração?”, sendo que as perguntas são selecionadas de acordo com a doença e a situação do paciente, ou seja, o questionário é personalizado. Para maioria dessas perguntas, espera-se respostas simples como “Sim” ou “Não”, mas para algumas outras é necessário informar a intensidade das dores como, por exemplo, “Muito intenso” ou “Mediano” para a pergunta sobre palpitação no coração. Por ser destinado ao acompanhamento de portadores de doenças crônicas, a intensidade das dores é importante para alertar os profissionais da saúde sobre a criticidade da situação. Como o acompanhamento desses pacientes é geralmente longo, a coleta dessas informações precisa ser realizada regularmente, a fim de manter os profissionais da saúde sempre atualizados sobre o tratamento dos pacientes [26].

Tradicionalmente o usuário responde o questionário da PBB através de uma aplicação no dispositivo móvel. Contudo, as perguntas e respostas são apresentas ao

usuário em modo texto. O MHNSS pode ser acionado para fornecer ao usuário uma opção facilitada de acessibilidade para responder às perguntas.

Para utilizar o MHNSS, a aplicação PBB disponibiliza através da interface de programação do *middleware* as questões selecionadas e algumas respostas de referência para cada pergunta. Essas respostas de referência auxiliam o gerenciador de diálogo a identificar respostas válidas para cada pergunta, principalmente quando for necessário realizar o PLN. Em posse dessas informações, o sistema de diálogo interage com o usuário através da fala até obter respostas válidas para todas as perguntas, ou seja, ele conversa com o usuário de várias formas e quantas vezes for necessário até conseguir as respostas. Ao final do questionário, o serviço informa à aplicação PBB as respostas coletadas. Na Figura 4.1 tem-se o exemplo de um diálogo gerenciado pelo MHNSS para o preenchimento de um questionário enviado pela aplicação PBB.

01. [sys]: Olá! Bem vindo ao Sistema de Diálogo para Assistência Remota de Saúde.
02. [sys]: Para avaliarmos o andamento do seu estado de saúde, precisamos que você responda as seguintes perguntas:
03. [sys]: Você está com febre?
04. [usr]: **ERROR: Please retry your query.**
05. [sys]: Desculpe, mas devido a um erro interno vou precisar repetir a pergunta. Você está com febre?
06. [usr]: Não
07. [sys]: Você vomitou alguma vez hoje?
09. [sys]: Você quis dizer: NÃO?
10. [usr]: Exatamente
11. [sys]: Está com falta de ar? Caso a resposta seja SIM, qual a intensidade da falta de ar?
12. [usr]: Bastante
13. [sys]: Não entendi. Vou repetir a pergunta. Está com falta de ar?
14. [usr]: Bastante
15. [sys]: Ainda não entendi. Algumas sugestões de respostas para esta pergunta são: Muito Intenso ou Intenso ou Mediano ou Pouco ou Nenhuma.
16. [sys]: Está com falta de ar??
17. [usr]: Intensamente.
18. [sys]: Ok! Obrigado por responder ao nosso questionário.

Figura 4.1: Questionário Preenchido com o MHNSS

Neste exemplo o sistema de diálogo busca fazer três perguntas ao usuário (linhas 3, 7 e 11). Para cada uma delas é demonstrado um comportamento diferente do sistema. Na primeira pergunta (linha 3), o MHNSS se recuperou de erro interno provavelmente ocorrido na captura da resposta do usuário (por exemplo, erro de comunicação com o serviço de conversão do áudio) e, por isso, foi necessário refazer

a pergunta (linha 5). Na segunda pergunta (linha 7) o sistema obteve como resposta uma expressão mais elaborada (linha 8), mas com a ajuda do PLN, como demonstrado no exemplo da Figura 3.8, ele conseguiu compreender a resposta e confirmou isso com o usuário (linhas 9 e 10). Já na terceira pergunta (linha 11) o sistema não foi capaz de compreender a resposta de imediato e, por isso, precisou interagir mais vezes com o usuário até obter uma resposta satisfatória para a pergunta (linhas 13 a 16). As linhas 1, 2 e 17 são expressões definidas pela aplicação que requisitou o serviço, que no caso foi a PBB, e são utilizadas para introduzir, comentar e concluir o diálogo da forma que a aplicação achar melhor para a experiência do usuário. Após conclusão do diálogo, o MHNSS envia as respostas obtidas para a aplicação PBB, que continuará com o processo de acompanhamento de saúde do paciente.

4.2 Metodologia de Avaliação

A avaliação do MHNSS foi realizada com sujeitos que, voluntariamente, utilizaram a aplicação PBB e responderam ao questionário através do diálogo gerado com os recursos do *middleware* proposto. Os sujeitos foram recrutados na universidade (nos arredores do laboratório de pesquisa) e durante os testes foram encaminhados a uma sala com baixa circulação de pessoas e com acesso à Internet via rede local sem fio. A aplicação PBB foi disponibilizada através de um *smartphone* Motorola Razer I e sua comunicação com o servidor deu-se através de conexão Gigabit Ethernet sem fio. O servidor utilizado para os testes era uma máquina com processador Intel Core I5 2.5GHz e 6GB de memória Ram.

Os sujeitos utilizaram a aplicação individualmente, para que cada um tivesse seu primeiro contato com o sistema de diálogo apenas no momento do teste. A avaliação do MHNSS foi realizada de duas formas: mensurando aspectos de performance de um SDS [30] (quantitativa) e analisando a aceitação da aplicação móvel com interações baseada na fala pelo usuário [18] (qualitativa).

Na avaliação de performance do MHNSS, o serviço é analisado quantitativamente com relação a seus componentes de forma individual e também o sistema completo. A avaliação individual dos componentes aborda aspectos de

precisão do sistema com relação às respostas dos usuários. As métricas utilizadas nessa avaliação são:

- Exatidão da Palavra (*Word Accuracy*) - Conta a precisão do sistema em reconhecer as palavras ditas pelo usuário. Essa métrica está diretamente relacionada com o componente de Reconhecimento da Voz e é mensurada em porcentagem através da relação entre o número de palavras ditas e o número de palavras reconhecidas corretamente;
- Exatidão da Sentença (*Sentence Accuracy*) - Mede a porcentagem de sentenças (frases completas) que o sistema reconheceu corretamente. Nessa métrica avaliamos se o sistema foi capaz de iniciar e finalizar o reconhecimento no momento certo, capturando apenas o que o usuário quis dizer e desconsiderando ruídos do ambiente. Ela também está relacionada diretamente com o componente de Reconhecimento da Voz e é mensurada pela diferença entre o número de sentenças ditas e sentenças reconhecidas erroneamente;
- Compreensão da Sentença (*Sentence Understanding*) - Mede quantas vezes o componente Compreensão da Linguagem, através das técnicas de PLN, foi capaz de entender uma sentença e identificar a resposta do usuário. Essa medição é dada pelo número de sentenças consultadas em relação ao número de sentenças compreendidas.

Quanto a avaliação de performance do sistema completo, utilizamos as seguintes métricas:

- Número de Turnos (*Number of Turns*) - Mede quantas interações foram necessárias para completar todo o diálogo. Cada interação, seja por parte do usuário ou do sistema, é considerada como um turno do diálogo e é mensurado pela soma do número de turnos de ambos;
- Taxa de Correção (*Correction Rate*) - Mede quantos turnos do diálogo foram dedicados a corrigir alguma falha do sistema, seja falha de reconhecimento, de compreensão ou falha interna do serviço;
- Tempo de Transação (*Transaction Time*) - Mede quanto tempo demorou todo o diálogo;

- Taxa de Sucesso (*Transaction Success*) - Mensura quantas informações foram possíveis coletar do usuário com sucesso. Essa métrica avalia se para cada pergunta realizada do questionário uma resposta válida foi obtida. No entanto, as falhas nas coletas de respostas não indicam que houve falha no sistema, pois há casos em que o sistema de diálogo não consegue compreender as respostas do usuário sucessivas vezes e precisa pular a pergunta para não ficar preso a ela.

Os resultados para a avaliação de performance do MHNSS são obtidos de três formas. Primeiramente através de um log que contém cada interação ocorrida no diálogo, similar ao mostrado na Figura 4.1. Este log é gerado apenas durante o processo de avaliação e é enviado ao servidor após a utilização por cada sujeito. Através dele é possível mensurar várias métricas, como Número de Turnos, Taxa de Correção, Tempo de Transação e Compreensão da Sentença. Para auxiliar a medir essas métricas e outras como Exatidão da Palavra e Exatidão da Sentença, utilizamos o áudio do diálogo que é gravado durante a avaliação individual com cada sujeito. Esse áudio é gravado, com o consentimento dos avaliados, através do gravador de som de um notebook e é utilizado, por exemplo, para identificar se o usuário falou algo diferente do que o sistema reconheceu e registrou no log. Assim, é possível identificar também se houve fatores que interferiram no reconhecimento da voz, como um barulho no ambiente ou se o usuário começou a falar antes do sistema finalizar a pergunta. Outra forma de obtenção da avaliação é através do relatório gerado pelo *middleware* com as perguntas realizadas e as respostas coletas durante o questionário. Este relatório é enviado para a aplicação PBB como retorno à solicitação do serviço e também auxilia na avaliação da métrica Taxa de Sucesso.

Após os testes com o uso da aplicação, cada usuário responde um questionário qualitativo¹ para avaliarmos a opinião dele com relação ao serviço. Neste questionário o sujeito informa seu gênero, idade, escolaridade, profissão e se ele já havia utilizado *smartphone* antes. Através desses dados podemos fazer uma classificação dos perfis dos usuários e observar as necessidades e dificuldades de cada perfil. Juntamente com essas informações, o questionário pede que o sujeito avalie o sistema qualitativamente baseado em oito afirmações em que ele especifica o seu nível de concordância com as afirmações, através de respostas baseadas na escala Likert [18].

¹Este questionário não refere-se ao mesmo solicitado pela aplicação PBB, mas sim a um preenchido a punho após o seu uso.

Através dessa medição é possível identificar o quão positiva foi a experiência dos sujeitos com o serviço, a fim de medir o seu nível de aceitação com o uso do sistema de diálogo. A Figura 4.2 exibe as afirmações e possíveis respostas contidas no questionário qualitativo.

Foi fácil responder ao questionário.

Muito Fácil Fácil Regular Difícil Muito Difícil

Foi fácil utilizar a aplicação.

Muito Fácil Fácil Regular Difícil Muito Difícil

O sistema entendia o que você dizia.

Sempre Frequentemente Regularmente Difícilmente Nunca

O voz do sistema era de fácil compreensão.

Sempre Frequentemente Regularmente Difícilmente Nunca

O sistema executava/processava em tempo hábil.

Sempre Frequentemente Regularmente Difícilmente Nunca

O sistema coletou as informações sobre seu estado de saúde corretamente.

Sempre Frequentemente Regularmente Difícilmente Nunca

O sistema foi capaz de lidar com os erros.

Sempre Frequentemente Regularmente Difícilmente Nunca

É preferível utilizar esse sistema em vez de preencher o questionário manualmente.

Sempre Frequentemente Regularmente Difícilmente Nunca

Figura 4.2: Questionário Qualitativo

Ainda no questionário qualitativo, há um espaço dedicado a críticas e sugestões, em que o sujeito pode expressar livremente sua opinião sobre o sistema de diálogo e que serão consideradas para futuras versões do *middleware*.

4.3 Resultados e Discussão

O perfil dos sujeitos que utilizaram o MHNSS através do uso da aplicação PBB para avaliação foi bastante diversificado. Ao todo foram 10 sujeitos, com perfis apresentados na Tabela 4.1 e identificados de P1 a P10. Dentre eles, P3 e P6 alegaram nunca terem utilizado *smartphone* antes. Esse perfil representa a diversidade de público que pode utilizar os recursos do *middleware* junto às aplicações

do projeto MobileHealthNet, inclusive a PBB, desde pacientes com baixa escolaridade até profissionais da saúde com formação superior.

Sujeito	Gênero	Idade	Escolaridade	Profissão
P1	F	30	Médio	Vendedora
P2	F	37	Médio	Dona de Casa
P3	F	35	Médio	Vendedora
P4	M	26	Mestrado	Programador
P5	M	28	Superior	Estudante
P6	M	50	Médio	Serviços Gerais
P7	F	26	Superior	Estudante
P8	M	26	Superior	Estudante
P9	M	33	Superior	Professor
P10	F	54	Médio	Dona de Casa

Tabela 4.1: Perfil dos Sujeitos Participantes da Avaliação

Os resultados obtidos na avaliação quantitativa foram bastante positivos. De acordo com os dados coletados através dos logs e dos áudios gravados durante a avaliação, a média de Exatidão das Palavras reconhecidas dentre todos os sujeitos foi de 96,09% ($DP^2=0,05$). Este resultado aponta o componente Reconhecimento da Voz como favorável para fazer a coleta da voz do usuário em português com alta taxa de precisão. Mesmo quando a palavra era reconhecida erroneamente, percebeu-se que geralmente havia uma similaridade fonética com a palavra correta, como por exemplo quando o usuário falava “sim”, mas era reconhecido como “se”. Esses erros geravam turnos adicionais no diálogo para corrigi-los, mas não comprometiam a performance geral do sistema.

Quanto à média de Exatidão da Sentença, 96,17% ($DP=0,04$) das expressões ditas pelos sujeitos foram reconhecidas completas e corretas. As falhas foram ocasionadas devido ao sujeito em algumas casos começar a responder antes do sistema estar pronto para capturar o áudio, o que resultava em palavras reconhecidas a mais ou a menos. Essas falhas também resultavam em números adicionais de turnos para correção, mas eram mais fáceis para o componente Compreensão da Linguagem

²Desvio Padrão.

identificar a resposta correta através do PLN. A taxa de Compreensão da Sentença, que está diretamente relacionada com o sucesso no reconhecimento da resposta por parte do PLN, obteve uma média de 88,33% (DP=0,15).

Em condições ideais, o MHNSS necessita de, no mínimo, 21 turnos de diálogo para coletar as nove respostas solicitadas pela PBB. Nessa contagem considera-se interações por parte do sistema e por parte do usuário. Durante as avaliações, o Número de Turnos médio foi 23 (DP=1,63), o que não está muito longe do ideal. A Taxa de Correção foi de 2 turnos em média, o que demonstra exatamente a diferença entre o número de turnos ideal do número de turnos obtido.

O Tempo de Transação médio para que cada sujeito completasse o questionário foi de 123 segundos (DP=14,47), o que não é um tempo ruim, pois estimasse que o sistema necessite cerca de 110 segundos para esse processo. A Taxa de Sucesso obtida foi de 100%, ou seja, todos os sujeitos conseguiram responder às 9 perguntas do questionário completamente e as respostas coletadas foram confirmadas como as fornecidas por eles. Os resultados completos de cada sujeito podem ser visualizados nas Tabelas A.1 e A.2 em anexo.

Quanto a avaliação qualitativa, também obtivemos resultados bastante positivos. No questionário cada sujeito deveria especificar o seu nível de concordância com as afirmações sobre o sistema que acabara de utilizar. A Figura 4.3 exhibe quantos sujeitos elegeram aquele nível de concordância para as respectivas afirmações.

Analisando esses resultados qualitativos, verifica-se que todos os sujeitos alegaram que foi fácil ou muito fácil responder o questionário (Afirmação 1) e utilizar a aplicação (Afirmação 2) através da fala. A avaliação dos sujeitos quanto ao sistema entender o que eles diziam (Afirmação 3) foi fortemente positiva em mais da metade dos casos. Já quanto a voz do sistema (Afirmação 4), ela foi unanimemente avaliada de fácil compreensão. O desempenho do sistema foi dito como sempre rápido (Afirmação 5) por 8 sujeitos e que em 100% dos casos ele coletou todas as respostas corretamente (Afirmação 6). Por fim, de acordo com os sujeitos, o sistema foi capaz de lidar bem com os erros (Afirmação 7) e que é sempre preferível utilizar o sistema de diálogo para responder ao questionário (Afirmação 8) em 90% das avaliações. O resultado completo com a avaliação qualitativa de cada sujeito pode ser visualizado nas Tabelas B.1 e B.2 em anexo.

Afirmção 1	Muito Fácil	Fácil	Regular	Difícil	Muito Difícil
Foi fácil responder ao questionário.	8	2	0	0	0
Afirmção 2	Muito Fácil	Fácil	Regular	Difícil	Muito Difícil
Foi fácil utilizar a aplicação.	7	3	0	0	0
Afirmção 3	Sempre	Frequentemente	Regularmente	Difícilmente	Nunca
O sistema entendia o que você dizia.	6	3	1	0	0
Afirmção 4	Sempre	Frequentemente	Regularmente	Difícilmente	Nunca
A voz do sistema era de fácil compreensão.	10	0	0	0	0
Afirmção 5	Sempre	Frequentemente	Regularmente	Difícilmente	Nunca
O sistema executava/processava em tempo hábil.	8	2	0	0	0
Afirmção 6	Sempre	Frequentemente	Regularmente	Difícilmente	Nunca
O sistema coletou as informações sobre seu estado de saúde	10	0	0	0	0
Afirmção 7	Sempre	Frequentemente	Regularmente	Difícilmente	Nunca
O sistema foi capaz de lidar com os erros.	9	1	0	0	0
Afirmção 8	Sempre	Frequentemente	Regularmente	Difícilmente	Nunca
É preferível utilizar esse sistema em vez de preencher o questionário	9	1	0	0	0

Figura 4.3: Resultado da Avaliação Qualitativa

4.4 Conclusão

Neste capítulo apresentamos a avaliação do MHNSS através de seu uso a partir de uma aplicação móvel dedicada a atenção à saúde.

Na seção 4.1 discorre-se sobre o estudo de caso aplicado, onde a aplicação PBB, que tradicionalmente apresenta-se para o usuário em forma de textos, foi utilizada através da fala a partir dos recursos oferecidos pelo *middleware* MHNSS.

Na seção 4.2 a metodologia utilizada na avaliação foi apresentada, detalhando as métricas avaliadas e os meios como elas foram obtidas. Essas métricas foram escolhidas com base em outros trabalhos que avaliaram sistemas de diálogo e apoiam-se em aspectos qualitativos e quantitativos.

Por fim, a seção 4.3 descreve os resultados, os quais demonstram que para os perfis de usuários que fizeram parte da avaliação, o MHNSS apresentou ótimo desempenho no reconhecimento e na compreensão das palavras, além de um tempo médio e número de turnos próximos ao ideal. O *middleware* foi ainda eficaz em coletar

as informações requisitadas pela aplicação e os sujeitos não tiveram dificuldades em utiliza-la através da fala. Os resultados mostram, principalmente, que a proposta foi bem aceita pelos sujeitos, pois eles alegaram preferir utilizar o sistema de diálogo em vez de preencher o questionário digitando as respostas.

5 Conclusões e Trabalhos Futuros

As aplicações móveis são a base das funcionalidades encontradas nos dispositivos móveis atuais. Através delas, o usuário envia mensagens, acessa Internet, armazena contatos e se entretêm, bastando apenas ter a respectiva aplicação instalada no seu *smartphone* ou *tablet*. O acesso a elas é geralmente feito através do toque na tela do dispositivo móvel e sua exibição é baseada em textos e imagens, por isso, as aplicações móveis são dependentes do contato físico e visual do usuário para interação.

No projeto MobileHealthNet, as aplicações móveis são criadas para conduzir ações relacionadas à assistência médica e sua educação. Essas aplicações visam proporcionar a troca de experiência e a comunicação entre profissionais da saúde e pacientes de comunidades carentes e distantes, onde o custo de deslocamento para um posto de atendimento é alto devido as suas condições de acesso. Parte desses usuários possuem dificuldades para ler, escrever (digitar) e utilizar aplicações móveis de modo geral. Por isso, necessita-se que esses usuários utilizem as aplicações do MobileHealthNet de forma correta e sem dificuldades.

Partindo da necessidade geral das aplicações móveis em obter opções de acessibilidade para o usuário e da demanda do MobileHealthNet para proporcionar facilidade e efetividade no acesso de suas aplicações móveis, neste trabalho dissertativo propusemos a criação de um *middleware* que fornece meios para uma interação entre usuário e aplicação móvel através da fala. Este *middleware* visa fornecer soluções a vários problemas relativos a acessibilidade para aplicações móveis, dentre as quais destacam-se:

- Utilização de recursos de diálogo falado para proporcionar uma interação que não dependa de contato físico ou visual com a aplicação móvel;
- Adaptação da solução para executar em ambiente móvel e em dispositivos considerados de baixo custo;
- Concepção de um modelo de arquitetura e processamento que atenda a diversos tipos de aplicações móveis;

- Adesão do idioma português brasileiro como padrão, mesmo que a maioria das ferramentas e técnicas sejam destinadas ao inglês;
- Estabelecimento de uma flexibilidade na interação capaz de atender a diferentes perfis de usuários.

Para a concepção do *middleware*, o qual denominamos *MobileHealthNet Speech Service (MHNSS)*, desenvolveu-se um serviço Android que proporciona os recursos de diálogo necessários para que aplicações móveis estabeleçam uma interação com o usuário via voz, e um serviço de diálogo servidor que processa as partes do diálogo que demandam mais recursos de hardware.

Para avaliarmos esta proposta, um estudo de caso foi realizado em que uma aplicação do MobileHealthNet, a *Patient-Buddy-Build*, que utiliza os recursos do *middleware* MHNSS para efetuar o preenchimento de um questionário de acompanhamento remoto de pacientes portadores de fibrilação atrial, uma doença crônica. Como resultados, as avaliações quantitativas demonstraram um bom desempenho dos componentes do *middleware* e as avaliações qualitativas mostraram, principalmente, que a proposta foi bem aceita pelos sujeitos que participaram da avaliação.

5.1 Contribuições

Este trabalho dissertativo produziu as seguintes contribuições:

- Revisão da literatura científica que trata dos aspectos relativos a sistemas de diálogo falado;
- Levantamento e análise dos principais requisitos de sistemas de diálogo em ambientes móveis e dedicados à saúde;
- Concepção de uma arquitetura adequada para o prover acessibilidade para aplicações móveis a partir de recursos do diálogo falado;
- Desenvolvimento de um modelo de processamento de diálogo capaz de atender a diversos tipos de aplicações móveis, sem um domínio específico;

- Projeto e implementação do *middleware* baseado na arquitetura e no modelo de processamento proposto;
- Avaliação do *middleware* com usuários reais.

Os objetivos específicos definidos para este trabalho dissertativo foram alcançados através das contribuições supra-citadas, o que deu origem à seguinte publicação:

- Ferreira, Arikleyton; et al. MHNSS: A Middleware for Developing Speech-based Interaction Mobile Applications. In *Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems, IHC '14*, Foz do Iguaçu, Brazil, Outubro, 2014. Brazilian Computer Society.

Este artigo apresenta as principais contribuições oriundas desse trabalho dissertativo, incluindo os desafios pra a aplicação do MHNSS no contexto do MobileHealthNet, sua arquitetura e principais componentes, bem como a avaliação do *middleware* através do estudo de caso.

5.2 Trabalhos Futuros

Durante o desenvolvimento desta pesquisa, algumas questões não puderam ser respondidas a contento devido às limitações de escopo e tempo. Desta forma, como principais trabalhos futuros destacamos:

- Testar serviços de conversão de áudio gratuitos: apesar do *Dragon Mobile SDK* apresentar um ótimo desempenho e facilidades para uso em ambiente móvel, ele possui limite de consultas diárias ao serviço de conversão na sua versão gratuita de desenvolvedor. Assim, é necessário testar outros serviços que possam oferecer desempenho equivalente e que não haja limite de consultas, como o serviço de conversão nativo do Android, que atualmente já realiza reconhecimento da fala e leitura de textos em português brasileiro, o que não estava disponível durante o desenvolvimento do trabalho;

- Avaliar a proposta com outras aplicações móveis: para utilizar os recursos de diálogo, a aplicação móvel deve ser desenvolvida considerando esses recursos. Porém, durante a avaliação prática, tivemos acesso apenas ao código da aplicação *Patient-Buddy-Build*, pois ela faz parte do projeto MobileHealthNet. Mas como o MHNSS pode oferecer recursos para qualquer tipo de aplicação móvel, seria interessante fazer avaliações com outras aplicações;
- Avaliar a proposta com usuários analfabetos, com limitações motoras e deficientes visuais: Esses perfis de usuários ajudariam a avaliar o *middleware* na essência da sua proposta, que é proporcionar acessibilidade que não dependa de contato físico ou visual do usuário para utilizar a aplicação móvel. Assim, agregaria-se mais valor à avaliação do *middleware*;
- Avaliar a proposta em conexão de Internet 3G/4G: Durante a avaliação da proposta com usuários reais, utilizamos apenas conexão local sem fio para a comunicação com os servidores e serviços. A conexão 3G foi testada durante a implementação da camada de comunicação do MHNSS, mas não foi utilizada durante as avaliações. Assim, este é outro ponto a ser avaliado que valorizaria os resultados do trabalho.

Além desses trabalhos futuros, destacamos algumas possíveis linhas de pesquisa que dariam sequência aos propósitos dos MHNSS. São elas:

- O desenvolvimento de um sistema de autoria de diálogos que gerasse de forma automática ou semiautomática o código necessário para a utilização dos serviços disponibilizados pelo MHNSS, dando-lhe autonomia para gerenciar diálogos completos, assim como foi proposto para a realização de questionários;
- Evoluir o gerenciamento do diálogo no MHNSS para um baseado em *frames*, possibilitando que o diálogo entre usuário e sistema seja menos restrito a um fluxo pré-definido, agregando ainda mais independência de domínio ao *middleware*;
- Estender a capacidade de compreensão do MHNSS, adicionando outras técnicas de processamento e testando novos modelos de compreensão.

Referências Bibliográficas

- [1] J. Allen. *Natural language understanding (2nd ed.)*. Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, USA, 1995.
- [2] R. Batista and F. Silva. Uma infraestrutura de comunicação para colaboração em redes sociais móveis. In *SBSC 2012: Simpósio Brasileiro de Sistemas Colaborativos*, São Paulo, SP, Brasil, Outubro 2012.
- [3] L. Besacier, E. Barnard, A. Karpov, and T. Schultz. Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, 56(0):85 – 100, 2014.
- [4] T. Bickmore and T. Giorgino. Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, 39(5):556 – 571, 2006.
- [5] D. Dahl. *Practical Spoken Dialog Systems*. Springer London, 2004.
- [6] L. David, R. Vasconcelos, L. Alves, R. André, G. Baptista, and M. Endler. A communication middleware for scalable real-time mobile collaboration. *IEEE 21st International WETICE, Track on Adaptive and Reconfigurable Service-oriented and component-based Applications and Architectures (AROSA)*, 2012.
- [7] R. de Carvalho Batista. Uma infraestrutura de comunicação centrada em dados para redes sociais móveis. Master's thesis, Universidade Federal do Maranhão, 2013.
- [8] G. Demiris. The diffusion of virtual communities in health care: concepts and challenges. *Patient Education and Counseling*, 62(2):178–188, Aug. 2006.
- [9] D. C. Dugdale, R. Epstein, and S. Z. Pantilat. Time and the patient–physician relationship. *Journal of General Internal Medicine*, 14(S1):34–40, 1999.
- [10] E. S. Furtado, D. Chagas, I. I. Bittencourt, and A. Façanha. Acessibilidade e inclusão digital — grandes desafios de pesquisa em interação humano-computador no brasil. Technical report, Comissão Especial de Interação Humano-Computador da Sociedade Brasileira de Computação, 2014.

- [11] J. F. Gonçalves. Desenvolvimento de uma infraestrutura de segurança para o middleware mobilehealthnet. Master's thesis, Universidade Federal do Maranhão, 2013.
- [12] J. Goncalves, A. Teles, and F. Silva. Um modelo de segurança e privacidade para redes sociais móveis aplicadas a área da saúde. In *XII Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais*, Curitiba, PR, Brasil, Outubro 2012.
- [13] M. Gonzalez and V. L. S. de Lima. Recuperação de informação e processamento de linguagem natural. In *Proceedings of XXIII Congresso da Sociedade Brasileira de Computação*, pages 347–395, Maceió, AL, Brasil, Julho 2003.
- [14] O. M. Group. Data distribution service for real-time systems specification - versão 1.2. <http://www.omg.org/spec/DDS/1.2/PDF/>, July 2001.
- [15] A. Grover, M. Plauche, E. Barnard, and C. Kuun. Hiv health information access using spoken dialogue systems: Touchtone vs. speech. In *International Conference on Information and Communication Technologies and Development (ICTD)*, pages 95–107, Doha, Qatar, April 2009.
- [16] A. Latham, K. Crockett, D. McLean, and B. Edmonds. A conversational intelligent tutoring system to automatically predict learning styles. *Computers & Education*, 59(1):95 – 109, 2012. {CAL} 2011 The {CAL} Conference 2011.
- [17] J. M. Leimeister, M. Daum, and H. Krcmar. Mobile virtual healthcare communities: An approach to community engineering for cancer patients. In *European Conference on Information Systems (ECIS)*, pages 1626–1637, Gdansk, Poland, June 2002.
- [18] S. Love. *Understanding mobile human-computer interaction*. Elsevier, Amsterdam, 2005.
- [19] R. López-Cózar and Z. Callejas. Multimodal dialogue for ambient intelligence and smart environments. In H. Nakashima, H. Aghajan, and J. Augusto, editors, *Handbook of Ambient Intelligence and Smart Environments*, pages 559–579. Springer US, 2010.

- [20] M. McTear. Spoken language understanding for conversational dialog systems. In *Spoken Language Technology Workshop, 2006. IEEE*, pages 6–6, Palm Beach, Aruba, Dec 2006.
- [21] M. F. McTear. *Spoken Dialogue Technology: Toward the Conversational User Interface*. Springer London, 2004.
- [22] R. Moore. Presence: A human-inspired architecture for speech-based human-machine interaction. *Computers, IEEE Transactions on*, 56(9):1176–1188, Sept 2007.
- [23] J. Morais, N. Neto, and A. Klautau. Tecnologias para o desenvolvimento de sistemas de diálogo falado em português brasileiro. In *7th Brazilian Symposium in Information and Human Language Technology*, São Carlos, Brazil, September 2009.
- [24] S. S. Narayanan, S. Ananthkrishnan, R. Belvin, E. Ettelaie, S. Gandhe, S. Ganjavi, P. G. Georgiou, C. M. Hein, S. Kadambe, K. Knight, D. Marcu, H. E. Neely, N. Srinivasamurthy, D. Traum, and D. Wang. The transonics spoken dialogue translator: An aid for english-persian doctor-patient interviews. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) Fall Symposium*, pages 97–103, Washington, D.C, 2004.
- [25] A. Neustein and J. Markowitz. *Mobile Speech and Advanced Natural Language Solutions*. Springer London, 2013.
- [26] V. Pinheiro, M. Endler, and E. Hermann. Patient-buddy-build: Customized mobile monitoring for patients with chronic diseases. In *Proceedings of The Fifth International Conference on eHealth, Telemedicine, and Social Medicine(eTELEMED)*, Nice, France, 2013.
- [27] I. Pitt and A. Edwards. *Design of Speech-based Devices: A Practical Guide*. Springer Publishing Company, 2012.
- [28] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon. Design and evaluation of a smart home voice interface for the elderly: Acceptability and objection aspects. *Personal Ubiquitous Comput.*, 17(1):127–144, Jan. 2013.
- [29] S. G. Pulman, J. Boye, M. Cavazza, C. Smith, and R. S. de la Cámara. ‘how was your day?’. In *Proceedings of the 2010 Workshop on Companionable Dialogue*

- Systems*, CDS '10, pages 37–42, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
- [30] N. Rajput and A. A. Nanavati. *Speech in Mobile and Pervasive Environments*. Wiley Publishing, 1st edition, 2012.
- [31] A. Rizzo, K. Sagae, E. Forbell, J. Kim, B. Lange, J. Buckwalter, J. Williams, T. Parsons, P. Kenny, D. Traum, J. Difiede, and B. Rothbaum. 'simcoach: An intelligent virtual human system for providing healthcare information and support'. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Orlando, Florida, Novembro 2011.
- [32] J. Sherwani, N. Ali, S. Mirza, A. Fatma, Y. Memon, M. Karim, R. Tongia, and R. Rosenfeld. Healthline: Speech-based access to health information by low-literate users. In *International Conference on Information and Communication Technologies and Development (ICTD)*, pages 1–9, Bangalore, India, 2007.
- [33] G. Skantze and A. Hjalmarsson. Towards incremental speech generation in conversational systems. *Computer Speech & Language*, 27(1):243 – 262, 2013. Special issue on Paralinguistics in Naturalistic Speech and Language.
- [34] A. J. Stent. *Dialogue Systems As Conversational Partners: Applying Conversation Acts Theory to Natural Language Generation for Task-oriented Mixed-initiative Spoken Dialogue*. PhD thesis, The University of Rochester, 2001. AAI3023749.
- [35] D. Suendermann. *Advances in Commercial Deployment of Spoken Dialog Systems*. Springer New York, 2011.
- [36] A. K. Syrdal and Y. jun Kim. Dialog speech acts and prosody: Considerations for tts. In *Proceedings of Speech Prosody*, Campinas, Brazil, May 2008.
- [37] A. Teles, J. Gonçalves, V. Pinheiro, F. J. da Silva e Silva, and M. Endler. Infraestrutura e aplicações de redes sociais móveis para colaboração em saúde. In *Congresso Brasileiro de Informática em Saúde, CBIS '12*, Curitiba, Brazil, November 2012.
- [38] A. S. Teles, D. Pinheiro, J. Gonçalves, R. Batista, V. Pinheiro, F. J. da Silva e Silva, and M. Endler. Redes sociais móveis: Conceitos, aplicações e aspectos de

- segurança e privacidade. In *SBRC '13: Anais dos Minicursos do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 51–100. SBC, Brasilia, DF, Brasil, Maio 2013.
- [39] A. S. Teles, D. N. Pinheiro, J. F. Gonçalves, R. C. Batista, F. J. S. Silva, V. Pinheiro, E. Haeusler, and M. Endler. Mobilehealthnet: A middleware for mobile social networks in m-health. In *Proceedings of MobiHealth 2012*, Paris, France, November 2012. Springer.
- [40] M. Turunen, J. Hakulinen, O. Ståhl, B. Gambäck, P. Hansen, M. C. R. Gancedo, R. S. de la Cámara, C. Smith, D. Charlton, and M. Cavazza. Multimodal and mobile conversational health and fitness companions. *Computer Speech & Language*, 25(2):192 – 209, 2011. Language and speech issues in the engineering of companionable dialogue systems.
- [41] W. Velicer, J. Prochaska, J. Fava, R. Laforge, and J. Rossi. Interactive versus noninteractive interventions and dose-response relationships for stage-matched smoking cessation programs in a managed care setting. *Health Psychol*, 18(1):21–8, 1999.
- [42] W. Zhenyu, Z. Chunhong, J. Yang, and W. Hao. Towards cloud and terminal collaborative mobile social network service. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing, SOCIALCOM '10*, pages 623–629, Washington, DC, USA, August 2010. IEEE Computer Society.

Apêndices

A Resultados da Avaliação dos Sujeitos

Sujeito		P1		P2		P3		P4		P5	
Exatidão da Palavra	Palavras ditas:	11	100%	10	90%	10	90%	9	100%	11	90,91%
	Palavras erradas:	0		1		1		0		1	
Exatidão da Sentença	Sentenças ditas:	10	90%	10	100%	10	100%	9	100%	10	100%
	Sentenças erradas:	1		0		0		0		0	
Compreensão da Sent.	Consultadas:	3	100%	3	100%	3	66,67%	2	100%	4	75%
	Compreendias:	3		3		2		2		3	
Taxa de Sucesso	Coletadas:	9	100%	9	100%	9	100%	9	100%	9	100%
	Sucedidas:	9		9		9		9		9	
Número de Turnos	Turnos do Sistema:	13	23	13	23	13	23	12	21	13	23
	Turnos do Usuário:	10		10		10		9		10	
Tempo de Transação	Início:	16:20:01	00:02:23	16:27:52	00:02:03	16:37:10	00:02:24	16:54:31	00:01:46	17:09:47	00:02:01
	Fim:	16:22:24		16:29:55		16:39:34		16:56:17		17:11:48	
Taxa de Correção	Falha do Sistema:	0	2	2	2	0	2	0	0	0	2
	de Compreensão:	2		0		2		0		2	

Tabela A.1: Resultados da Avaliação dos Sujeitos P1 a P5

Sujeito		P6		P7		P8		P9		P10	
Exatidão da Palavra	Palavras ditas:	13	100%	9	100%	10	90%	11	100%	10	100%
	Palavras erradas:	0		0		1		0		0	
Exatidão da Sentença	Sentenças ditas:	12	91,67%	9	100%	10	100%	10	90%	10	90%
	Sentenças erradas:	1		0		0		1		1	
Compreensão da Sent.	Consultadas:	4	75%	2	100%	3	66,67%	3	100%	3	100%
	Compreendias:	3		2		2		3		3	
Taxa de Sucesso	Coletadas:	9	100%	9	100%	9	100%	9	100%	9	100%
	Sucedidas:	9		9		9		9		9	
Número de Turnos	Turnos do Sistema:	15	27	12	21	13	23	13	23	13	23
	Turnos do Usuário:	12		9		10		10		10	
Tempo de Transação	Início:	17:23:51	00:02:09	17:30:56	00:01:40	15:15:13	00:01:54	15:39:03	00:01:58	19:11:59	00:02:12
	Fim:	17:26:00		17:32:36		15:17:07		15:41:01		19:14:11	
Taxa de Correção	Falha do Sistema:	0	6	0	0	0	2	0	2	0	2
	de Compreensão:	6		0		2		2		2	

Tabela A.2: Resultados da Avaliação dos Sujeitos P6 a P10

B Resultados Qualitativos dos Sujeitos

	P1	P2	P3	P4	P5
Foi fácil responder ao questionário.	Fácil	Muito Fácil	Fácil	Muito Fácil	Muito Fácil
Foi fácil utilizar a aplicação.	Muito Fácil	Fácil	Fácil	Muito Fácil	Muito Fácil
O sistema entendia o que você dizia.	Sempre	Sempre	Frequentemente	Sempre	Frequentemente
A voz do sistema era de fácil compreensão.	Sempre	Sempre	Sempre	Sempre	Sempre
O sistema executava/processava em tempo hábil.	Sempre	Sempre	Sempre	Sempre	Frequentemente
O sistema coletou as informações sobre seu estado de saúde corretamente.	Sempre	Sempre	Sempre	Sempre	Sempre
O sistema foi capaz de lidar com os erros.	Sempre	Sempre	Sempre	Sempre	Sempre
É preferível utilizar esse sistema em vez de preencher o questionário manualmente.	Sempre	Sempre	Sempre	Sempre	Sempre

Tabela B.1: Resultados Qualitativos dos Sujeitos P1 a P5

	P6	P7	P8	P9	P10
Foi fácil responder ao questionário.	Muito Fácil	Muito Fácil	Muito Fácil	Muito Fácil	Muito Fácil
Foi fácil utilizar a aplicação.	Fácil	Muito Fácil	Muito Fácil	Muito Fácil	Muito Fácil
O sistema entendia o que você dizia.	Regularmente	Sempre	Sempre	Sempre	Frequentemente
A voz do sistema era de fácil compreensão.	Sempre	Sempre	Sempre	Sempre	Sempre
O sistema executava/processava em tempo hábil.	Sempre	Sempre	Sempre	Frequentemente	Sempre
O sistema coletou as informações sobre seu estado de saúde corretamente.	Sempre	Sempre	Sempre	Sempre	Sempre
O sistema foi capaz de lidar com os erros.	Frequentemente	Sempre	Sempre	Sempre	Sempre
É preferível utilizar esse sistema em vez de preencher o questionário manualmente.	Frequentemente	Sempre	Sempre	Sempre	Sempre

Tabela B.2: Resultados Qualitativos dos Sujeitos P6 a P10