

VALÉRIA FARINAZZO MARTINS

**AVALIAÇÃO DE USABILIDADE PARA SISTEMAS DE TRANSCRIÇÃO
AUTOMÁTICA DE LAUDOS EM RADIOLOGIA**

São Paulo
2011

VALÉRIA FARINAZZO MARTINS

**AVALIAÇÃO DE USABILIDADE PARA SISTEMAS DE TRANSCRIÇÃO
AUTOMÁTICA DE LAUDOS EM RADIOLOGIA**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para a
obtenção do Título de Doutor em Engenharia

São Paulo
2011

VALÉRIA FARINAZZO MARTINS

**AVALIAÇÃO DE USABILIDADE PARA SISTEMAS DE TRANSCRIÇÃO
AUTOMÁTICA DE LAUDOS EM RADIOLOGIA**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para a
obtenção do Título de Doutor em Engenharia

Área de Concentração: Sistemas Eletrônicos

Orientador: Prof. Dr. Lincoln de Assis Moura Jr.

São Paulo
2011

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 13 de junho de 2011.

Assinatura do autor

Valéria Farinazzo Martins

Assinatura do orientador

João A. Moura Jr

FICHA CATALOGRÁFICA

Martins, Valéria Farinazzo

Avaliação de usabilidade para sistemas de transcrição automática de laudos em radiologia / V.F. Martins. -- ed.rev. -- São Paulo, 2011.

216 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Telecomunicações e Controle.

1.Usabilidade de software 2.Interface homem-computador
I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle II.t.

DEDICATÓRIA

*Aos meus pais Maria Izolete Farinazzo Martins e
Segundo Garcia Martins*

AGRADECIMENTOS

Ao Prof. Dr. Lincoln de Assis Moura Jr. pelo grande apoio, incentivo, paciência e orientação em todo o desenvolvimento deste trabalho e pela contribuição para a minha carreira acadêmica e profissional. Sem ele, eu não teria chegado até aqui.

À Universidade Presbiteriana Mackenzie, por ter me fornecido subsídios para a finalização deste trabalho e apoio financeiro para as publicações de trabalhos relacionados a esta pesquisa, sobretudo ao Prof. Dr. Arnaldo Rabello de Aguiar Vallim Filho.

Ao Hospital Edmundo Vasconcelos, sobretudo ao Prof. Dr. Claudio Campi de Castro por sempre ser tão prestativo e apoiar a minha pesquisa, desde a fase de qualificação.

À empresa MacSym Tecnologia Médica, em especial ao Sr. Carlos Kawai, pela concessão da licença do software de transcrição automática, por tempo determinado, e pelas informações cedidas.

À minha amiga Ana Grasielle Dionísio Corrêa que participou, efetivamente, como especialista de usabilidade nas inspeções e testes de usabilidade. Agradeço também por todo o apoio e amizade em muitos momentos.

Ao prof. Pericles Prado Turnes Jr., pelo auxílio e críticas, sempre construtivas, na análise dos dados das avaliações. E pelo apoio e amizade sempre presentes.

Aos meus amigos e colegas da Universidade Presbiteriana Mackenzie pela disponibilidade em participarem dos testes de usabilidade e pelas sugestões dadas neste trabalho.

Ao Prof. Dr. Leandro Zerbinatti pelo material cedido para os testes de usabilidade.

A todas as pessoas que, de uma forma direta ou indireta, contribuíram para que fosse possível que eu alcançasse este objetivo tanto no plano acadêmico e profissional quanto pessoal.

*“...Que ninguém se engane,
só se consegue a simplicidade
através de muito trabalho.”
Clarice Lispector*

RESUMO

Este trabalho relaciona elementos das áreas de Computação e Saúde para comporem a elaboração de uma metodologia para avaliação de usabilidade de sistemas de transcrição automática de laudos na área de Radiologia. Inicialmente, é apresentado um estudo realizado sobre a área de Interface do Usuário Baseada em Voz que identifica requisitos para os sistemas que trabalham com comunicação mediada por voz assim como as iniciativas no sentido de se criar uma metodologia para sua avaliação. Em seguida é realizado um estudo dos sistemas de transcrição automática de laudos, no qual os principais requisitos são caracterizados e classificados. Os dois estudos acima citados foram integrados para a elaboração de uma metodologia para a avaliação de Sistemas de Transcrição Automática de Laudos em Radiologia. A metodologia foi, então, validada previamente através de inspeções e testes de usabilidade realizados fora do ambiente hospitalar, através do uso de um sistema de transcrição automática de laudos em Radiologia. Posteriormente, a metodologia foi também aplicada a um hospital da cidade de São Paulo. Como resultado principal foi proposto um guia bastante detalhado para se avaliar os Sistemas de Transcrição Automática de Laudos em Radiologia, cada vez mais presentes em hospitais e clínicas no país, além dos relatos das experiências obtidas com a aplicação desta metodologia em um caso real.

Palavras-chaves: Avaliação de Usabilidade, *Voice User Interface*, Sistema de Transcrição Automática de Laudos em Radiologia, Sistema de Ditado.

ABSTRACT

This work combines knowledge from Computer Science and Health Science in order to propose an evaluation methodology for Automatic Transcription System of Radiology Reports. At first a study regarding a Voice User Interface is presented, this interface identifies the requirements for Spoken Language Dialogue Systems and it can also be used as a tool for an evaluation methodology. Following, a study of automatic transcription systems is presented; in this study the main requirements are listed and classified. Both studies were integrated to allow a new methodology to evaluate an Automatic Transcription System in Radiology. This methodology was previously validated through some inspections and usability tests outside the hospital environment and afterward the methodology was used in a hospital in São Paulo city. As a main result it was proposed a very detailed guide for evaluating the Automatic Transcription Systems Reports in Radiology, increasingly found in hospitals and clinics in this country, apart from reports on experiences gained in applying this methodology in a real case.

Keywords: Usability Evaluation, *Voice User Interface*, Automatic Transcription System of Radiology Reports, Dictation System.

LISTA DE ILUSTRAÇÕES

Figura 1: Componentes de um Sistema Interativo	34
Figura 2: Exemplo de um diálogo entre o sistema e um usuário.....	36
Figura 3: Arquitetura de um sistema de compreensão de linguagem falada.....	41
Figura 4: O ponto de finalização determina onde as ondas de voz começam e terminam	41
Figura 5: A extração de características	42
Figura 6: O módulo de reconhecimento de voz.....	43
Figura 7: Módulo de compreensão da Linguagem Natural.....	44
Figura 8: Gerenciador de diálogo	45
Figura 9: Taxonomia de Aspectos de Qualidade.....	70
Figura 10: Etapas de um Sistema de Transcrição Automática de Laudo	83
Figura 11: Dispositivo para a entrada de voz – SpeechMike Pro 5274	84
Figura 12: Interface gráfica de um sistema de transcrição automática de laudo.....	85
Figura 13: Tela Inicial do Aplicativo SmlInitialTraining	110
Figura 14: Tela de Treinamento do Aplicativo SmlInitialTraining	111
Figura 15: Tela Principal do Neoc 6.1	112
Figura 16: Tela do Sistema Philips Speech Magic 2.1	113
Figura 17: Ambiente no qual são realizados os laudos dos exames de Ressonância Magnética através do sistema de transcrição automática.	114
Figura 18: SpeechMike Pro - Dispositivo de Entrada de Áudio.....	123
Figura 19: Decibelímetro Utilizado nos Testes.....	123
Figura 20: Testes de Usabilidade com Voluntários.	149

LISTA DE TABELAS

Tabela 1: Perfil dos Participantes.....	149
Tabela 2: Interferência Sonora do Ambiente.	150
Tabela 3: Tempo de duração de cada teste de usabilidade.....	150
Tabela 4: Parte do Questionário sobre Experiência de Uso do Sistema.....	155
Tabela 5: Parte do Questionário sobre Experiência de Uso do Sistema – Carga Cognitiva	157
Tabela 6: Parte do Questionário sobre Experiência de Uso do Sistema – Adequação do <i>Feedback</i>	158
Tabela 7: Questão sobre a percepção do áudio pelo próprio usuário	159
Tabela 8: Formulário de Observação do <i>Delay</i> do Sistema.	161
Tabela 9: Formulário de Observação da Interferência Sonora do Ambiente.....	162
Tabela 10: Formulário de Observação do Tempo para o Laudo ficar Pronto.....	162
Tabela 11: Resultado da Observação do <i>Delay</i> do Sistema.	165
Tabela 12: Formulário de Observação da Interferência Sonora do Ambiente.....	166
Tabela 13: Resultado da Observação do Tempo para o Laudo ficar Pronto.....	167

LISTA DE GRÁFICOS

Gráfico 1: Porcentagem de Erro de Teste de um Laudo	125
Gráfico 2: Porcentagem de Erro do Sistema Com x Sem Treinamento Prévio	126
Gráfico 3: Porcentagem de Erro do Sistema Com ou Sem Interferência Sonora.....	126
Gráfico 4: Diferença da porcentagem de Erro do Sistema entre Dois Especialistas	127
Gráfico 5: Resultado da Porcentagem de Erro em 10 Testes com o Mesmo Laudo	129
Gráfico 6: Porcentagem de Erro causado por Interferência Sonora no Ambiente – Teste com o Equipamento Head Set Philips	131
Gráfico 7: Porcentagem de Erro causado por Interferência Sonora no Ambiente – Teste com o Equipamento SpeechMike Pro	131
Gráfico 8: Porcentagem de Erro causado por Interferência Sonora no Ambiente – Teste com o Equipamento SpeechMike Pro	132
Gráfico 9: Diferença da Porcentagem de Erro entre os dois Equipamentos	135
Gráfico 11: Naturalidade da Fala de Oito Voluntários	153
Gráfico 12: Porcentagem de Erro de Teste com Seis Usuários	155

LISTA DE QUADROS

Quadro 1: Diferenças entre GUI e VUI. (Fonte: autor)	38
Quadro 2: Lista de Heurísticas Revisadas de Usabilidade.....	56
Quadro 3: Características de Qualidade Relacionadas ao Diálogo.....	72
Quadro 4: Características de Qualidade relacionadas à Comunicação, Tarefa e Serviço	73
Quadro 5: Classificação de Medidas (1)	75
Quadro 6: Classificação de medidas (2)	76
Quadro 7: Agrupamento dos Requisitos em Classes.....	93
Quadro 8: Classificação da complexidade de se avaliar cada requisito.....	94
Quadro 9: <i>Template</i> utilizado para realizar a avaliação dos requisitos.	98
Quadro 10: Avaliação da Satisfação do Cliente.....	99
Quadro 11: Avaliação da Modalidade Adequada.....	99
Quadro 12: Avaliação da Diversidade e Percepção Humana.	99
Quadro 13: Avaliação da Minimização da Sobrecarga de Memória.....	100
Quadro 14: Avaliação do <i>Feedback</i> do Sistema.	100
Quadro 15: Avaliação da Eficiência do Sistema de Ajuda.....	100
Quadro 16: Avaliação da Visibilidade do Sistema.	101
Quadro 17: Avaliação da Precisão.....	101
Quadro 18: Avaliação da Integração com Sistemas já Existentes.	102
Quadro 19: Avaliação da Recuperação de Erros.	102
Quadro 20: Avaliação da Naturalidade da Fala do Usuário.	103
Quadro 21: Avaliação do Tempo para o Laudo ficar Pronto.	104
Quadro 22: Avaliação do Tamanho do Vocabulário.....	104
Quadro 23: Avaliação da Aderência do Dicionário.....	105
Quadro 24: Avaliação da Interferência Sonora do Ambiente.	105
Quadro 25: Avaliação da Capacidade do Sistema de Resolver Homônimos.....	105
Quadro 26: Avaliação da Qualidade de Áudio do Sistema.....	106
Quadro 27: Avaliação da Qualidade das Entradas de Áudio.	106
Quadro 28: Avaliação do Tempo de <i>Feedback</i> do Sistema.....	106
Quadro 29: Avaliação do Tratamento e Prevenção de Erros.....	107
Quadro 30: <i>Template</i> para Casos de Testes.....	116
Quadro 31: Caso de Teste para Avaliação da Precisão.....	116
Quadro 32: Caso de Teste para Avaliação da Recuperação de Erros do Sistema por Inspeção.....	117
Quadro 33: Caso de Teste para Avaliação da Interferência Sonora.....	117
Quadro 34: Caso de Teste para Avaliar a Resolução de Ambiguidade para Homônimos.....	118
Quadro 35: Caso de Teste para Avaliar Tempo de Resposta do Sistema.....	118
Quadro 36: Caso de Teste para Avaliar a Qualidade das Entradas de Áudio.....	119
Quadro 37: Caso de Teste para Avaliar a Qualidade do Sistema de Áudio.....	119
Quadro 38: Caso de Teste para Avaliar a Visibilidade do Sistema por Inspeção.	120
Quadro 39: Caso de Teste para Avaliar a Adequação do Sistema de Ajuda por Inspeção	120

Quadro 40: Caso de Teste para Avaliar Modalidade Apropriada.	121
Quadro 41: Caso de Teste para Avaliar a Naturalidade da Fala do Usuário por Inspeção.	121
Quadro 42: Obtenção do Tamanho do Vocabulário.	121
Quadro 43: Resultados e Análise do Caso de Teste para Avaliação da Precisão.	128
Quadro 44: Resultados e Análise do Caso de Teste para Avaliação da Recuperação de Erros do Sistema por Inspeção.	130
Quadro 45: Resultados e Análise do Caso de Teste para Avaliar a Interferência Sonora.	133
Quadro 46: Resultados e Análise do Caso de Teste para Avaliar a Resolução de Ambiguidade para Homônimos.	134
Quadro 47: Resultados e Análise do Caso de Teste para Avaliar Tempo de Resposta do Sistema.	135
Quadro 48: Resultados e Análise do Caso de Teste para Avaliar a Qualidade das Entradas de Áudio.	136
Quadro 49: Resultados e Análise do Caso de Teste para Avaliar a Qualidade do Sistema de Áudio.	137
Quadro 50: Resultados e Análise do Caso de Teste para Avaliar a Visibilidade do Sistema por Inspeção.	138
Quadro 51: Resultados e Análise do Caso de Teste para Avaliar a Adequação do sistema de ajuda por Inspeção.	139
Quadro 52: Resultados e Análise do Caso de Teste para Avaliar Modalidade Apropriada.	140
Quadro 53: Resultados e Análise do Caso de Teste para Avaliar a Naturalidade da Fala do Usuário por Inspeção	142
Quadro 54: Obtenção do Tamanho do Vocabulário.	142
Quadro 55: Caso de Teste para Avaliar a Naturalidade da Fala do Usuário por Observação.	146
Quadro 56: Caso de Teste para Avaliação da Recuperação de Erros do Sistema por Observação.	146
Quadro 57: Caso de Teste da Avaliação da Carga Cognitiva do Sistema.	146
Quadro 58: Caso de Teste para Avaliar a Adequação do <i>Feedback</i> por Observação.	147
Quadro 59: Caso de Teste para Avaliação do Tempo de <i>Feedback</i> do Sistema.	147
Quadro 60: Caso de Teste para Avaliar a Qualidade do Sistema de Áudio.	147
Quadro 61: Caso de Teste para Avaliar a Naturalidade da Fala do Usuário por Observação.	154
Quadro 62: Caso de Teste para Avaliação da Recuperação de Erros do Sistema por Observação.	156
Quadro 63: Caso de Teste da Avaliação da Carga Cognitiva do Sistema.	157
Quadro 64: Caso de Teste para Avaliar a Adequação do <i>Feedback</i> por Observação.	158
Quadro 65: Caso de Teste para Avaliar a Qualidade do Sistema de Áudio.	159
Quadro 66: Formulário de Observação da Naturaliza da Fala do Usuário	161
Quadro 67: Formulário de Observação da Recuperação de Erros	162
Quadro 68: Formulário de Observação da Sobrecarga Cognitiva do Usuário.	163
Quadro 69: Formulário de Satisfação do Usuário. (Fonte: autor)	165
Quadro 70: Formulário de Observação da Naturaliza da Fala do Usuário.	166
Quadro 71: Resultado da Observação da Recuperação de Erros.	167

Quadro 72: Formulário de Observação da Sobrecarga Cognitiva do Usuário.....	168
Quadro 73: Medição de Ruído	212
Quadro 74: Laudos utilizados nos testes	214

LISTA DE ABREVIATURAS

GUI	<i>Graphical User Interface</i>
HIS	Hospital Information System
IHC	<i>Interface Homem-Computador</i>
IVR	Interactive Voice Response
NVA	Nonverbal Audio
PACS	Picture Archiving and Communication System
RIS	Radiology Information System
SUI	Speech User Interface
TAT	Turnaround Time
TI	Tecnologia da Informação
TTS	Text to Speech
VUI	<i>Voice User Interface</i>
WER	Word Error Rate
WIMP	Windows, Icons, Mouse and Pointers
WYSIWYG	<i>What you see is what you get</i>

SUMÁRIO

ABSTRACT	9
LISTA DE ILUSTRAÇÕES	10
LISTA DE TABELAS	11
LISTA DE GRÁFICOS.....	12
LISTA DE QUADROS	13
LISTA DE ABREVIATURAS.....	16
SUMÁRIO.....	17
1. INTRODUÇÃO.....	20
1.1 Considerações iniciais.....	20
1.2 Objetivos da Pesquisa.....	23
1.3 Motivações e Justificativas	25
1.4 Métodos da Pesquisa	27
1.5 Organização do Trabalho	28
2. INTERFACE DO USUÁRIO BASEADA EM VOZ	30
2.1 Interface Homem-Computador (IHC).....	30
2.1.1 Requisitos de Interface	30
2.1.2 Tipos de Interfaces	32
2.1.3 Independência de diálogo.....	34
2.2 Definição.....	35
2.3 Principais Diferenças entre Interfaces Gráficas e interfaces do usuário baseadas em voz	37
2.4 Arquitetura de um Sistema Baseado em Interface de Voz	39
2.4.1 Módulo de Ponto de Finalização.....	41
2.4.2 Módulo de Extração de Características	42
2.4.3 Módulo de Reconhecimento de Voz	42
2.4.4 Módulo de Compreensão da Linguagem Natural	44
2.4.5 Módulo Gerenciador de Diálogo	45
2.5 Sintetização de Voz.....	45
2.6 Vantagens e Dificuldade no Uso de VUI.....	46
2.7 Questões Técnicas.....	49
2.8 Considerações Finais	50
3. AVALIAÇÃO DE USABILIDADE	51
3.1 Definição.....	51
3.2 Métodos Tradicionais de Avaliação de Usabilidade de Interfaces.....	54
3.2.1 Métodos de Inspeção de Usabilidade	54
3.2.2 Testes de Usabilidade	58
3.2.3 Comparação entre os Métodos de Avaliação de Usabilidade.....	60
3.3 Considerações Finais	60
4. USABILIDADE DE INTERFACES DO USUÁRIO BASEADAS EM VOZ	62

4.1	Contextualização	62
4.2	Avaliação de VUI – Ideia inicial	63
4.3	Requisitos não funcionais para Interface do Usuário Baseada em Voz	63
4.3.1	Requisitos não-funcionais relacionados à representação da informação	64
4.3.2	Requisitos não-funcionais relacionados à entrada de dados.....	66
4.4	Características da Qualidade em Interface do Usuário Baseada em Voz	69
4.5	Métricas para a Avaliação de Usabilidade para Interface do Usuário Baseada em Voz	73
4.6	Considerações Finais	77
5.	INTERFACE DO USUÁRIO BASEADA EM VOZ PARA SISTEMAS DE INFORMAÇÃO RADIOLÓGICA	78
5.1	Considerações Iniciais.....	78
5.2	Sistemas de Transcrição Automática de Laudos.....	80
5.3	Funcionamento de um Sistema de Transcrição Automática de Laudos.....	83
5.4	Considerações Financeiras	85
5.5	Requisitos de um Sistema de Transcrição Automática de Laudo.....	86
5.6	Considerações Finais	89
6.	METODOLOGIA DE AVALIAÇÃO DE INTERFACES DE VOZ PARA A ÁREA DA SAÚDE – SISTEMAS DE TRANSCRIÇÃO AUTOMÁTICA DE LAUDO	90
6.1	Introdução.....	90
6.2	Objetivos da Metodologia Utilizada	91
6.3	Agrupamento dos Requisitos em Classes.....	91
6.4	Classificação dos Requisitos quanto à Dificuldade de Avaliação	93
6.5	Requisitos Seleccionados	94
6.6	Método de Avaliação para cada Requisito Seleccionado.....	97
6.7	Considerações Finais	107
7.	APLICAÇÃO DA METODOLOGIA A SISTEMAS JÁ IMPLANTADOS.....	108
7.1	Descrição dos sistemas utilizados.....	109
7.2	Planejamento da Inspeção do Sistema Stand-alone.....	115
7.3	Planejamento das Observações dos Usuários Finais em Ambiente de Produção	159
7.4	Discussões sobre os resultados dos testes.....	168
8.	CONCLUSÕES.....	173
8.1	Aspectos Gerais do Trabalho	173
8.2	Contribuições do Trabalho.....	176
8.3	Trabalhos Futuros	182
	REFERÊNCIAS BIBLIOGRÁFICAS	184
	APÊNDICES E ANEXOS	194
	APÊNDICE A - QUESTIONÁRIO DE SATISFAÇÃO DE USUÁRIOS DE SISTEMAS DE TRANSCRIÇÃO DE LAUDOS EM RADIOLOGIA	195
	APÊNDICE B - CARTA DE APRESENTAÇÃO PARA OS PARTICIPANTES DO TESTE	198
	APÊNDICE C – AUTORIZAÇÃO PARA O USO DE IMAGEM E SOM	200
	APÊNDICE D – PLANO DO OBSERVADOR.....	202
	APÊNDICE E - RESULTADOS DA APLICAÇÃO DO QUESTIONÁRIO	205

APÊNDICE F - PLANO DE OBSERVAÇÃO PARA O USO DO SISTEMA DE TRANSCRIÇÃO AUTOMÁTICA DE LAUDOS EM RADIOLOGIA.....	208
ANEXO A - TABELA DE REFERÊNCIA DE INTERFERÊNCIA SONORA	211
DESEJÁVEL PARA AMBIENTES	211
ANEXO B - LAUDOS UTILIZADOS NOS TESTES E INSPEÇÕES.....	213
DE USABILIDADE.....	213

1. INTRODUÇÃO

Este trabalho relaciona elementos das áreas de Computação e Saúde para comporem a elaboração de uma metodologia para avaliação de usabilidade de sistemas de transcrição automática de laudos na área de Radiologia.

Serão apresentados, neste capítulo, as considerações iniciais deste trabalho, os objetivos, motivações e justificativas para a elaboração da Tese, a metodologia utilizada e a organização do trabalho.

1.1 Considerações iniciais

Interface Homem-Computador (IHC) pode ser definida como a parte de um sistema computacional com a qual uma pessoa entra em contato físico, perspectivo e conceitualmente (ORTH, 2005), ou seja, a interface pode ser entendida como um sistema de comunicação que possui dois componentes, um sendo o componente físico – *hardware* e *software* – em que o usuário é capaz de percebê-lo e interagi-lo; e outro conceitual, em que ele interpreta, processa e raciocina.

No início da computação, em que apenas os especialistas utilizavam sistemas computacionais, pouco se explorou efetivamente da área de Interface Homem-Computador. Porém, desde a popularização do acesso aos equipamentos e a consequente proliferação de usuários não especialistas, há uma preocupação, cada vez mais acentuada, de se criar interfaces que se julguem mais “amigáveis”.

Nos últimos anos, a área de IHC tem se desenvolvido e um grande esforço tem sido empregado em relação a questões de projeto e avaliação de interfaces não convencionais, tais como em Realidade Virtual, Realidade Aumentada, Interfaces por Gestos e Interface do Usuário Baseada em Voz, do inglês *Voice User Interface*.

A origem das Interfaces do Usuário Baseadas em Voz se deu nos anos 50, quando pesquisadores vislumbraram a voz como forma de comunicação entre o humano e a máquina. Afinal, desde que a fala constitui a principal interface para a comunicação entre pessoas, era de se esperar que essa forma natural de interação entre seres humanos fosse pensada como interação entre homens e máquinas.

Os sistemas de diálogo de voz surgiram de pesquisas em Inteligência Artificial dos anos 50 (ALLEN; PERRAULT, 1980; PRICE, 1990; KAMM et al, 1997), a partir do desenvolvimento das “conversational interfaces”. Porém, somente a partir dos anos 90, houve um crescimento significativo na criação e desenvolvimento de interfaces do usuário baseadas em voz, especialmente para uso via telefone, para fins comerciais, mas também científicos. Isto se deve ao maior poder computacional dos dispositivos digitais, aos avanços de tecnologia de reconhecimento e síntese de voz, processamento de linguagens naturais e a modelagem de diálogo entre homem e máquina.

Interfaces do usuário baseadas em voz usam tecnologia de reconhecimento e síntese de voz para fornecer a seus usuários acesso à informação, permitindo-os executar transações e oferecendo suporte à comunicação (MCTEAR, 2002; COHEN; GIANGOLA; BALOGH, 2004). O estado da arte em tecnologia de voz já permite que sistemas automáticos sejam desenvolvidos para trabalhar em condições reais (SAN-SEGUNDO et al, 2005). Empresas como a Nuance (2010), a IBM ViaVoice (IBM, 02/03/2009) e a Philips Speech (2010) – compradas recentemente pela Nuance (2010) – e a Microsoft (MICROSOFT, 10/04/2010) têm investido no desenvolvimento de sistemas de voz para domínios restritos¹.

Apesar deste avanço, pouco se tem falado a respeito do processo de desenvolvimento de aplicações baseadas em comandos de voz. As particularidades dessas aplicações devem ser mais bem destacadas e os impactos dessas no processo de desenvolvimento da aplicação devem ser estudados de maneira mais aprofundada.

¹ O termo “domínios restritos” está sendo empregado aqui como sub-áreas do conhecimento, que trabalham com um conjunto finito de palavras em seu vocabulário. Como exemplo, a área de Radiologia é uma sub-área da Saúde, e os seus especialistas trabalham com uma fatia bem menor de termos do que os 100 mil termos da área Médica e 60 milhões de possibilidades de diagnósticos presentes no SNOMED (07/03/09).

Outro ponto-chave são os processos de avaliação destas interfaces do ponto de vista de usabilidade. A criação de diretrizes (*guidelines*) que facilitem todas as fases da análise de requisitos, a posterior criação destes sistemas e de sua avaliação é crucial para o aumento da qualidade destes sistemas.

Ainda vale destacar que poucos trabalhos tem se voltado para o estudo de metodologias de avaliação para esta classe de aplicações. O conhecimento atual sobre avaliação de *Voice User Interface* (VUI) advém de pequenas contribuições de avaliações que foram desenvolvidas para projetos específicos, e que tentam generalizar e propor recomendações para tais classes de aplicações, como PARADISE (WALKER et al, 1997), EAGLES (GIBBON; MOORE; WINSKI, 1997) e DISC (DYBKJAER; BERNSEN, 2000).

Analisando, especialmente, a área da Saúde, a utilização de reconhecimento de voz em sistemas da Saúde de propósito geral, tais como em emergência, não tem sido eficiente devido ao grande vocabulário do domínio - sabe-se que um profissional de saúde utiliza, no seu cotidiano, mais de 100 mil itens no seu vocabulário. Além disto, as possibilidades diagnósticas presentes no SNOMED (07/03/09), por exemplo, são mais de 60 milhões. Ou seja, a informação disponível na área da Saúde é extremamente variada.

Assim, a tecnologia de reconhecimento de voz tem sido usada para propósitos mais específicos, tais como em sistemas de ditado como, por exemplo, sistemas de transcrição automática de laudos²; sistemas com unidades de resposta audível baseados em voz (IVR); sistemas para controlar equipamentos médicos, e sistemas para interpretação de linguagem (DURLING; LUMSDEN, 2008). Isto significa que o vocabulário é consideravelmente menor, proporcionando uma precisão mais alta no reconhecimento dos termos específicos.

Porém, muito pouco tem sido explorado em questão da avaliação destes sistemas de voz para uso em domínio restrito na área da Saúde. Não há trabalhos suficientes na literatura que estabeleçam os requisitos específicos da área que devem ser atendidos de forma a tornar o uso de reconhecimento de voz efetivo e eficaz.

² O termo “laudo radiológico” ou laudo em Radiologia é um documento, em que são mencionadas a técnica utilizada e as possíveis intercorrências durante o procedimento. A descrição dos achados deve ser redigida de forma sucinta e completa, e quando o diagnóstico é seguro, terminar com uma impressão/conclusão (FENELON, 2003). Exemplos de laudos podem ser visualizados na Anexo B deste trabalho.

Os sistemas que utilizam reconhecimento de voz na área da Saúde, não são, na sua grande maioria, analisados seguindo métodos mais completos e complexos e bem-estabelecidos de avaliação de usabilidade. Além disto, algumas avaliações de usabilidade para esta classe de aplicações são focadas apenas no desempenho do reconhecimento de voz num contexto particular do ambiente, enquanto outras avaliações se voltam para a usabilidade de um sistema todo, em que o reconhecimento de voz é apenas um dos aspectos (DURLING; LUMSDEN, 2008).

Sendo assim, torna-se necessário o desenvolvimento de uma metodologia para analisar a usabilidade de sistemas de transcrição automática de laudos que possa abordar não somente falhas no reconhecimento da voz, mas que possa tratar questões mais amplas, tais como a satisfação do usuário.

1.2 Objetivos da Pesquisa

O objetivo deste trabalho de pesquisa é propor uma metodologia de avaliação da usabilidade de interfaces do usuário baseadas em voz, em aplicações da Saúde, mais especificamente, em sistemas de transcrição automática de laudos em Radiologia.

A elaboração desta metodologia tem o intuito de poder ser utilizada como um guia para hospitais e clínicas médicas que desejem comprar sistemas de transcrição automática de laudos em Radiologia, o próprio desenvolvimento deste sistema ou customizações e, que estejam preocupados, tanto com a usabilidade quanto pela aceitabilidade destes sistemas por seu corpo de médicos radiologistas.

Tal plano deve incluir o estudo e o estabelecimento de métricas de usabilidade a partir dos atuais requisitos de interface - criados principalmente para atender às interfaces gráficas (GUI) -, de requisitos próprios para interfaces do usuário baseadas em voz e o estudo e a metrificação de requisitos específicos, definidos e estudados, de sistemas de transcrição automática de laudos em radiologia. Assim, os objetivos específicos deste trabalho são:

1. Estudar o estado da arte das tecnologias envolvidas na construção de interfaces de usuário baseadas em voz, inclusive em relação às suas limitações.

2. Conhecer todo o processo de criação de interfaces do usuário baseadas em voz: análise dos requisitos, projeto, implementação, testes e implantação.
3. Estudar os requisitos próprios de VUI, em contraposição aos requisitos de GUI.
4. Entender as particularidades da área da Saúde, principalmente em Radiologia, no que tange às interfaces do usuário baseadas em voz.
5. Estudar aplicações de VUI de um modo geral, VUI na área da Saúde e mais especificamente na Radiologia.
6. Buscar, na literatura, os requisitos mencionados para a avaliação de VUIs e VUI em Sistemas de Informação Radiológica (RIS), desenvolvendo uma análise quanto a sua completude e suficiência.
7. Estabelecer outros requisitos para VUI em RIS que forem julgados pertinentes.
8. Classificar os requisitos propostos, assim como sua importância no processo de avaliação e dificuldades de avaliação.
9. Escolher e justificar o uso dos requisitos classificados dentro da metodologia.
10. Estudar os métodos de inspeção de usabilidade e os métodos de testes de usabilidade utilizados na área de Interface Homem-Computador.
11. Escolher o método de avaliação apropriado para cada requisito escolhido, levando em consideração fatores como eficiência, tempo e custo.
12. Estabelecer métricas a fim de que se possam analisar os requisitos de maneira quantitativa.
13. Aplicar a metodologia proposta a um conjunto de sistemas de VUI em Radiologia.
14. Planificar todos os testes necessários para a validação da metodologia.
15. Analisar os resultados encontrados a partir da aplicação da metodologia.
16. Apresentar pontos fortes e fracos do uso da metodologia.
17. Apresentar trabalhos futuros que possam ser desenvolvidos a partir desta pesquisa.
18. Estabelecer considerações finais sobre o trabalho, incluindo a discussão sobre as limitações do mesmo.

1.3 Motivações e Justificativas

Um dos principais problemas apontados na literatura (WHITE, 2005; KANAL et al, 2001; BHAN et al, 2008; HUNDT et al, 1999; DURLING; LUMSDEN, 2008; GOPAKUMAR et al, 2008; KIMBERLY, 2006) é a demora da disponibilização dos laudos de Radiologia devido ao tempo gasto entre o envio de laudos gravados e seu retorno de forma textual para a avaliação do radiologista (TAT – *Tournaround Time*).

Este problema pode se agravar consideravelmente se o laudo pertencer a um paciente que necessite de uma investigação radiológica imediata para receber intervenção apropriada. Como os laudos transcritos podem consumir bastante tempo, os médicos necessitam esperar mais do que o necessário antes de definir uma conduta adequada para o caso em estudo.

Os sistemas de transcrição automática de laudos (que utilizam VUI) têm sido cogitados como solução para diminuir este tempo (TAT) e também como redutor dos custos globais dos departamentos de Radiologia. Porém, para se verificar a eficácia de se utilizar sistemas de transcrição automática de laudos, é preciso avaliar se os requisitos gerais de VUI e também os requisitos específicos da área estão sendo tratados de maneira correta nos produtos disponíveis no mercado.

Dos trabalhos pesquisados até o momento na literatura, pouco se tem abordado sobre avaliação de sistemas de transcrição automática de laudos. Quando isto é realizado, apenas a precisão (acurácia) do reconhecimento de voz ou a detecção de erros (KANAL et al, 2001; KIMBERLY, 2006; VOLL; ATKINS; FORSTER, 2008; PAULETT; LANGLOTZ, 2009; McGURK et al, 2008) tem sido avaliada, ou seja, os trabalhos na área de avaliação ainda são muito incipientes. Não há suficiência de trabalhos e, tampouco, de conteúdo sobre uma metodologia de como avaliar tais sistemas.

Por outro lado, a maioria dos sistemas de VUI encontrados na literatura trata de sistemas de consultas a determinadas informações vindas de uma base de dados. São, assim, sistemas relativamente simples de reservas de passagens aéreas, reserva de hotéis, consulta de horários de vôos, e similares. Desta maneira, possuem um vocabulário pequeno, com uma interação entre o usuário e a máquina bastante grande (diálogo). Muitos pesquisadores (WALKER; PASSNNEAU; BOLAND, 2001; MÖLLER, 2002; MÖLLER 2005; LARSEN, 2003; HARTIKAINEN; SALONEN; TURUNEN, 2004) têm proposto métodos e metodologias para avaliar a usabilidade destes sistemas mais simples. Porém, avaliar os sistemas de transcrição automática de laudos é uma tarefa relativamente diferente, desde que:

- O vocabulário é mais extenso, implicando numa taxa de reconhecimento mais baixa.
- O vocabulário deve ser específico para a área de Radiologia.
- O diálogo entre o usuário e o sistema é bem menor, visto que o sistema deve apenas transcrever (gerar um texto) a partir da fala do usuário, sem haver consultas a serem realizadas e informadas ao usuário.
- O radiologista não deseja ser interrompido quando está ditando um laudo.
- O tratamento e a prevenção de erros são bastante diferentes. O sistema deve ser apto a apontar, de alguma maneira, as palavras, no texto, que foram mal-compreendidas, mas não deve perguntar novamente ao usuário ou permitir confirmação.
- A satisfação do usuário está diretamente ligada à precisão (acurácia) do texto transcrito e também ao tempo para a realização da tarefa.
- A qualidade das mensagens de saída e a adequação das frases de saída devem ser substituídas pela precisão do texto.

Há, também, algumas outras considerações importantes aplicáveis aos sistemas de transcrição automática de laudos:

1. Falhas no laudo causadas por erros no reconhecimento das palavras podem ser desastrosas para os pacientes, pois podem significar um diagnóstico equivocado.

2. Como criar uma metodologia que leve em consideração os principais requisitos para este tipo de sistema.
3. Os desafios encontrados para avaliar estes tipos de sistemas são:
 - a) Quais requisitos de VUI devem ser considerados para se avaliar um sistema de laudo automático;
 - b) Quais os requisitos específicos dos sistemas de laudo automático que devem ser considerados para esta avaliação;
 - c) Quais, dentre os muitos requisitos apresentados, são ditos fundamentais e viáveis de se avaliar;
 - d) Como medir cada requisito apontado como fundamental destes sistemas;
 - e) Como avaliar estes sistemas de uma maneira viável, com custos e tempos aceitáveis para a instituição da saúde;

Sendo assim, há que se propor uma metodologia para avaliação destes sistemas que considere estas questões particulares.

1.4 Métodos da Pesquisa

Esta pesquisa foi desenvolvida a partir do interesse crescente desenvolvido na participação de projetos na área de interfaces não-convencionais, especificamente em projetos no âmbito profissional em VUI e também no interesse na aplicabilidade de IHC na área da Saúde. A investigação do uso de VUI na área radiológica incentivou e motivou um estudo mais aprofundado sobre o assunto.

Assim, o levantamento bibliográfico teve o objetivo de desvendar, coletar e analisar as principais pesquisas sobre os assuntos relacionados que contribuíram para o avanço do tema.

A seguir, o método exploratório visou um aprofundamento no conhecimento de sistemas de transcrição automática de laudos, através de visitas técnicas, conversas informais e experimentação do sistema.

Ao final, foi realizada uma pesquisa de campo por corresponder à coleta direta de informação no ambiente em que a técnica pode ser usada. Segundo Marconi e Lakatos (2007), pesquisa de campo é aquela que é realizada fora do laboratório, no próprio ambiente no qual as ocorrências são geradas.

Desta forma, a fim de se alcançar os objetivos gerais e específicos propostos neste trabalho, as seguintes atividades foram realizadas:

1. Revisão bibliográfica dos temas envolvidos no projeto, incluindo: Interface Homem-Computador, interfaces do usuário baseadas em voz, sistemas de informação radiológica com VUI, e métodos tradicionais de avaliação de usabilidade;
2. Identificação dos requisitos genéricos para interfaces do usuário baseadas em voz;
3. Identificação dos requisitos de interface do usuário baseada em voz para sistemas em Saúde, especialmente, para sistemas de informação radiológica;
4. Geração de uma metodologia de avaliação de interfaces do usuário baseadas em voz, adaptada às peculiaridades desta classe de interfaces;
5. Aplicação da Metodologia proposta a um conjunto restrito, porém significativo, de sistemas baseados no diálogo por voz, em domínios restritos da área da Saúde, a fim de validar a Metodologia.

Percebeu-se que um nome mais apropriado para esta tese poderia envolver o termo “metodologia”, pois, na verdade, trata-se de uma metodologia proposta pela autora para avaliar os sistemas de transcrição de laudos em Radiologia.

1.5 Organização do Trabalho

Este trabalho de pesquisa está organizado em capítulos, conforme explicitado a seguir:

O Capítulo 1, INTRODUÇÃO, aborda as considerações sobre a pesquisa, tais como seus objetivos, metodologia e justificativas.

No Capítulo 2, INTERFACE DO USUÁRIO BASEADA EM VOZ, trata de um estudo aprofundado sobre o tema, apontando suas diferenças em relação às interfaces gráficas do usuário, além de sua arquitetura e principais funções.

O Capítulo 3, AVALIAÇÃO DE USABILIDADE DE INTERFACES, resgata as principais metodologias clássicas para avaliação as interfaces do usuário. Estas metodologias foram, principalmente, criadas para avaliação das interfaces gráficas.

No Capítulo 4, USABILIDADE DE INTERFACES DO USUÁRIO BASEADAS EM VOZ, são apresentados os principais requisitos que devem ser analisados quando se avalia uma VUI, algumas heurísticas para sua avaliação e características de qualidade para estes sistemas.

O Capítulo 5, INTERFACE DO USUÁRIO BASEADA EM VOZ PARA SISTEMAS DE INFORMAÇÃO RADIOLÓGICA, versa sobre o uso de VUI aplicado à área da Saúde, principalmente no que se refere ao seu uso em transcrições de laudos em Radiologia, tratando, inclusive sobre seus requisitos.

O Capítulo 6, METODOLOGIA DE AVALIAÇÃO DE INTERFACES DE VOZ PARA A ÁREA DA SAÚDE – SISTEMAS DE TRANSCRIÇÃO AUTOMÁTICA DE LAUDO, expõe a metodologia de avaliação, que compõe os resultados parciais.

O Capítulo 7, APLICAÇÃO DA METODOLOGIA A SISTEMAS JÁ IMPLANTADOS, aborda os passos utilizados para que seja possível avaliar sistemas de transcrição automática de laudo sobre uma metodologia proposta pela autora.

No Capítulo 8, CONSIDERAÇÕES FINAIS, são destacadas as contribuições da pesquisa e as limitações dos resultados obtidos, trazendo algumas indicações de trabalhos futuros.

Os Apêndices e Anexos apresentam as ferramentas de coleta de dados, a tabela de referência sonora desejável para ambientes e os laudos utilizados para os testes.

2. INTERFACE DO USUÁRIO BASEADA EM VOZ

Este capítulo faz um levantamento bibliográfico de Interface Homem-Computador (IHC) e daí dos conceitos relacionados à Interface do Usuário Baseada em Voz (VUI). Também versa sobre as principais diferenças entre um projeto de Interfaces Gráficas do Usuário (GUI) e Interfaces do Usuário Baseadas em Voz.

2.1 Interface Homem-Computador (IHC)

Segundo Myers (1995), a pesquisa em IHC estuda como as pessoas projetam, implementam e usam sistemas computacionais interativos e como os computadores afetam o indivíduo, as organizações e a sociedade. Isto abrange não só a facilidade de uso, mas também novas técnicas de interface para dar suporte às tarefas dos usuários, fornecendo um melhor acesso à informação e criando formas mais poderosas de comunicação. Ela envolve dispositivos de entrada e saída (E/S) e técnicas de interação que os usam; como a informação é apresentada e requisitada; como os processos computacionais são controlados e monitorados; todas as formas de ajuda, documentação e treinamento; as ferramentas usadas para projetar, construir, testar e avaliar as interfaces do usuário; e os processos que os desenvolvedores seguem para criar interfaces.

2.1.1 Requisitos de Interface

Liang (1987), Fischer (1990) e Rocha e Baranauskas (2003), estabeleceram que para se obter uma boa interface, é necessário que certos requisitos sejam atendidos:

- **Diversidade** – A interface deve fazer todo o tratamento do usuário, ou seja, suportar todas as classes de usuários, identificar cada usuário e adaptar-se a ele, desenvolvendo uma linguagem adequada ao usuário em questão;

- Complacência – A interface deve dar suporte ao usuário, quando este fizer a recuperação de seus erros, assim como prever que o usuário se esqueça de informações já apresentadas.
- Eficiência – A interface deve minimizar o esforço do usuário para executar uma dada tarefa. Deve existir a garantia de que o sistema auxilia ao usuário a executar determinada ação mais rapidamente do que o faria manualmente.
- Conveniência – A interface deve ser capaz de proporcionar acesso fácil a todas as operações.
- Flexibilidade – A interface deve fornecer várias maneiras para se executar uma dada operação.
- Consistência – A interface deve ter comportamentos e apresentação física bem definidos por regras e conhecidos pelo usuário, como por exemplo: empregar sempre a mesma codificação; empregar caracteres de teclado sempre na mesma função; e mostrar as mensagens de estado do sistema em local fixo.
- Prestimosidade – A interface deve fornecer ajuda sempre que requisitada ou quando perceber que o usuário se encontra em dificuldades.
- Imitação – A interface deve explorar os aspectos de comunicação não orientados a comandos, tais como o uso de exemplos, explanações, analogias, comparações e descrições.
- Naturalidade – A interface deve envolver o usuário de uma maneira bastante natural, não exigindo terminologia não referente à tarefa.
- Satisfação – A interface deve realizar o que o usuário espera, não o frustrando. Seu tempo de resposta deve ser suficientemente pequeno a ponto de não desmotivar o usuário e deve permitir que o usuário obtenha ajuda em qualquer ponto de sua interação.
- Passividade – A interface deve permitir que o usuário detenha o controle da interação.

Foley e Van Dan (1990) e Nielsen (1993) destacam as principais metas a serem atingidas, ao se desenvolver uma dada IHC:

- Aumento da Velocidade de Aprendizado – Significa a diminuição do tempo em que um novo usuário leva para atingir um certo grau de proficiência com um sistema.
- Aumento da Velocidade de Uso – Ou seja, uma diminuição do tempo em que um usuário experiente requer para executar alguma tarefa específica dentro do sistema, seja porque a interface utiliza *shortcut* ou porque evita retrabalho.
- Redução da Taxa de Erros – Significa uma redução do número de erros do usuário por interação. Alguns erros de interação que podem ocorrer uma interface são: digitação incorreta (sem validação), seleção incorreta de um item em um formulário, valores corretos em campos errados, campos no formulário que confundem o usuário (SALVADOR et al, 2006).
- Facilidade de uma Rápida Recordação de como Utilizar a Interface – Permitindo que um usuário inativo, possa retornar rapidamente a utilizar o sistema.
- Aumento da Atratividade para Potenciais Usuários e Compradores – Isto é, aumentar a percepção do conjunto de usuários.

2.1.2 Tipos de Interfaces

Segundo a classificação tradicional, há quatro tipos mais comuns de interface: WYSIWYG, icônica, de manipulação direta e interface gráfica com o usuário, que são consideradas como “intrinsecamente gráficas”.

- WYSIWYG (*What you see is what you get*): Neste tipo de interface, a representação com que o usuário interage nas telas de interface é basicamente a mesma que a imagem final criada pela aplicação. A maioria das aplicações interativas possui algum componente nesta representação. Por exemplo, um editor de texto que imprime uma letra em negrito também a mostra neste formato em vídeo.

- Icônica: Aqui, os objetos, atributos ou relações que podem ser operados são representados de forma pictórica (ícones). Um ícone é um símbolo caracterizado pela semelhança ou por analogia entre o símbolo escolhido e o que ele representa.
- De Manipulação Direta: A manipulação direta trata os objetos, atributos, relações e operações de uma forma visual, seja através de gráficos ou textos. Esta forma de interface faz com que o usuário esteja mais integrado com a semântica da aplicação. Nas interfaces de manipulação direta, não há operações escondidas, nem sintaxes para serem aprendidas (ROCHA; BARANAUSKAS, 2003). Tal estilo de interface apresenta uma metáfora da sua aplicação, em que o usuário pode manipular os objetos. A metáfora utilizada é o ponto fundamental na manipulação direta (CURTIS, 1994). Ela deve ser suficientemente simples para que o usuário a compreenda rapidamente, e ao mesmo tempo bastante sofisticada para oferecer de forma natural, intuitiva e consistente todas as operações necessárias à tarefa do usuário. O desenvolvimento desta metáfora pode ser complicado, até mesmo porque, muitas vezes, é difícil identificar os objetos a manipular. Entre as interfaces de manipulação direta estão: seleção de menus, linguagem de comandos, linguagem natural e caixa de diálogo.
- Interface Gráfica com o Usuário (GUI): Conforme Vavassori (1995), as formas de IHC evoluíram de comandos textuais para mecanismos de interação avançados, tais como gráficos e objetos tridimensionais, oferecendo suporte a uma ampla variedade de características. Vários benefícios foram trazidos com a interface gráfica:
 - Acesso Direto – Um objeto gráfico pode ser considerado como um objeto atômico com facilidade de ser focado. Utilizando-se variações de forma, cor e tamanho, pode-se tornar fácil atrair o usuário para a área de interesse.
 - Largura de Banda – A quantidade de informação contida em um gráfico é muito grande se comparada a outras formas de representação.
 - Riqueza de Expressão – Se tratados de forma eficiente, os gráficos desenvolvem um contexto apropriado para representar uma informação ao usuário.

- Consistência – Com o surgimento e utilização das GUIs, facilitou-se a transposição de consistência de um sistema para outro, reduzindo o tempo de aprendizado e a ansiedade do usuário.
- Comandos Naturais – As interfaces gráficas trabalham a intuição natural do usuário, o que torna o seu aprendizado relativamente fácil.
- Multi-tarefa – Permite que diversas aplicações compartilhem a tela, simultaneamente, através de janelas diferentes.

Porém, há, atualmente, vários outros tipos de interface que não se enquadram nestas quatro categorias supracitadas, as chamadas interfaces não-convencionais, entre as quais é possível citar as interfaces *touch-screen* e as interfaces do usuário baseadas em voz (que serão vistas na próxima seção).

2.1.3 Independência de diálogo

Independência de diálogo consiste na clara separação entre interface (componentes de diálogo) e aplicação (componentes de computação), como mostra a Figura 1 (VAVASSORI,1995).

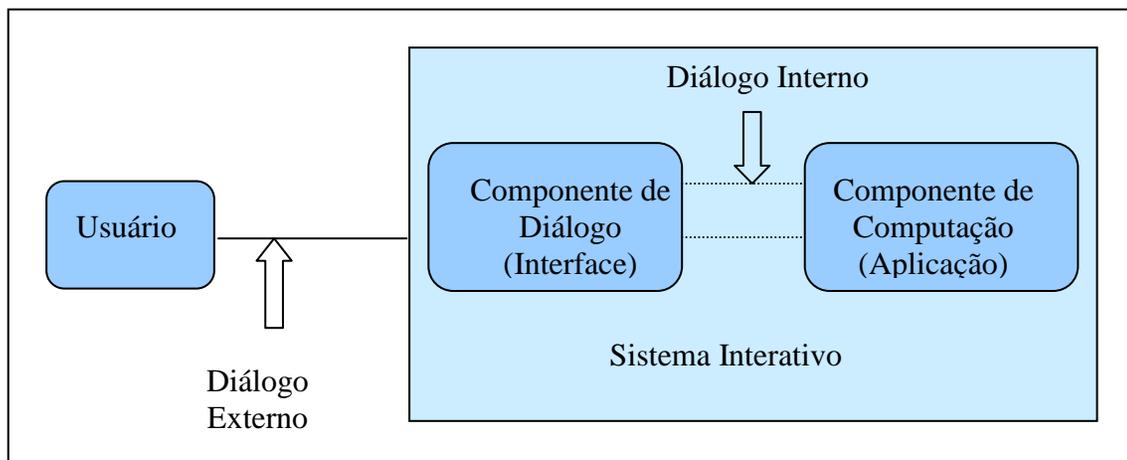


Figura 1: Componentes de um Sistema Interativo

Como se sabe, a interface é responsável pela comunicação com o usuário, o que é chamado de diálogo externo. Já a aplicação é responsável pela execução das tarefas requisitadas pelo usuário através da interface. Entre a interface e a aplicação é estabelecido o diálogo interno.

Esta abordagem de desenvolvimento permite que a construção dos componentes do sistema seja feita em separado. Assim, uma equipe composta por psicólogos, artistas visuais, cientistas da computação e usuários pode cuidar de desenvolver a interface, enquanto que uma equipe integrada por analistas, programadores e usuários deve cuidar do desenvolvimento da aplicação. Este enfoque também possibilita uma manutenção mais simplificada, já que é possível alterar um componente sem influenciar o outro.

Entretanto, a independência de diálogo possui algumas desvantagens: se o diálogo interno não for bem estruturado pode não ser capaz de efetivar todas as trocas de informação entre os dois componentes; a sobrecarga na comunicação interna em tempo de execução pode gerar uma queda de desempenho; e é necessário saber como se efetuará a gerência da comunicação interface-aplicação, já que há dois componentes se comunicando. Segundo Liang (1987), Fischer (1990) e Rocha e Baranauskas (2003), para se obter uma boa interface, é necessário que certos requisitos sejam atendidos.

2.2 Definição

Interface do usuário baseada em voz, em inglês *Voice User Interface* (VUI) ou *Speech User Interface* (SUI), consiste na interação de uma pessoa com um sistema através de voz, utilizando uma aplicação de linguagem falada (SHNEIDERMAN, 2000).

As VUIs tiveram sua origem em pesquisas de Inteligência Artificial, especialmente do desenvolvimento de “conversational interfaces”, na década de 50. Mas somente depois dos anos 90 é que esta tecnologia contou com uma significativa melhoria (MCTEAR, 2002; COHEN, GIANGOLA, BALOGH, 2004).

Sistemas com interfaces do usuário baseadas em voz são sistemas capazes não somente de reconhecer a voz do usuário, mas compreender o que ele diz e responder a estas entradas, geralmente, em tempo real (LAI, 2000).

Dependendo do propósito do sistema em questão, é possível ter aplicações que somente reconhecem o que o usuário está dizendo, qualquer que seja este usuário, chamados de sistemas independentes do usuário; outros sistemas precisam apenas garantir que aquele usuário seja quem diz ser (autenticador de voz) – chamados de sistemas dependentes do usuário, ou ambos os requisitos em conjunto, ou seja, é necessário que o sistema entenda o que o usuário autenticado diz.

Este tipo de interface inclui elementos tais como: *prompts* ou mensagens do sistema, gramáticas e lógica de diálogo ou fluxo de chamada (*call flow*). Os *prompts* são todas as mensagens de voz pré-gravadas ou sintetizadas que devem ser executadas durante o diálogo com o usuário. Gramáticas definem todas as palavras, sentenças ou frases que podem ser ditas pelo usuário em resposta a um *prompt*. A lógica de diálogo define todas as ações a serem tomadas pelo sistema em determinado ponto da interação, tais como um acesso à base de dados (COHEN; GIANOLA; BALOGH, 2004).

A seguir é apresentado um exemplo de um sistema de atendimento automático de chamadas telefônicas (Figura 2):

SISTEMA: [Nome da empresa] Bom dia. Com quem você deseja falar?
USUÁRIO: Eu gostaria de falar com Valéria Farinazzo.
SISTEMA: Desculpe-me, você gostaria de falar com Valéria Farinazzo?
USUÁRIO: Sim.
SISTEMA: Estamos transferindo sua chamada para Valéria Farinazzo. Obrigado.

Figura 2: Exemplo de um diálogo entre o sistema e um usuário

Neste exemplo, uma locutora gravou todas as possíveis falas do sistema. Após o *prompt* “Bom dia. Com quem você gostaria de falar?” o sistema ouve, usando uma gramática que ajusta entradas do usuário com o nome da pessoa da empresa; como o reconhecimento não atingiu uma taxa considerada segura, o sistema pede a confirmação “Desculpe-me, você gostaria de falar com” seguido do nome da pessoa que possivelmente o sistema entendeu. Dada a confirmação, o sistema reproduz a mensagem “Estamos transferindo sua chamada para..” acompanhado do nome da pessoa e conecta o áudio das duas pessoas.

Interfaces audíveis são as que interagem somente através do som – tipicamente, uma entrada falada (sonora) do usuário e uma saída falada (sonora) ou não-falada do sistema. As saídas não-faladas (Nonverbal audio - NVA, em inglês) podem ser na forma de *earcons* (ícones audíveis, ou sons projetados para comunicar um significado específico – por exemplo, o aviso de que um amigo está online), música de fundo ou ambiental (para criar uma sensação) ou outros sons de fundo (CHAMBERLAIN et al, 2006; OLIVEIRA NETO; SALVADOR, 2007).

Uma das principais questões relacionadas à voz para interfaces é que ela é mais transiente do que estática, ou seja, o usuário deve obter a informação no momento em que esta informação é fornecida - não há uma tela para visualizar a informação. Não existe a oportunidade de rever as saídas do sistema ao seu próprio ritmo. O ritmo é totalmente controlado pelo sistema. Se a interface não for muito evidente em todos os seus caminhos, o usuário terá que aprender através de tentativa-e-erro, tornando o sistema insatisfatório (DYBKJAER; BERNSEN, 2001). Portanto, a natureza efêmera das interfaces audíveis exige um esforço cognitivo significativo do usuário. O projetista desse tipo de interfaces deve seguir diretrizes que diminuam a carga cognitiva do usuário e que proveja mecanismos para que o ritmo da interação seja controlado por ele (OLIVEIRA NETO; SALVADOR, 2007).

2.3 Principais Diferenças entre Interfaces Gráficas e interfaces do usuário baseadas em voz

Ao ser detectado que num projeto de Interface Gráfica com o Usuário, há uma separação clara da lógica da aplicação do modelo em que a aplicação é mostrada, ou seja, uma separação entre a parte funcional e a forma de apresentação desta, então pode-se perceber que é plausível estender a aplicação para trabalhar sobre uma interface de voz.

Porém, é possível perceber que a natureza dos projetos baseados em componentes para aplicações gráficas e de voz podem ser bastante diferentes.

O Quadro 1 reflete esta diferença e o texto a seguir explica estas diferenças (HUNT; WALKER, 2000):

CARACTERÍSTICA	GUI	VUI
Visibilidade	X	
Transiência		X
Assimetria de Banda-Larga		X
Temporalidade		X
Concorrência		X

Quadro 1: Diferenças entre GUI e VUI. (Fonte: autor)

- Visibilidade: Interfaces gráficas são visíveis para o usuário e a maioria das funcionalidades está aparente para o usuário para inspeção visual ou para manipulação (através de menus, por exemplo). Em contraste, voz é invisível. Assim, o usuário deste tipo de aplicação pode apresentar maior dificuldade para determinar as ações que ele pode realizar e como invocá-las. Também, uma entrada de voz com erro no reconhecimento pode levar a ações não desejadas no sistema, se estas falhas não forem previstas no projeto. Sendo assim, sob o ponto de vista do desenvolvedor, a invisibilidade dos sistemas de voz é mais difícil de projetar e controlar e isto motiva o uso de monitores visuais.
- Transiência: entradas e saídas de voz são transientes. Uma vez que se disse ou se ouviu algo, esta informação não está mais disponível. Por outro lado, as interfaces gráficas podem ser persistentes e boas práticas de projeto permitem que as informações visuais estejam disponíveis ao usuário no momento que ele delas necessite.
- Assimetria de Banda-Larga: entradas de voz são tipicamente mais rápidas do que entradas digitadas, entretanto saídas de voz podem ser mais lentas do que saídas gráficas de leitura, particularmente em circunstâncias que permitam uma varredura visual.

- **Temporalidade:** eventos de teclado e mouse são discretos, quase eventos instantâneos em que significado e intenção podem ser tratados por um evento simples ou podem ser tratados por uma seqüência de eventos discretos. Entradas e voz não são nem instantâneos nem discretos desde que a elocução pode gastar muitos segundos para serem ditas e consiste de dados contínuos que são transformados em uma seqüência de palavras através do reconhecedor de voz. Apesar do resultado final do reconhecimento de voz ser efetivamente um evento instantâneo, ele pode ser protelado por um tempo perceptível quando o usuário para de falar.
- **Concorrência:** comunicação apenas por voz tende a ser serial e com canal único. A maioria das pessoas encontra dificuldade para ouvir e falar simultaneamente ou ouvir mais de uma voz, de maneira eficiente, ao mesmo tempo. Por outro lado, saídas visuais permitem concorrência de múltiplos canais de dados que podem ser processados juntos pelo usuário ou a atenção do usuário pode simplesmente mudar. Além disto, muitos aspectos de saída visual são projetados para responder em tempo real às entradas do usuário (por exemplo, digitar uma letra e ela aparecer na tela).

2.4 Arquitetura de um Sistema Baseado em Interface de Voz

De acordo com Zukerman e Litman (2001), é possível construir sistemas baseados em comandos de voz interativos e em tempo-real, em que as entradas do usuário sejam capturadas por um reconhecedor automático de voz e as saídas do sistema sejam enviadas através de um sintetizador de voz ou através de mensagens previamente gravadas. Assim, para atender as necessidades do usuário, a máquina necessita “entender” o que o usuário diz, desempenhar um processo de computação/transação, e responder ao usuário de tal forma que dê prosseguimento à conversação e cumprimento dos objetivos do usuário.

O processo de compreensão da linguagem falada pode ser, por exemplo, a transformação de uma fala em texto, através do reconhecimento da fala. Este reconhecimento é o processo de extração automática da informação linguística do sinal da voz (LAMEL; MINKER; PAROUBEK, 2000; LAI, 2000; SAN-SEGUNDO et al, 2005, COHEN; GIANGOLA; BALOGH, 2004; MCTEAR, 2002).

O reconhecimento de voz se estabelece quando o usuário emite um comando: o sistema interpreta o que foi dito e a saída é a concretização da ação derivada desse comando de voz - por exemplo, a transformação da voz em texto ou uma requisição de um serviço para o sistema operacional ou aplicativo utilizado.

Sendo assim, para converter a fala em um texto ou em um comando computável, é necessário o processamento de vários passos. A fala gera vibrações – ondas analógicas – que são captadas e traduzidas por um conversor analógico-digital (ADC) em dados digitais. Estas ondas são medidas em freqüências; filtradas a fim de que sejam removidos os ruídos; e separadas em diferentes faixas de freqüência. Além disso, o ADC também padroniza o som, ajustando-o a um nível de volume constante.

A seguir, o sinal é fracionado em fonemas. O reconhecedor, então, contrapõe esses segmentos aos fonemas conhecidos do idioma desejado.

No próximo passo, o motor de reconhecimento examina os fonemas dentro do contexto de outros fonemas próximos a eles. Isso é realizado utilizando um modelo estatístico, tal como o Modelo de Markov (HUANG; ACERO; HON, 2001; SONNENBERG; BECK, 1993), comparando uma grande coleção de palavras, frases e sentenças conhecidas. Por fim, o motor de reconhecimento de voz determina o que o usuário provavelmente estava dizendo e o transforma em texto ou comandos para o computador.

Um sistema com interface do usuário baseada em voz típico é composto por módulos conforme mostrado na Figura 3 (LAMEL; MINKER; PAROUBEK, 2000; LAI, 2000; SAN-SEGUNDO et al, 2005, COHEN; GIANGOLA; BALOGH, 2004; MCTEAR, 2002).



Figura 3: Arquitetura de um sistema de compreensão de linguagem falada

Fonte: Adaptado de (Cohen; Giangola; Balogh, 2004)

Nas sub-seções a seguir, serão detalhados e explicados cada um dos módulos que compõe a Figura 3 acima, de maneira mais detalhada.

2.4.1 Módulo de Ponto de Finalização

O funcionamento do primeiro módulo que compõe a arquitetura em questão é mostrado na Figura 4, onde o *ponto de finalização* (*endpointing*) detecta o início e o fim da fala, ou seja, ele determina quando as ondas, representando as vibrações da fala do usuário, começam e depois aguarda um período de silêncio, suficiente para indicar que a fala terminou.



Figura 4: O ponto de finalização determina onde as ondas de voz começam e terminam

Fonte: Adaptado de (Cohen; Giangola; Balogh, 2004)

As ondas são, então, empacotadas e enviadas para o próximo módulo, no qual será realizada a *extração de características* (*feature extraction*).

2.4.2 Módulo de Extração de Características

O módulo de *extração de características* (Figura 5) transforma as ondas captadas na fala em uma sequência de *vetores de características* (*feature vectors*). Um *vetor de característica* é uma lista de valores numéricos representando características mensuráveis do discurso que são úteis para o reconhecimento. Os números tipicamente representam características do discurso relacionadas a uma quantidade de energia em várias frequências.



Figura 5: A extração de características
 Fonte: Adaptado de (Cohen; Giangola; Balogh, 2004)

Sistemas típicos dividem as ondas do *ponto de finalização* em uma sequência de *vetores de características*, com um vetor para cada pequeno período de tempo (ex. um *vetor de característica* para cada segmento de 10 milissegundos sucessivos do discurso).

2.4.3 Módulo de Reconhecimento de Voz

O reconhecimento de voz é o processo pelo qual o computador pode converter um sinal acústico em um texto, livrando o usuário da limitação da interface WIMP – *windows, icons, menus e pointers*.

O módulo *reconhecedor* (Figura 6) usa a sequência dos *vetores de características* – que são as representações numéricas de fonemas - para determinar as palavras que foram ditas pelo usuário.

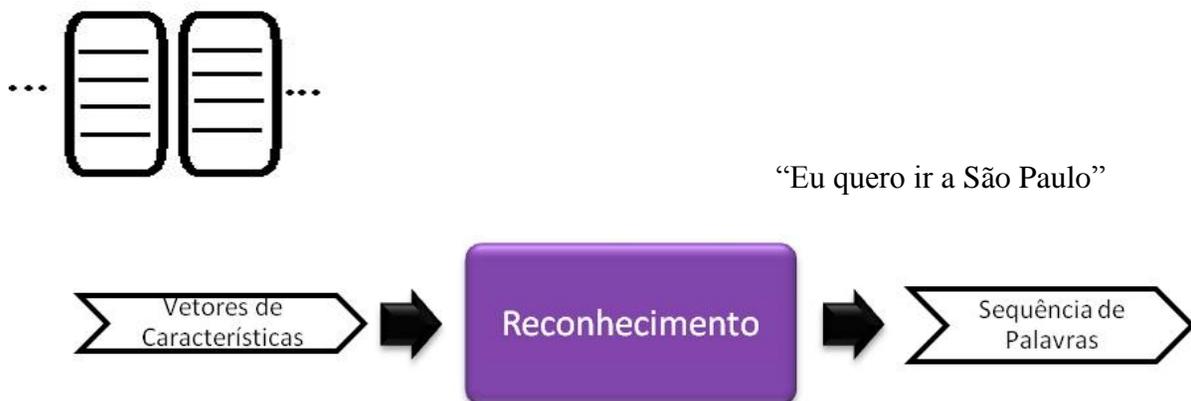


Figura 6: O módulo de reconhecimento de voz
 Fonte: Adaptado de (Cohen; Giangola; Balogh, 2004)

Para criar o modelo de reconhecimento, três componentes são necessários: um modelo acústico, um dicionário e uma gramática.

- **Modelo Acústico:** é uma representação interna do *reconhecedor* da pronúncia de cada possível fonema, ou som básico, na linguagem. Por exemplo, em português, um modelo acústico pode representar o som geralmente associado com a letra X. Modelos acústicos da maioria dos sistemas atuais são criados por um processo de treinamento, ou seja, muitos exemplos de sentenças e frases faladas, rotuladas com a sequência de palavras efetivamente pronunciadas, são alimentados no sistema. Baseado em um conjunto de exemplos, um modelo estatístico para cada fonema é criado, representando a variedade de maneiras que o fonema pode ser pronunciado. As *características* que são modeladas são as mesmas como nos *vetores de características* criados pelo módulo de *extração de características*.
- **Dicionário:** Um dicionário consiste em uma lista de palavras e suas pronúncias. A pronúncia indica para o *reconhecedor* quais os modelos acústicos para a sequência do conjunto para criar um modelo de palavra.

- Gramática: é a definição de todas as palavras que o usuário pode dizer para o sistema e ser entendido. Ela inclui uma definição de todas as possíveis seqüências de palavras que o *reconhecedor* pode tratar. Gramáticas diferentes podem ser ativadas em momentos diferentes durante a conversação.

2.4.4 Módulo de Compreensão da Linguagem Natural

Após o reconhecimento, ocorre a compreensão da linguagem natural. A função deste módulo é atribuir um significado às palavras ditas. Há um número de maneiras para representar o significado, uma delas, bastante comum, é como um conjunto de *slots* com valores. Um *slot* é definido para cada item da informação que é relevante para a aplicação. Por exemplo, uma informação relevante para uma aplicação de uma companhia aérea, pode incluir a cidade de origem do usuário do sistema, cidade de destino, data da viagem, e o horário de partida predileto. O sistema analisa a *string* da palavra passada pelo módulo de reconhecimento de voz e atribui valores aos *slots* apropriados. Por exemplo, na Figura 7, o usuário diz, “Eu quero ir a São Paulo”, informando que a cidade destino é “São Paulo”. O sistema define o valor do *slot* <destino> com “São Paulo”.



Figura 7: Módulo de compreensão da Linguagem Natural
Fonte: Adaptado de (Cohen; Giangola; Balogh, 2004)

2.4.5 Módulo Gerenciador de Diálogo

O módulo *gerenciador de diálogo* (Figura 8) é o principal componente dentro do sistema com interface do usuário baseada em voz. Ele é o elo entre o sistema e o usuário, utilizando voz, sendo responsável por gerenciar os demais módulos do sistema. Determina o que o sistema deve fazer dentre as várias possibilidades. Por exemplo, este módulo é responsável por gerenciar o módulo de reconhecimento, quando o usuário entrar, por voz, com uma informação, gerenciar o acesso à base de dados e gerenciar a sintetização de voz, a fim de fornecer a saída ao usuário.

Nos sistemas comerciais atuais, o gerenciamento de diálogo é o resultado de um programa escrito explicitamente para controlar o fluxo da aplicação (frequentemente com ferramentas especiais fornecidas pelos fornecedores de plataformas, ou em linguagens para fins especiais como VoiceXML).



Figura 8: Gerenciador de diálogo
 Fonte: Adaptado de (Cohen; Giangola; Balogh, 2004)

2.5 Sintetização de Voz

A tecnologia de sintetização de Voz (TTS – do inglês *Text to Speech*) é o processo que converte texto em voz. O sintetizador recebe um texto na forma digital e faz sua vocalização. Um programa de síntese de voz é útil para vocalizar informações resultantes de consultas à base de dados e em situações em que o usuário não pode desviar a atenção para ler algo ou não tem acesso ao texto escrito; um sistema com interface do usuário baseada em voz pode usar um módulo para sintetização de voz ou utilizar mensagens pré-gravadas quando não houver variação da informação a ser prestada ao usuário.

Embora a tecnologia TTS ainda não reproduza com fidelidade a qualidade da voz humana gravada - vale a pena destacar que, até o momento, os sintetizadores de voz não conseguem representar entonação - ela tem melhorado muito nos últimos anos. Tipicamente, o uso da voz humana pré-gravada está atrelado aos *prompts* e ao envio de mensagens para os usuários. No entanto, determinadas aplicações, como leitores de e-mail e notícias, tem dados muito dinâmicos. Nesses casos, uma vez que os textos das mensagens não podem ser previstos, pode-se usar a tecnologia TTS para criar os discursos de saída (COHEN; GIANGOLA; BALOGH, 2004).

As medidas primárias da qualidade destas mensagens que usam voz sintetizada são as seguintes:

- Inteligibilidade: quão bem o ouvinte pode entender o que é dito.
- Naturalidade: quanto o discurso sintetizado soa como fala humana real.
- Precisão: a precisão do que é sintetizado (por exemplo, fazer a escolha correta entre “*doctor*” e “*drive*” quando o texto de entrada incluir a sequência “*Dr*”).
- Capacidade de Escutar: quão bem os usuários toleram ouvir uma fala extensa sem se cansar.

2.6 Vantagens e Dificuldade no Uso de VUI

Atualmente, é possível usar sistemas que interagem através da fala em aplicações que permitem verificar informações de vôos, passagens e companhias aéreas, realizar consultas em instituições financeiras (como bancos e cartões de crédito), alugar carros, verificar as condições do tráfego e de tempo, localizar endereços, acessar o horário de trens e ônibus, encontrar os locais e horários de sessões de cinemas, e fazer reservas em restaurantes. A esta lista, em constante crescimento, podem ser adicionados muitos outros serviços, como o acesso à secretária de voz do telefone celular (OLIVEIRA NETO; SALVADOR, 2007).

Para as empresas, os principais benefícios em utilizar esses sistemas são (COHEN; GIANGOLA; BALOGH, 2004):

- Economia: embora Cohen, Giangola e Balogh (2004) digam que o custo de sistemas baseados em interfaces de voz usualmente se paga em questão de meses, há que se entender que este fator pode estar fortemente relacionado ao padrão financeiro do país. Para países desenvolvidos que pagam altos salários para seus funcionários, isso pode ser realmente incontestável, já para países como o Brasil, em que os salários são relativamente baixos, este fator pode ser contestado. Por outro lado, a diminuição do tempo de duração das ligações é outro fator financeiro importante e pode ser reduzido significativamente através do uso de VUI. No Brasil, as leis estabelecidas para o atendimento ao cliente por telefone têm contribuído para a implantação de automatizações nos *Call Centers* e também podem impulsionar a implantação de VUIs.
- Disponibilidade: as empresas querem estar disponíveis para os seus clientes em todo lugar a qualquer hora (24x7). Em certos casos, o sistema é o complemento de um serviço já desenvolvido para a Web; portanto, a empresa pode alcançar os usuários que não possuam acesso à Web (ou que momentaneamente não podem acessar a Rede).
- Ampliar uma marca: através da pronúncia e tom de voz, várias sensações podem ser passadas ao usuário, como polidez, firmeza e emoção. Isso pode ser usado a favor da construção da imagem da empresa ou marca de um produto.
- Satisfação do cliente: vários estudos mostram um alto grau de satisfação dos clientes que usam sistemas baseados em voz; porém ainda há uma grande resistência de pessoas que não gostam de serem atendidas por máquinas.

Para os usuários, existem várias outras vantagens em se usar bons sistemas baseados em voz sobre outras formas de interação (COHEN; GIANGOLA; BALOGH, 2004):

- Intuição e eficiência: os sistemas de boa qualidade que interagem com voz utilizam as habilidades natas dos usuários. As tarefas podem ser feitas de maneira mais simples e mais eficiente do que, por exemplo, em aplicações *touchtone*. Numa aplicação bancária, por exemplo, o usuário pode falar algo como “Eu quero meu

extrato bancário dos últimos 15 dias”, que é muito mais simples do que navegar numa série de menus.

- Não precisa de treinamento: exatamente por utilizar as habilidade natas dos usuários, os sistemas se tornam fáceis e amigáveis e dispensam, na maioria das vezes, o treinamento inicial do usuário no sistema.
- Onipresente: o telefone é onipresente e a maioria das pessoas tem celular. Dessa forma, sistemas que fornecem acesso por voz se tornam disponíveis em todo lugar, mesmo que o usuário esteja longe de um computador.
- Experiência agradável: ao ser eficiente, o sistema baseado em comandos de voz bem projetado possibilita uma experiência agradável ao usuário, aumentando sua usabilidade.
- Rapidez: falar é muito mais rápido do que digitar.
- Liberdade para as mãos e olhos: algumas atividades, como dirigir, ocupam as mãos e os olhos do usuário. As aplicações baseadas em comandos de voz se tornam uma solução ideal para esse tipo de situação, até mesmo porque a interface dos telefones celulares ainda é bastante pobre.

Porém, os sistemas baseados em comandos de voz não são indicados para todos os tipos de aplicações. Por exemplo, uma aplicação que necessite mover uma janela pela tela, obviamente, não deve utilizar a fala como entrada para realizar a operação.

Segundo Deng e Huang (2004), para que os sistemas baseados em comandos de voz tenham uma aceitabilidade maciça pela sociedade, há alguns desafios a serem vencidos:

- Diminuir a diferença entre o que a tecnologia atualmente oferece em termos de interface e o que os usuários desejam de um sistema de reconhecimento de voz, essencialmente num ambiente multimodal³.

³ De acordo com Suhm, Myers e Waibel (2001), o termo “multimodal” deveria se referir ao uso de mais de uma modalidade, indiferente da natureza das modalidades. Entretanto, muitos pesquisadores têm usado o termo “multimodal” referindo-se especificamente às modalidades que são comumente utilizadas na comunicação entre pessoas, tais como fala, gestos e escrita a mão.

- Construir sistemas robustos em todos os possíveis ambientes acústicos: os sistemas de reconhecimento de voz trabalham bem em ambientes silenciosos, mas quando o usuário está inserido num ambiente com barulho – um restaurante, uma rua ou um shopping center, geralmente não pode usar este sistema eficazmente, elevando, imensamente, a taxa de erros de reconhecimento. Este foi por muitos anos, o principal problema estudado por pesquisadores de reconhecimento de voz, tanto das universidades quanto das empresas.
- Necessidade de se trabalhar com sistemas para linguagem natural, com estilo livre: até o momento, quando usuários interagem com sistemas baseados em comandos de voz, eles estão cientes de que seu “partner” é uma máquina. Esta máquina falha no reconhecimento de voz se o usuário tenta introduzir um estilo natural e casual na conversação.

2.7 Questões Técnicas

De acordo com Alapetite, Boje e Morten (2009), quando se desenvolve uma aplicação com interface do usuário baseada em voz, há algumas questões que não devem ser esquecidas ou menosprezadas para o sucesso da aplicação:

- O vocabulário afeta o reconhecimento de voz pelo seu tamanho e pela cobertura do seu domínio. Assim, vocabulários extensos com uma boa cobertura do domínio são atrativos, exatamente porque são capazes de reconhecer mais palavras. Porém, vocabulários menores proporcionam um aumento de precisão no reconhecimento. Além disto, sistemas que fazem a transcrição funcionam melhor para domínios restritos, tais como domínio radiológico.
- Os usuários influenciam o reconhecimento de voz por sua clareza e sua consistência na pronúncia das palavras. Sistemas dependentes do usuário possuem uma taxa de reconhecimento da voz mais alta do que os sistemas que são independentes do usuário, porém, necessitam de sessões de treinamento – baseadas no fato de que o sistema adapta o modelo acústico ao usuário – e pode ser mais sensível às variações de barulho, microfones e voz (por exemplo, se o

usuário estiver com um resfriado). Além disto, haveria a necessidade de treinamento para pessoas não-nativas na língua do sistema, assim como considerar as taxas de reconhecimento de voz de crianças e idosos.

- Interferência sonora afeta o reconhecimento da voz de duas formas: a) distorções no sinal da voz causam maior dificuldade para distinguir as palavras ditas; b) na presença de barulho, usuários tendem a alterar sua voz e, assim, causar distorção no sinal da fala.
- Todos os sistemas de reconhecimento de voz estão baseados em princípios de padrões estatísticos. Entretanto, apesar de suas semelhanças, sistemas diferem entre si na sua parametrização do sinal de voz, o modelo acústico de cada fonema e o modelo de linguagem utilizado na escolha de palavras mais de acordo com as palavras ditas anteriormente. Assim, muitos sistemas causam diferenças em relação aos erros de reconhecimento, mesmo quando eles têm taxas de reconhecimento similares.

2.8 Considerações Finais

Este capítulo versou sobre as definições de interface homem-máquina e interface do usuário baseada em voz. Esta última consiste numa tendência tanto nas interfaces unimodais, como também como um dos modos que compõe as interfaces multimodais e ubíquas. Estas definições são de extrema importância para darem base para que se entenda os capítulos 4 e 5.

A seguir, serão apresentadas as principais metodologias clássicas para avaliação de interfaces homem-computador.

3. AVALIAÇÃO DE USABILIDADE

A avaliação de usabilidade deve ser utilizada em todo o processo de desenvolvimento de uma aplicação computacional. Esta avaliação está ligada à satisfação e facilidade do usuário em utilizar tais sistemas.

Este capítulo apresenta uma revisão bibliográfica dos principais métodos de avaliação de usabilidade, criados, principalmente, para interfaces gráficas do usuário.

3.1 Definição

A avaliação de usabilidade consiste de uma das três avaliações possíveis para um sistema. É possível pensar em avaliar: i) tecnicamente os componentes do sistema; ii) a usabilidade do sistema; e iii) o sistema e seus componentes sobre o ponto-de-vista dos clientes. Obviamente estas três formas não estão completamente desassociadas; por exemplo, um sistema com uma boa integração de componentes considerados excelentes pode ter uma usabilidade baixa, assim como um cliente pode preferir um sistema com componentes não considerados tão eficientes por questões de custo e compatibilidade de plataformas (BERNSEN; DYBKJAER, 2001).

Usabilidade é um requisito de qualidade de software que compreende aspectos relacionados à eficiência e eficácia no uso de software; avaliar, então, é o processo de assegurar a usabilidade de uma interface e garantir que os requisitos de usuários sejam atendidos (NIELSEN, 1993; SOMMERVILLE, 2001).

A fase de avaliação de um sistema deve ocorrer durante todo o processo de desenvolvimento, utilizando seus resultados como melhorias a serem aplicadas gradativamente na interface. Assim, a avaliação de usabilidade é cada vez mais usada no desenvolvimento de software, especialmente para avaliar os projetos de interação do usuário. Há também um crescimento dos resultados vindos da avaliação de usabilidade de sistemas já existentes (ROCHA, BARANAUSKAS, 2003).

Várias definições sobre usabilidade são propostas, na literatura, incluindo a dada pela ISO (ISO 9241-11): usabilidade é a capacidade que um sistema interativo oferece a seu usuário, em um determinado contexto de operação, para a realização de tarefas, de maneira eficaz (que é a precisão e a completude com que usuários conseguem atingir seus objetivos), eficiente (que é a relação entre a precisão e a inteireza com que usuários atingem certos objetivos e os recursos gastos para atingi-los) e satisfatória (que é o conforto e as atitudes positivas dos usuários concernentes ao uso do sistema).

Já para Lauesen (2005), Avouris (2001), Sommerville (2001) e Nielsen (1989; 1993), a usabilidade consiste de fatores como:

- Funcionalidade: o sistema pode dar suporte às tarefas que o usuário tem na vida real.
- Fácil aprendizado: quão fácil é o aprendizado sobre o sistema para vários grupos de usuários.
- Eficiente: o objetivo é alcançado através do sistema por um usuário freqüente.
- Satisfação subjetiva: o quanto o usuário está satisfeito com aquele determinado sistema.
- Compreensão: quão fácil é entender o que o sistema executa. Este fator é particularmente importante para situações não usuais que podem ocorrer quando há falhas ou erros no sistema.
- Adequação a padrões específicos: o quanto o sistema consegue estar de acordo com um padrão de projeto de interfaces.

- Adaptação ao nível de experiência dos usuários: o quanto o sistema é adaptativo às necessidades de usuários com níveis diferentes de conhecimento sobre o sistema, que podem variar entre iniciante, intermediário e avançado.

Enquanto a ISO e outros autores (LAUESEN, 2005; AVOURIS, 2001; SOMMERVILLE, 2001; NIELSEN, 1993) são conceitualmente claros, é difícil usar estas definições na prática. Quando a avaliação é realizada através de estudos empíricos, os pesquisadores necessitam decidir sobre indicadores (métricas) para cada fator. Por exemplo, uma métrica de memorização pode determinar que um operador que estivesse familiarizado com o trabalho deveria estar habilitado a usar 80 por cento das funcionalidades do sistema depois de uma sessão de três horas de treinamento.

Porém, é complexo decidir como estes fatores contribuirão para uma conclusão sobre a usabilidade de um sistema. Uma alternativa é definir usabilidade de um sistema como a ausência de obstáculos que impedem os usuários de completar suas tarefas. Com base nesta abordagem, usabilidade pode ser medida em termos dos problemas de usabilidade que são identificados quando os usuários estão usando um sistema em seu trabalho (SKOV; STAGE, 2005).

A avaliação de usabilidade envolve um conjunto de atividades. Segundo Rubin e Chisnell (2008), esta avaliação pode ser realizada em termos de cinco atividades: (1) determinações básicas, (2) planejamento do processo, (3) criação de situações de teste, (4) condução dos testes e (5) interpretação dos dados.

A usabilidade é importante por diversos fatores, tais como: economia de tempo na manipulação do sistema, maior número de pessoas podem usar o sistema, pessoas podem usar diversos sistemas computacionais, sem ter que se especializar em um único sistema.

3.2 Métodos Tradicionais de Avaliação de Usabilidade de Interfaces

A avaliação sistemática de um projeto de interface do usuário pode ser um processo caro, envolvendo cientistas cognitivos e projetistas gráficos. Pode significar que se tenha que trabalhar com um número estatisticamente significativo de experimentos com usuários típicos em laboratórios construídos especialmente para este propósito. Uma avaliação de interface de usuários desta proporção pode ser economicamente não realística (nem viável) para sistemas desenvolvidos por pequenas organizações, com recursos limitados (SOMMERVILLE, 2001).

Assim, projetistas e profissionais de IHC procuram métodos mais rápidos e baratos de avaliação de interfaces em substituição aos testes de laboratório. Em virtude disso, as técnicas de avaliação denominadas métodos de inspeção de usabilidade prometem oferecer informação de usabilidade mais barata e rápida que os tradicionais testes de usabilidade. Os métodos mais populares incluem avaliação heurística e percurso cognitivo.

Os métodos de avaliação serão aqui classificados em duas dimensões, de acordo com Rocha e Baranauskas (2003):

- Inspeção de Usabilidade: não envolve os usuários e pode ser aplicada em qualquer fase de desenvolvimento de um sistema, estando este implementado ou não.
- Teste de Usabilidade: envolve, necessariamente, o usuário e inclui métodos experimentais ou empíricos, métodos observacionais e técnicas de questionamento. Neste caso, é necessário que o sistema possua alguma implementação real.

3.2.1 Métodos de Inspeção de Usabilidade

A inspeção de usabilidade consiste de um conjunto de métodos baseados em avaliadores humanos que inspecionam aspectos relacionados à usabilidade de uma interface de usuário (ROCHA, BARANAUSKAS, 2003; SOMMERVILLE, 2001).

Muito do trabalho de inspeção consiste em classificar e contar o número de problemas de usabilidade (aspectos da interface do usuário que podem ocasionar uma usabilidade reduzida ao usuário final do sistema) apresentado pela interface. A estes problemas de usabilidade são dados graus de severidade para que os problemas mais graves sejam tratados com maior prioridade que problemas menos importantes (ROCHA, BARANAUSKAS, 2003).

A seguir, serão abordados dois dos métodos de inspeção de usabilidade mais utilizados: avaliação heurística e percurso cognitivo.

a) Avaliação Heurística

A maioria dos métodos de inspeção tem um efeito significativo na interface final somente se for usada durante o ciclo de vida do projeto, o que, efetivamente, não ocorre na maioria dos casos, seja porque são caros, difíceis ou demorados (ROCHA, BARANAUSKAS, 2003). Pesquisando este problema, Nielsen (1993) propõe o que ele chama de engenharia econômica de usabilidade. A avaliação heurística é o principal método desta proposta.

Este tipo de avaliação envolve um pequeno conjunto de avaliadores examinando a interface e julgando suas características, em face de reconhecidos princípios de usabilidade, definidos por uma heurística.

A avaliação heurística é realizada, primeiramente, de maneira individual. Cada avaliador, durante a sessão de avaliação, percorre a interface diversas vezes, inspecionando os diferentes componentes de diálogo e ao verificar problemas, estes são relacionados às heurísticas violadas. Estas heurísticas (Quadro 2) são, na verdade, regras gerais que objetivam descrever propriedades comuns de interfaces usáveis (NIELSEN, 1994).

Ao final desta etapa inicial, as listas elaboradas pelos avaliadores são consolidadas em uma única lista. Tipicamente, uma sessão de avaliação dura cerca de 2 horas, mas, dependendo do tamanho ou complexidade da interface, é recomendável dividi-la em várias sessões abordando cenários específicos.

Adicionalmente, o avaliador também pode considerar heurísticas específicas da categoria do produto que está sendo analisado.

HEURÍSTICAS DE NIELSEN

1. Visibilidade do status do sistema

- manter *feedback* sobre o que está acontecendo no sistema

2. Compatibilidade do sistema com o mundo real

- vocabulário pertinente ao ambiente do usuário.
- não utilizar termos específicos à área computacional.

3. Controle do usuário e liberdade

- quando o usuário chega a funções do sistema por engano, é necessário ter saídas claras destes estados indesejados, sem ter que percorrer um extenso diálogo. Prover funções de *undo* e *redo*.

4. Consistência e padrões

- seguir convenções de plataforma computacional ajudam o usuário a se familiarizar com a interface.

5. Prevenção de erros

- sempre que houver uma mensagem de erro, verificar se este erro não poderia ser evitado.

6. Reconhecimento ao invés de lembrança

- não sobrecarregar o usuário, fazendo-o lembrar informações de uma parte da aplicação em outros pontos.

7. Flexibilidade e eficiência de uso

-prover mecanismos ao usuário experiente para que ele possa evitar extensos diálogos e mensagens de informações que ele não deseje.

8. Estética e design minimalista

- diálogos devem ser objetivos, sem mostrar ao usuário informações pouco relevantes.

9. Ajudar os usuários a reconhecer, diagnosticar e corrigir erros

- informar ao usuário qual é o erro de forma clara, sugerindo uma solução.

10. Suporte e documentação

- as informações sobre suporte (*help*) e documentação devem ser fáceis de serem encontradas, focalizadas na tarefa do usuário e não muito extensas.

Quadro 2: Lista de Heurísticas Revisadas de Usabilidade

Fonte: Baseada em (NIELSEN, 1993)

Estas heurísticas são bastante gerais, podendo ser adaptadas dependendo do tipo de sistema que se venha a trabalhar. Por exemplo, Salvador, Oliveira Neto e Guimarães (2009) compuseram outras heurísticas para avaliar interface do usuário baseadas em voz e Sutcliffe e Gault (2004) propuseram outras heurísticas para avaliar sistemas de Realidade Virtual.

b) Percurso Cognitivo

Percurso Cognitivo é um método de inspeção de usabilidade que tem como foco central a avaliação do projeto quanto à sua facilidade de aprendizagem, sobretudo por exploração (ROCHA, BARANAUSKAS, 2003).

Ao invés de investir tempo em treinamento formal ou leitura de manuais de utilização do sistema, os usuários preferem aprender a usar um sistema enquanto executam suas tarefas, adquirindo conhecimento sobre as características do sistema à medida que necessitem destas funcionalidades. Isto garante que o custo da aprendizagem de uma determinada funcionalidade seja em parte determinado pelo seu benefício imediato ao usuário.

O percurso cognitivo é um processo de revisão. Os revisores avaliam a interface proposta no contexto de uma ou mais tarefas do usuário. Este processo pode ser dividido em duas fases:

- Fase preparatória: definição das tarefas, sequências de ações para cada tarefa, universo de usuários e da interface a ser analisada.
- Fase de análise: Os revisores analisam se os usuários farão a ação correta para atingir o resultado esperado; se os usuários perceberão a ação correta disponível; se os usuários irão associar a ação correta com o efeito desejado; se a ação correta for executada os usuários perceberão que foi realizado um progresso em relação à tarefa desejada.

3.2.2 Testes de Usabilidade

O teste com usuários é um método fundamental de usabilidade. Embora seja considerado um método bastante caro e demorado, alguns autores afirmam que os testes de usabilidade têm acelerado muitos projetos e levado a uma significativa redução em seus custos (GOLD; LEWIS, 1985; RUBIN; CHISNELL, 2008).

O objetivo do teste de usabilidade pode ajudar no desenvolvimento do sistema ou pode visar a avaliação da qualidade global desta interface.

Num teste de usabilidade, dois problemas devem receber atenção especial:

- **Confiabilidade:** o grau de certeza de que o mesmo resultado será obtido se o teste for repetido.
- **Validade:** se os resultados de teste refletem os aspectos de usabilidade que se deseja testar.

Basicamente um teste de usabilidade é um processo composto por quatro passos:

- **Preparação:** nesta fase, busca-se garantir que tudo esteja pronto antes do usuário chegar.
- **Introdução:** os usuários são apresentados à situação de teste e colocados à vontade.
- **Teste:** somente um avaliador deve se comunicar com o usuário e é importante que sejam evitados comentários sobre o desempenho, ou observações sobre o usuário e ajudas ao ele.
- **Sessão final:** Depois do tempo definido para completar as tarefas, os participantes são convidados a fazer comentários ou sugestões gerais ou a responder a um questionário específico.

Há um número mais simples e menos dispendioso de técnicas de avaliação de interface do usuário que podem ser identificadas para projetos de interface de usuário:

- Questionários: que colocam informações sobre o que os usuários pensam sobre a interface.
- Observação de usuário no trabalho com o sistema e “pensando em voz alta” sobre como eles estão tentando usar o sistema, de acordo com algumas tarefas que eles tenham que desempenhar.
- Vídeo *Snapshots* do uso típico de sistema: o usuário tem suas ações gravadas em vídeo – estando ciente de que a gravação está acontecendo.
- A inclusão, no software, de programas que colem informações sobre as maiores facilidades e erros encontrados pelos usuários.

A aplicação de um questionário para inspecionar usuários da aplicação é um meio relativamente barato de avaliar uma interface. As questões devem ser tão precisas quanto possível a fim de conseguir respostas claras. Por exemplo, questões do tipo “por favor, comente sobre a usabilidade desta interface” não devem ser utilizadas.

Já a avaliação baseada em observação envolve simplesmente observar como os usuários estão utilizando o sistema, procurando pontos como as facilidades no uso e os erros cometidos. Isto pode ser complementado com sessões de “pensando alto”, nas quais os usuários falam sobre o que eles estão tentando realizar, como eles entendem o sistema e como eles estão tentando usar o sistema de acordo com os seus objetivos.

Equipamentos de vídeo de custo relativamente baixo permitem registrar a observação direta de sessões do usuário para posterior análise. A análise completa de vídeo é cara e requer um conjunto de avaliação de equipamentos específicos com muitas câmeras focadas no usuário e na tela. Entretanto, a gravação de vídeo de operações de um usuário selecionado pode ser útil para detectar problemas.

Código de instrumentação para coleta estatística permite que as interfaces sejam melhoradas várias vezes. As operações mais comuns podem ser detectadas. A interface pode ser reorganizada de tal forma que seja mais rápida, por exemplo, através do uso de menus do tipo *pop-up*.

Nenhuma destas abordagens relativamente simples para avaliar a interface com o usuário é completamente segura e são, provavelmente, insuficientes para detectar todos os problemas

de interface do usuário. Entretanto, estas técnicas podem ser usadas por um grupo de voluntários, antes do sistema ser liberado, sem um grande gasto de recursos. A maioria dos grandes problemas de projeto de interface do usuário pode, então, ser descoberta e corrigida.

3.2.3 Comparação entre os Métodos de Avaliação de Usabilidade

Um estudo abrangente (DESURVIRE, 1994), comparando testes de usabilidade, avaliação heurística e percurso cognitivo, apresenta os seguintes resultados:

- Os resultados dos métodos de inspeção são melhores quando os avaliadores são especialistas em avaliação. Mesmo assim, não substituem o teste de usabilidade. Em experimentos relatados por Desurvire (1994), os melhores avaliadores, usando o método de melhor desempenho, não detectaram, em média, 56% dos problemas encontrados nos testes de usabilidade.
- A avaliação heurística permite uma avaliação global da interface facilitando a identificação de melhorias na interface. Foi a mais eficaz na detecção de erros e principalmente na identificação da maioria de erros sérios. Além disso, é a de menor custo.
- O teste de usabilidade é o mais eficaz para detectar erros, mas o mais caro. O custo de um teste de usabilidade é da ordem de 50 vezes o custo dos métodos de inspeção. Todos os problemas sérios são encontrados, mas há perda de valor na detecção de consistências.

3.3 Considerações Finais

Este capítulo apresentou os métodos tradicionalmente utilizados para a avaliação de usabilidade de interfaces de sistemas computacionais. Todos os métodos apresentam falhas ou custo muito elevados e, se combinados, podem oferecer um resultado mais eficaz.

O próximo capítulo aplicará os conceitos de usabilidade para uma interface não-convencional – a interface do usuário baseada em voz – que apresenta certas características especiais não

levadas em consideração pelos métodos tradicionais de avaliação, visto que estes métodos foram criados, principalmente, para interfaces gráficas do usuário.

4. USABILIDADE DE INTERFACES DO USUÁRIO BASEADAS EM VOZ

Este capítulo expõe o escopo de usabilidade para interfaces do usuário baseadas em voz. Sendo assim, define e propõe requisitos gerais e específicos para esta classe de aplicações. Por fim, versa sobre aspectos de qualidade destes sistemas, além de listar métricas para este fim.

4.1 Contextualização

A usabilidade está se tornando uma questão cada vez mais importante no desenvolvimento e na avaliação de sistemas de reconhecimento de voz. Muitas companhias começam a gastar um volume alto de recursos para conhecerem exatamente as características destes tipos de sistemas que são atraentes aos usuários e, também, saberem como avaliar se seus sistemas possuem estas características. Apesar destas importantes questões sobre usabilidade, durante os últimos anos, o que mais se tem pesquisado são os componentes deste tipo de sistema, e não, sobretudo, os aspectos de usabilidade.

Estes aspectos têm sido frequentemente negligenciados no desenvolvimento e na avaliação de interfaces do usuário baseadas em voz. Por outro lado, tem crescido a cultura de que esta usabilidade é um fator tão importante quanto à qualidade técnica dos componentes destes sistemas e, que a qualidade desta interface constitui um importante parâmetro de competição (DYBKJAER; BERNSEN, 2001).

Além disto, o conhecimento atual sobre avaliação de VUI advém de pequenas contribuições em avaliações que foram desenvolvidas para projetos específicos, e que tentam generalizar e propor recomendações para tais aplicações, como o *framework* PARADISE (WALKER et al, 1997), o projeto EAGLES (Expert Advisory Group on Language Engineering Standards) (GIBBON; MOORE; WINSKI, 1997) e EU DISC (DYBKJAER; BERNSEN, 2000).

4.2 Avaliação de VUI – Ideia inicial

Segundo Dybkjaer e Bernsen (2001), para fazer uma avaliação de um sistema com reconhecimento de voz, torna-se necessário o uso de *templates* que contenham as seguintes questões:

- O que está sendo avaliado: por exemplo, *feedback* adequado.
- Qual a parte do sistema que está sendo avaliado: por exemplo, o gerenciamento do diálogo.
- Tipo de avaliação: por exemplo, qualitativo.
- Método de avaliação: por exemplo, observação do usuário.
- Sintomas a procurar: por exemplo, se o sistema de ajuda é consistente.
- Importância da avaliação: por exemplo, crucial.
- Dificuldade da avaliação: por exemplo, fácil.
- Custo da avaliação: por exemplo, alto.
- Ferramentas de suporte: ferramenta para auxiliar na contagem do tempo para que a tarefa seja completada.

A ideia é fornecer ferramental suficiente ao avaliador para que, seguindo este *template*, seja possível avaliar a interface do usuário baseada em voz de maneira eficaz e eficiente. Há que se apontar, sobretudo, que a importância dos critérios para a avaliação de um VUI depende da aplicação e de usuários ou grupo de usuários deste sistema.

4.3 Requisitos não funcionais para Interface do Usuário Baseada em Voz

Há diferenças substanciais na análise dos requisitos de sistemas de reconhecimento de voz dos já conhecidos sistemas interativos tradicionais (que utilizam o paradigma WIMP). Talvez a diferença mais importante esteja relacionada ao fato das entradas de dados em sistemas de comandos de voz serem transientes, enquanto as entradas para GUIs são persistentes

(DYBKJAER; BERNSEN, 2001). Os requisitos não-funcionais desejáveis para uma boa interface podem ser classificados em duas categorias: requisitos relacionados à representação da informação e os relacionados à entrada de dados (PRESSMAN, 2004). Esta classificação é mais aplicada às GUIs; entretanto, uma análise de como estes requisitos pode ser aplicada à interface do usuário baseada em voz pode e deve ser realizada.

4.3.1 Requisitos não-funcionais relacionados à representação da informação

Aqui são apresentados os requisitos relacionados, basicamente, à forma com que a interação tem que assumir para ser utilizada pelo usuário, tais como (DYBKJAER; BERNSEN, 2001; SALVADOR, OLIVEIRA NETO; KAWAMOTO, 2008):

- **Consistência:** é uma das principais características de usabilidade de interface (NIELSEN, 2000a). Controla o comportamento não esperado do sistema, reduzindo a frustração do usuário. Para interfaces do usuário baseada em voz que usam adaptação do seu diálogo e conteúdo, estes requisitos não seriam respeitados, uma vez que o sistema pode ser modificado baseando-se no perfil dinâmico de seus usuários.
- **Modalidade apropriada:** a maioria das tarefas orientadas a sistemas de reconhecimento de voz usam apenas a voz para entrada e saída de informações. Porém, este tipo de sistema não é, claramente, indicado para todos os tipos de aplicações, principalmente, quando, o usuário precisa entrar com códigos de segurança, por exemplo, em sistemas bancários. Assim, há casos em que é conveniente integrar voz com outras formas de interação.
- **Adequação do *feedback*:** é importante em qualquer comunicação. Assim, a interação com o computador requer um *feedback* planejado (FOLEY, VAN DAN, 1990). Um *feedback* adequado significa que o usuário possa se sentir no controle durante a interação. O usuário deve se sentir confiante de que o sistema compreendeu e que está atendendo aos seus objetivos. Há *feedbacks* em três níveis; nível de *hardware* – indica se as entradas do usuário tiveram sucesso (para entrada de voz, indica que o sistema entendeu o que o usuário está dizendo); nível

de sequência – indica que uma unidade de linguagem de comando foi aceita (em interface do usuário baseada em voz significa que o sistema entendeu a ação solicitada pelo usuário); nível funcional: indica que o sistema está trabalhando no problema (mensagens como “por favor, aguarde um momento”) ditas ao usuário.

- **Diversidade e Percepção Humana:** a interface deve dar suporte a todas as classes de usuários, identificando cada usuário e se adaptando a ele, executando adaptações de conteúdo e de apresentação através do Modelo do Usuário. Algumas estratégias podem ser usadas para isto: por exemplo, fornecendo aos usuários mais experientes informações mais detalhadas e *barge-in*⁴; e proporcionando aos usuários novatos, informações mais concisas e superficiais e sentenças no final de diálogos (KOMATANI et al, 2003).
- **Minimização do Esforço para Lembrar:** a interface deve minimizar o esforço cognitivo do usuário para executar suas tarefas. Diálogos de iniciativa mista e sentenças adicionais no final de diálogo podem ser fornecidas para guiar o usuário a utilizar o sistema de maneira adequada.
- **Frases de saída adequadas:** o conteúdo das saídas do sistema deve ser correto, relevante e suficientemente informativo, sem prover sobrecarga de informação ao usuário. A forma de expressão do sistema deve ser clara e não-ambígua e a linguagem deve fornecer uma terminologia apropriada e familiar ao usuário.
- **Qualidade da saída de voz:** sob o ponto-de-vista de usuários, a qualidade da saída de voz de um sistema está ligada a questões de clareza e inteligibilidade (entonação correta, emoção, ritmo da fala apropriado e prazer de se ouvir). Há três classes de saída de voz num sistema: a) as que gravam as frases inteiras do sistema (quando as informações não são dinâmicas); b) as que concatenam palavras e frases gravadas, ou; c) as que utilizam *text-to-speech*.

⁴ *Barge-in* é conhecido como a interrupção no sistema de telefonia (sistemas automáticos) causada por usuários experientes que estão acostumados a navegar pelo sistema de maneira mais rápida para os próximos *prompts* do sistema. É uma característica importante porque permite que usuários experientes consigam se mover rapidamente através do sistema a fim de obter as informações desejáveis (COHEN; GIANGOLA; BALOGH, 2004).

4.3.2 Requisitos não-funcionais relacionados à entrada de dados

Dybkaer e Bersen (2001) e Salvador, Oliveira Neto e Kawamoto (2008) definiram um conjunto de critérios de avaliação de usabilidade para sistemas de comandos de voz que estão relacionados ao acesso do usuário ao sistema:

- Reconhecimento adequado da entrada: sob o ponto-de-vista de usuários, um reconhecimento adequado de voz significa que o sistema raramente compreende erroneamente uma entrada do usuário. Porém, isto está ligado a muitos fatores do ambiente (se o ambiente é ruidoso ou não) e também a fatores do usuário: sexo, idade, sotaque, voz grave ou aguda, e qualidade da voz recebida pelo sistema.
- Naturalidade da fala do usuário: é preciso, através da linguagem de saída do sistema, conseguir gerenciar as entradas do usuário enquanto este sente a naturalidade da fala. Se as limitações impostas pela tarefa são satisfeitas e a linguagem de saída do sistema controla, de maneira adequada, a linguagem de entrada do usuário, os usuários podem sentir que o diálogo seja natural.
- Iniciativa de diálogo adequado: para dar suporte a uma interação natural, é necessário que o sistema escolha, de forma razoável, a iniciativa do diálogo estabelecido entre ele e o usuário. Isto está ligado ao nível de conhecimento do usuário sobre o sistema. Diálogos direcionados ao sistema podem funcionar de maneira apropriada para tarefas em que o sistema simplesmente requer uma série de pedaços específicos de informação do usuário, especialmente se o usuário é novo no sistema. A fim de satisfazer usuários experientes, o sistema terá que ser hábil ao lidar com grandes pacotes de informação que são naturais para estes usuários.
- Naturalidade da estrutura do diálogo: é importante que a estrutura fixada pelo desenvolvedor no diálogo seja natural para o usuário, refletindo as expectativas esperadas por ele, especialmente no diálogo direcionado pelo sistema em que o usuário não deva interferir na estrutura do diálogo. Estruturas de diálogo não naturais frequentemente levam os usuários a tentar tomar a iniciativa de maneira que o sistema não está preparado para atender.

- Suficiência de orientação de interação: orientação suficiente de interação é essencial para que o usuário se sinta no controle durante a interação. A fala é inapropriada para fornecer instruções complexas do manual do usuário para usuários com pouca experiência. Há usuários experientes enquanto que outros são novatos. Sendo assim, é necessário considerar questões vinculadas a *turn-taking*⁵ x *barge-in*; facilidades de sistema de ajuda, e saídas para comportamentos não óbvios do sistema.
- Suficiência da cobertura de tarefas e domínio: coberturas suficientes de tarefas e domínios são também cruciais à interação natural. Mesmo que o usuário não esteja familiarizado com o sistema de reconhecimento de voz, normalmente, ele prefere ter expectativas detalhadas sobre as informações ou os serviços que o sistema possa fornecer.
- Suficiência da capacidade de raciocínio do sistema: os usuários, mesmo quando cientes de que estão falando com um interlocutor primitivo (máquina), tendem a assumir que o sistema seja capaz de executar pedaços de raciocínio que humanos fazem sem sequer pensar e que são partes inseparáveis do diálogo natural sobre a tarefa.
- Mecanismos de ajuda: a interface deve fornecer ajuda quando requisitada ou quando perceber que o usuário se encontra em dificuldades. Para interface do usuário baseada em voz, um diálogo deve prover a lista de possíveis opções que o usuário pode realizar na aplicação quando ele perceber que o usuário não tomará a iniciativa do diálogo. Estratégias de confirmação de diálogo também podem ser utilizadas.
- Prevenção de erros: é desejável que uma boa interface seja capaz de prevenir possíveis erros realizados pelos usuários. O que pode ser feito, no caso de VUI, é tentar guiar o usuário a alcançar seus objetivos rapidamente, por exemplo, passando o controle do diálogo para o sistema, se for percebido que o usuário tem alguma dificuldade, ou fornecer sentenças adicionais ao final de cada diálogo alertando o usuário para os próximos passos que ele pode tomar no sistema.

⁵ *Turn-Taking* é um dos mecanismos básicos para todos os tipos de diálogo e conversas envolvendo mais do que duas pessoas. É um mecanismo crucial na interação homem-computador permitindo que cada interlocutor tenha a sua vez de interagir (BOSCH; OOSTDIJK; RUITER, 2004).

- Tratamento de erros: uma boa interface deve ser capaz de fornecer correção das entradas rapidamente, aumentando a produtividade de usuários e estimulando que eles explorem o sistema. As VUIs podem atender a este requisito adotando diálogos de iniciativa mista, técnicas de confirmação de diálogo e, no caso de sistemas de telefonia, transferindo a chamada para um atendente humano quando o usuário se encontra em sérias dificuldades. É possível dividir este tratamento de erros em quatro modalidades.
 - Reparo de iniciativa do sistema: é necessário quando o sistema não pode entender ou não pode ter certeza se entendeu corretamente a entrada do usuário. Então, o sistema pode pedir para o usuário repetir a entrada, falar mais alto, modificar a forma em que a entrada está sendo feita, ou repetir o que o sistema entendeu da entrada do usuário e pedir a ele que confirme ou a corrija. Caso isto não resolva o problema, o sistema pode transferir a chamada para um atendente humano ou modificar a interação para uma forma mais simples.
 - Reparo de iniciativa do usuário: alguns sistemas necessitam que o usuário utilize palavras-chave especificamente definidas. O problema é que isto não é natural e se torna difícil para o usuário recordar. Outra possibilidade é utilizar o princípio “eraser” em que o usuário simplesmente repete suas entradas até que o sistema receba sua mensagem.
 - Esclarecimento iniciado pelo sistema: é necessário quando a entrada do usuário é inconsistente ou ambígua, então o sistema solicita um esclarecimento ao usuário.
 - Esclarecimento iniciado pelo usuário: é necessário se o sistema produzir saídas inconsistentes ou ambíguas, ou se o usuário não estiver familiarizado com os termos usados na comunicação.
- Número de problemas de interação: a falta de cooperação na saída do sistema pode ser diagnosticada a partir da ocorrência de problemas de comunicação nas interações reais ou simuladas entre o usuário e o sistema. O problema de se capturar e analisar os dados é que isto é uma atividade com custo alto, especialmente porque uma grande quantidade de dados é necessária para resolver

a maioria dos problemas de comunicação causados no sistema. Para resolver estes problemas de interação de maneira mais eficaz, seria necessário aplicar alguma metodologia de avaliação já na fase de projeto do sistema.

- Satisfação do usuário: a satisfação do usuário é uma medida subjetiva da usabilidade, construída com preferências pessoais e fatores contextuais. É possível aplicar questionários e promover entrevistas com o usuário para obter esta medida.

4.4 Características da Qualidade em Interface do Usuário Baseada em Voz

As avaliações de usabilidade são, normalmente, divididas em objetivas e subjetivas. As primeiras estão relacionadas à efetividade do usuário em relação ao sistema, enquanto que as subjetivas coletam as opiniões do usuário quanto ao sistema, usualmente através de questionários. As avaliações objetivas ainda podem ser divididas em quantitativas e qualitativas.

Os fatores que compõem as avaliações subjetivas e objetivas de usabilidade foram encontrados em vários trabalhos na literatura (WALKER; PASSONNEAU; BOLAND, 2001; MÖLLER, 2005; LARSEN, 2003; DYBKJAER; BERNSEN; MINKER, 2004a; DYBKJAER; BERNSEN; MINKER, 2004b; MÖLLER, 2002; DYBKJAER; BERNSEN, 2001; HARTIKAINEN; SALONEN; TURUNEN, 2004). Möller (2002) faz um estudo bastante aprofundado sobre a questão da qualidade de serviços de telefonia baseados em VUI, estabelecendo uma nova taxonomia a ser aplicada a esta classe de sistemas. Embora trate apenas de sistemas que utilizam um aparelho telefônico para a realização de interações entre o usuário e o sistema, esta taxonomia pode ser ajustada, facilmente, aos sistemas que utilizam VUI de uma maneira geral. Na Figura 9 há uma boa visão da interrelação dos vários fatores que compõem a qualidade destes tipos de sistemas.

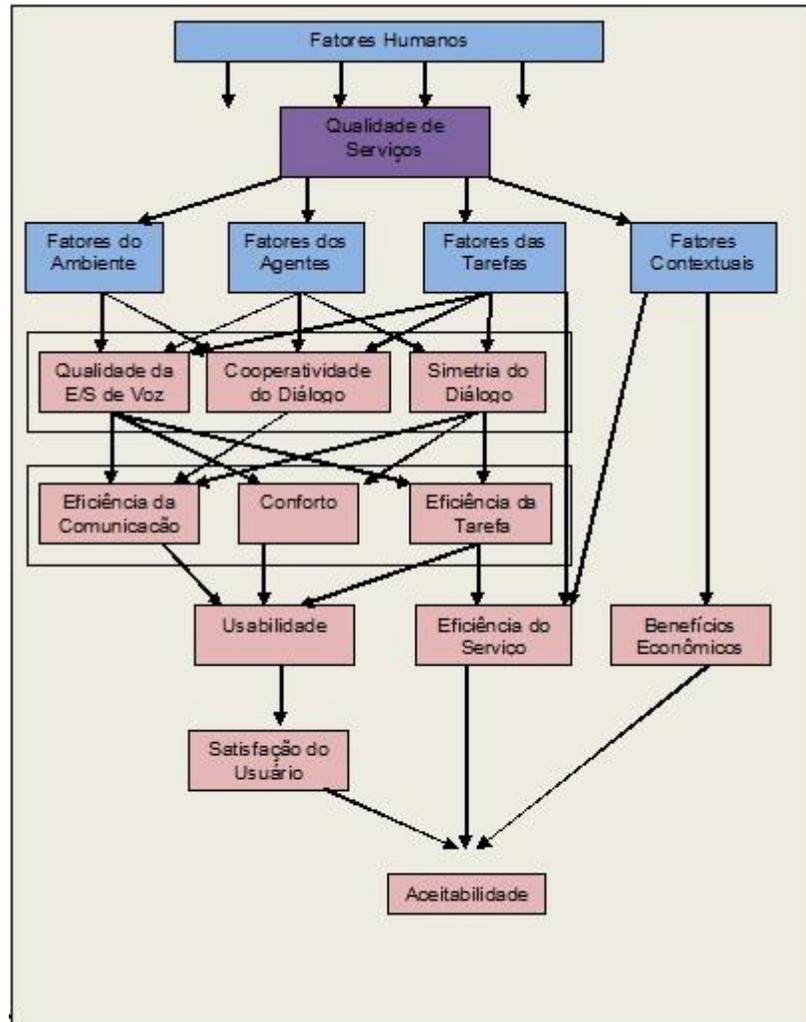


Figura 9: Taxonomia de Aspectos de Qualidade

Fonte: Adaptada de (Möller, 2002)

Na figura supracitada, há vários fatores que contribuem para a qualidade de VUI, entre eles é possível destacar:

- Fatores Humanos correspondem às atitudes, objetivos, motivações, flexibilidade, experiência, conhecimento do domínio e da tarefa pelos usuários.
- Fatores do Ambiente são o canal de transmissão utilizado, interferência sonora barulho detectada e acústica do ambiente utilizado.
- Fatores dos Agentes indicam a flexibilidade e estratégia do diálogo entre o sistema e o usuário.

- Fatores das Tarefas: cobertura da tarefa e do domínio, flexibilidade e dificuldade na execução da tarefa.
- Fatores Contextuais: custo, disponibilidade e acesso ao sistema; Também: economia de tempo.

De acordo com a Figura 9, é possível discutir que a aceitabilidade de uma aplicação VUI está ligada a três pilares fundamentais: a usabilidade – que leva à satisfação do cliente, a eficiência do serviço e aos benefícios econômicos com a aquisição da aplicação. Obviamente, a eficiência da tarefa, que leva à eficiência do serviço, é requisito fundamental e está ligada à eficiência do serviço.

Sendo assim, para que um sistema seja aceito, ele precisa atender a requisitos como precisão do reconhecimento da voz, tempo de resposta adequado e minimização da sobrecarga de memória, mas também qualidade do sistema de áudio e de entrada da voz e atender às condições mínimas estabelecidas de interferência sonora do ambiente (vide Anexo A deste trabalho); também é primordial estudar as questões financeiras que podem inviabilizar a implementação de tal sistema.

Com base nos estudos encontrados na literatura, foi possível coletar as principais características da qualidade para aplicações VUI que constam da Quadro 3: e Quadro 4:. Estas características referem-se ao diálogo estabelecido entre o sistema e o usuário e também à comunicação, tarefa e serviço.

As características de qualidade relacionadas ao diálogo se referem à informação que deve ser apresentada ao usuário pelo sistema, deixando-o confortável, ou seja, estão ligadas a informações claras, verdadeiras, consistentes e que permitem um diálogo simples e verdadeiro entre o usuário e o sistema.

Já as características de qualidade relacionadas à comunicação, tarefa e serviço se referem a fatores como cordialidade na comunicação do sistema com o usuário, facilidade do usuário executar suas tarefas e se sentir satisfeito com o desempenho do sistema e a um serviço de qualidade, estando, por exemplo, sempre disponível.

Percebe-se que estas características de qualidade são desejáveis no fornecimento de serviços de um modo geral, mesmo aqueles que envolvem atendimento humano, tais como os *Call Centers* que ainda utilizam atendimento com pessoas.

QUALIDADE	ASPECTO	CARACTERÍSTICA DA QUALIDADE
Cooper.do Diálogo	Capacidade de ser informativa	– Informação exata, específica, completa, clara, concisa e adequada ao <i>feedback</i> do sistema
	Verdade e Evidência	– Informação com credibilidade, consistente, confiável e raciocínio do sistema
	Relevância	– Adequação do <i>feedback</i> do sistema,, compreensão e raciocínio do sistema, naturalidade da Interação
	Forma	– Expressão clara, não-ambígua, consistente e concisa. Transparência da Interação
	Conhecimento adquirido	– Congruência com o Conhecimento da Tarefa/ Domínio e com a experiência do usuário – Adequação da Adaptação do Usuário e da Inferência – Orientação de Interação
	Gerenciamento da Meta-Comunicação	– Adequação no gerenciamento do reparo e do esclarecimento – Capacidade de ajuda e re repetição
Simetria do Diálogo	Iniciativa	– Flexibilidade, orientação e naturalidade da Interação
	Controle da Interação	– Capacidade de controle, de <i>barge-in</i> , e de cancelamento
	Assimetria do <i>Partner</i>	– Transparência da Interação, da cobertura da Tarefa/Domínio; – Orientação da Interação – Naturalidade da Interação – Carga cognitiva requerida do usuário
Qualid.a E/S da Voz	Qualidade da Saída de Voz	– Inteligibilidade – Naturalidade da Voz – Esforço necessário para o usuário escutar
	Qualidade da Entrada de Voz	– Compreensão percebida do Sistema – Raciocínio do sistema

Quadro 3: Características de Qualidade Relacionadas ao Diálogo

Fonte: Adaptada de (Möller, 2002)

QUALIDADE	ASPECTO	CARACTERÍSTICA DA QUALIDADE
Eficiência da Comunicação	Velocidade	<ul style="list-style-type: none"> – Ritmo da interação – Tempo de resposta
	Concisão	<ul style="list-style-type: none"> – Comprimento e duração da interação
	Suavidade	<ul style="list-style-type: none"> – Adequação do <i>feedback</i>, compreensão e raciocínio do sistema – Adequação no Gerenciamento do Reparo e de esclarecimento – Naturalidade, orientação e transparência da interação – Congruência com a experiência do usuário
Conforto	Personalidade do Agente	<ul style="list-style-type: none"> – Polidez – Amigabilidade – Naturalidade do Comportamento
	Carga Cognitiva	<ul style="list-style-type: none"> – Facilidade de Comunicação – Concentração requerida do usuário – Stress / Ansiedade
Eficiência da Tarefa	Sucesso da Tarefa	<ul style="list-style-type: none"> – Adequação da Cobertura da Tarefa/Domínio – Validade, precisão e confiabilidade dos Resultados da Tarefa
	Facilidade da Tarefa	<ul style="list-style-type: none"> – Capacidade de ajudar – Orientação da tarefa – Transparência da cobertura da Tarefa/Domínio
Eficiência do Serviço	Adequação do Serviço	<ul style="list-style-type: none"> Adequação do acesso, da modalidade e da tarefa Disponibilidade Funcionalidade e utilidade do serviço
	Valor Agregado	<ul style="list-style-type: none"> – Melhoria do Serviço – Interface comparável
Usabilidade	Facilidade de Uso	<ul style="list-style-type: none"> – Operabilidade, inteligibilidade e capacidade de aprendizagem do serviço

Quadro 4: Características de Qualidade relacionadas à Comunicação, Tarefa e Serviço

Fonte: Adaptada de (Möller, 2002)

4.5 Métricas para a Avaliação de Usabilidade para Interface do Usuário Baseada em Voz

Assim como a classificação dada à avaliação de usabilidade, as métricas de usabilidade também são, normalmente, divididas em métricas objetivas e métricas subjetivas. As métricas objetivas ainda podem ser divididas em quantitativas e qualitativas. Tanto as métricas objetivas quanto as subjetivas podem ser utilizadas para avaliar aplicações VUI, podendo ser utilizadas em conjunto, dentro de um *template* para avaliação, tal como o proposto por Dybkjaer e Bernsen (2004b).

A medida de desempenho mais comum em aplicações VUI é a taxa de erro de reconhecimento, do inglês, word error rate (WER), que é a métrica relacionada à precisão (acurácia).

Na literatura, foram encontrados vários trabalhos que citam possíveis métricas e medidas para avaliar a usabilidade de VUIs, agrupados entre avaliações subjetivas e objetivas. Dentre os vários trabalhos sobre avaliação subjetiva para Interfaces do Usuário Baseadas em Voz estão (WALKER; PASSONNEAU; BOLAND, 2001; MÖLLER, 2002; MÖLLER, 2005; LARSEN, 2003; DYBKJAER; BERNSEN, 2001; HARTIKAINEN; SALONEN; TURUNEN, 2004). Alguns autores citam a avaliação subjetiva como a mais importante, por colherem a opinião da satisfação do usuário. Os trabalhos encontrados foram baseados na experiência adquirida pelos autores na construção e avaliação de sistemas de reconhecimento de voz; outros na modificação de métodos utilizados em outras áreas do conhecimento para medir qualidade, tal como o trabalho de Hartikainen, Salonen e Turunen (2004) que consiste em fazer uma modificação no método SERVQUAL (PARASURAMAN; ZEITHAML; BERRY, 1988). Este método produz uma medida subjetiva entre a diferença entre expectativas e percepções dos usuários em cinco dimensões de qualidade de serviço para todos os serviços. Outro trabalho encontrado estabelece um método para avaliar a qualidade dos serviços de telefonia baseados em voz através do método utilizado para avaliar a qualidade dos serviços de telefonia baseados no atendimento humano (MÖLLER, 2002; 2005).

Entre as medidas objetivas (quantitativas e qualitativas), assim como as subjetivas mais utilizadas estão (Quadro 5: e Quadro 6:) (WALKER; PASSONNEAU; BOLAND, 2001; MÖLLER, 2005; MÖLLER, 2002; LARSEN, 2003; DYBKJAER; BERNSEN; MINKER, 2004a; DYBKJAER; BERNSEN; MINKER, 2004b; PARASURAMAN; ZEITHAML; BERRY, 1988; HARTIKAINEN; SALONEN; TURUNEN, 2004):

QUALIDADE	ASPECTO	CARACTERÍSTICA DA QUALIDADE
Cooper.do Diálogo	Capacidade de ser informativo	<ul style="list-style-type: none"> – # perguntas do usuário; – # ajuda requisitada pelo usuário;
	Verdade e Evidência	<ul style="list-style-type: none"> – # perguntas corretamente/ incorretamente/ parcialmente/ falhas sendo feita;
	Relevância	<ul style="list-style-type: none"> – # barge-in realizados pelos usuário;
	Forma	<ul style="list-style-type: none"> – # <i>turns</i> do sistema; – número de palavras por <i>turns</i> do sistema;
	Conhecimento adquirido	<ul style="list-style-type: none"> – # solicitações de ajuda; – # cancelamento realizado pelo usuário; – # barge-in realizado pelo usuário; – # <i>prompts</i> por <i>time-out</i>;
	Gerenciamento da Meta-Comunicação	<ul style="list-style-type: none"> – # mensagens de erro do sistema ; – # solicitações de ajuda; – # cancelamento realizado pelo usuário; – taxa de correção; – recuperação implícita;
Simetria do Diálogo	Iniciativa	<ul style="list-style-type: none"> – # perguntas do usuário; – # perguntas do sistema; – taxa de correção;
	Controle da Interação	<ul style="list-style-type: none"> – # <i>barge-in</i> realizados pelo usuário; – # solicitações de ajuda; – # cancelamento realizado pelo usuário; – taxa de correção; – # <i>prompts</i> por <i>time-out</i>;
	Assimetria do Partner	<ul style="list-style-type: none"> – # <i>barge-in</i> realizados pelo usuário; – # <i>prompts</i> por <i>time-out</i>;
Qualid. E/S da Voz	Qualidade da Saída de Voz	<ul style="list-style-type: none"> – desempenho de TTS (Inteligibilidade, Naturalidade, Precisão, Capacidade de escutar);
	Qualidade da Entrada de Voz	<ul style="list-style-type: none"> – precisão da palavra, taxa de erro reconhecimento; – precisão da sentença, taxa de erro da sentença; – # erros por sentença; – # erro da palavra por sentença; – taxa de erro de entendimento; – # rejeições por parte do sistema; – conteúdo da informação; – # mensagens de erro do sistema;

Quadro 5: Classificação de Medidas (1)

Fonte: Adaptada de (Möller, 2002)

QUALIDADE	ASPECTO	CARACTERÍSTICA DA QUALIDADE
Eficiência da Comunicação	Velocidade	<ul style="list-style-type: none"> – duração do <i>turn</i>; – <i>delay</i> da resposta do sistema; – <i>delay</i> da resposta do usuário; – # <i>prompts</i> por <i>time-out</i>; – # <i>barge-in</i> realizado pelo usuário; Tempo e <i>turns</i> usado para confirmações Variação do Tempo de Resposta;
	Concisão	<ul style="list-style-type: none"> – duração do diálogo – # <i>turns</i> (# <i>turns</i> do sistema, # <i>turns</i> do usuário);
	Suavidade	<ul style="list-style-type: none"> – # mensagens de erro do sistema; – # cancelamento realizado pelo usuário; – # ajuda solicitada; – # rejeições por parte do sistema; – # <i>barge-in</i> realizados pelo usuário; – # <i>prompts</i> por <i>time-out</i>; – polidez para o diálogo; – polidez para transferir a chamada para um atendente humano, quando o sistema não é capaz de entender a entrada; – comportamento esperado; – diferenças culturais tratadas pelo sistema.
Conforto	Personalidade do Agente	<ul style="list-style-type: none"> – adequação às preferências do usuário;
	Carga Cognitiva	<ul style="list-style-type: none"> – # <i>prompts</i> por <i>time-out</i>; – <i>delay</i> de resposta do usuário; – cobertura do vocabulário e gramática do usuário;
Eficiência da Tarefa	Sucesso da Tarefa	<ul style="list-style-type: none"> – sucesso da tarefa; – solução da tarefa; – precisão da solução; – qualidade da solução (# tarefas e sub-tarefas completadas);
	Facilidade da Tarefa	<ul style="list-style-type: none"> – # solicitações de ajuda;
Eficiência do Serviço	Adequação do Serviço	cobertura da tarefa/domínio;
	Valor Agregado	- # tempo diminuído da tarefa antes executada;
Usabilidade	Facilidade de Uso	<ul style="list-style-type: none"> - habilidade para se adaptar a novos usuários; - habilidade para se adaptar ao mesmo usuário; - habilidade para gerenciar interface multi-modal; - número de problemas de interação.

Quadro 6: Classificação de medidas (2)

Fonte: Adaptada de (Möller, 2002)

4.6 Considerações Finais

Através deste capítulo é possível perceber a complexidade de se avaliar um sistema que utiliza reconhecimento de voz para a interação com o usuário. Muitas outras métricas e medidas teriam que ser utilizadas se o objetivo desta pesquisa fosse ainda avaliar tecnicamente e/ou sobre o ponto-de-vista dos clientes os componentes do sistema. Apenas estão sendo considerados aqui, como objetivo da pesquisa, os requisitos, métricas e medidas referentes à usabilidade do sistema. Devido à quantidade de requisitos e métricas encontradas, torna-se inviável a utilização de todo este conjunto, sendo necessária a escolha das mais relevantes de acordo com o sistema em questão.

O próximo capítulo abordará VUI para sistemas de informação radiológica, sobretudo para sistemas de transcrição automática de laudos em radiologia.

5. INTERFACE DO USUÁRIO BASEADA EM VOZ PARA SISTEMAS DE INFORMAÇÃO RADIOLÓGICA

O uso de VUI em sistemas de informação de propósito geral na área da Saúde não é apropriado devido às peculiaridades intrínsecas às informações médicas. Porém, para um propósito mais específico, estes sistemas têm sido cogitados como possíveis soluções para a entrada e consulta destas informações.

Este capítulo versa sobre o uso de VUI aplicado à área da Saúde, principalmente no dia que se refere ao seu uso em transcrições de laudos em radiologia, tratando, inclusive sobre seus requisitos.

5.1 Considerações Iniciais

Segundo Rodrigues (1996), um dos grandes desafios para a comunidade científica no campo da informática aplicada à Saúde é a implementação de soluções de grande impacto funcional, aliadas a um baixo custo operacional, que possam contribuir para a melhoria da qualidade do setor, facilitar o acesso e expedir serviços oferecidos aos pacientes e prover dados organizados rapidamente disponíveis para os profissionais de saúde, sejam eles médicos, pesquisadores, epidemiologistas, estatísticos, administradores etc.

Apesar da Tecnologia da Informação ter avançado rapidamente no que concerne à conectividade física para a comunicação de informação, os aspectos referentes à integração de sistemas e utilização da informação pelos profissionais da Saúde não avançaram na mesma proporção. No Brasil, isto se deve principalmente aos seguintes fatores: computadores disponíveis em pequena quantidade nos serviços de saúde, dificuldade de convergência a modelos comuns para aquisição, representação e troca das informações (padrões), e poucos sistemas clínicos eficazes nas instituições de saúde. Adicionalmente, são raros os profissionais qualificados e experientes na área (SHORTLIFFE, 1990; AMARAL; SALVADOR, 2005).

Uma área de grande interesse, atualmente, é a busca de modelos para a implementação de prontuários eletrônicos. Diversos centros de pesquisa e desenvolvimento no mundo, financiados por instituições acadêmicas ou empresas de desenvolvimento de software, estão buscando soluções tecnológicas e metodológicas para a construção de seus sistemas de registro eletrônico de saúde (AMARAL; SALVADOR, 2005).

Neste ponto, uma das tecnologias promissoras para aumentar a usabilidade desta classe de sistemas é o reconhecimento de voz. Esta tecnologia tem sido utilizada, na Área da Saúde, em aplicações como (DURLING; LUMSDEN, 2008):

- Sistemas de ditado: são sistemas que visam facilitar a transcrição de linguagem falada em texto para sua inclusão em documentos médicos como, por exemplo, em prontuários eletrônicos do paciente. Este tipo de aplicação deve utilizar um vocabulário extenso, sendo, na maioria, sistemas dependentes do usuário que são utilizados para a entrada de dados clínicos gerais, tais como em radiologia, patologia, pediatria e salas de cirurgia. Um estudo realizado por Rodger e Pendharkar (2007) mostrou que a precisão do sistema utilizado variou de 85% a 99%, dependendo de fatores, tais como idade do usuário, sexo, familiaridade com sistemas de VUI. Já Mohr et al (2003) estudaram a comparação de qualidade e precisão de documentos médicos gerados por transcrição humana e por sistemas baseados em reconhecimento de voz. Mohr et al (2003) concluíram que os níveis de precisão da tecnologia de reconhecimento de voz, naquele momento, forçavam a edição em grande proporção dos documentos transcritos e, portanto, esta tecnologia não deveria ser proferida como redução assegurada nos custos da entrada de dados.
- Sistemas com unidades de resposta audível baseados em voz (IVR): o reconhecimento de voz também tem sido utilizado para documentação médica por pacientes a fim de fornecer informações adicionais que ajudem no tratamento de doenças crônicas. Um exemplo é dado por Levin e Levin (2006) que avaliaram um sistema de Unidade de Resposta Audível baseada em voz para monitorar os níveis de dor do paciente. Os pacientes podiam fazer uma chamada para o serviço de monitoramento da dor ou configurar o sistema, em intervalos especificados, para

que ele pudesse reportar informações tais como os níveis de dor e a sua localização. Para este sistema, a taxa de erro do reconhecedor foi de apenas 2%.

- Sistemas para controlar equipamentos médicos: conforme relatado por Durling e Lumsden (2008), há uma considerável falta de pesquisas sobre controle de equipamentos médicos através de reconhecimento de voz. Um exemplo a ser citado é dado por Sebastian (2004) que utiliza voz para controlar comandos de um *scanner* de ultrassonografia, tais como “Zoom in” e “Zoom out”.
- Sistemas para interpretação de linguagem: estes sistemas de reconhecimento de voz são utilizados quando tradutores humanos não estão disponíveis a fim de facilitar a comunicação com pacientes que tenham proficiência limitada em inglês (SELIGMAN, DILINGER; 2006).

5.2 Sistemas de Transcrição Automática de Laudos

Os sistemas de transcrição humana têm sido usados para gerar laudos de radiologia manualmente, de acordo com os seguintes passos: as imagens são interpretadas por um radiologista e um laudo verbal é ditado e gravado. Este áudio é digitado por uma pessoa, resultando num laudo escrito. Este laudo é considerado preliminar, mas, em algumas práticas, ele já pode ser compartilhado por médicos nesta fase. O radiologista então finaliza o laudo transcrito depois de tê-lo revisado (usualmente sem revisar as imagens) e supõe-se que o texto esteja correto. O laudo final e as imagens são então disponibilizados para os médicos. Este é um processo que pode levar várias horas (VORBECK et al, 2000; BHAN et al, 2008).

Entre as principais desvantagens em relação ao uso de sistemas de transcrição humana, é possível destacar (HUNDT et al, 1999; DURLING; LUMSDEN, 2008; WHITE, 2005; GOPAKUMAR et al, 2008):

- Demora na devolução dos laudos digitalizados. Segundo Gopakumar et al (2008), este tempo, chamado de *turnaround time* (TAT), é de cerca de 72 horas. Os serviços tradicionais de transcrição são conhecidos por, às vezes, produzirem acúmulos de fitas de ditado esperando para serem transcritas por secretárias, ou transcrições esperando para serem revistas e aprovadas por médicos.
- Alguns pacientes necessitam de uma investigação radiológica imediata para fornecer intervenção apropriada. Entretanto, como os laudos transcritos podem consumir bastante tempo, os médicos necessitam esperar mais do que o necessário antes de aplicar a terapia correta ao paciente.
- Questões éticas da Área da Saúde, sobretudo ligadas à privacidade. Segundo Salvador e Almeida Filho (2006), várias leis e normas brasileiras e internacionais, tais como Constituição Federal Brasileira, Código Penal Brasileiro, Conselho Federal de Medicina e Organização Pan-Americana de Saúde tratam da questão da privacidade da informação em um meio eletrônico. Os profissionais envolvidos na transcrição são obrigados a não revelar as informações fornecidas em confidência sem autorização prévia do paciente, exceto quando estas informações possam ser utilizadas em função da necessidade de cuidado ao paciente, por justa causa ou dever legal.
- Adicionalmente, uma consequência não esperada é o fato de que secretárias necessitam assumir mais uma obrigação: de compartilhar o tempo com suas obrigações com o tempo nas transcrições.

A fim de encurtar o tempo necessário para escrever laudos, a tecnologia de reconhecimento de voz aparece como uma alternativa para o serviço de transcrição. A possibilidade de usar transcrição de interpretações radiológicas controladas por voz foi descrita pela primeira vez em 1981 (Robbins et al, 1987) , mas a tecnologia disponível naquela época era insuficiente e economicamente inviável para trabalhar. Em 1987 e 1988, Robbins et al (1987) descreveram o primeiro sistema de reconhecimento de voz para a escrita de laudos. Neste momento, a tecnologia de reconhecimento de voz era de natureza discreta, o que significava que os usuários deveriam aprender a falar devagar, inserindo pausas entre as palavras para serem entendidos pelo sistema. O sistema desenvolvido por Robbins et al (1987) tinha outra grande

desvantagem relacionada ao tamanho reduzido de vocabulário (900 palavras na primeira versão e 5000 na segunda), que era irreal para a área radiológica (HUNDT et al, 1999; DURLING; LUMSDEN, 2008; WHITE, 2005).

Atualmente, os sistemas de reconhecimento de voz têm trabalhado de forma contínua, permitindo que seus usuários falem de maneira mais natural (DURLING; LUMSDEN, 2008).

Alapetite, Boje e Morten (2009) relataram suas expectativas em relação ao aumento da qualidade dos registros do paciente, com o uso de sistemas de reconhecimento de voz, desde que os médicos agora chequem e validem os registros escritos imediatamente após suas conclusões, que ainda estão recentes na memória.

Outro fator importante para o uso de sistemas de reconhecimento de voz é a possibilidade de integrá-los aos sistemas já existentes, incluindo Picture Archiving and Communication System (PACS), Sistemas de Informação Radiológica (RIS) e Sistemas de Informação Hospitalares (HIS)

Portanto, os principais propósitos de se utilizar reconhecimento de voz nos departamentos de radiologia são o de diminuir o TAT e, conseqüentemente, os custos operacionais; conseguir uma qualidade mais alta dos prontuários do paciente e; integrar o sistema de reconhecimento de voz com os sistemas já existentes (PACS, RIS e HIS) (WHITE, 2005; ALAPETITE; BOJE; MORTEN, 2009).

Há a necessidade de se ressaltar, conforme apontado por Ichikawa et al (2007), que a taxa de reconhecimento de voz para a transcrição automática de laudos está intrinsecamente ligada ao ambiente (silencioso ou barulhento), assim como a quantidade de trabalho deste ambiente. Assim, em áreas ambulatoriais há menos barulho do que na área de internação, o que aumentaria a taxa de reconhecimento de voz para o primeiro caso.

Porém, outros autores têm relatado na literatura algumas desvantagens de se utilizar a tecnologia de reconhecimento de voz no departamento de radiologia, entre as quais podem ser citadas (WHITE, 2005; GUTIERREZ; MULLINS; NOVELLINE, 2005):

- O tempo que os radiologistas levam para fazer seu trabalho é maior porque eles tem que ler e corrigir os laudos.

- O uso de tecnologias como VUI e PACS têm aumentado a velocidade das informações, porém tem diminuído o contato face-a-face dos profissionais da saúde com os radiologistas.
- De acordo com White (2005), outra desvantagem de se utilizar novas tecnologias seria o aumento da dependência dos radiologistas em relação às tecnologias eletrônicas. Assim, diariamente, o radiologista visualiza imagens através de sistemas PACS, dita laudos utilizando sistemas de reconhecimento de voz, faz conferência através da Web, utiliza HIS para verificar os valores dos resultados de exames de pacientes. Assim, o radiologista gasta parte do seu tempo inserido em aprender novas tecnologias e fazer seus trabalhos utilizando-as.

Outra solução apontada por Ichikawa et al (2007), propõe o uso de reconhecimento de voz para a geração de arquivos de texto, mas passando pela correção de erros de pessoas para fazerem a transcrição (utilizando os arquivos de voz também gerados), ou seja, uma abordagem mista entre as duas frentes defendidas pelos autores supracitados. Obviamente esta solução diminui o TAT, mas não torna o laudo disponível em tempo real.

5.3 Funcionamento de um Sistema de Transcrição Automática de Laudos

As principais etapas, que compõe um sistema automático de transcrição de laudo em radiologia atualmente comercializados, são mostradas na Figura 10.



Figura 10: Etapas de um Sistema de Transcrição Automática de Laudo. (Fonte: autor)

São elas:

- Entrada de Voz: utilização de um equipamento, tal como mostrado na Figura 11, para a entrada de voz do radiologista.
- Captura do Laudo: um servidor capaz de armazenar os arquivos de áudio gerados durante o ditado do radiologista.
- Reconhecimento de Voz: um módulo de reconhecimento de voz com o vocabulário e gramática próprias da área de radiologia, fornecido, atualmente, pela empresa Nuance (2010).
- Correção e Edição de Laudo: verificação de possíveis erros de reconhecimento de voz que possam ter acontecido na fase anterior. Estes erros podem ser simplesmente a falta de pontuação ou erros que comprometam o entendimento do lado.
- Laudo Disponível: já corrigido, o laudo é disponibilizado para aqueles que possam ter acesso a ele.



Figura 11: Dispositivo para a entrada de voz – SpeechMike Pro 5274

Fonte: <http://www.solitonit.com/speechHardware.htm>

A interface gráfica relativa a estes sistemas de laudos pode ser vista a seguir, conforme a Figura 12.

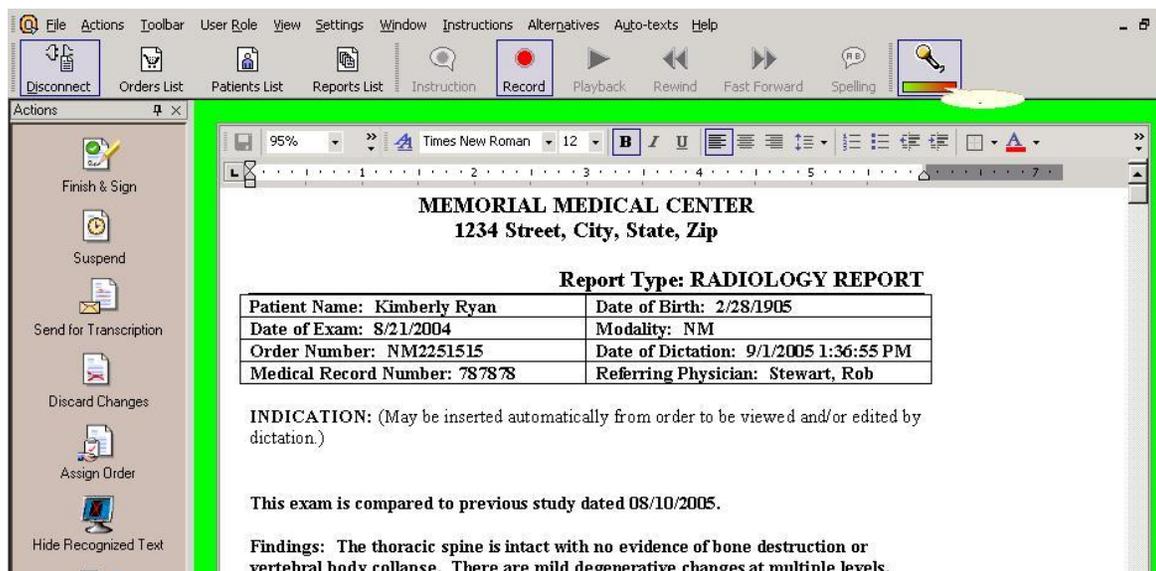


Figura 12: Interface gráfica de um sistema de transcrição automática de laudo

Fonte: <http://www.medquist.com>)

5.4 Considerações Financeiras

Os sistemas de reconhecimento de voz necessitam de um investimento de capital que podem ser substanciais. É necessário um investimento em (WHITE, 2005):

- Computadores *desktops*: equipados com sistemas de som que são relativamente baratos.
- Servidores: que armazenarão todas as gravações. São relativamente caros.
- Infra-estrutura de rede: embora geralmente já exista na instituição, pode ter que ser realizado uma atualização.
- Software e atualizações: são geralmente mais caros e é necessário pensar sobre as atualizações que deverão ser realizadas. Além disto, geralmente, estes sistemas são vendidos por números de licenças (uma licença para cada voz).

- Treinamento: depois de toda a implantação do sistema, é necessário fazer um treinamento a todos que utilizarão o sistema, a fim de que sua produtividade seja incrementada.
- Suporte: custos com falhas de equipamento e software e ajudas devem ser considerados.

5.5 Requisitos de um Sistema de Transcrição Automática de Laudo

Através da revisão bibliográfica sobre Interface Homem-Computador, VUI, Sistemas de Informação Radiológica, foi possível listar os seguintes requisitos no uso de VUI em RIS em sistemas de transcrição automática de laudo:

1. Precisão (Acurácia): é um dos requisitos mais importantes já que informações erradas podem comprometer a qualidade do laudo, alterar um diagnóstico e comprometer um tratamento.
2. Tamanho do vocabulário: é também um requisito bastante importante, pois o vocabulário não pode ser grande o bastante para baixar o nível de reconhecimento das palavras, nem tão pequeno que não contemple as palavras do domínio da aplicação.
3. Dicionário específico para RIS: o sistema deve contemplar palavras utilizadas diariamente num laudo de radiologia.
4. Interferência Sonora: o ambiente hospitalar, dependendo da área, pode ser um ambiente de muito ruído (ICHIKAWA et al, 2007). Porém, isto não pode influenciar grandemente a eficiência do reconhecimento.
5. Reconhecimento contínuo: o usuário deve ditar o laudo sem se preocupar com pausas entre palavras, ou seja, deve falar de maneira natural e contínua.

6. Desejável separação entre o teclado e o sistema de ditado: o usuário deveria poder fazer o ditado através de um aparelho específico (como o mostrado na Figura 11), através do celular ou de um telefone comum, permitindo que a aplicação seja ubíqua.
7. Ajuda: caso o usuário tenha algum problema na execução do sistema, é preciso que haja uma ajuda e que o sistema seja capaz de identificar quando o usuário está em apuros. Também é desejável que o usuário possa acionar a ajuda, dizendo alguma palavra específica, tal como “ajuda” ou “help”.
8. Minimização da sobrecarga de memória: o sistema não deve exigir que o usuário se recorde de termos específicos do sistema ou deve proporcionar a lista destes termos para que o usuário não se frustre.
9. Recuperação de erros: o usuário deve ser capaz de corrigir seu ditado a qualquer momento.
10. Tempo de resposta do *feedback*: o tempo de resposta para o usuário não deve ser alto deixando o usuário pensando que o sistema não o compreendeu ou que algo errado foi realizado. Segundo Nielsen (1993), um décimo de segundo (0,1s) é o limite para o usuário pensar que o sistema está reagindo instantaneamente, o que significa que nenhum *feedback* especial é necessário; um segundo (1.0s) é o limite para que o fluxo de pensamento do usuário não seja interrompido, mesmo que o usuário perceba uma certa demora; e dez segundos (10s) é o limite para manter a atenção do usuário focalizada no diálogo. Às vezes, *feedbacks* especiais são necessários para contextualizar uma navegação mais demorada do usuário. É necessário salientar que estas medidas foram realizadas para sistemas *stand-alone* ou páginas WEB. Porém, para sistemas de reconhecimento de voz, este tempo não foi encontrado na literatura, precisando ser tomado a partir de observação do usuário no uso do sistema.

11. Visibilidade do sistema: o sistema deve proporcionar ao usuário do sistema, o que está se passando, se o sistema está consultando determinada informação, se não entendeu uma entrada ou se está processando determinada ação; a transição entre diálogos em uma interação por voz deve ser perceptível ao usuário, pois as suas habilidades e necessidades, bem como o contexto de uso podem ser modificadas durante a interação.
12. Uso da arquitetura Cliente-Servidor: para que os radiologistas estejam livres para se moverem de uma estação a outra dentro do hospital ou clínica.
13. Integração com os sistemas já existentes: PACS, HIS e RIS.
14. Qualidade das entradas de dados: o sistema deve possuir periféricos de entrada, tal como microfones que não causem interferências às entradas de voz.
15. Qualidade das mensagens de saída: está ligada à inteligibilidade e o quão agradável é seu uso. É possível produzir mensagens de saída através de sintetização de voz, por mensagens pré-gravadas ou um texto exibido na tela. Embora a sintetização de voz produza mensagens bastante inteligíveis, elas não são muito agradáveis de serem escutadas. Já as mensagens pré-gravadas não podem ser utilizadas em todos os sistemas, sendo bastante restritas e inflexíveis.
16. Qualidade do sistema de áudio: a voz gravada deve ser compreensível e muito semelhante à voz original.
17. Tempo de retorno: tempo para o laudo ficar pronto que deve ser menor do que os sistemas de transcrição humana.
18. Modalidade apropriada: se o sistema não compreender o que o usuário diz, é necessário fornecer outra forma de entrada de dados que não a voz, tal como a possibilidade de soletrar as letras ou entrar via teclado.
19. Adequação do *Feedback* (sincronização da fala e escrita na tela): as frases ditas pelo radiologista aparecerão, normalmente, na tela, para que este possa fazer o acompanhamento do que está ditando e o que o sistema está reconhecendo. Se o *delay* for muito alto, isto poderá causar um efeito de sobrecarga de trabalho para o radiologista, que estará ditando uma frase e visualizando outra.

20. Diversidade e Percepção Humana: o sistema deve permitir que usuário experientes possam realizar suas tarefas de maneira mais eficiente que usuários iniciantes, seja através de teclas de atalho ou uso de *barge-in* - por exemplo, sendo capaz de clicar com o mouse e recuperar padrões de laudos já disponibilizados.
21. Satisfação do Cliente: está ligada à porcentagem de erros gerados pelo sistema e facilidade de uso, assim como o entusiasmo do usuário para utilizar o sistema novamente.
22. Naturalidade da fala do usuário: o usuário deve poder falar de maneira natural e contínua, da mesma forma que faria se estivesse gravando o áudio.
23. Resolução de ambiguidade para homônimos: o sistema deveria ser capaz de diferenciar palavras com o mesmo fonema, porém com escrita diferente.

5.6 Considerações Finais

Neste capítulo foi possível verificar o uso de reconhecimento de voz para aplicações efetivas na área da Saúde, sobretudo em aplicações de propósito específico, tais como na radiologia. Também, aqui, foram descritas as principais características e os requisitos para sistemas de transcrição automática de laudos em radiologia.

Uma metodologia será proposta, utilizada para medir estes principais requisitos, aqui estabelecidos. O próximo capítulo tratará de abordar esta metodologia.

6. METODOLOGIA DE AVALIAÇÃO DE INTERFACES DE VOZ PARA A ÁREA DA SAÚDE – SISTEMAS DE TRANSCRIÇÃO AUTOMÁTICA DE LAUDO

O objetivo deste capítulo é apresentar a metodologia de avaliação para sistemas de transcrição automática de laudos em radiologia como parte dos resultados obtidos a partir de estudos aprofundados em temas discutidos nos capítulos anteriores. Será mostrado como os requisitos estão sendo agrupados em seis classes e também sua dificuldade de avaliação. Então, são listados e explicados os requisitos selecionados e é proposta uma metodologia para avaliação de cada requisito selecionado.

6.1 Introdução

Conforme visto no Capítulo 2, a avaliação de VUIs é diferente da avaliação de GUIs, principalmente no que se refere à transiência natural da voz, o que afeta fatores importantes de usabilidade, tais como transparência, aprendizado, sobrecarga cognitiva, manipulação de erros e controle do usuário.

Em geral, a maioria dos métodos de avaliação, testes, padrões e métricas de usabilidade pode ser aplicada a VUIs. Porém, há requisitos, conforme vistos no capítulo 4 e 5, que são específicos para aplicações VUI, ou, mais diretamente, para sistemas de transcrição automática de laudos em radiologia. Há requisitos de VUIs que não são aplicáveis a estes sistemas, principalmente aqueles referentes à iniciativa do diálogo ou comunicação. Isto leva a se pensar numa metodologia específica para esta classe de aplicações.

6.2 Objetivos da Metodologia Utilizada

A metodologia proposta a seguir para a avaliação dos requisitos de sistemas de informação em radiologia (sistemas de transcrição automática de laudos) que utilizam VUI tem como objetivos:

1. Utilizar testes de usabilidade e inspeção de usabilidade de maneira complementar a fim de propiciar um menor custo e um menor tempo de avaliação.
2. Poder ser aplicada a sistemas já implementados.
3. Servir como um guia para se avaliar a usabilidade desta classe de sistemas.
4. Investigar a viabilidade de se avaliar determinados requisitos.
5. Agrupar os requisitos propostos segundo classes de características.
6. Propor métricas para avaliar estes requisitos.

Esta metodologia sugere que se utilizem os métodos para avaliação da usabilidade baseados em inspeção sempre que seja possível avaliar a interface sem envolver os usuários finais, a fim de diminuir os custos das sessões de testes de usabilidade. Devem-se utilizar testes de usabilidade quando for verificado que a inspeção não seja suficiente.

Embora, na literatura apontada por Nielsen (1989, 1993) seja desejável utilizar de três a quatro avaliadores para se compor a lista de inspeção de usabilidade, esta metodologia utilizará dois avaliadores (a própria autora e um especialista em VUI convidado). O motivo para esta abordagem se deu porque as inspeções utilizariam muito mais horas – cerca de 10 vezes mais - do que Nielsen (1989, 1993) discute em seu trabalho – em torno de 2 horas. Também por esta razão, a 2ª especialista não fez parte de todas as inspeções.

6.3 Agrupamento dos Requisitos em Classes

Möller (2005), em seu trabalho, encontrou 36 diferentes métricas objetivas e as dividiu em cinco categorias, relacionadas a:

- a. diálogo e comunicação;
- b. meta-comunicação;
- c. capacidade de trabalho colaborativo (*cooperativeness*);
- d. tarefa, e
- e. entrada de dados.

Porém, para o domínio estabelecido na área da Saúde, a partir dos requisitos e métricas listados no Capítulo 4 e 5, para se avaliar sistemas de transcrição automática de laudo, as seguintes classes são, aqui, propostas:

- Classe 1 – Desempenho: requisitos que estejam ligados ao correto funcionamento da aplicação, em termos de reconhecimento da voz e interferências sofridas pela aplicação que possam degradar o seu desempenho.
- Classe 2 - Facilidade de Uso: requisitos que estejam ligados às facilidades que o sistema pode fornecer ao usuário, a fim de que sua tarefa possa ser executada de maneira eficaz e eficiente, diminuindo a carga cognitiva do usuário.
- Classe 3 - Hardware e Integração: requisitos que estejam ligados ao desempenho físico, com dispositivos que permitam uma boa entrada de áudio, assim como à facilidade de integração deste sistema a outros tais como PACS e HIS.
- Classe 4 - Fatores Humanos: requisitos que estejam ligados ao prazer do usuário em usar o sistema e sua vontade de continuar a utilizar o sistema.
- Classe 5 - *Feedback*: requisitos que estejam ligados às mensagens apresentadas ao usuário sobre o *status* do sistema, assim como o tempo de resposta entre o ditado do usuário e sua transcrição na tela.
- Classe 6 – Tratamento de Erro e Sistema de Ajuda: requisitos que estejam relacionados à capacidade tanto do sistema corrigir erros encontrados, quanto do sistema corrigir um ditado, seja em tempo real, ou tempo posterior.

O quadro a seguir mostra o agrupamento sugerido de acordo com a definição das classes:

CLASSE	REQUISITO
Classe 1: Desempenho	Precisão, Tamanho do vocabulário, Dicionário específico para RIS, Interferências sonoras, Naturalidade da fala do usuário (Reconhecimento Contínuo), Resolução de ambiguidade para homônimos
Classe 2: Facilidade de Uso	Minimização da sobrecarga de memória, Modalidade apropriada, Tempo para o laudo ficar pronto
Classe 3: Hardware e Integração	Integração com sistemas já existentes, Qualidade do sistema de áudio, Qualidade das entradas de dados
Classe 4: Fatores Humanos	Diversidade e Percepção humana, Satisfação do usuário
Classe 5: <i>Feedback</i>	Tempo de <i>feedback</i> do sistema, Visibilidade do sistema, Adequação do <i>feedback</i>
Classe 6: Tratamento de erros e Sistema de Ajuda	Tratamento e Recuperação de Erros, Sistema de Ajuda

Quadro 7: Agrupamento dos Requisitos em Classes. (Fonte: autor)

6.4 Classificação dos Requisitos quanto à Dificuldade de Avaliação

A necessidade de se classificar os requisitos para os sistemas de transcrição automática de laudos segundo sua complexidade em serem avaliados, gerou a criação de 3 classes de dificuldade: Grau 1 – Baixa Complexidade; Grau 2: Complexidade Média; e Grau 3: Alta Complexidade.

Assim, foram classificados todos os requisitos listados no item 6.3 em três graus, conforme mostrado na Quadro 8:.

COMPLEXIDADE	REQUISITOS
Grau 1	Precisão (Acurácia); Tamanho do vocabulário; Interferências sonoras; Reconhecimento contínuo; Tempo para o laudo ficar pronto;
Grau 2	Adequação do Sistema de Ajuda, Tratamento e Recuperação de erros; Qualidade das entradas de dados; Modalidade apropriada; Naturalidade da fala do usuário;
Grau 3	Minimização da sobrecarga de memória; Tempo de resposta do <i>feedback</i> ; Adequação do <i>feedback</i> ; Qualidade do sistema de áudio; Diversidade e Percepção Humana; Satisfação do cliente; Resolução de ambiguidade para homônimos

Quadro 8: Classificação da complexidade de se avaliar cada requisito. (Fonte: autor)

6.5 Requisitos Selecionados

Dos 19 requisitos estudados, foram selecionados os seguintes para compor a metodologia de avaliação:

1. Precisão (Acurácia): significa a porcentagem de acerto no reconhecimento das palavras ditas pelo usuário e que serão transcritas. Trata-se de um requisito crucial para os sistemas de transcrição automática de laudo, pois enganos no reconhecimento das palavras podem trazer consequências desastrosas ao paciente.
2. Tamanho do vocabulário: o vocabulário não pode ser muito pequeno de tal forma que não cubra o domínio da aplicação, nem tão amplo que diminua a taxa de reconhecimento (é fundamental investigar qual é, aproximadamente, este tamanho e se os sistemas analisados tem variação deste tamanho).
3. Dicionário específico para RIS: o vocabulário deve cobrir o domínio da aplicação, no caso, os termos utilizados na área de radiologia (torna-se prescindível investigar como este dicionário é formado, e verificar se o sistema não reconhece palavras além do que necessitaria).

4. Interferência sonora: num ambiente hospitalar, é necessário investigar qual a contribuição do ruído do ambiente tanto na redução da taxa de reconhecimento quanto no aumento da carga cognitiva do usuário final.
5. Naturalidade da Fala do Usuário (Reconhecimento contínuo): este requisito está ligado à capacidade do sistema de transcrição automática de laudo conseguir reconhecer o ditado feito pelo usuário final, referente a um laudo. Deve ser lembrado que o usuário está construindo as frases em tempo-real, podendo haver vários problemas da fala natural, tais como concordância verbal e nominal, inserção de palavras irrelevantes ao texto, tais como “ééé”, “hummm” que terão que ser tratadas pelo sistema.
6. Ajuda: este requisito está ligado à facilidade com que usuários, principalmente iniciantes, terão para aprenderem e acessarem o sistema de ajuda para conseguirem resolver seus problemas com o uso do sistema e conseguirem finalizar seus objetivos (ditar um laudo de maneira eficiente).
7. Tratamento e Recuperação de erros: está relacionado a como o sistema deve agir para se recuperar de um erro. Isto, em um sistema de transcrição automática de laudo, estaria ligado, por exemplo, a como o sistema age quando não reconhece uma palavra que o usuário ditou.
8. Adequação do *feedback*: tratando-se de um sistema de tempo-real, em que o usuário está ditando o laudo, existem alguns agravantes: a) o sistema não deve fornecer *feedbacks* que atrapalhem a capacidade de raciocínio do usuário; b) deve estar presente, de maneira que o usuário, em tempo oportuno, saiba de erros que tenham ocorrido no sistema enquanto ele ditava o laudo (tal como uma falha no reconhecimento de alguma palavra).
9. Tempo de resposta do *feedback*: as transcrições devem ocorrer em tempo-real, sem que o *delay* possa interferir na carga cognitiva do usuário final.

10. Visibilidade do sistema: este requisito está relacionado à maneira como o sistema terá para mostrar ao usuário seu *status* atual. Por exemplo, se o sistema encontrar algum erro de reconhecimento ou tiver algum problema no acesso ao dicionário, como será realizada a visibilidade do sistema.
11. Integração com os sistemas já existentes: este é um requisito bastante desejável, já que evita re-trabalho dentro das unidades que compõe o Sistema de Informação Hospitalar e o PACS.
12. Qualidade das entradas de dados: este requisito está relacionado a bons dispositivos de captação de voz (microfones) que não causem problemas de interferência sonora ou elétrica que promovam distorções na fala.
13. Qualidade do sistema de áudio: apesar do radiologista realizar a correção do laudo imediatamente após tê-lo ditado, é notoriamente importante que haja a gravação do áudio para que seja possível realizar comparações, em tempo oportuno, entre o que foi transcrito e o áudio inicial, a fim de sanar eventuais dúvidas. Sendo assim, este áudio tem que ser inteligível a quem necessite rever estas informações.
14. Tempo para o laudo ficar pronto: o tempo para o laudo estar disponível para consulta através do uso de um sistema de transcrição automática de laudo deve ser menor do que através de outros métodos: escrita à mão, digitação pelo próprio radiologista, uso do serviço de transcrição humana.
15. Modalidade apropriada: este requisito está ligado ao uso de modalidades de interação complementares à VUI. Para um sistema de transcrição automática de laudo, o uso de monitor, teclado e mouse são não somente desejáveis, como fundamentais.
16. Diversidade e Percepção Humana: o uso de outras modalidades de interação, além da voz, deveriam tornar o sistema mais utilizável para usuários experientes, permitindo encurtar caminhos e agilizar a tarefa.

17. Satisfação do Cliente: este requisito está ligado ao prazer de se utilizar o sistema de transcrição automática de voz. É um requisito totalmente subjetivo que geralmente é medido através da aplicação de questionários, tais como o apresentado no Apêndice A.

Os requisitos que não serão analisados no contexto deste trabalho são:

1. Minimização da sobrecarga de memória: apesar de ser um requisito bastante importante quando se avalia a usabilidade de praticamente todas as aplicações com interfaces, é considerado um requisito bastante difícil de ser medido.
2. Resolução de ambiguidade para homônimos: é um requisito bastante difícil de se avaliar, com necessidade de certa inteligibilidade do sistema para detectar a semântica das sentenças.

6.6 Método de Avaliação para cada Requisito Selecionado

A fim de tentar diminuir os custos com os testes de usabilidade, será utilizada a avaliação por inspeção de usabilidade para as medidas objetivas baseada em heurísticas. Já para as medidas subjetivas, serão utilizados testes e questionários.

O desenvolvimento de um questionário válido e confiável não é uma tarefa trivial. A grande maioria dos questionários para avaliação de interfaces do usuário foi desenvolvida para interfaces gráficas do usuário, tendo que ser repensada para o uso em VUIs. Os questionários que serão utilizados nesta metodologia foram elaborados com base na pesquisa desenvolvida por Hartikainen, Salonen e Turunen (2004), em que é proposta a aplicação de dois questionários aos usuários: questionário de expectativas – aplicado antes do uso do sistema, e questionário de percepção – aplicado depois do uso do sistema (Apêndice A). A elaboração do questionário também está baseada na metodologia Questionnaire User Interface Satisfaction (QUIS) proposta em (CHIN; DIEHL; NORMAN, 1988).

Por outro lado, as heurísticas utilizadas foram baseadas num adaptação de Nielsen (1993), no estudo de requisitos não funcionais para VUI e nas boas práticas de desenvolvimento de VUI apontados por (DYBKJAER; BERSEN, 2001), (SALVADOR et al., 2008) e (KOMATANI et al., 2003), de acordo com as peculiaridades dos sistemas de transcrição automática de laudos. Utilizando os requisitos para sistemas de transcrição de laudos utilizando VUI selecionados na sessão anterior, foram pesquisadas métricas na literatura (seção 4.5) e, também, propostas, pela autora, novas métricas para sua avaliação de acordo com as peculiaridades dos sistemas de transcrição automática de laudos. Um *template* baseado em Dybkjaer e Bernsen (2001) é proposto para melhor organizar as avaliações (Quadro 9:).

REQUISITO	
Tipo de Avaliação	
Métodos de Avaliação	
Importância	
Dificuldade de Avaliação	
Sintomas a procurar/ Métricas a utilizar	

Quadro 9: *Template* utilizado para realizar a avaliação dos requisitos. (Fonte: autor)

Assim, foram abordadas as métricas de acordo com o caráter objetivo ou subjetivo (quantitativo ou qualitativo) dos requisitos selecionados, conforme é possível visualizar a seguir.

6.6.1 Requisitos Subjetivos

Os seguintes requisitos medidos de maneira subjetiva compõem a metodologia proposta (Quadro 10: a 16).

SATISFAÇÃO DO CLIENTE	
Tipo de Avaliação	Subjetiva
Métodos de Avaliação	Questionário
Importância	Alta
Dificuldade de Avaliação	Grau 3
Sintomas a procurar/ Métricas a utilizar	Facilidade de uso, valor agregado, sucesso da tarefa.

Quadro 10: Avaliação da Satisfação do Cliente. (Fonte: autor)

MODALIDADE APROPRIADA	
Tipo de Avaliação	Subjetiva
Métodos de Avaliação	Inspeção, Observação e questionário
Importância	Alta
Dificuldade de Avaliação	Grau 2
Sintomas a procurar/ Métricas a utilizar	Verificar se as modalidades complementares de interação facilitam a eficiência do uso do sistema e podem diminuir o tempo para o cumprimento dos objetivos do usuário.

Quadro 11: Avaliação da Modalidade Apropriada. (Fonte: autor)

DIVERSIDADE E PERCEPÇÃO HUMANA	
Tipo de Avaliação	Subjetiva
Métodos de Avaliação	Observação e questionário
Importância	Alta
Dificuldade de Avaliação	Grau 3
Sintomas a procurar/ Métricas a utilizar	Verificar se usuários com graus de experiência diferentes conseguem realizar suas tarefas de maneira apropriada. Para isto o sistema deve fornecer formas diferentes de acesso às funcionalidades.

Quadro 12: Avaliação da Diversidade e Percepção Humana. (Fonte: autor)

MINIMIZAÇÃO DA SOBRECARGA DE MEMÓRIA	
Tipo de Avaliação	Subjetiva
Métodos de Avaliação	Observação
Importância	Alta
Dificuldade de Avaliação	1
Sintomas a procurar/ Métricas a utilizar	Verificar qual a diferença de desempenho entre o usuário ditar o laudo enquanto olha para a tela e quando não olha.

Quadro 13: Avaliação da Minimização da Sobrecarga de Memória. (Fonte: autor)

ADEQUAÇÃO DO <i>FEEDBACK</i>	
Tipo de Avaliação	Subjetiva
Métodos de Avaliação	Inspeção e Questionário
Importância	Alta
Dificuldade de Avaliação	Grau 3
Sintomas a procurar/ Métricas a utilizar	a) o sistema não deve fornecer <i>feedbacks</i> que atrapalhem a capacidade de raciocínio do usuário; b) deve estar presente, de maneira que o usuário, em tempo oportuno, saiba de erros que tenham ocorrido no sistema enquanto ele ditava o laudo (tal como uma falha no reconhecimento de alguma palavra)

Quadro 14: Avaliação do *Feedback* do Sistema. (Fonte: autor)

ADEQUAÇÃO DO SISTEMA DE AJUDA	
Tipo de Avaliação	Subjetiva
Métodos de Avaliação	Inspeção e Questionário
Importância	Média
Dificuldade de Avaliação	Grau 1
Sintomas a procurar/ Métricas a utilizar	Verificar se o sistema de ajuda satisfaz as necessidades do usuário, caso esteja com dúvidas.

Quadro 15: Avaliação da Eficiência do Sistema de Ajuda. (Fonte: autor)

VISIBILIDADE DO SISTEMA	
Tipo de Avaliação	Subjetiva
Métodos de Avaliação	Inspeção e Questionário
Importância	Alta
Dificuldade de Avaliação	Grau 2
Sintomas a procurar/ Métricas a utilizar	Verificar se o sistema é capaz de informar ao usuário o que está acontecendo em determinado momento.

Quadro 16: Avaliação da Visibilidade do Sistema. (Fonte: autor)

6.6.2 Requisitos Objetivos

Os seguintes requisitos, medidos de maneira objetiva, compõem a metodologia proposta (Quadro 17: a Quadro 29).

PRECISÃO	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção de usabilidade
Importância	Alta
Dificuldade de Avaliação	Grau 1
Sintomas a procurar/ Métricas a utilizar	Dentro da quantidade total de palavras que compõe o ditado, qual a porcentagem de erro de reconhecimento da voz, utilizando a variação de três questões: <ul style="list-style-type: none"> a) Interferência sonora no ambiente ou não b) Treino da voz ou não c) Frases do ditado geradas em tempo-real, podendo haver vários problemas da fala natural ou leitura de um texto já escrito. d) Tipo de equipamento utilizado para captura do áudio.

Quadro 17: Avaliação da Precisão. (Fonte: autor)

INTEGRAÇÃO COM SISTEMAS JÁ EXISTENTES	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção
Importância	Média
Dificuldade de Avaliação	Grau 1
Sintomas a procurar/ Métricas a utilizar	Verificar se o sistema tem integração com o PACS e com o HIS e qual este nível de integração

Quadro 18: Avaliação da Integração com Sistemas já Existentes. (Fonte: autor)

RECUPERAÇÃO DE ERROS	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção
Importância	Alta
Dificuldade de Avaliação	Grau 3
Sintomas a procurar/ Métricas a utilizar	Verificar como o sistema age perante os seguintes erros: <ul style="list-style-type: none"> a) O usuário usa palavras que não estão no dicionário da aplicação b) O sistema não consegue reconhecer o que o usuário dita c) O sistema não consegue acessar o dicionário d) O sistema não consegue integração com os demais sistemas e) O usuário erra o laudo e deseja refazê-lo.

Quadro 19: Avaliação da Recuperação de Erros. (Fonte: autor)

NATURALIDADE DA FALA DO USUÁRIO	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção e Observação
Importância	Alta
Dificuldade de Avaliação	Grau 3
Sintomas a procurar/ Métricas a utilizar	<p>1. Escolhendo usuários de ambos os sexos:</p> <p>a) Através da leitura de laudos, verificar a diferença de precisão entre o usuário falar mais pausadamente ou mais rapidamente;</p> <p>b) Através da leitura de laudos, verificar a diferença de precisão entre o usuário falar com tom baixo ou alto;</p> <p>c) Usuário lê um laudo e usuário fala um laudo sem ler;</p> <p>2. Escolhendo usuários de ambos os sexos com sotaque diferente da cidade de São Paulo:</p> <p>a) Através da leitura de laudos, verificar a diferença de precisão entre o usuário falar mais pausadamente ou mais rapidamente;</p> <p>b) Através da leitura de laudos, verificar a diferença de precisão entre o usuário falar com tom baixo ou alto;</p>

Quadro 20: Avaliação da Naturalidade da Fala do Usuário. (Fonte: autor)

TEMPO PARA O LAUDO FICAR PRONTO	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção e observação
Importância	Alta
Dificuldade de Avaliação	Grau 1
Sintomas a procurar/ Métricas a utilizar	<p>Verificar quanto tempo é necessário para o laudo ficar pronto, ou seja, desde o ditado até a correção por parte do radiologista, em detrimento aos seguintes casos:</p> <ul style="list-style-type: none"> a) O radiologista é rápido para datilografar o laudo e fazer sua correção; b) O radiologista é lento para datilografar o laudo e fazer sua correção; c) O radiologista grava o áudio, envia para o serviço de transcrição humana e espera o laudo voltar para fazer sua correção;

Quadro 21: Avaliação do Tempo para o Laudo ficar Pronto. (Fonte: autor)

TAMANHO DO VOCABULÁRIO	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção
Importância	Alta
Dificuldade de Avaliação	Grau 1
Sintomas a procurar/ Métricas a utilizar	<p>Verificar o número de palavras que pertencem ao vocabulário do sistema, inicialmente.</p> <p>Verificar o número de palavras que foram inseridas ao vocabulário do sistema, pelo usuário ou grupo de TI.</p>

Quadro 22: Avaliação do Tamanho do Vocabulário. (Fonte: autor)

DICIONÁRIO ESPECÍFICO PARA RIS	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção
Importância	Alta
Dificuldade de Avaliação	Grau 2
Sintomas a procurar/ Métricas a utilizar	Através da inserção de laudos falados, verificar qual a porcentagem de termos técnicos que não são reconhecidos pelo sistema.

Quadro 23: Avaliação da Aderência do Dicionário. (Fonte: autor)

INTERFERÊNCIA SONORA	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção
Importância	Alta
Dificuldade de Avaliação	Grau 2
Sintomas a procurar/ Métricas a utilizar	Através de um decibelímetro, medir a interferência sonora e inserir um laudo, verificando a precisão do reconhecimento; fazer isso para diferentes graus de interferência sonora.

Quadro 24: Avaliação da Interferência Sonora do Ambiente. (Fonte: autor)

RESOLUÇÃO DE AMBIGUIDADE PARA HOMÔNIMOS	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção
Importância	Baixa
Dificuldade de Avaliação	Grau 2
Sintomas a procurar/ Métricas a utilizar	Através de uma lista de palavras homônimas, inserir um laudo falado e verificar se o sistema é capaz de escolher a palavra correta, pelo contexto.

Quadro 25: Avaliação da Capacidade do Sistema de Resolver Homônimos. (Fonte: autor)

QUALIDADE DO SISTEMA DE ÁUDIO	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção e Análise de Dados
Importância	Alta
Dificuldade de Avaliação	Grau 2
Sintomas a procurar/ Métricas a utilizar	Gravar a voz de cada usuário e verificar a interferência sonora encontrada no áudio.

Quadro 26: Avaliação da Qualidade de Áudio do Sistema. (Fonte: autor)

QUALIDADE DAS ENTRADAS DE DADOS	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção
Importância	Alta
Dificuldade de Avaliação	Grau 1
Sintomas a procurar/ Métricas a utilizar	Inserir laudos falados utilizando dois tipos de equipamentos diferentes: um microfone comum e um SpeechMike da Philips – aparelho utilizado em hospitais. Medir a diferença de precisão.

Quadro 27: Avaliação da Qualidade das Entradas de Áudio. (Fonte: autor)

TEMPO DE <i>FEEDBACK</i> DO SISTEMA	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção
Importância	Alta
Dificuldade de Avaliação	Grau 1
Sintomas a procurar/ Métricas a utilizar	Verificar o tempo entre a pronúncia de uma palavra ou frase e a escrita na tela.

Quadro 28: Avaliação do Tempo de *Feedback* do Sistema. (Fonte: autor)

TRATAMENTO E RECUPERAÇÃO DE ERROS	
Tipo de Avaliação	Objetiva
Métodos de Avaliação	Inspeção
Importância	Alta
Dificuldade de Avaliação	Grau 3
Sintomas a procurar/ Métricas a utilizar	Verificar, através de simulação, erros que poderiam acontecer durante o ditado do laudo e como seria possível corrigi-los pelo sistema.

Quadro 29: Avaliação do Tratamento e Prevenção de Erros. (Fonte: autor)

6.7 Considerações Finais

Neste capítulo foi proposta uma metodologia de avaliação de Sistemas de Transcrição Automática de Laudos em Radiologia, levando em consideração requisitos gerais de interfaces e também requisitos específicos de aplicações VUI e, mais diretamente, requisitos que devem ser considerados para os sistemas de ditado em Saúde.

O próximo capítulo tem como foco todo o plano de avaliação, em detalhes, realizados para a validação da metodologia aqui proposta.

7. APLICAÇÃO DA METODOLOGIA A SISTEMAS JÁ IMPLANTADOS

Serão abordadas, neste Capítulo, todas as fases de preparação dos testes e de inspeção de usabilidade de sistemas já implantados em hospitais, utilizando a metodologia detalhada no Capítulo 6.

Este Capítulo tem por objetivo verificar quão bem funciona um sistema de transcrição de laudos de radiologia, utilizando-se a metodologia proposta no Capítulo 6, no qual foram propostos os requisitos e as métricas para esse tipo de sistema.

Sabe-se que praticamente todos os sistemas de transcrição de laudos em radiologia que estão implantados atualmente nos hospitais e clínicas do Brasil têm o mesmo motor de reconhecimento de voz, fornecido pela Nuance (2010). Isso significa que, em termos de taxa de reconhecimento da voz, os sistemas não devem diferir, numa versão inicial, muito entre si, a não ser por versões diferentes. Porém, existem customizações que são realizadas, geralmente pela equipe de TI dos hospitais, que podem aumentar a eficiência ou alterar as características desses sistemas. Outro ponto a considerar é a capacidade de aprendizado do sistema em relação aos erros de reconhecimento e palavras novas que são inseridas ao vocabulário inicial do sistema.

Serão utilizados aqui dois sistemas de transcrição automática de laudos em radiologia, que serão descritos a seguir. A fim de facilitar os testes, economizar tempo e custos, além de incomodar, o mínimo possível, os médicos radiologistas, optou-se por ser realizada a maioria dos testes com usuários participantes interessados na área tecnológica. Outro ponto importante para a redução de tempo e de custos é o uso de inspeção de usabilidade, por ser realizado por especialistas em usabilidade, ao invés de usuários finais.

O *corpus* utilizado para a pesquisa corresponde a um conjunto de radiologia de tórax. Percebe-se que estes dados são bastante formatados e não correspondem a uma fala natural de um Radiologista.

7.1 Descrição dos sistemas utilizados

A empresa Nuance (2010) é, atualmente, a principal empresa fornecedora do motor de reconhecimento de voz para o tipo de sistema em estudo. Até o ano de 2005, aproximadamente, existiam várias outras empresas que forneciam motores de reconhecimento de voz, tais como: IBM ViaVoice (IBM, 02/03/2009), Philips Speech Pearl (Nuance, 2010), Dragon Systems (Nuance, 2010) - que terminaram por serem adquiridas pela Nuance.

Várias empresas, tais como Philips Healthcare (Philips, 2010) e Agfa (Agfa, 2010), que vendem sistemas radiológicos, incorporam o motor de reconhecimento de voz da Nuance em suas soluções. O que torna um sistema mais atrativo que outro são, então, as facilidades – customizações - que podem ser incorporadas – programadas - ao sistema radiológico.

A seguir serão descritos dois sistemas testados neste trabalho. O primeiro utiliza a máquina de reconhecimento de voz da Nuance, sem facilidades programadas nos sistemas de radiologia. O segundo sistema, também da empresa Philips (2010), na versão 2.1, pertence a um hospital de São Paulo em que poucos usuários costumam utilizá-lo.

7.1.1 Sistema Philips de Uso Local

A fim de baratear os custos com os testes nos hospitais, diminuir o tempo de testes e também o incômodo para o setor de TI dos hospitais que teriam que disponibilizar o ambiente de testes do reconhecedor de voz, optou-se pela instalação do sistema Speech Magic™ versão 6.1 (2010) em uma máquina fora do ambiente hospitalar.

Embora hajam alguns inconvenientes no uso desta abordagem, tais como o sistema não estar instalado como cliente-servidor e não tendo condições, por isso, de um aprendizado mais eficaz da máquina de reconhecimento de voz – e, ainda, não permitir a entrada de novas palavras -, para vários tipos de testes, esta abordagem já é suficiente. Foi possível testar vários requisitos do sistema, como o *delay* do sistema em resposta às entradas do usuário – requisito tempo de *feedback*, além de ser possível entender o funcionamento do sistema.

O primeiro componente do sistema, treinamento da voz do usuário – chamado de SmlInitialTraining -, leva entre 2 minutos - treinamento mínimo -, a 8 minutos para finalizar a introdução. Porém, mais textos podem ser lidos a fim de melhorar o seu desempenho. É um sistema bastante intuitivo, mesmo para quem nunca o utilizou. A Figura 13 e a Figura 14 mostram a interface deste aplicativo.

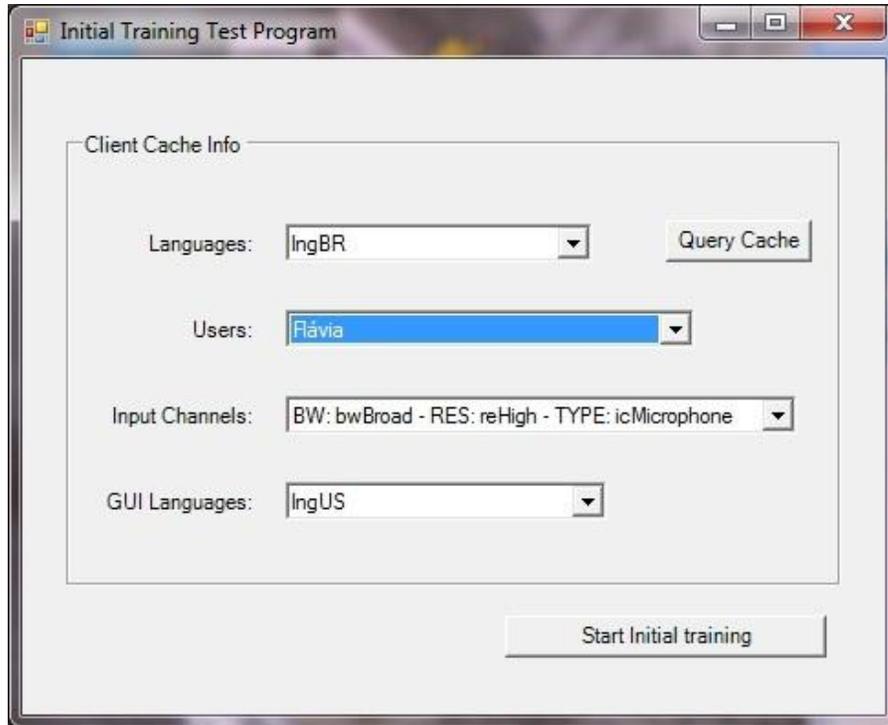


Figura 13: Tela Inicial do Aplicativo SmlInitialTraining

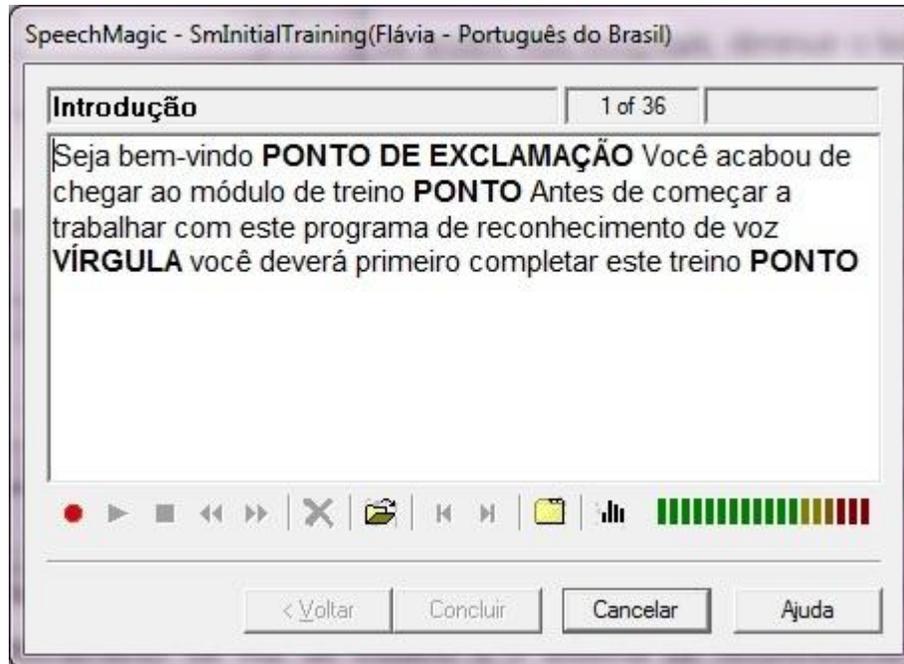


Figura 14: Tela de Treinamento do Aplicativo SmInitialTraining

É possível, neste aplicativo, selecionar a linguagem a ser utilizada, o tipo de usuário, sendo “Flávia” a única voz feminina presente nas opções, o tipo de canal de entrada e também a linguagem para as telas (GUI Languages). Porém, a interface de telas, apresenta as linguagens misturadas entre versões em inglês, português e espanhol, gerando certo desconforto cognitivo.

O segundo componente do sistema, chamado Neoc Interactive 6.1, é a interface que permite a gravação dos laudos e visualização dos textos. É uma interface bastante intuitiva, estando no idioma espanhol. A tela inicial deste componente é apresentada na Figura 15.

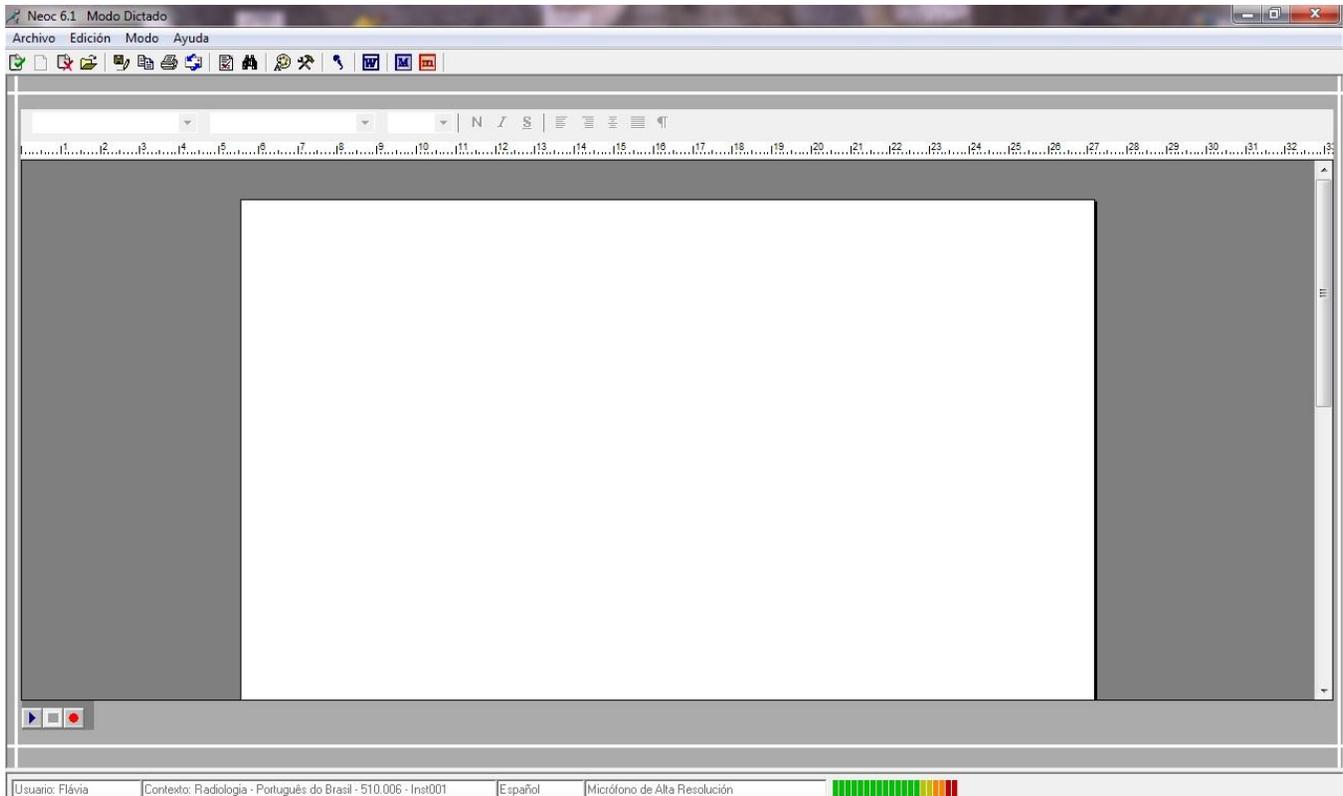


Figura 15: Tela Principal do Neoc 6.1

Segundo a documentação cedida pela empresa MacSym (MacSym, 2010), o módulo de reconhecimento de voz do SpeechMagic™ (2010) possui características como:

- O sistema é capaz de reconhecer a fala natural do usuário. Isso significa que o usuário não deva falar muito pausadamente nem muito rapidamente, tampouco alterar o timbre de sua voz.
- O sistema é capaz de se adaptar a sotaques, mesmo de pessoas estrangeiras;
- O sistema vai ignorar hesitações e interjeições que o usuário tenha, quando está pensando na próxima frase a dizer, assim como ignorará o som ambiente – ou seja, interferências sonoras.
- O sistema é capaz de colocar pontuação – ponto final e vírgula – automaticamente, caso o usuário não o faça.
- O sistema é capaz de tratar ambiguidades de números e palavras, tais como “um” e “1”, “segundo” e “2º”.

- O sistema não é capaz, por si só, de uniformizar siglas e nomes, assim “QSE” se ditado desta maneira, produzirá a sigla “QSE”, enquanto que “quadrante superior externo” produzirá o texto completo e não a sigla.
- As unidades de medida devem ser pronunciadas por extenso. Assim, “g/cm²” deve ser pronunciada por “grama por centímetro quadrado” e não “g por cm quadrado”.

É possível utilizar o NEOC Interactive 6.1 sem ter realizado o treinamento inicial da voz do usuário, porém, segundo a documentação, haverá diferença de acertos no reconhecimento da voz.

7.1.2 Sistema Philips Speech Magic 2.1

O sistema Philips Speech Magic versão 2.1 é o que está instalado no hospital foco do estudo. Embora seja uma versão bastante antiga, visto que a última versão se encontra em 7.

Sua interface é bastante parecida com a versão descrita no item anterior, como pode ser visualizada na Figura 16.

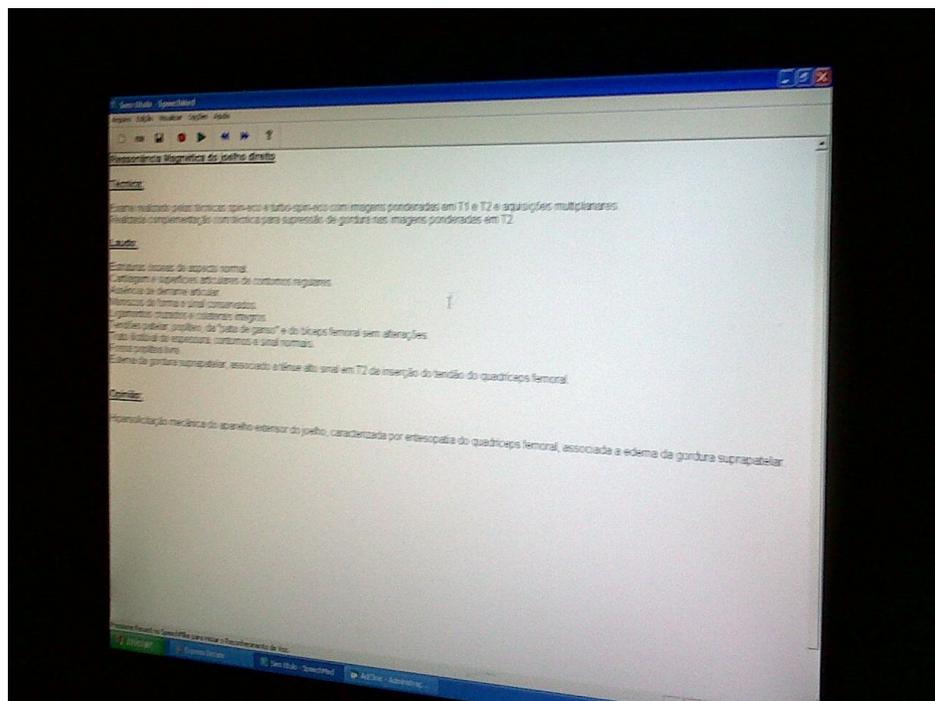


Figura 16: Tela do Sistema Philips Speech Magic 2.1

Apesar de haverem 8 médicos radiologistas e 4 licenças para o uso do sistema, apenas um médico – entrevistado – utiliza a ferramenta.

O sistema não está interligado ao PACS ou RIS, o que contribui para o aumento do tempo para o laudo ficar pronto.

O ambiente físico pode ser visualizado na Figura 17. É possível perceber que o médico necessita visualizar três telas ao mesmo tempo, e em uma delas, podem estar concentradas quatro imagens do exame em questão.



Figura 17: Ambiente no qual são realizados os laudos dos exames de Ressonância Magnética através do sistema de transcrição automática. . (Fonte: autor)

7.1.3 Planejamento dos Testes e das Inspeções de Usabilidade

A aplicação da avaliação de Usabilidade dos Sistemas de Transcrição Automática de Laudos em Radiologia ocorreu de duas maneiras:

- Analisando todos os requisitos possíveis com um sistema que está sendo utilizado *stand-alone* e não está implantado num hospital. Estes requisitos foram analisados utilizando as técnicas de questionários de satisfação, observação e inspeção de usabilidade.

- Analisando os demais requisitos que não poderiam ser analisados fora do ambiente de produção – ou seja, no hospital – com usuários reais – radiologistas – utilizando técnica de observação e questionários de satisfação.

Já para a segunda fase da avaliação, foi selecionado um usuário que utiliza o sistema com mais frequência.

A metodologia utilizada foi adaptada de Nielsen (1993).

7.2 Planejamento da Inspeção do Sistema *Stand-alone*

A inspeção da usabilidade do sistema de transcrição automática de laudos em radiologia foi realizada de acordo com as seguintes etapas: criação das heurísticas; escolha dos especialistas; preparação das inspeções; geração das listas; e concatenação dos resultados. Estas etapas serão detalhadas nas próximas subseções.

7.2.1 Criação das heurísticas

Esta seção tem por objetivo propor as heurísticas e métricas para que os especialistas verifiquem a conformidade do sistema com o que foi estabelecido. De todas as heurísticas e métricas detalhadamente descritas no Capítulo 6 deste trabalho, apenas as com metodologia de avaliação por inspeção constarão desta parte da avaliação.

Nesta seção foi elaborado cada Caso de Teste, apoiado na Metodologia proposta no Capítulo 6. Um caso de teste é descrito na literatura de maneira bastante padronizada, tais como abordado em (IEEE 829-2008; DELAMARO; MALDONADO; JINO, 2007; LEE, 2003). Porém, para casos de teste de usabilidade não foi encontrado, na literatura, nenhum *template*. Assim, a autora definiu um *template* que se adequasse a suas necessidades de padronização dos testes, conforme é possível verificar na Quadro 30.

Nome da Métrica	
Método de Avaliação	
Participantes	
Material Necessário	
Roteiro	
Resultados	
Análise	

Quadro 30: Template para Casos de Testes. (Fonte: autor)

Estas heurísticas e métricas são detalhadas a seguir, utilizando o *template* proposto.

Nome da Métrica	Avaliação da Precisão
Método de Avaliação	Inspeção de Usabilidade
Participantes	1 especialista – a autora - para o teste a e b e 2 especialistas para o teste c.
Material Necessário	10 laudos diferentes de radiologia, 1 decibelímetro, 1 equipamento SpeechMike™
Roteiro	<p>a) Através da leitura de laudos, verificar a diferença de precisão com ou sem treinamento prévio do sistema;</p> <p>b) Através da leitura de laudos, verificar a diferença de precisão para ambiente com ou sem interferência sonora;</p> <p>c) Através da leitura de um único laudo, verificar a média e o desvio padrão da taxa de reconhecimento de um especialista.</p> <p>d) Através da leitura de laudos, já com treinamento da voz, verificar a taxa de reconhecimento de especialistas.</p>
Resultados	
Análise	

Quadro 31: Caso de Teste para Avaliação da Precisão. (Fonte: autor)

Nome da Métrica	Recuperação de Erros
Método de Avaliação	Inspeção
Participantes	1 especialista
Material Necessário	1 texto com 77 palavras, com várias palavras fora do vocabulário específico da área de Radiologia, 1 equipamento SpeechMike™
Roteiro	Verificar, através de inspeção, como o sistema age perante os seguintes erros: a) O usuário usa palavras que não estão no dicionário da aplicação b) O usuário erra o laudo e deseja refazê-lo.
Resultados	
Análise	

Quadro 32: Caso de Teste para Avaliação da Recuperação de Erros do Sistema por Inspeção. (Fonte: autor)

Nome da Métrica	Interferência Sonora
Método de Avaliação	Inspeção
Participantes	1 especialista
Material Necessário	10 laudos de radiologia, 1 decibelímetro, <i>Headset</i> Philips™ SHM 3300, 1 equipamento SpeechMike™
Roteiro	a) Através da leitura de laudos, verificar a diferença de precisão para ambiente com ou sem interferência sonora utilizando o HeadSet; b) Através da leitura de laudos, verificar a diferença de precisão para ambiente com ou sem interferência sonora utilizando o SpeechMike; c) Verificar a diferença entre os dois.
Resultados	
Análise	

Quadro 33: Caso de Teste para Avaliação da Interferência Sonora. (Fonte: autor)

Nome da Métrica	Resolução de Ambiguidade para Homônimos
Método de Avaliação	Inspeção
Participantes	1 especialista
Material Necessário	10 conjuntos de palavras com homônimos, 1 equipamento SpeechMike™
Roteiro	Através da inserção de palavras homônimas num texto, verificar como o sistema reage. As palavras e frases utilizadas foram (Recanto das Letras, 2010): 1. “Acerca do paciente” e “Cerca de um milhão” 2. “O paciente apresenta Arteriosclerose” e “O paciente apresenta Aterosclerose” 3. “Câmara” e “Câmera” 4. “o Comprimento é de 1 metro” e “o Comprimento do exame” 5. “Emergir” e “Imergir” 6. “Acidente” e “Incidente” 7. “Ratificar” e “Retificar” 8. “Cerca de um laudo” e “Cerca de 1 metro” 9. “Ponto de referência” e “baixo contraste ‘.’” 10. “Característica afim” e “A fim de caracterizar”.
Resultados	
Análise	

Quadro 34: Caso de Teste para Avaliar a Resolução de Ambiguidade para Homônimos. (Fonte: autor)

Nome da Métrica	Tempo de <i>Feedback</i> do sistema
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	10 laudos, 1 cronômetro, 1 equipamento SpeechMike™
Roteiro	Através da leitura de um laudo, verificar o <i>delay</i> do sistema para mostrar a palavra ou frase no texto.
Resultados	
Análise	

Quadro 35: Caso de Teste para Avaliar Tempo de Resposta do Sistema. (Fonte: autor)

Nome da Métrica	Qualidade das entradas de áudio
Método de Avaliação	Inspeção
Participantes	1 especialista
Material Necessário	10 laudos de radiologia, 1 decibelímetro, Headset Philips TM SHM 3300, 1 equipamento SpeechMike™
Roteiro	Através da leitura de laudos, verificar a diferença de precisão para os dois equipamentos de entrada de áudio, com baixa interferência sonora.
Resultados	
Análise	

Quadro 36: Caso de Teste para Avaliar a Qualidade das Entradas de Áudio. (Fonte: autor)

Nome da Métrica	Qualidade dos sistemas de áudio
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	2 laudos, 1 decibelímetro, Headset Philips™ SHM 3300, 1 equipamento SpeechMike™
Roteiro	Analisar o ruído do áudio captado através de uma aplicação ou através de investigação informal.
Resultados	
Análise	

Quadro 37: Caso de Teste para Avaliar a Qualidade do Sistema de Áudio. (Fonte: autor)

Nome da Métrica	Visibilidade do Sistema e Adequação do <i>Feedback</i>
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	10 laudos, 1 equipamento SpeechMike™
Roteiro	<p>Verificar:</p> <p>A aplicação informa o que está acontecendo ao usuário; por exemplo, se uma palavra não foi reconhecida durante o ditado, ou se o sistema está apresentando dificuldade no reconhecimento – ou seja, baixo reconhecimento. Também, se há <i>delay</i> no sistema, seria conveniente informar isso ao usuário.</p>
Resultados	
Análise	

Quadro 38: Caso de Teste para Avaliar a Visibilidade do Sistema por Inspeção. (Fonte: autor)

Nome da Métrica	Adequação do Sistema de Ajuda
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	Sistema Neoc.
Roteiro	<p>Verificar se o sistema ajuda ao usuário a resolver suas dúvidas.</p> <p>Se possui adequação do vocabulário em relação ao conhecido pelo usuário.</p> <p>Se possui um sistema de busca por palavras-chave que facilite que o usuário encontre o desejado.</p>
Resultados	
Análise	

Quadro 39: Caso de Teste para Avaliar a Adequação do Sistema de Ajuda por Inspeção. (Fonte: autor)

Nome da Métrica	Modalidade Apropriada
Método de Avaliação	Inspeção
Participantes	1 especialista
Material Necessário	1 laudo, <i>Headset Philips</i> ™ SHM 3300, 1 equipamento <i>SpeechMike</i> ™
Roteiro	Analisar as facilidades de uso dos dois equipamentos.
Resultados	
Análise	

Quadro 40: Caso de Teste para Avaliar Modalidade Apropriada. (Fonte: autor)

Nome da Métrica	Naturalidade da Fala do Usuário
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	10 laudos de radiologia, 1 equipamento <i>SpeechMike</i> ™
Roteiro	Através de ditado de laudos: <ul style="list-style-type: none"> a) Verificar o que acontece se houver interjeições do usuário, tais como “éeee”, “hummm”, “aaaa”; b) Verificar o que acontece se o usuário não pontuar explicitamente o texto.
Resultados	
Análise	

Quadro 41: Caso de Teste para Avaliar a Naturalidade da Fala do Usuário por Inspeção. (Fonte: autor)

Nome da Métrica	Tamanho do Vocabulário
Método de Avaliação	Investigação
Participantes	1 especialista
Material Necessário	Questionário ao representante
Roteiro	Não aplicável
Resultados	
Análise	

Quadro 42: Obtenção do Tamanho do Vocabulário. (Fonte: autor)

7.2.2. Escolha dos Especialistas

Embora Nielsen (1993) estabeleça que uma inspeção deva ter entre 3 e 5 especialistas, foram utilizadas duas especialistas em usabilidade para esta inspeção. As duas especialistas encontram-se em fase de doutoramento na área de Interfaces Homem-Computador, são do sexo feminino, com experiência acadêmica e em empresa, com mais de 30 anos de idade.

7.2.3. Preparação das inspeções

Segundo Nielsen (1993), uma inspeção de usabilidade dura cerca de 2 horas e deve ser realizada individualmente. Porém, é possível perceber, através dos resultados dos testes, na próxima seção, que estas inspeções demoraram bastante tempo, cerca de 20 horas. Esta foi uma das principais razões do não uso da 2ª especialistas em todas as inspeções e também da não utilização de mais de dois especialistas.

Os materiais e softwares necessários para esta inspeção foram:

- Laptop Philips: Processador Core 2 Duo, Memória RAM de 4GB, HD de 320 GB.
- SpeechMike Pro™ Philips: utilizado para as entradas de áudio, tal como apresentado na Figura 18.
- Headset Philips™ SHM 3300.
- Decibelímetro: utilizado para medir a interferência Sonora do ambiente de teste, conforme mostrado na Figura 19.
- Cronômetro: utilizado para medir o tempo dos testes e tempo de resposta do sistema.
- 10 Laudos de Radiologia do Tórax. Os laudos utilizados constam do Anexo B.
- Software SpeechMagic™ (2010)⁶ da empresa Nuance (2010).
- Software SmlInitialTraining: leva cerca de 2 minutos para ser completada a versão suficiente para executar.

⁶ Esta licença foi gentilmente cedida pela empresa MacSym Tecnologia Médica (Macsym, 2010) por um prazo de 60 dias

- Software Neoc Interactive 6.1: pode ser utilizado sem a execução prévia do SmlInitialTraining, porém apresentará uma precisão menor.
- Casos de Testes descritos na seção anterior.



Figura 18: SpeechMike Pro - Dispositivo de Entrada de Áudio. (Fonte: autor)



Figura 19: Decibelímetro Utilizado nos Testes. (Fonte: autor)

7.2.4. Geração das Listas e Concatenação dos Resultados

Esta subseção tem por objetivo gerar os resultados de cada caso de teste determinado para inspeção, assim como a análise dos resultados.

Os dez laudos utilizados nessa inspeção são padronizados; muitos termos aparecem de maneira repetitiva nos textos, conforme pode ser visto no Anexo B.

A fim de que um teste de inspeção não influenciasse os demais, realizados na sequência, os especialistas não olhavam os problemas de reconhecimento de voz antes de finalizarem aquela bateria de testes.

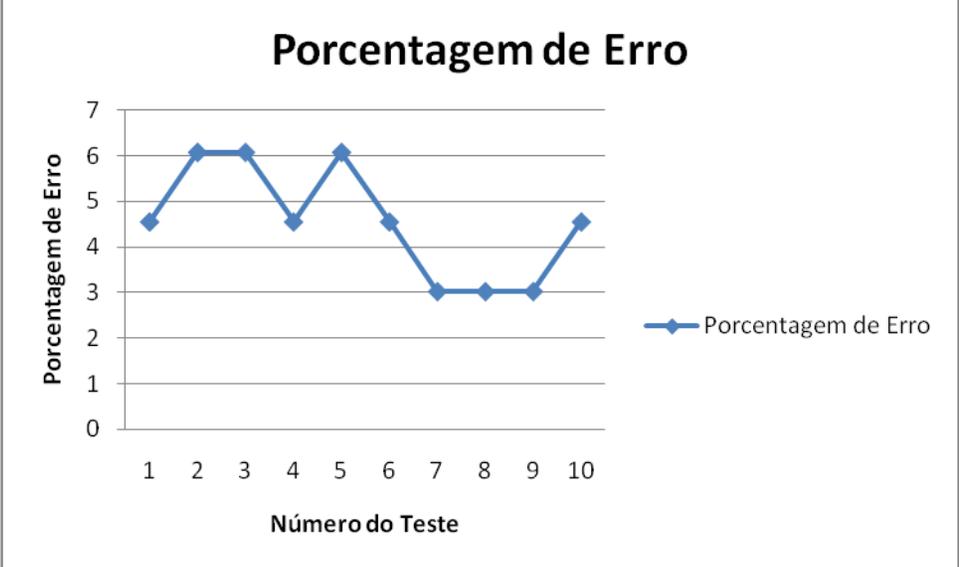
A porcentagem de erros de reconhecimento de voz foi calculada pela diferença de sentido entre palavras entre o texto original - laudo escrito - e o texto gerado pelo sistema de reconhecimento. Para tanto foi utilizada a função de comparação do editor de texto WordTM e também a inspeção manual da especialista.

Os textos têm uma média de 56.4 palavras por laudo, sendo que o menor laudo tem 52 palavras e o maior, 66.

Os testes foram realizados no mesmo ambiente, com interferência sonora equivalente.

A seguir são apresentados os resultados dos testes de inspeção realizados.

Nome da Métrica	Avaliação da Precisão
Método de Avaliação	Inspeção de Usabilidade
Participantes	1 especialista – a autora - para o teste a e b e 2 especialistas para o teste c.
Material Necessário	10 laudos diferentes de radiologia, 1 decibelímetro, 1 equipamento SpeechMike TM
Roteiro	a) Através da leitura de laudos, verificar a diferença de precisão com ou sem treinamento prévio do sistema; b) Através da leitura de laudos, verificar a diferença de precisão para ambiente com ou sem interferência sonora;

	<p>c) Através da leitura de um único laudo, verificar a média e o desvio padrão da taxa de reconhecimento de um especialista.</p> <p>d) Através da leitura de laudos, já com treinamento da voz, verificar a taxa de reconhecimento de especialistas.</p>																						
Resultados	<p>A fim de poder validar os testes aqui determinados, um teste anterior para medir a diferença de precisão no reconhecimento de voz com um especialista por 10 vezes com o mesmo laudo foi realizado.</p> <p>a) O texto referente ao laudo utilizado neste teste tem 66 palavras. O teste foi repetido 10 vezes. Este teste demorou cerca de 21 minutos.</p> <div data-bbox="496 800 1455 1367"><table border="1"><thead><tr><th>Número do Teste</th><th>Porcentagem de Erro</th></tr></thead><tbody><tr><td>1</td><td>4,5</td></tr><tr><td>2</td><td>6,0</td></tr><tr><td>3</td><td>6,0</td></tr><tr><td>4</td><td>4,5</td></tr><tr><td>5</td><td>6,0</td></tr><tr><td>6</td><td>4,5</td></tr><tr><td>7</td><td>3,0</td></tr><tr><td>8</td><td>3,0</td></tr><tr><td>9</td><td>3,0</td></tr><tr><td>10</td><td>4,5</td></tr></tbody></table></div> <p>Gráfico 1: Porcentagem de Erro de Teste de um Laudo</p> <p>b) Os resultados da precisão sem treinamento e com treinamento da voz estão a seguir. Este teste demorou cerca de 40 minutos.</p>	Número do Teste	Porcentagem de Erro	1	4,5	2	6,0	3	6,0	4	4,5	5	6,0	6	4,5	7	3,0	8	3,0	9	3,0	10	4,5
Número do Teste	Porcentagem de Erro																						
1	4,5																						
2	6,0																						
3	6,0																						
4	4,5																						
5	6,0																						
6	4,5																						
7	3,0																						
8	3,0																						
9	3,0																						
10	4,5																						

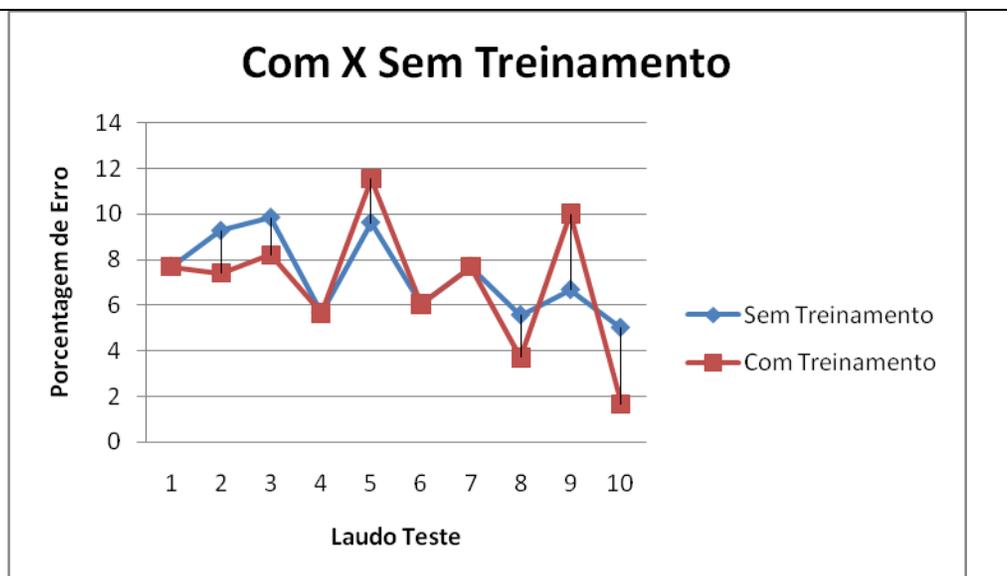


Gráfico 2: Porcentagem de Erro do Sistema Com x Sem Treinamento Prévio

b) Os resultados dos testes com Interferência sonora alta entre 75 e 81 db e baixa entre 40 e 45 db podem ser visualizados no gráfico abaixo. Estes testes levaram cerca de 30 minutos.

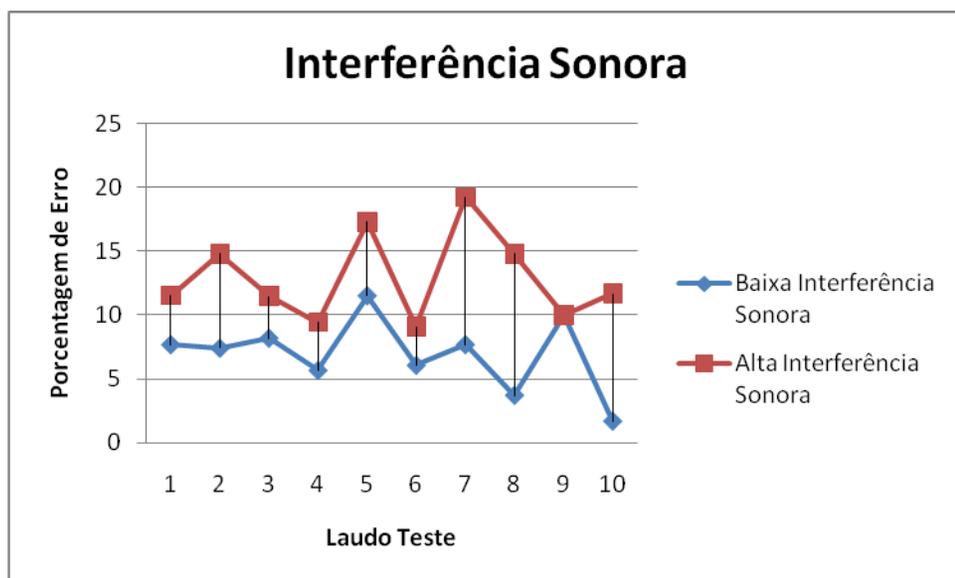


Gráfico 3: Porcentagem de Erro do Sistema Com ou Sem Interferência Sonora

d) Os resultados dos testes da precisão do reconhecimento de voz com treinamento prévio da voz, realizadas por duas especialistas estão a seguir. Estes testes levaram cerca de 50 minutos.

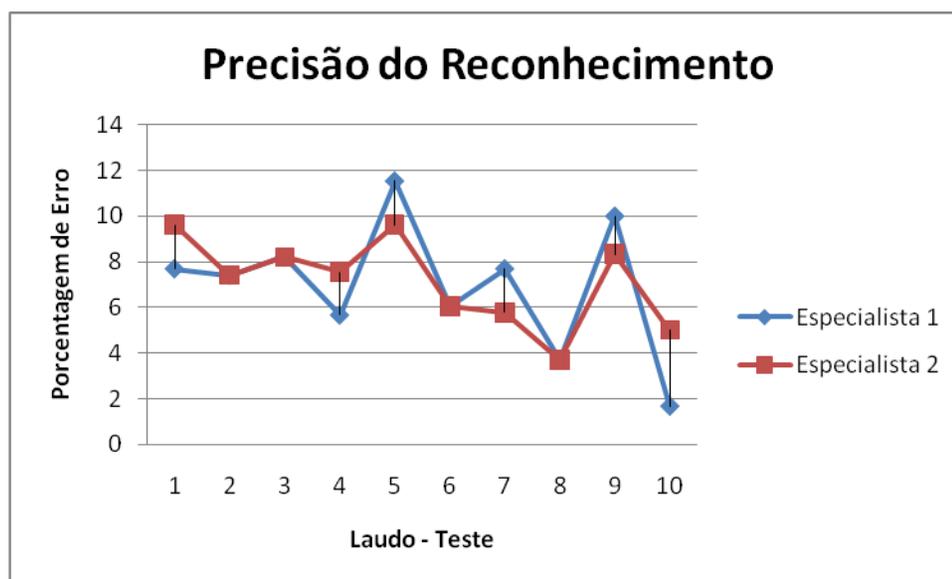


Gráfico 4: Diferença da porcentagem de Erro do Sistema entre Dois Especialistas

Análise

a) A média de erro entre todos os 10 testes foi de 4.54%, com desvio padrão de 1.237%, o que indica que houve uma certa homogeneidade nos dados, analisados a partir do Gráfico 1 Porcentagem de Erro.

b) No Gráfico 2 representando a diferença da taxa de reconhecimento da voz com ou sem treinamento prévio da voz é possível perceber que não houve diferença significativa na taxa de reconhecimento sem o treinamento e com o treinamento, como era pensado. No caso dos testes sem treinamento prévio, verifica-se que a média foi de 7.3% de erros por laudo, enquanto o desvio padrão foi de 1.79%. Já para os testes com treinamento prévio da voz, a média de erros foi de 6.96%, enquanto o desvio padrão foi de 2.87%. O sistema já se mostrou bastante eficiente sem treinamento, com taxa de reconhecimento de 92.7% em média.

c) No Gráfico 3 referente à Interferência Sonora, é possível verificar uma piora significativa na taxa de reconhecimento proporcionada pela

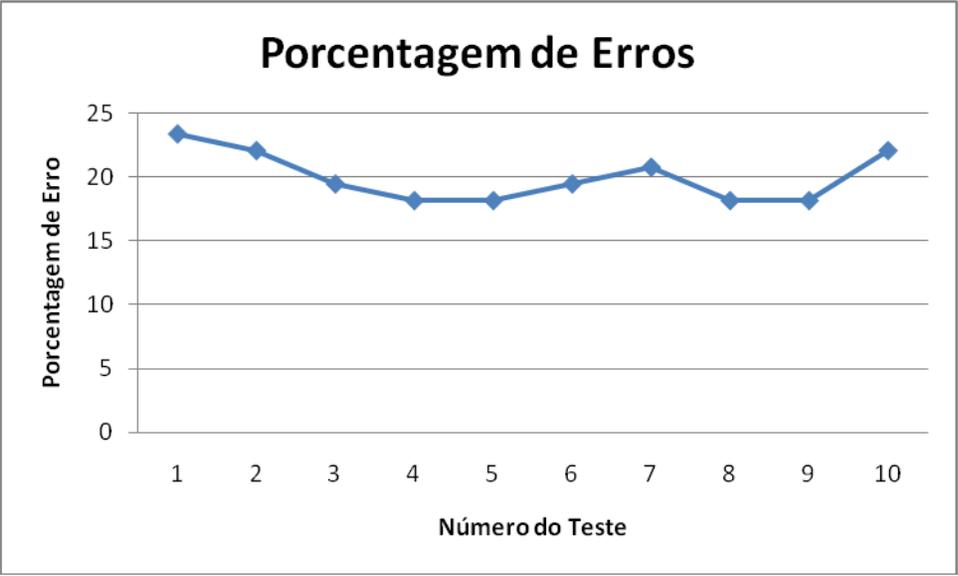
interferência sonora do ambiente – ruído. Com interferência sonora considerada baixa (vide Anexo A), entre 40 e 45 db, a média de erros de reconhecimento do sistema ficou em 6.96%, com desvio padrão de 2.87%. Já para um ambiente considerado com alta interferência sonora, entre 75 e 81 db, a taxa de erro de reconhecimento foi, em média, de 12.94%, com desvio padrão de 3.4%. Isso significa que o ambiente interfere significativamente na qualidade do reconhecimento e deve ser levado em consideração quando for realizada a implantação do sistema.

d) O gráfico 4 da Precisão do Reconhecimento mostra que os erros no reconhecimento foram, com uma exceção, sempre menores que 10%, com média de 6.96% de erros por laudo, com desvio padrão de 2.87%. Já os erros de reconhecimento de voz foram para a Especialista 2 sempre menores do que 10%, sendo que a média foi de 7.12% e o desvio padrão de 1.95%.

De acordo com o fabricante do motor de reconhecimento (NUANCE, 2010), reconhecimento de voz fica em torno de 90% inicialmente, podendo chegar a 98%. Assim, se utilizado um software de reconhecimento de voz com aprendizado, estas taxas, já baixas, podem ser decrementadas consideravelmente. Estes testes foram realizados em software sem a possibilidade de inserção de palavras novas ao vocabulário.

A palavra “quilovoltagem” não pertence ao vocabulário inicial da aplicação e por isso foi contabilizada como erro de reconhecimento de voz nos 10 testes.

Quadro 43: Resultados e Análise do Caso de Teste para Avaliação da Precisão. (Fonte: autor)

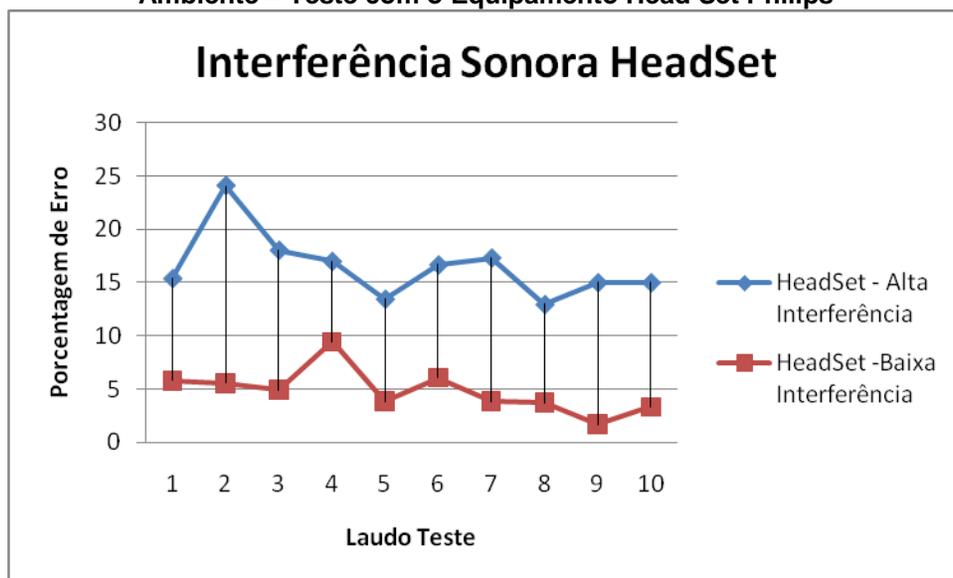
Nome da Métrica	Recuperação de Erros																						
Método de Avaliação	Inspeção																						
Participantes	1 especialista																						
Material Necessário	1 texto com 77 palavras, com várias palavras fora do vocabulário específico da área de Radiologia, 1 equipamento SpeechMike™																						
Roteiro	<p>Verificar, através de inspeção, como o sistema age perante os seguintes erros:</p> <p>a) O usuário usa palavras que não estão no dicionário da aplicação</p> <p>b) O usuário erra o laudo e deseja refazê-lo.</p>																						
Resultados	<p>a) O texto referente ao laudo utilizado neste teste tem 77 palavras. O teste foi repetido 10 vezes. Este teste demorou cerca de 17 minutos.</p> <p>Gráfico 5: Resultado da Porcentagem de Erro em 10 Testes com o Mesmo Laudo</p>  <table border="1"> <caption>Dados do Gráfico 5: Resultado da Porcentagem de Erro em 10 Testes com o Mesmo Laudo</caption> <thead> <tr> <th>Número do Teste</th> <th>Porcentagem de Erro</th> </tr> </thead> <tbody> <tr><td>1</td><td>23</td></tr> <tr><td>2</td><td>22</td></tr> <tr><td>3</td><td>19</td></tr> <tr><td>4</td><td>18</td></tr> <tr><td>5</td><td>18</td></tr> <tr><td>6</td><td>19</td></tr> <tr><td>7</td><td>21</td></tr> <tr><td>8</td><td>18</td></tr> <tr><td>9</td><td>18</td></tr> <tr><td>10</td><td>22</td></tr> </tbody> </table>	Número do Teste	Porcentagem de Erro	1	23	2	22	3	19	4	18	5	18	6	19	7	21	8	18	9	18	10	22
Número do Teste	Porcentagem de Erro																						
1	23																						
2	22																						
3	19																						
4	18																						
5	18																						
6	19																						
7	21																						
8	18																						
9	18																						
10	22																						
Análise	<p>a) É possível, através de comparação entre o Gráfico 1 e o Gráfico 5, verificar que houve uma piora significativa na porcentagem de erros quando se utiliza um texto com palavras fora do vocabulário</p>																						

	<p>específico de Radiologia. Enquanto, no primeiro caso, a média era de 4.54%, com desvio médio de 1.237%, no segundo teste foi de 20%, com desvio médio de 1.95%. Isso era esperado, visto que o vocabulário para o sistema é específico para a área de Radiologia.</p> <p>b) Não há atalhos de voz para corrigir um erro, o que pode causar muito incômodo ao usuário que, a princípio, pode se sentir perdido em como desfazer um ditado errado. Segundo o fornecedor da Neoc (NUNCE, 2010), o usuário deve terminar o ditado e depois corrigir, no texto, as palavras que estiverem erradas.</p>
--	--

Quadro 44: Resultados e Análise do Caso de Teste para Avaliação da Recuperação de Erros do Sistema por Inspeção. (Fonte: autor)

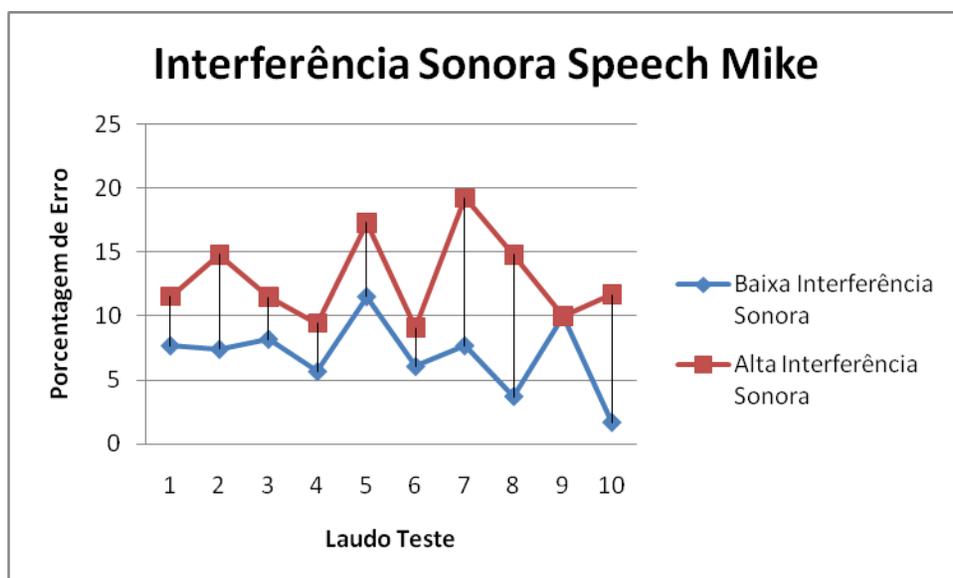
Nome da Métrica	Interferência Sonora
Método de Avaliação	Inspeção
Participantes	1 especialista
Material Necessário	10 laudos de radiologia, 1 decibelímetro, <i>Headset Philips</i> ™ SHM 3300, 1 equipamento <i>SpeechMike</i> ™
Roteiro	<p>a) Através da leitura de laudos, verificar a diferença de precisão para ambiente com ou sem interferência sonora utilizando o HeadSet;</p> <p>b) Através da leitura de laudos, verificar a diferença de precisão para ambiente com ou sem interferência sonora utilizando o SpeechMike;</p> <p>c) Verificar a diferença entre os dois.</p>
Resultados	<p>Cada teste durou cerca de 18 minutos, totalizando 36 minutos.</p> <p>a) Teste com baixa e alta interferência com o equipamento Head Set Philips.</p>

Gráfico 6: Porcentagem de Erro causado por Interferência Sonora no Ambiente – Teste com o Equipamento Head Set Philips



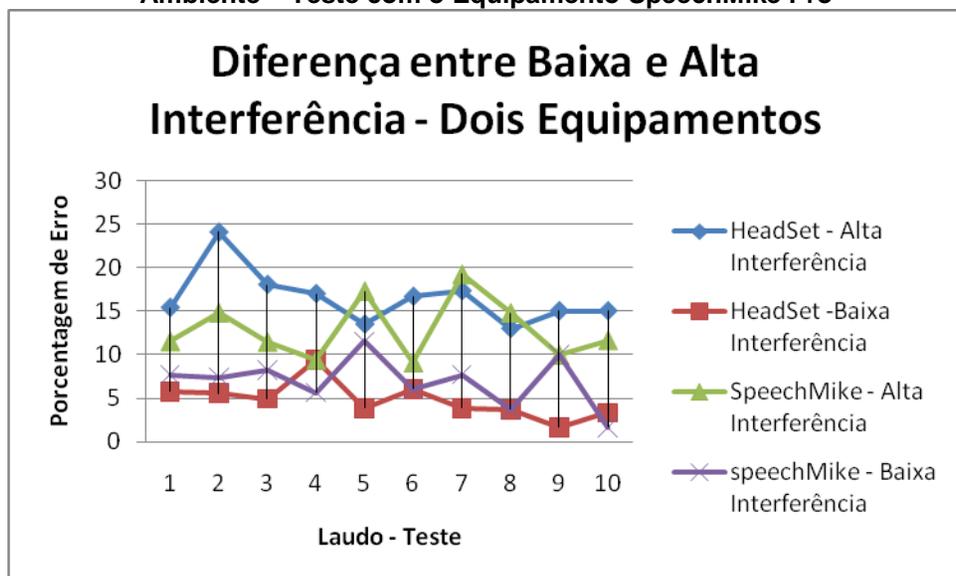
b) Teste com baixa e alta interferência com o equipamento SpeechMike.

Gráfico 7: Porcentagem de Erro causado por Interferência Sonora no Ambiente – Teste com o Equipamento SpeechMike Pro



c) Diferença entre os dois equipamentos, com baixa e alta interferência sonora.

Gráfico 8: Porcentagem de Erro causado por Interferência Sonora no Ambiente – Teste com o Equipamento SpeechMike Pro



Análise

a) No gráfico referente à Interferência Sonora captada pelo equipamento HeadSet, é possível verificar uma piora significativa na taxa de reconhecimento proporcionada pela interferência sonora do ambiente – ruído. Com interferência sonora considerada baixa (ABNT – NBR 10152/1987, 2010) (vide Anexo A), entre 40 e 45 db, a média de erros de reconhecimento do sistema ficou em 4.81%, com desvio padrão de 2.01%. Já para um ambiente considerado com alta interferência sonora, entre 75 e 81 db, a taxa de erro de reconhecimento foi, em média, de 16.48%, com desvio padrão de 3.12%. Isso significa que o ambiente interferiu significativamente na qualidade do reconhecimento e deve ser levado em consideração quando for realizada a implantação do sistema.

b) No gráfico referente à Interferência Sonora do equipamento SpeechMike, é possível verificar uma piora significativa na taxa de reconhecimento proporcionada pela interferência sonora do

	<p>ambiente – ruído. Com interferência sonora entre 40 e 45 db, a média de erros de reconhecimento do sistema ficou em 6.96%, com desvio padrão de 2.87%. Já para um ambiente considerado com alta interferência sonora, entre 75 e 81 db, a taxa de erro de reconhecimento foi, em média, de 12.94%, com desvio padrão de 3.4%. Isso significa que o ambiente interfere significativamente na qualidade do reconhecimento e deve ser levado em consideração quando for realizada a implantação do sistema.</p> <p>c) Comparando os dois equipamentos - Gráfico 8 - é possível verificar que o equipamento HeadSet se comporta melhor do que o SpeechMike para ambientes com pouca interferência sonora, mas este quadro se inverte quando há muita interferência sonora.</p>
--	---

Quadro 45: Resultados e Análise do Caso de Teste para Avaliar a Interferência Sonora. (Fonte: autor)

Nome da Métrica	Resolução de Ambiguidade para Homônimos
Método de Avaliação	Inspeção
Participantes	1 especialista
Material Necessário	10 conjuntos de palavras com homônimos, 1 equipamento SpeechMike™
Roteiro	<p>Através da inserção de palavras homônimas num texto, verificar como o sistema reage.</p> <p>As palavras e frases utilizadas foram (Recanto das Letras, 2010):</p> <ol style="list-style-type: none"> 1. “Acerca do paciente” e “Cerca de um milhão” 2. “O paciente apresenta Arteriosclerose” e “O paciente apresenta Aterosclerose” 3. “Câmara” e “Câmera” 4. “o Comprimento é de 1 metro” e “o Cumprimento do exame” 5. “Emergir” e “Imergir” 6. “Acidente” e “Incidente”

	<ol style="list-style-type: none"> 7. “Ratificar” e “Retificar” 8. “Cerca de um laudo” e “Cerca de 1 metro” 9. “Ponto de referência” e “baixo contraste ‘.’” 10. “Característica afim” e “A fim de caracterizar”.
Resultados	<p>Este teste demorou cerca de 5 minutos.</p> <p>As respostas que o sistema deu às entradas mostradas acima foram:</p> <ol style="list-style-type: none"> 1. “Acerca do paciente.” e “Cerca de 1 milhão” 2. “o paciente apresenta arteriosclerose” e “A paciente apresenta a aterosclerose” 3. ”Câmara” e “Câmera” 4. “O comprimento é de 1 m” e “o comprimento do exame” 5. “Emergir” e “Emergir” 6. “Acidente” e “Incidente” 7. “Ratificar” e “Retificar” 8. “Cerca de o laudo” e “Cerca de 1m” 9. “Ponto de referência” e “baixo contraste ‘.’” 10. “Característica afim” e “A fim de caracterizar”.
Análise	O sistema foi capaz de distinguir entre os homônimos em 70% dos casos estudados, falhando nos itens 4, 5 e 8.

Quadro 46: Resultados e Análise do Caso de Teste para Avaliar a Resolução de Ambiguidade para Homônimos. (Fonte: autor)

Nome da Métrica	Tempo de <i>Feedback</i> do Sistema
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	10 laudos, 1 cronômetro, 1 equipamento SpeechMike™
Roteiro	Através da leitura dos laudos, verificar o <i>delay</i> do sistema para mostrar a palavra ou frase no texto.
Resultados	Os testes foram realizados no mesmo ambiente, com interferência sonora equivalente.

	<p>Especialista 1: o <i>delay</i> chega a ser de 5 segundos, com média de 4.5 segundos.</p> <p>Especialista 2: o <i>delay</i> chega a ser de 4.8 segundos, com média de 4.2 segundos.</p>
Análise	Devido à diferença da velocidade de fala de cada especialista, é possível ter um <i>delay</i> maior um menor do sistema.

Quadro 47: Resultados e Análise do Caso de Teste para Avaliar Tempo de Resposta do Sistema. (Fonte: autor)

Nome da Métrica	Qualidade das entradas de áudio
Método de Avaliação	Inspeção
Participantes	1 especialista
Material Necessário	10 laudos de radiologia, 1 decibelímetro, Headset Philips TM SHM 3300, 1 equipamento SpeechMike™
Roteiro	Através da leitura de laudos, verificar a diferença de precisão para os dois equipamentos de entrada de áudio, com baixa interferência sonora.
Resultados	Os testes foram realizados mesmo ambiente para os dois equipamentos de entrada de áudio, com interferência sonora variando entre 42 e 50 db.

Gráfico 9: Diferença da Porcentagem de Erro entre os dois Equipamentos

	<h3 style="text-align: center;">Diferença de Exatidão entre Equipamentos</h3> <table border="1" style="margin-left: auto; margin-right: auto;"> <caption>Dados do Gráfico: Porcentagem de Erro por Laudo Teste</caption> <thead> <tr> <th>Laudo Teste</th> <th>Speech Mike (%)</th> <th>HeadSet (%)</th> </tr> </thead> <tbody> <tr><td>1</td><td>7.5</td><td>5.5</td></tr> <tr><td>2</td><td>7.0</td><td>5.5</td></tr> <tr><td>3</td><td>7.5</td><td>4.5</td></tr> <tr><td>4</td><td>5.5</td><td>9.5</td></tr> <tr><td>5</td><td>11.5</td><td>4.0</td></tr> <tr><td>6</td><td>6.0</td><td>6.0</td></tr> <tr><td>7</td><td>7.5</td><td>4.0</td></tr> <tr><td>8</td><td>4.0</td><td>4.0</td></tr> <tr><td>9</td><td>10.0</td><td>1.5</td></tr> <tr><td>10</td><td>1.5</td><td>3.5</td></tr> </tbody> </table>	Laudo Teste	Speech Mike (%)	HeadSet (%)	1	7.5	5.5	2	7.0	5.5	3	7.5	4.5	4	5.5	9.5	5	11.5	4.0	6	6.0	6.0	7	7.5	4.0	8	4.0	4.0	9	10.0	1.5	10	1.5	3.5
Laudo Teste	Speech Mike (%)	HeadSet (%)																																
1	7.5	5.5																																
2	7.0	5.5																																
3	7.5	4.5																																
4	5.5	9.5																																
5	11.5	4.0																																
6	6.0	6.0																																
7	7.5	4.0																																
8	4.0	4.0																																
9	10.0	1.5																																
10	1.5	3.5																																
Análise	<p>Através do gráfico é possível perceber que, ao contrário do que afirmado pelo revendedor do SpeechMike, o Headset teve um comportamento melhor, isto é, menor porcentagem de erro na maioria dos casos, com exceção do teste número 4.</p> <p>A média da porcentagem de erro para o HeadSet é de 2.87%, com desvio padrão de 2.09%. Já a média da porcentagem de erro para o SpeechMike está em 6.96%, com desvio padrão de 4.81%. Isso mostra que, com o uso do SpeechMike, além de apresentar menor taxa de reconhecimento, ainda apresenta maior dispersão dos dados.</p>																																	

Quadro 48: Resultados e Análise do Caso de Teste para Avaliar a Qualidade das Entradas de Áudio.
(Fonte: autor)

Nome da Métrica	Qualidade dos sistemas de áudio
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	2 laudos, <i>Headset Philips</i> ™ SHM 3300, 1 equipamento <i>SpeechMike</i> ™
Roteiro	Analisar o ruído do áudio captado através de uma aplicação ou através de avaliação informal.

Resultados	Os resultados foram retirados de avaliação informal com os avaliadores do sistema, fazendo com que o áudio gravado fosse reproduzido. O caráter de inteligibilidade na reprodução dos textos gravados foi mantido.
Análise	Os resultados foram satisfatórios

Quadro 49: Resultados e Análise do Caso de Teste para Avaliar a Qualidade do Sistema de Áudio. (Fonte: autor)

Nome da Métrica	Visibilidade do Sistema e Adequação do <i>Feedback</i>
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	10 laudos, 1 equipamento SpeechMike™
Roteiro	<p>Verificar:</p> <p>A aplicação informa o que está acontecendo ao usuário; por exemplo, se uma palavra não foi reconhecida durante o ditado, ou se o sistema está apresentando dificuldade no reconhecimento – ou seja, baixo reconhecimento. Também, se há <i>delay</i> no sistema, seria conveniente informar isso ao usuário.</p>
Resultados	<p>Lista de problemas encontrados pelos dois especialistas:</p> <ul style="list-style-type: none"> - O sistema não informa quando o reconhecimento da fala é baixo, inserindo palavras que poderiam acontecer naquele contexto. - O sistema não informa quando palavras novas são ditas, inserindo palavras que poderiam acontecer naquele contexto. - O sistema não informa que o sistema está <i>delay</i> alto. Fazendo com que o usuário pense que realizou alguma entrada com problema ou que o sistema não está funcionando.

Análise	<p>Melhorias possíveis:</p> <ul style="list-style-type: none"> - Quando o reconhecimento for abaixo de certo percentual, grifar ou indicar de outra forma a palavra que foi inserida a fim de que o usuário possa visualizar mais facilmente o possível erro de reconhecimento. - Inserir um mecanismo de visualização do <i>delay</i> tal como uma ampulheta.
---------	--

Quadro 50: Resultados e Análise do Caso de Teste para Avaliar a Visibilidade do Sistema por Inspeção.
(Fonte: autor)

Nome da Métrica	Adequação do Sistema de Ajuda
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	Sistema Neoc.
Roteiro	<p>Verificar se o sistema ajuda ao usuário a resolver suas dúvidas.</p> <p>Se possui adequação do vocabulário em relação ao conhecido pelo usuário.</p> <p>Se possui um sistema de busca por palavras-chave que facilite que o usuário encontre o desejado.</p>
Resultados	Os testes demoraram cerca de 15 minutos.

Análise	<p>a) O sistema de ajuda está em espanhol, o que dificulta o entendimento para usuários que não tenham proficiência na língua;</p> <p>b) O sistema de ajuda não possui sistema de busca remissiva ou qualquer outra forma de busca do problema;</p> <p>c) O sistema de ajuda é composto apenas por três partes: truques, Web e Sobre o Neoc. Para ser um sistema de ajuda, estas três partes são insuficientes;</p> <p>d) O componente “truques” é bastante superficial;</p> <p>e) O componente “Web” que chama a página http://www.speechrecognition.philips.com/ não consegue fazer a conexão;</p> <p>f) O componente “Neoc” possui níveis superficiais e aprofundados do uso do sistema, tal como abrir um documento e integrar o sistema com sistemas externos.</p>
---------	---

Quadro 51: Resultados e Análise do Caso de Teste para Avaliar a Adequação do sistema de ajuda por Inspeção. (Fonte: autor)

Nome da Métrica	Modalidade Apropriada
Método de Avaliação	Inspeção
Participantes	1 especialista
Material Necessário	1 laudo, <i>Headset Philips</i> ™ SHM 3300, 1 equipamento <i>SpeechMike</i> ™
Roteiro	Analisar as facilidades de uso dos dois equipamentos.
Resultados	Não aplicável

Análise	<p>O equipamento SpeechMike possui uma ergonomia mais favorável para a gravação de laudos: botão de “Record” acessível desde a mão. Já para o HeadSet é necessário ativar e desativar a gravação no próprio sistema de transcrição de laudos.</p> <p>O botão “Record” no equipamento SpeechMike dá uma maior liberdade ao usuário para parar e iniciar uma gravação, evitando interferências no laudo quando o usuário precise sair da aplicação.</p> <p>Além disso, o equipamento SpeechMike possui mouse acoplado e botões de volume, o que facilita também o acesso a outras aplicações do <i>desktop</i> ou mesmo o acesso a partes do texto digitado e botões do menu, sem ocupar a outra mão do usuário.</p> <p>Por outro lado, a diferença de custos entre os dois equipamentos é bastante importante. O equipamento SpeechMike custa cerca de R\$ 1.200,00 e o HeadSet Philips custa cerca de R\$ 35,00.</p>
---------	--

Quadro 52: Resultados e Análise do Caso de Teste para Avaliar Modalidade Apropriada. (Fonte: autor)

Nome da Métrica	Naturalidade da Fala do Usuário
Método de Avaliação	Inspeção
Participantes	2 especialistas
Material Necessário	10 laudos de radiologia, 1 equipamento SpeechMike TM
Roteiro	<p>Através de ditado de laudos:</p> <ul style="list-style-type: none"> a) Verificar o que acontece se houver interjeições do usuário, tais como “éeee”, “hummm”, “aaaa”; b) Verificar o que acontece se o usuário não pontuar explicitamente o texto.

Resultados	<p>Os testes levaram cerca de 40 minutos, para cada especialista.</p> <p>a) Os resultados da fala natural, em que é bastante usual a inserção do usuário de interjeições, está a seguir (texto não lido).</p> <p>Técnica de exame realizada equipamento de alto potência com técnica de alta quilovolt falange 21 I fornecem escala de contraste longa de laudo de estruturas ósseas da e tecidos moles da parede torácica sem alterações do vascularização pulmonar normal meia, de configuração e topografia e dimensões normais na aorta alongada e 0 Marco mede normal do parênquima pulmonar de transparência normal do</p> <p>b) Os resultados dos testes realizados foram semelhantes ao apresentado a seguir.</p> <p>Texto Original:</p> <p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste). Laudo: Estruturas ósseas e tecidos moles da parede torácica, sem alterações. Seios e cúpulas diafragmáticas livres. Vascularização pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca aumentada. Aorta alongada. Arco médio normal. Parênquima pulmonar de transparência normal.</p> <p>Texto Transcrito:</p> <p>técnica exame realizado em equipamento de alta potencia com técnica de alta quilovolt imagem de fornece escala de contraste longa mais de contraste laudo estruturas ósseas e tecidos moles da parede torácica sem alterações no vascularização pulmonar normal, de configuração e topografia e dimensões normais e imagem cardíaca aumentada. Aorta alongada e do arco médio normal do parênquima pulmonar de transparência normal</p>
------------	--

Análise	<p>a) As interjeições comumente usadas quando se está ditando frases criadas naquele momento foram transformadas em palavras sem sentido dentro do texto – grifadas em negrito. Ao contrário do que está especificado no manual, as interjeições não são ignoradas pelo sistema.</p> <p>b) Através desta inspeção, foi observado que o sistema, ao contrário do que indica o manual de utilização, não é capaz, quase nunca, de inserir pontuação adequada, seguindo as pausas que o usuário faz no texto. Portanto, se o usuário não fizer a pontuação explícita do texto quando dita, terá que fazê-lo depois no próprio texto transcrito. Ter que inserir pontuação explícita no ditado torna a naturalidade da fala menor.</p> <p>Além disso, o sistema inseriu algumas palavras que servem como conexões entre as frases, tais como “no”, “e”, “do”, destacados no texto, em resultados.</p>
---------	---

Quadro 53: Resultados e Análise do Caso de Teste para Avaliar a Naturalidade da Fala do Usuário por Inspeção

Nome da Métrica	Tamanho do Vocabulário
Método de Avaliação	Investigação
Participantes	1 especialista
Material Necessário	Questionário ao representante do sistema
Roteiro	Não aplicável
Resultados	21.500 palavras (fonte: MACSYM, 2010)
Análise	O tamanho do vocabulário original não é, geralmente, suficientemente adequado às necessidades, sendo necessário incluir novas palavras. Em um dos hospitais visitados, a inclusão foi de, aproximadamente, 1000 outras palavras.

Quadro 54: Obtenção do Tamanho do Vocabulário. (Fonte: autor)

7.2.5. Planejamento do Teste de Usabilidade por Observação do Sistema *Stand-alone*

Utilizando uma metodologia para a elaboração dos testes por Observação adaptada de Diah et al (2010), Nielsen (1993) e de Mitchell (2007), este planejamento é composto pelas seguintes atividades executadas de maneira consecutiva:

- Planejar o teste de usabilidade;
- preparar os materiais de teste;
- criar as tarefas;
- selecionar os participantes;
- tratar questões éticas relativas aos testes;
- conduzir os testes de usabilidade;
- questionar o participante; e
- analisar os dados do teste de usabilidade.

A seguir será detalhada cada uma destas atividades:

7.2.5.1. Planejar o teste de usabilidade

Esta etapa tem por objetivo verificar quão bem funciona o sistema de transcrição automática de radiologia para um sistema *stand-alone*, desejando-se verificar o quanto o sistema está adaptado às métricas estabelecidas no Capítulo 6 deste trabalho.

Os testes foram realizados em um ambiente não hospitalar, através de participantes não pertencentes à área médica. Isso é aceitável, visto que os requisitos aqui abordados se referem a questões de usabilidade de um sistema de ditado, não especificamente de laudos.

Aqui serão validadas, através da aplicação de casos de testes de usabilidade, as métricas propostas no Capítulo 6 - que versa sobre a metodologia.

Depois de preparado o material, definidas as tarefas, selecionados os participantes, entre as outras tarefas que compõe este tópico, foram realizados dois testes-piloto a fim de verificar possíveis incongruências no teste - um realizado pela autora e outro por um participante voluntário. O detalhamento da aplicação do plano de teste encontra-se na seção 7.3.6. “Conduzir os testes de usabilidade”.

7.2.5.2. Preparar os materiais de teste

Esta subseção detalha a configuração dos equipamentos utilizados durante os testes com o software SpeechMagic (2010), assim como os materiais necessários:

- Laptop Philips: Processador Core 2 Duo, Memória RAM de 4GB, HD de 320 GB.
- SpeechMike Pro™ Philips.
- *Headset* Philips™ SHM 3300.
- Câmera Fotográfica.
- Decibelímetro: utilizado para medir a interferência Sonora do ambiente de teste.
- Cronômetro: utilizado para medir o tempo dos testes e tempo de resposta do sistema.
- 10 Laudos de Radiologia do Tórax.
- Questionário de pré-teste: para verificar o perfil do usuário – consta do Apêndice A deste trabalho.
- Carta de Apresentação dos Testes de Usabilidade: conforme apresentado no Apêndice B.
- Questionários de pós-teste – também chamado de questionário de percepção ou de usabilidade: Conforme apresentados no Apêndice A, baseado na metodologia

Questionnaire User Interface Satisfaction (QUIS) proposta em (CHIN; DIEHL; NORMAN, 1988).

- Formulários de Observação do Usuário: baseado em Mitchel (2007).
- Software SpeechMagic™ (2010)⁷ da empresa Nuance. Ele é composto de duas partes: treinamento da voz do usuário, chamado SmlInitialTraining e o sistema de reconhecimento de voz para radiologia propriamente dito – Interactive Dictation.
- Software SmlInitialTraining: leva cerca de 2 minutos para ser completada a versão suficiente para executar o reconhecimento de voz.
- Software Neoc Interactive 6.1: pode ser utilizado sem a execução prévia do SmlInitialTraining, porém apresentará uma precisão menor.

7.2.5.3. Criar as tarefas

As tarefas preparadas para os participantes executarem fazem parte do contexto de atividades que o radiologista teria que executar no seu dia-a-dia. Sendo assim, foram selecionados laudos escritos, sem identificação do paciente, para os participantes ditarem.

O *template* proposto pela autora no Quadro 30 foi utilizado para os testes de usabilidade que estão detalhados nesta subseção.

A seguir serão destacados os casos de teste para testar a Naturalidade da Fala do Usuário, Recuperação de Erros do Sistema, Carga Cognitiva e Adequação do *Feedback* do Sistema.

Nome da Métrica	Naturalidade da Fala do Usuário
Método de Avaliação	Observação
Participantes	6 usuários, 3 do sexo masculino e 3 do sexo feminino para o teste “a” e 2 usuários, um dos sexo feminino e um do sexo masculino, com sotaque diferente da cidade de São Paulo para o teste “b”
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™

⁷ Esta licença foi gentilmente cedida pela empresa MacSym Tecnologia Médica (Macsym, 2010) por um prazo de 60 dias

Resultados	
Análise	

Quadro 55: Caso de Teste para Avaliar a Naturalidade da Fala do Usuário por Observação. (Fonte: autor)

Nome da Métrica	Recuperação de Erros
Método de Avaliação	Observação e Questionário
Participantes	6 participantes
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™
Resultados	
Análise	

Quadro 56: Caso de Teste para Avaliação da Recuperação de Erros do Sistema por Observação. (Fonte: autor)

Nome da Métrica	Carga Cognitiva
Método de Avaliação	Observação e Questionário
Participantes	6 participantes
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™
Roteiro	
Resultados	
Análise	

Quadro 57: Caso de Teste da Avaliação da Carga Cognitiva do Sistema. (Fonte: autor)

Nome da Métrica	Adequação do <i>Feedback</i> e Visibilidade do Sistema
Método de Avaliação	Observação e Questionário
Participantes	6 participantes
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™
Resultados	
Análise	

Quadro 58: Caso de Teste para Avaliar a Adequação do *Feedback* por Observação. (Fonte: autor)

Nome da Métrica	Tempo de <i>Feedback</i> do sistema
Método de Avaliação	Questionário
Participantes	6 participantes
Material Necessário	1 laudo, 1 cronômetro, 1 equipamento SpeechMike™
Resultados	
Análise	

Quadro 59: Caso de Teste para Avaliação do Tempo de *Feedback* do Sistema. (Fonte: autor)

Nome da Métrica	Qualidade dos sistemas de áudio
Método de Avaliação	Questionário
Participantes	6 usuários
Material Necessário	1 laudo, 1 equipamento SpeechMike™
Resultados	
Análise	

Quadro 60: Caso de Teste para Avaliar a Qualidade do Sistema de Áudio. (Fonte: autor)

7.2.5.4. Selecionar os participantes

A escolha dos participantes se deu levando em consideração as seguintes questões:

- Necessidade de se ter diferentes timbres de voz, tanto masculino quanto feminino. Assim foram selecionados seis participantes, três do sexo feminino, três do sexo masculino.

- Necessidade de se ter voz de pessoa adulta, os participantes devem ter entre 20 e 50 anos.
- Necessidade de se ter, ao menos, dois profissionais em inspeção de software. A autora do trabalho e outra profissional de usabilidade participaram para a inspeção do software.
- Necessidade de se ter pessoas com sotaque diferente das pessoas da cidade de São Paulo. Assim, foram utilizados dois usuários que tinham sotaque bastante acentuado: uma pessoa nascida e criada em João Pessoa e outra pessoa nascida e criada em Concepción, no Chile.
- Os participantes não são da área de Radiologia, já acostumados a trabalhar no software. Os participantes não são da área médica.

Para cada participante, foi realizado um primeiro contato pessoal e, posteriormente, enviada uma carta de apresentação especificando o intuito do teste, como seria realizada a seção e o tempo necessário. Esta carta, na íntegra, consta do Apêndice B deste trabalho.

Embora Nielsen (2000b) descreva, em seu estudo, que um grupo com 5 participantes seja suficiente para encontrar 80% das características desejáveis de serem descobertas em um sistema e que quanto mais usuários forem adicionados, menos informações adicionais são encontradas em oposição ao custo para se fazer os testes, foi verificada a necessidade de se trabalhar com 8 participantes, pois se desejava verificar características específicas da voz de homens e mulheres e também de pessoas com sotaque. Assim, o trabalho englobou um grupo de 6 participantes, sendo 3 homens e 3 mulheres, e um grupo de 2 participantes, um de cada sexo, com sotaques diferentes das pessoas nativas na cidade de São Paulo.

A Tabela 1 mostra o perfil dos participantes deste teste de usabilidade.

Tabela 1: Perfil dos Participantes. (Fonte: autor)

Sexo	Feminino	Masculino	
	50%	50%	
Experiência Sistema de Transcrição Automática	Nunca Utilizei	Funcionam regularmente	
	0.875	0.125	
Sotaque Pronunciado	Sim	Não	
	0.25	0.75	
Nível Escolaridade	Graduação	Mestrado	Doutorado
	0.125	0.625	0.25

Também participaram destes testes duas especialistas em usabilidade cuja função foi a de observarem os participantes em suas atividades, fazendo as coletas de dados necessárias: preenchimento de formulário e captação de imagens e sons.

A seguir são mostradas algumas fotos tiradas na fase de testes de usabilidade com alguns voluntários.



Figura 20: Testes de Usabilidade com Voluntários. (Fonte: autor)

A interferência sonora nestes testes de usabilidade foi medida, no início de cada teste, e consta da Tabela 2.

Tabela 2: Interferência Sonora do Ambiente. (Fonte: autor)

Participante	Interferência Sonora Inicial
1	48 a 51
2	50 a 60
3	48 a 52
4	44 a 50
5	45 a 60
6	52 a 61
7	40 a 50
8	42 a 50

O tempo do teste total foi, em média, de 45 minutos. O tempo de cada teste consta da Tabela 3.

Tabela 3: Tempo de duração de cada teste de usabilidade. (Fonte: autor)

Participante	Tempo Total (Minuto)
1	48
2	44
3	45
4	44
5	50
6	49
7	43
8	44
Média	45,875

7.2.5.5. Tratar as questões éticas

Em relação às questões éticas que devem ser tratadas num teste de usabilidade, foram utilizadas aquelas indicadas por Pernice e Nielsen (2009) e Mitchell (2007), vindas do Human Factors and Ergonomics Society Code of Ethics de 1993 (28/12/2010). Assim, foram realizadas as seguintes ações:

- O primeiro contato foi realizado pessoalmente, explicando as razões para a realização dos testes.
- Envio, por correio eletrônico, de uma Carta de Apresentação dos Testes de Usabilidade, conforme apresentada no Apêndice B, e marcação de um horário para a realização do teste.
- Coleta de autorização para o uso de imagem e voz dos participantes, conforme apresentada no Apêndice C desde trabalho.

7.2.5.6. Conduzir os testes de usabilidade

As sessões foram compostas por quatro partes, conforme descritas por Rubin e Chisnell (2008) e tiveram duração média de 45 minutos:

- Introdução à sessão de testes: explica os motivos para a realização dos testes e como o usuário deve agir. Deve durar cerca de 5 minutos.
- Questionário de pré-teste: coleta informações relevantes sobre o perfil do usuário que possam interferir no resultado do teste. Hartikainen, Salonen e Turunen (2004) utilizam a nomenclatura de “questionário de expectativas”. Deve durar cerca de 2 minutos.
- Tarefas: quais as tarefas que o usuário deve executar. Deve durar cerca de 30 minutos. Os voluntários fizeram um treinamento mínimo prévio necessário de sua voz que é de 8 frases, totalizando cerca de 2 minutos de treino, em média.

- Questionário de pós-teste: coleta as informações referentes à impressão do usuário sobre as tarefas que ele executou. Hartikainen, Salonen e Turunen (2004) utilizam a nomenclatura de “questionário de percepção”. Deve durar cerca de 3 minutos.

A descrição de cada um destes passos é encontrada na próxima subseção deste trabalho.

7.2.5.7. Questionar os participantes

Nessa fase foi passado um questionário de pós-teste para cada um dos participantes a fim de verificar a satisfação de utilização do sistema e também verificar parte de questões de usabilidade.

7.2.5.8. Analisar os dados

Nessa subseção é possível visualizar, em detalhe, todos os casos de teste realizados através de observação dos participantes, assim como seus resultados dos testes e análise dos dados coletados. Os quadros e tabelas são apresentados a seguir.

Nome da Métrica	Naturalidade da Fala do Usuário
Método de Avaliação	Análise dos dados e Observação
Participantes	6 usuários, 3 do sexo masculino e 3 do sexo feminino para o teste “a” e 2 usuários, um do sexo feminino e um do sexo masculino, com sotaque diferente da cidade de São Paulo
Material Necessário	5 laudos de radiologia, 1 equipamento SpeechMike™
Roteiro	<p>a) Através da leitura de laudos:</p> <ul style="list-style-type: none"> i) Verificar a diferença de precisão entre o usuário falar mais pausadamente ou mais rapidamente; ii) Verificar a diferença de precisão entre o usuário falar com tom baixo

	<p>ou alto;</p> <p>iii) Usuário lê um laudo e usuário fala um laudo sem ler;</p> <p>b) Através da leitura de laudos:</p> <p>i) Verificar a diferença de precisão entre o usuário falar mais pausadamente ou mais rapidamente;</p> <p>ii) Verificar a diferença de precisão entre o usuário falar com tom baixo ou alto;</p>																																																															
Resultados	<p>O laudo continha 54 palavras.</p> <p>Os usuários 1, 3, 5 e 8 são mulheres. Os usuários 7 e 8 são, respectivamente, um homem com sotaque e uma mulher com sotaque.</p> <p>Os resultados podem ser visualizados no Gráfico 10.</p> <p style="text-align: center;">Gráfico 10: Naturalidade da Fala de Oito Voluntários</p> <table border="1"> <caption>Dados estimados do Gráfico 10: Naturalidade da Fala</caption> <thead> <tr> <th>Usuário</th> <th>Sem Pontuação</th> <th>Com Pontuação</th> <th>Pausado</th> <th>Acelerado</th> <th>Tom Baixo</th> <th>Tom Alto</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>18</td> <td>3</td> <td>73</td> <td>12</td> <td>100</td> <td>3</td> </tr> <tr> <td>2</td> <td>8</td> <td>5</td> <td>15</td> <td>8</td> <td>10</td> <td>7</td> </tr> <tr> <td>3</td> <td>7</td> <td>5</td> <td>57</td> <td>22</td> <td>100</td> <td>7</td> </tr> <tr> <td>4</td> <td>12</td> <td>10</td> <td>5</td> <td>2</td> <td>5</td> <td>5</td> </tr> <tr> <td>5</td> <td>37</td> <td>5</td> <td>15</td> <td>10</td> <td>100</td> <td>3</td> </tr> <tr> <td>6</td> <td>12</td> <td>10</td> <td>18</td> <td>12</td> <td>10</td> <td>10</td> </tr> <tr> <td>7</td> <td>17</td> <td>24</td> <td>57</td> <td>20</td> <td>10</td> <td>7</td> </tr> <tr> <td>8</td> <td>24</td> <td>12</td> <td>37</td> <td>15</td> <td>100</td> <td>10</td> </tr> </tbody> </table>	Usuário	Sem Pontuação	Com Pontuação	Pausado	Acelerado	Tom Baixo	Tom Alto	1	18	3	73	12	100	3	2	8	5	15	8	10	7	3	7	5	57	22	100	7	4	12	10	5	2	5	5	5	37	5	15	10	100	3	6	12	10	18	12	10	10	7	17	24	57	20	10	7	8	24	12	37	15	100	10
Usuário	Sem Pontuação	Com Pontuação	Pausado	Acelerado	Tom Baixo	Tom Alto																																																										
1	18	3	73	12	100	3																																																										
2	8	5	15	8	10	7																																																										
3	7	5	57	22	100	7																																																										
4	12	10	5	2	5	5																																																										
5	37	5	15	10	100	3																																																										
6	12	10	18	12	10	10																																																										
7	17	24	57	20	10	7																																																										
8	24	12	37	15	100	10																																																										
Análise	<p>a) Embora esteja na especificação do sistema, fornecido pelo fornecedor, que o sistema faz a inserção de pontos finais e vírgula, o sistema não foi capaz de fazê-lo, colocando conjunções tais como “com”, “e”, que, em muitas vezes afetavam o sentido das frases.</p>																																																															

	<p>b) A precisão de reconhecimento da voz, pela leitura do texto com pontuação explícita foi maior entre as mulheres, mas mesmo entre todos os participantes, exceto os que tinham um sotaque bastante pronunciado (usuário 7 e 8) foi superior a 90%.</p> <p>c) Ficou claro que o sotaque influencia enormemente a taxa de reconhecimento, necessitando que estes usuários façam um treinamento mais intensivo e que o sistema aprenda à medida que o usuário utilize o sistema de forma contínua.</p> <p>d) O uso do sistema para uma fala natural do usuário sempre se mostra melhor do que a inserção dos dados através de um ditado mais pausado ou mais rápido do que o habitual.</p> <p>e) O uso do sistema para um tom de voz mais baixo ou mais alto também afeta o reconhecimento, devendo o usuário sempre utilizar seu tom normal de voz. Entre as mulheres, o uso de um tom mais baixo de voz impediu a entrada do laudo, visto que o sistema foi incapaz de captar a entrada de áudio.</p> <p>f) O uso de pontuação explícita no texto influencia a taxa de reconhecimento da voz.</p>
--	--

Quadro 61: Caso de Teste para Avaliar a Naturalidade da Fala do Usuário por Observação

Nome da Métrica	Recuperação de Erros
Método de Avaliação	Observação e Questionário
Participantes	6 participantes
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™
Roteiro	Verificar, através de observação, como o sistema age perante os seguintes erros: a) O usuário usa palavras que não estão no dicionário da aplicação – um laudo

será modificado para atingir este objetivo.

b) O sistema não consegue reconhecer o que o usuário dita

c) O usuário erra o laudo e deseja refazê-lo.

Resultados

O texto continha 77 palavras, com várias palavras fora do vocabulário específico da área de Radiologia.

Os usuários 1, 3, 5 são mulheres.

Os resultados podem ser visualizados no Gráfico 11.

Gráfico 11: Porcentagem de Erro de Teste com Seis Usuários



Os usuários também responderam a perguntas do questionário a respeito da recuperação de erros (Apêndice E) e esta questão apareceu até nas questões abertas como críticas ou melhorias.

Tabela 4: Parte do Questionário sobre Experiência de Uso do Sistema

	Experiência de uso do sistema				
	Concordo Fortemente	Concordo	Indeciso	Discordo	Discordo Fortemente
Quando erro no ditado, é fácil fazer o sistema reconhecer o que falo e se adaptar.		2		4	

Pontos Negativos da Aplicação: O sistema não possui mecanismo de reparo de erros durante o ditado.

Sugestões ou reclamações sobre a aplicação: Tratamento de erro – algum mecanismo para reparar algum erro, solicitado pelo usuário

Análise	<p>a) Em todos os testes, para todos os participantes, quando o usuário errou o laudo, ele se sentiu bastante desconfortável em relação a como promover a correção. Interjeições foram freqüentes quando o usuário se conscientizou do erro. Também perguntas ao observador de como promover a correção foram freqüentes.</p> <p>b) Quando o usuário olhava a tela enquanto ditava e o sistema demorava em apresentar o texto – que acontecia com atraso de até 5 segundos – o usuário gaguejava ou tentava recomeçar a fala.</p> <p>c) É possível verificar, através deste gráfico, que houve uma considerável diminuição de precisão com o uso de um texto com palavras fora do vocabulário específico de radiologia. O que era esperado.</p>
---------	---

Quadro 62: Caso de Teste para Avaliação da Recuperação de Erros do Sistema por Observação. (Fonte: autor)

Nome da Métrica	Carga Cognitiva
Método de Avaliação	Observação e Questionário
Participantes	6 participantes
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™
Orientações aos Participantes	<p>Se você não fez o treinamento inicial da fala, faça-o agora;</p> <p>a) Leia um laudo sem a pontuação.</p> <p>b) Leia um laudo e faça a pontuação correta.</p>
Roteiro	<p>Verificar, através de observação, como o usuário age:</p> <p>a) Há aumento de carga cognitiva significativa?</p>
Resultados	As três últimas perguntas do questionário de satisfação (pós-teste) que constam do Apêndice A referiam-se a esta questão de aumenta da carga cognitiva. As respostas estão a seguir:

Tabela 5: Parte do Questionário sobre Experiência de Uso do Sistema – Carga Cognitiva					
Experiência de uso do sistema					
	Concordo Fortemente	Concordo	Indeciso	Discordo	Discordo Fortemente
As formatações, tais como “ponto final”, “vírgula”, “negrito”, são de fácil assimilação, não me fazendo pensar muito para utilizá-las.	3	1		2	
Quando faço hesitações, tais como "mmh", "Aaah", o sistema ignora, captando somente palavras pertencentes a meu vocabulário para o laudo.	1	2	1		2
É fácil fazer a formatação do texto enquanto dito um laudo, tal como mudar de linha e pontuar.	3	1		2	
Análise	<p>a) Um terço dos participantes entendeu que ter que fazer a recordação da palavra-chave para inserir a pontuação, tais como: nova linha, ponto e vírgula – aumentava a atenção que ele deveria ter no ditado, mas a maioria dos usuários concordou que considera fazer a formatação do texto enquanto dita o laudo, uma tarefa bastante fácil.</p> <p>b) Metade dos participantes teve problemas de se lembrar da palavra-chave para a pontuação, necessitando perguntar ao observador como fazer.</p>				

Quadro 63: Caso de Teste da Avaliação da Carga Cognitiva do Sistema. (Fonte: autor)

Nome da Métrica	Adequação do <i>Feedback</i>
Método de Avaliação	Observação e Questionário
Participantes	6 participantes
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™
Roteiro	<p>a) o sistema não deve fornecer <i>feedbacks</i> que atrapalhem a capacidade de raciocínio do usuário:</p> <ul style="list-style-type: none"> • O sistema interrompe o ditado; • O sistema inicia processos de alertas;

	<p>b) deve estar presente, de maneira que o usuário, em tempo oportuno, saiba de erros que tenham ocorrido no sistema enquanto ele ditava o laudo:</p> <ul style="list-style-type: none"> • O sistema permite, de alguma forma, que ocorreu uma falha no reconhecimento de alguma palavra; 																		
Resultados	<p>Uma das perguntas do questionário de satisfação (pós-teste) que consta do Apêndice A referia-se a esta questão de adequação do <i>feedback</i>. Também esta questão apareceu em sugestões de melhoria do sistema, uma questão aberta. As respostas estão a seguir:</p> <p>Tabela 6: Parte do Questionário sobre Experiência de Uso do Sistema – Adequação do <i>Feedback</i></p> <table border="1" data-bbox="354 695 1523 947"> <thead> <tr> <th colspan="6" style="background-color: #d9e1f2;">Experiência de uso do sistema</th> </tr> <tr> <th></th> <th>Concordo Fortemente</th> <th>Concordo</th> <th>Indeciso</th> <th>Discordo</th> <th>Discordo Fortemente</th> </tr> </thead> <tbody> <tr> <td>A aplicação informa o que está acontecendo ao usuário (por exemplo, se uma palavra não foi reconhecida durante o ditado)</td> <td style="text-align: center;">1</td> <td></td> <td></td> <td></td> <td style="text-align: center;">5</td> </tr> </tbody> </table> <p>Sugestões ou reclamações sobre a aplicação: Tratamento de erros – prevenção: Marcar as palavras que foram escritas com certo grau de indecisão aumentaria o grau de confiabilidade na aplicação.</p>	Experiência de uso do sistema							Concordo Fortemente	Concordo	Indeciso	Discordo	Discordo Fortemente	A aplicação informa o que está acontecendo ao usuário (por exemplo, se uma palavra não foi reconhecida durante o ditado)	1				5
Experiência de uso do sistema																			
	Concordo Fortemente	Concordo	Indeciso	Discordo	Discordo Fortemente														
A aplicação informa o que está acontecendo ao usuário (por exemplo, se uma palavra não foi reconhecida durante o ditado)	1				5														
Análise	<p>a) De acordo com o questionário, os usuários percebem o problema da não identificação pelo sistema de palavras erradas.</p> <p>b) O sistema não informa, de maneira nenhuma, falhas no reconhecimento da VOZ.</p>																		

Quadro 64: Caso de Teste para Avaliar a Adequação do *Feedback* por Observação. (Fonte: autor)

Nome da Métrica	Qualidade dos sistemas de áudio			
Método de Avaliação	Questionário			
Participantes	6 usuários			
Material Necessário	1 laudo, 1 equipamento SpeechMike™			
Roteiro	Analisar se o usuário é capaz de reconhecer o que ele mesmo ditou.			
Resultados	Tabela 7: Questão sobre a percepção do áudio pelo próprio usuário			
	Experiência de uso do sistema			
		Concordo Fortemente	Concordo	Indeciso
	6			
Análise	Esta medida subjetiva foi altamente confirmada como verdadeira entre todos os participantes do experimento, mostrando que o áudio gravado e usado para a transcrição é altamente fiel, na percepção do usuário, ao som original.			

Quadro 65: Caso de Teste para Avaliar a Qualidade do Sistema de Áudio. (Fonte: autor)

7.3 Planejamento das Observações dos Usuários Finais em Ambiente de Produção

Este teste de usabilidade será executado apenas analisando o usuário final – radiologista – em seu ambiente de trabalho. Assim, observador apenas anotará questões observáveis nas tarefas sendo executadas, não fazendo intervenções com o usuário.

Utilizando uma metodologia para a elaboração dos testes por Observação adaptada de Diah et al (2010), Nielsen (1993) e de Mitchell (2007), este planejamento é composto pelas seguintes atividades executadas de maneira consecutiva:

- Planejar o teste de usabilidade;
- preparar os materiais de teste;

- criar as tarefas;
- selecionar os participantes;
- tratar questões éticas relativas aos testes;
- conduzir os testes de usabilidade;
- questionar o participante; e
- analisar os dados do teste de usabilidade.

A seguir será detalhada cada uma destas atividades:

7.3.1. Planejar o teste de usabilidade

Estes testes de usabilidade serão realizados em um hospital da cidade de São Paulo que possui um sistema de transcrição automática de laudos em radiologia.

O detalhamento da aplicação do plano de teste encontra-se na seção 7.2.3.6. “Conduzir os testes de usabilidade”.

7.3.2. Preparar os materiais

Esta subseção detalha a configuração dos equipamentos utilizados durante os testes com o software Speech Magic 2.1 (Philips, 2010), assim como os materiais necessários:

- PC Intel Core 2, CPU 6.300, 1.86 GHz, 1.97 GB de RAM, com Sistema Operacional XP 2002.
- SpeechMike Pro™ Philips, modelo 6274.
- Câmera Fotográfica.
- Decibelímetro: utilizado para medir a interferência Sonora do ambiente de teste.
- Cronômetro: utilizado para medir o tempo dos testes e tempo de resposta do sistema.
- Questionário de pré-teste: para verificar o perfil do usuário – consta do Apêndice A deste trabalho.

- Questionários de pós-teste – também chamado de questionário de percepção ou de usabilidade: Conforme apresentados no Apêndice A, baseado na metodologia Questionnaire User Interface Satisfaction (QUIS) proposta em (CHIN; DIEHL; NORMAN, 1988).
- Formulários de Observação do Usuário: baseado em Mitchel (2007), que constam do Apêndice F deste trabalho. Basicamente, serão observadas características como: Tempo de *Delay* do sistema; naturalidade da fala (maneira de se falar, pontuação e interjeições; tempo para o laudo ficar pronto; sobrecarga cognitiva; recuperação de erros e interferência sonora. Estes formulários são apresentados a seguir.

Tabela 8: Formulário de Observação do *Delay* do Sistema. (Fonte: autor)

DELAY DO SISTEMA						
	Medição 1	Medição 2	Medição 3	Medição 4	Medição 5	Média
(Segundos)						

NATURALIDADE DA FALA	
O usuário fala de maneira natural, ou seja, nem muito pausado nem muito rápido	() Sim () Não
Observações:	
O usuário insere pontuação explicitamente	() Sim () Não
Observações:	
O sistema coloca pontuação automaticamente	() Sim () Não
Observações:	
O usuário fez interjeições durante a gravação do laudo	() Sim () Não
Observações:	
O sistema ignorou as interjeições do usuário	() Sim () Não
Observações:	
O usuário falou em tom normal	() Sim () Não
Observações:	

Quadro 66: Formulário de Observação da Naturaliza da Fala do Usuário

Tabela 9: Formulário de Observação da Interferência Sonora do Ambiente. (Fonte: autor)

INTERFERÊNCIA SONORA						
	Medição 1	Medição 2	Medição 3	Medição 4	Medição 5	Média
(db)						

Tabela 10: Formulário de Observação do Tempo para o Laudo ficar Pronto. (Fonte: autor)

TEMPO PARA O LAUDO FICAR PRONTO	
Tempo de ditado:	
Observações:	
Tempo para correções:	
Observações:	
Tempo para impressão:	
Observações:	
Tempo de verificação e assinatura:	
Observações:	
Tempo Total	

RECUPERAÇÃO DE ERROS	
O usuário refez parte do texto depois do ditado	() Sim () Não
Observações:	
Quantas palavras foram editadas	
Observações:	
O usuário inseriu palavras novas no sistema	() Sim () Não
Observações:	
O usuário formatou o texto	() Sim () Não
Observações:	

Quadro 67: Formulário de Observação da Recuperação de Erros

SOBRECARGA COGNITIVA	
O usuário olhou para a tela enquanto ditava o laudo	() Sim () Não
Observações:	
O usuário fez hesitações quando o texto apresentava <i>delay</i> para aparecer na tela	() Sim () Não
Observações:	
O usuário dita o laudo e já faz a formatação explícita (ditada) do laudo	() Sim () Não
Observações:	

Quadro 68: Formulário de Observação da Sobrecarga Cognitiva do Usuário. (Fonte: autor)

7.3.3. Criar as tarefas

Neste teste não serão criadas tarefas para os participantes executarem, serão apenas observadas as tarefas do cotidiano do radiologista e verificar como as tarefas estão sendo executadas.

7.2.3.4. Selecionar os participantes

Conforme determinado pela banca de qualificação de doutorado deste trabalho, seriam observados dois radiologistas em um hospital e um radiologista em outro hospital. Estes usuários finais seriam selecionados quando forem realizadas as visitas aos dois hospitais, de forma aleatória. Porém, houve enorme dificuldade em conseguir realizar os testes em um dos hospitais, optando-se, então, pelos testes em apenas um dos hospitais.

7.3.5. Tratar as questões éticas relativas aos testes

De maneira similar ao descrito para os usuários voluntários, as questões éticas envolveram:

- O primeiro contato foi realizado pessoalmente, explicando as razões para a realização dos testes.

- Envio, por correio eletrônico, de uma Carta de Apresentação dos Testes de Usabilidade, conforme apresentada no Apêndice B, e marcação de um horário para a realização do teste.
- Coleta de autorização para o uso de imagem e voz dos participantes, conforme apresentada no Apêndice C deste trabalho.

7.3.6. Conduzir o teste de usabilidade

Os testes de usabilidade por observação serão realizados em tempo médio de 1 hora.

7.3.7. Questionar o participante

Nessa fase foi passado um questionário de pós-teste para os participantes fim de verificar a satisfação de utilização do sistema e também verificar parte de questões de usabilidade. Este questionário consta do Apêndice A deste trabalho e o resultado é apresentado a seguir.

Experiência de uso do sistema					
	Concordo Fortemente	Concordo	Indeciso	Discordo	Discordo Fortemente
Esta aplicação tem uma apresentação gráfica agradável e legível	X				
É fácil entender o que eu preciso fazer na aplicação	X				
A interface gráfica é bastante intuitiva	X				
A aplicação informa o que está acontecendo ao usuário (por exemplo, se uma palavra não foi reconhecida durante o ditado)					X
Foi fácil aprender a utilizar a aplicação	X				
Quanto à velocidade de resposta (transcrição do ditado em texto) é suficiente			X		
Mesmo sem experiência no uso deste sistema, é fácil utilizá-lo	X				
Acredito que a aplicação funciona de maneira adequada quanto ao reconhecimento da minha pronúncia	X				
Há mecanismos no sistema, que acesso através de mouse ou teclado, que proporcionam um melhor desempenho na transcrição (tais como	X				

acessar frases já prontas através de um menu, enquanto dito)					
Quando erro no ditado, é fácil fazer o sistema reconhecer o que falo e se adaptar.			X		
O sistema possui atalhos para que eu possa realizar minha tarefa de maneira mais eficiente		X			
As formatações, tais como “ponto final”, “vírgula”, “negrito”, são de fácil assimilação, não me fazendo pensar muito para utilizá-las.		X			
Quando faço hesitações, tais como "mmh", "Aaah", o sistema ignora, captando somente palavras pertencentes a meu vocabulário para o laudo.					X
É fácil fazer a formatação do texto enquanto dito um laudo, tal como mudar de linha e pontuar.		X			
O sistema de ajuda conseguiu esclarecer minhas dúvidas sobre o programa.		X			
Consigo entender perfeitamente o som gravado do ditado que fiz.		X			
Considerações sobre a aplicação					
Pontos positivos da aplicação: eficiente e de fácil aprendizado					
Pontos negativos da aplicação: muita interferência de ruído da sala					
Sugestões ou reclamações sobre a aplicação: melhorar sistema de correção de texto					

Quadro 69: Formulário de Satisfação do Usuário. (Fonte: autor)

7.3.8. Analisar os dados do teste de Usabilidade

Nessa subseção é possível visualizar, em detalhe, todos os resultados das observações casos de teste realizados através de observação do participante, assim como seus resultados dos testes e análise dos dados coletados.

Tabela 11: Resultado da Observação do Delay do Sistema. (Fonte: autor)

<i>DELAY DO SISTEMA</i>						
	Medição 1	Medição 2	Medição 3	Medição 4	Medição 5	Média
(Segundos)	3.2	1.5	4.5	3.3	4.5	3.4
Observação: estas medidas foram tomadas a cada 2 minutos, quando o sistema estava ativo.						

NATURALIDADE DA FALA	
O usuário fala de maneira natural, ou seja, nem muito pausado nem muito rápido	(X) Sim () Não
Observações:	Segundo o usuário, ele fala um pouco mais devagar que no sistema similar utilizado em outro hospital
O usuário insere pontuação explicitamente	(X) Sim () Não
Observações:	Percebe-se que esta já é uma prática comum para os radiologistas que utilizam o sistema de transcrição manual de laudos, ou seja, o uso de digitadoras.
O sistema coloca pontuação automaticamente	() Sim (X) Não
Observações:	
O usuário fez interjeições durante a gravação do laudo	() Sim (X) Não
Observações:	
O sistema ignorou as interjeições do usuário	(X) Sim () Não
Observações:	Em todas as observações, isso não interferiu na taxa de reconhecimento
O usuário falou em tom normal	(X) Sim () Não
Observações:	

Quadro 70: Formulário de Observação da Naturaliza da Fala do Usuário. (Fonte: autor)

Tabela 12: Formulário de Observação da Interferência Sonora do Ambiente. (Fonte: autor)

INTERFERÊNCIA SONORA						
	Medição 1	Medição 2	Medição 3	Medição 4	Medição 5	Média
(db)	49.2	72	55.4	52.1	51	55.94
Observação: estas medidas foram tomadas a cada 2 minutos, independentes do sistema estar ativo ou não. Havia uma interferência sonora bastante pronunciada devido à reformas no prédio. Outro médico estava ditando laudos na mesma sala, o que também contribuía para o aumento de interferência sonora no ambiente.						

Tabela 13: Resultado da Observação do Tempo para o Laudo ficar Pronto

TEMPO PARA O LAUDO FICAR PRONTO	
Tempo de ditado: média	
Observações:	Este tempo é bastante diferente, dependendo do tamanho do laudo e do uso ou não de macros ⁸ e também se o usuário é habituado a ditar o laudo todo ou frases espaçadas, intercaladas com correções.
Tempo para correções:	
Observações:	Este tempo também é bastante influenciado, sendo percebido menor, se o médico utiliza a abordagem de falar uma frase e já fazer a correção, ou quando dita o laudo todo e depois faz a correção.
Tempo para impressão:	
Observações:	Tempo equivalente à impressão de um documento Word.
Tempo de verificação e assinatura:	
Observações:	A verificação é realizada no tempo de correção.
Tempo Total (média)	5 min

RECUPERAÇÃO DE ERROS	
O usuário refez parte do texto depois do ditado	(X) Sim () Não
Observações:	Boa parte da intervenção do usuário no texto se referia a problemas de espaçamento desta versão do software.
Quantas palavras foram editadas, em média	
Observações:	Nos três laudos observados, as taxas de erro foram: 9%,14% e 18%. A taxa menor é entendida porque foram utilizadas macros e pouco foi a parte de ditado propriamente dito.
O usuário inseriu palavras novas no sistema	() Sim (X) Não
Observações:	A amostra é muito pequena para afirmar que não é necessário inserir palavras novas no sistema, no seu uso diário.
O usuário formatou o texto	(X) Sim () Não
Observações:	Ele formatou o texto depois de terminar de ditar o laudo

Quadro 71: Resultado da Observação da Recuperação de Erros. (Fonte: autor)

⁸ O termo “macro” utilizado aqui, se refere a textos pré-digítados que fazem parte, comumente, de um laudo. Podendo, assim, serem inseridos no texto, economizando ditados.

SOBRECARGA COGNITIVA	
O usuário olhou para a tela enquanto ditava o laudo	() Sim (X) Não
Observações:	O médico já é orientado a não olhar a tela enquanto faz o laudo. Além disso, neste momento, o médico já está olhando para duas outras telas, com quadro imagens em uma delas.
O usuário fez hesitações quando o texto apresentava <i>delay</i> para aparecer na tela	() Sim (X) Não
Observações:	O usuário não olhava na tela do sistema de reconhecimento enquanto falava.
O usuário dita o laudo e já faz a formatação explícita (ditada) do laudo	(X) Sim () Não
Observações:	Os médicos já estão acostumados a este padrão de formatação explícita, pois é este o padrão utilizado gravar os laudos que serão digitados por uma terceira pessoa.

Quadro 72: Formulário de Observação da Sobrecarga Cognitiva do Usuário. (Fonte: autor)

7.4 Discussões sobre os resultados dos testes

As discussões sobre os resultados dos testes de usabilidade aplicados neste trabalho foram divididas em duas partes, conforme os sistemas utilizados: sistema *stand-alone* e o sistema implantado no Hospital Edmundo Vasconcelos.

7.4.1 Sistema *Stand-Alone*

Através de todos os testes realizados, foi possível concluir várias questões sobre o sistema utilizado em sua versão *stand-alone*, tanto referentes à inspeção de usabilidade quanto aos testes de usabilidade fora do ambiente hospitalar, que estão descritas a seguir.

O sistema de transcrição automática de laudos se mostrou bastante eficiente em relação à precisão do reconhecimento de voz, ficando acima de 93%, em média, mesmo incluindo, no grupo, pessoas com sotaque pronunciado.

A inspeção de usabilidade mostrou que a precisão da voz chegou a 95% de acerto, mesmo com um treinamento da voz mínimo – cerca de 2 minutos.

Isso prova que o sistema possui uma boa precisão do reconhecimento, podendo ainda ser melhorado, se forem incluídas palavras novas ao vocabulário inicial do sistema e também havendo um treinamento mais aprofundado para os usuários, sobretudo aqueles com sotaque pronunciado. Para se exemplificar o ganho que se teria, se a palavra “quilovoltagem” fosse incluída ao vocabulário inicial do sistema, haveria uma diminuição do erro, em média, de 25.65%, pelos testes efetuados.

Em questões de velocidade da fala do usuário, os testes apontaram que o sistema é sensível à mudança da velocidade da fala. O usuário deve utilizar sua velocidade habitual de fala, nem mais lento, nem mais rápido.

Já os testes com um tom mais baixo que o habitual entre os usuários mostraram uma característica interessante: nenhuma das mulheres do grupo conseguiu inserir o laudo utilizando um tom mais baixo da voz. Já, entre os homens, todos conseguiram inserir o laudo, tendo um não muito significativo decréscimo na taxa de reconhecimento da voz. Provavelmente isso é decorrente do timbre de voz mais grave.

Em relação ao treino prévio da voz, não houve diferença significativa na taxa de reconhecimento sem o treinamento e com o treinamento, como era pensado. O sistema já se mostrou bastante eficiente sem treinamento, com taxa de reconhecimento de quase 93% em média. Isso leva a crer que o sistema poderia ser utilizado mesmo sem ser realizado um treinamento mínimo da voz, com taxas aceitáveis de reconhecimento da voz.

O *delay* do sistema em apresentar o texto referente ao que estava sendo ditado, em tempo real, foi bastante alto, chegando a quase 5 segundos, em alguns momentos. Segundo o fornecedor do sistema, muitos especialistas sugerem que os médicos radiologistas não olhem para a tela enquanto estão ditando seu laudo, pois o *delay* interfere na cognição do usuário, inserindo sobrecarga à tarefa. Em testes com os voluntários, havia um sentimento de que o sistema não estava funcionando quando este *delay* era alto.

Os testes com interferência sonora mostraram que este é um requisito que influencia fundamentalmente a precisão do reconhecimento de voz, devendo, o ambiente, se necessário, ser modificado para estar de acordo com as normas de ruído estabelecidas pela ABNT – NBR 10152/1987 (2010).

Em relação ao requisito “Dicionário específico para RIS”, foi analisado que, segundo os testes realizados neste trabalho, a massa de dados é insuficiente para se afirmar que o dicionário inicialmente inserido no sistema de reconhecimento seja considerado aprovado ou reprovado. Para que este requisito pudesse ser caracterizado de maneira apropriada, uma análise de todos os laudos gerados por todos os radiologistas durante certo intervalo de tempo conseguiria medi-lo de maneira mais eficaz. Levando em consideração esta limitação, para os testes realizados, apenas uma palavra não estava contemplada no dicionário - “quilovoltagem”. Esta palavra apareceu em todos os laudos examinados e representou, em média, 1.78% das palavras presentes nos laudos.

O requisito “Resolução de Homônimos” foi testado de maneira superficial, pois não foi levado em consideração todas as palavras homônimas que poderiam aparecer no dicionário do sistema.

Houve dificuldade no uso de métodos automáticos de comparação das versões dos textos – original e gerada pelo ditado -, pois houve a necessidade de desconsiderar pontos tais como:

- Diferenças entre maiúsculas e minúsculas.
- Diferenças de acentuação.
- Diferenças na ortografia de algumas palavras no laudo e no que o sistema gera, tal como a palavra “costo-frênicos” e “costofrênicos”.
- A inclusão de uma ou mais palavras no texto eram tratadas apenas como 1 inclusão.

Os dois equipamentos utilizados para a entrada dos laudos – HeadSet Philips e SpeechMike Philips – se mostraram eficientes. Era esperada, segundo informação do fornecedor, uma grande diferença de qualidade sonora, porém, pelos testes efetuados, a precisão no reconhecimento se mostrou melhor com o HeadSet em condições de pouca interferência sonora. Já com muita interferência sonora, o SpeechMike foi melhor. A diferença expressiva de custo e a ergonomia entre os dois equipamentos devem ser levadas em consideração na aquisição de um ou outro equipamento.

O sistema de ajuda – *help* – foi considerado bastante superficial, entre os dois especialistas: o *help* online não funcionou e não havia um sistema de buscas de palavras nem de contexto.

Os pontos considerados mais críticos, tanto entre os especialistas quanto entre os voluntários se referem aos requisitos de “Falta de Visibilidade e Adequação do *Feedback*” e “Recuperação de Erros”.

Em relação à falta de visibilidade e adequação do *feedback*, a criticidade se deu pelo sistema nunca informar ao usuário palavras que, possivelmente, tenham sido reconhecidas com baixa taxa. Uma sugestão proposta foi que estas palavras fossem marcadas – tal como o uso de negrito ou sublinhado - para que o usuário conseguisse visualizá-las mais facilmente.

Já em relação à recuperação de erros, os usuários se sentiram perdidos em não conseguirem consertar erros de ditado de maneira mais automática, ou seja, pela própria voz como interface, ou seja, os usuários não conseguiam retroagir no ditado. Uma sugestão seria utilizar o mecanismo intrínseco no software DosVox (Projeto Dosvox, 2010), que permite ao usuário voltar e apagar parte do texto já ditado.

7.4.2 Sistema Implantado no Hospital Edmundo Vasconcelos

Uma das razões apresentadas por um dos médicos da equipe de Radiologistas do hospital para o não uso da ferramenta refere-se ao fato dele digitar mais rapidamente do que o sistema consegue transcrever. Este médico também fala muito rapidamente, o que ocasionaria uma queda de *performance* no sistema.

Atualmente este hospital gera, no departamento de Ressonância Magnética, uma média de 120 laudos por dia, através de 8 médicos radiologistas. Existem 2 digitadoras no departamento, que, segundo o médico entrevistado, possuem gargalo de atendimento, devido à escassez de pessoas.

Também, segundo informações do médico, quando o laudo precisa de urgência, é possível fazê-lo em 20 minutos utilizando o serviço de digitação, contra 5 minutos utilizando o sistema de transcrição automática em Radiologia.

O tempo normal para que um laudo seja liberado no hospital é de 3 dias, sendo comum que atrasem até 5 dias.

Outro ponto importante mencionado pelo médico é que, quando o serviço de digitação é utilizado, o médico precisa reaver as imagens médicas para confirmar o laudo; já no uso do sistema de transcrição automática, isso se torna desnecessário devido ao imediatismo com que o laudo fica pronto para uso.

Entre os pontos fracos na observação do uso deste sistema, percebeu-se que a taxa de reconhecimento de voz foi bastante inferior ao apresentado pelo sistema *stand-alone*, conforme foi apresentado no Quadro 71. Como este médico utiliza o sistema de transcrição automática de radiologia de outro hospital, em versão mais atualizada, pode-se concluir que esta versão instalada no Edmundo Vasconcelos capta mais as interferências sonoras externas do que as versões mais recentes – conforme foi explicitamente dito pelo médico.

Também foi percebido e confirmado pelo médico que, nesta versão, o sistema possui grande dificuldade em reconhecer epônimos, tal como “T2” que é reconhecido como “tendões”.

Segundo o médico entrevistado e observado, que trabalha em outro hospital com o mesmo sistema, em versão diferente, as principais melhorias no uso do sistema, em versão superior são:

- Maior taxa de reconhecimento da voz.
- Menor interferência sonora do ambiente.
- Eliminou o problema da versão 2.1 de espaçamento entre palavras.
- É integrado ao sistema PACS e RIS.
- O sistema aprende mais rápido os epônimos.

8. CONCLUSÕES

Este trabalho apresentou uma metodologia para a avaliação de usabilidade de Sistemas de Transcrição Automática de Laudos em Radiologia, que contribui para as necessidades de mercado na escolha de um sistema mais “amigável” ao usuário final. Para tanto, esta metodologia leva em consideração muitos aspectos que vão além da taxa de reconhecimento destes sistemas, tratando questões como tamanho do vocabulário utilizado, ambiente utilizado e satisfação do usuário.

Este capítulo apresenta as considerações finais obtidas com o desenvolvimento de trabalho. O item 8.1 traz os aspectos gerais do trabalho versando sobre a metodologia para a avaliação de usabilidade para Sistemas de Transcrição Automática de Laudos em Radiologia, proposta nesta Tese de Doutorado; o item 8.2 expõe as contribuições significativas deste trabalho; e o item 8.3 descreve os trabalhos futuros passíveis de serem implementados.

8.1 Aspectos Gerais do Trabalho

Este trabalho apresentou uma revisão bastante extensa de temas relacionados à pesquisa, como Interface Homem-Computador, Interface do Usuário Baseada em Voz, Avaliação de Interfaces, Requisitos de Usabilidade para VUI e Sistemas de Transcrição Automática de Laudos. De acordo com este amplo levantamento bibliográfico, foi verificado que não há, ainda, na literatura, trabalhos suficientes sobre como se avaliar, sob o ponto de vista da usabilidade, tais aplicações de maneira metodológica.

Em entrevistas informais com responsáveis pela área de TI e também usuários finais – médicos radiologistas - de dois hospitais-alvo da pesquisa, foi discutida a resistência do médico para lidar com tais sistemas no começo de sua implantação, mas a grande aceitabilidade deste sistema com o passar do tempo. Assim, verifica-se que tais sistemas tendem a ocupar, cada vez mais, o centro radiológico de hospitais em todo o país. Isto permite dizer que a criação de um *guideline* que ajude na escolha de tais sistemas é de suma importância. Atualmente, somente a taxa de reconhecimento de voz é levada em

consideração quando os hospitais e clínicas escolhem um sistema de transcrição automática de laudos. Existem outros tantos requisitos que deveriam ser levados em consideração quando se escolhe comprar um sistema desta classe, tal como a diminuição da carga cognitiva do usuário, já preocupado com sua atividade de realizar um laudo radiológico.

Como apontado no Capítulo 3, mesmo para aplicações de VUI de domínios não específicos, tal metodologia ainda não está solidamente estabelecida. Os trabalhos apontados por (WALKER et al, 1997; GIBBON; MOORE; WINSKI, 1997; DYBKJAER; BERNSEN, 2000) advém de contribuições para se avaliar projetos específicos em domínios restritos.

Um estudo aprofundado sobre VUI foi realizado no Capítulo 2, de onde surgiram vários dos artigos publicados (OLIVEIRA NETO; SALVADOR, 2007; SALVADOR; OLIVEIRA NETO; KAWAMOTO, 2008; SALVADOR, OLIVEIRA NETO; GUIMARAES, 2009), além de um capítulo de livro nacional (OLIVEIRA NETO; SALVADOR; KAWAMOTO; 2010) e dois capítulos de livros internacionais (SALVADOR; OLIVEIRA NETO; PAIVA, 2010; KAWAMOTO; SALVADOR; OLIVEIRA NETO, 2010).

Houve uma preocupação de captar os requisitos específicos para aplicações VUI, tratados no item 4.3, assim como os requisitos estabelecidos para sistemas de transcrição automática de laudos no item 5.5, que deu origem aos requisitos estabelecidos na metodologia descrita no Capítulo 6 (SALVADOR; MOURA JR., 2010a; SALVADOR; MOURA JR., 2010b).

Através desta revisão bibliográfica, visitas em hospitais e entrevistas com profissionais da área de TI na Saúde e de trabalhos práticos realizados pela autora na sua vida profissional na área de VUI, foi possível elencar, analisar, escolher e propor requisitos de usabilidade para a geração de uma metodologia de avaliação de Sistemas de Transcrição Automática de Laudos. Esta metodologia utiliza dois principais meios de se avaliar interfaces: avaliações com especialistas e avaliações com o usuário final.

Esta metodologia foi composta das seguintes partes:

- Identificação e escolha dos requisitos a serem utilizados.
- Classificação destes requisitos quanto ao seu grupo apropriado e grau de dificuldade.

- Identificação de uma ou mais métricas associadas a cada requisito, a fim de gerar dados quantitativos.
- Geração de *templates* que facilitem e organizem a avaliação de cada requisito.
- Revisão dos requisitos.

Esta metodologia consiste em se determinar como será avaliado cada um dos requisitos selecionados e classificados, segundo características como: tipo e método de avaliação, dificuldade e importância da avaliação e métricas que devem ser seguidas.

Com o intuito de validar a metodologia, foram realizadas as coletas dos dados das avaliações realizadas, junto a especialistas através do uso de avaliação por inspeção, sem intervenção do usuário. Também foram realizadas avaliações junto a voluntários não especialistas em Radiologia, através de avaliação por testes de usabilidade. Outra avaliação por testes de usabilidade se deu com um usuário final – um médico radiologista que trabalha em um hospital-alvo da pesquisa que conta com o sistema já implantado.

Os dados foram, então, analisados a fim de ser possível estabelecer os pontos fortes e fracos de tal metodologia e que melhorias pudessem ser apontadas. Esta fase de preparação para os testes e análise consta do Capítulo 7 deste trabalho.

Entre os principais pontos fortes do sistema que foram percebidos e estudados é possível destacar:

- Todos os testes mostraram um alto nível de reconhecimento do sistema, mesmo tendo havido uma fase de treinamento da voz do usuário bastante reduzido.
- O usuário percebeu o sistema como fácil de ser utilizado.
- O usuário apresentou-se bastante interessado no uso da tecnologia.
- O sistema permite que os laudos sejam disponibilizados mais rapidamente.
- Pode diminuir os custos com digitadoras.
- Quando o serviço de digitação é utilizado, o médico precisa reaver as imagens médicas para confirmar o laudo; já no uso do sistema de transcrição automática, isso se torna desnecessário devido ao imediatismo com que o laudo fica pronto para uso.

Por outro lado, foram considerados os seguintes pontos críticos dos sistemas avaliados:

- O *delay* do sistema é um fator crítico para o sistema, pois causa aumento da carga cognitiva.
- A interferência sonora do ambiente é outro fator crítico para o sucesso do sistema e deve ser modificado para estar de acordo com as normas de ruído estabelecidas pela ABNT – NBR 10152/1987 (2010).
- Em relação à falta de visibilidade e adequação do *feedback*, a criticidade se deu pelo sistema nunca informar ao usuário palavras que, possivelmente, tenham sido reconhecidas com baixa taxa. Uma sugestão proposta foi que estas palavras fossem marcadas – tal como o uso de negrito ou sublinhado - para que o usuário conseguisse visualizá-las mais facilmente.
- Sobre a recuperação de erros, os usuários se sentiram perdidos em não conseguirem consertar erros de ditado de maneira mais automática, ou seja, pela própria voz como interface, ou seja, os usuários não conseguiam retroagir no ditado.

8.2 Contribuições do Trabalho

Os Sistemas de Transcrição Automática de Laudos em Radiologia têm sido cogitados como solução para diminuir o tempo para o laudo ficar pronto e também como redutor dos custos globais do departamento de Radiologia. Porém, para se verificar a eficácia de se utilizar tais sistemas, é preciso avaliar se os requisitos gerais de VUI e também os requisitos específicos da área de Radiologia estão sendo tratados de maneira correta nos produtos disponíveis no mercado. Nesse intuito, este trabalho propôs uma metodologia para avaliação de tais sistemas de uma maneira bastante aprofundada e completa.

O trabalho visou a redução de custos com testes de usabilidade, por ser conhecido na literatura de IHC, que este é um custo que pode ser impeditivo da avaliação de muitos sistemas. Assim, procurou-se utilizar, ao máximo possível, as técnicas de inspeção que não utilizam usuários finais e que diminuem significativamente o tempo e os custos das avaliações.

Também, para que os hospitais não necessitassem disponibilizar tempo e incômodos de seus radiologistas, optou-se por utilizar a inspeção realizada por especialistas em usabilidade e, por outro lado, usuários voluntários foram selecionados para os testes de usabilidade, quando a inspeção não fosse o método mais apropriado. Somente a observação dos usuários finais – radiologistas – e o preenchimento, por eles, de um questionário de satisfação foram as técnicas utilizadas para se analisar a usabilidade, quando da necessidade de estar “em campo”. Este questionário não tomou mais que 3 minutos do tempo do radiologista.

O trabalho resultou num conjunto de três componentes para avaliar sistemas de transcrição automática de laudos em Radiologia. Em cada um destes componentes, foi gerado um *guideline* de como testar cada requisito.

Na parte que versa sobre a inspeção de usabilidade, foi criado um *template* que serve como guia para cada requisito que deve ser testado. Este *template* contém: nome da métrica, método de avaliação, participantes, material necessário, roteiro, resultados e análise.

Na parte que versa sobre os testes de usabilidade, foi levemente modificado o *template* utilizado na fase anterior acrescido do item “Orientações aos Participantes”. O conjunto destes *templates* serve como guia para se testar a usabilidade junto a voluntários.

Já na terceira parte, que versa sobre testes de usabilidade por observação e questionários com usuários finais – traz vários formulários que devem ser preenchidos pelo observador quando fizer estes testes, trazendo um *checklist* do que observar.

Sendo assim, este trabalho serve de guia para a área de TI de hospitais e clínicas radiológicas quando se deseja avaliar a compra de sistemas de transcrição automática de laudos, cada vez mais comuns no mercado interno. Também pode ser utilizado para se verificar, quando se trabalhe com customizações de tais sistemas, a usabilidade que se deseja atingir e a que está atualmente vigente.

É desejável que a avaliação de usabilidade proposta através desta metodologia seja realizada, ou comandada, por especialistas em usabilidade, pois é necessária, mesmo para um especialista, uma quantidade bastante grande de horas para, principalmente, a avaliação de inspeção.

Alguns requisitos mostraram-se difíceis de serem testados, pois envolviam um conhecimento muito específico mesmo para especialistas em usabilidade, como é o caso do conhecimento profundo sobre áudio. Um destes requisitos aquele que se refere à “Qualidade do Sistema de Áudio”. Como não se deseja que o avaliador de um sistema como este tenha que adquirir conhecimentos mais aprofundados sobre áudio, optou-se por fazer uma avaliação informal deste requisito, ou seja, o som gravado pelo sistema era reproduzido e o usuário verificava se era inteligível.

Outra medição de requisito que obteve uma resposta aquém do desejável foi “Dicionário específico para RIS”, pois a massa de dados utilizada para os testes é insuficiente para se afirmar que o dicionário inicialmente inserido no sistema de reconhecimento de voz seja considerado aprovado ou reprovado. Para que este requisito pudesse ser caracterizado de maneira apropriada, uma análise de todos os laudos gerados por todos os radiologistas durante certo intervalo de tempo conseguiria medi-lo de maneira mais eficaz.

Também o requisito “Resolução de Homônimos” foi testado, mas seria necessário um teste mais aprofundado: identificação, dentro do dicionário do sistema, de todas as palavras homônimas e então testadas.

Não era o intuito do trabalho descobrir maneiras de automatizar os testes, então muitos requisitos foram extremamente cansativos de serem testados, tais como “Naturalidade da Fala”, necessitando de um tempo bastante grande e monótono para serem efetivados.

Embora Nielsen (1987) tenha abordado que a maioria das inspeções de usabilidade dure em torno de duas horas por participante, estas inspeções demoraram bem mais de vinte horas para serem efetivadas. Esta foi uma questão bastante decisiva na escolha de quando utilizar somente um avaliador e quando utilizar dois.

Algumas limitações nestas avaliações - nas inspeções de usabilidade e também nos testes de usabilidade com voluntários – foram decorrentes do uso de um sistema *stand-alone*, pois:

- Não utilizava a arquitetura cliente-servidor, ficava restrito ao aprendizado de novas palavras.
- Não estava conectado a outros sistemas hospitalares e de consultas.

- Não continha facilidades de acesso, tais como teclas de atalho ou palavras-chave para a ativação de máscaras – frases prontas – de Radiologia.

Outro ponto a ser considerado, é a quantidade de usuários voluntários e finais utilizada nos testes. Mesmo utilizando uma quantidade maior de usuários que Nielsen (1987) verse para os testes de usabilidade, foi percebido que, para a avaliação de VUI, devem-se levar em consideração muitas características intrínsecas da voz, tais como: sotaque, idade do usuário, timbre, velocidade da voz, além de características do próprio ambiente físico - interferência sonora do ambiente – e características das pessoas - resistência ao uso destes sistemas e *expertise* com o uso de computadores.

Um ponto bastante frustrante para a autora neste trabalho foi a dificuldade em conseguir permissão nos hospitais para se fazer a pesquisa de campo. Mesmo utilizando uma metodologia de avaliação que previa pouca interação com os médicos radiologistas, muitas barreiras foram criadas pelos hospitais e não puderam ser ultrapassadas. Um exemplo desta barreira foi conseguir contatos na área de TI nos hospitais de São Paulo que possuíam o sistema instalado. Mesmo depois de realizados os contatos, havia, em alguns hospitais, uma barreira, às vezes intransponível, para conseguir a continuidade do contato com pessoas de TI e também observar os usuários finais utilizando o sistema.

Conforme mencionado no Capítulo 7, o *corpus* utilizado para a pesquisa corresponde a um conjunto de radiologia de tórax, sendo percebido que estes dados são bastante formatados e não correspondem a uma fala natural de um Radiologista. Além disso, este corpus corresponde a uma faixa muito pequena de palavras utilizadas na área Radiológica. Os sistemas atualmente desenvolvidos utilizam cerca de 21.000 palavras, enquanto este corpus continha em torno de 60 palavras por texto. Além disso, versavam somente sobre uma região do corpo. Isso, conforme foi explicitado nos resultados do Capítulo 7, pode causar uma impressão não exata de precisão das palavras reconhecidas.

O objetivo inicial deste trabalho era de que a metodologia pudesse ser aplicada a, pelo menos, dois hospitais; porém teve que ser aplicada a apenas um hospital, por haver esta barreira supracitada. Um amenizador para este problema foi que o médico entrevistado no hospital foco da pesquisa também trabalha no outro hospital que não foi possível fazer a pesquisa de

campo; assim, este médico pode dar uma contribuição significativa das diferenças entre os dois sistemas, a partir do seu ponto-de-vista.

Embora tenham sido encontrados alguns pontos fracos supracitados nesta metodologia para a avaliação de usabilidade de Sistemas de Transcrição Automática de Laudos em Radiologia e também as barreiras impeditivas para um melhor trabalho de campo, é possível dizer que ela contribui significativamente para as necessidades de mercado na escolha de um sistema mais “amigável” ao usuário final, pois leva em consideração os requisitos gerais de VUI, mas agora adaptados aos sistemas de ditado e também aspectos ligados a um sistema específico de Radiologia, focando sempre na satisfação do usuário final.

É interessante lembrar que não existe, ainda, na literatura, uma metodologia similar que leve em consideração requisitos gerais de VUI e de sistemas de ditado em Radiologia. Também é possível destacar que esta metodologia não possui custos financeiros e nem de tempo impeditivos, o que a torna factível.

Obviamente, se o desejo de escolha na compra de um Sistema de Transcrição de Laudos em Radiologia, ou mesmo na melhoria da usabilidade de um sistema já existente, parte da área gerencial ou estratégica do hospital ou clinica, os recursos das equipes de TI e de usuários finais poderão ser mais facilmente incorporados ao processo de avaliação, tornando a metodologia de maior sucesso e sem as barreiras supracitadas.

Vale ainda a pena ressaltar que, sendo a autora da área de IHC, existe um desejo sempre presente que os sistemas disponíveis no mercado, tenham sempre atendidos critérios de usabilidade focados no usuário final. É desejo da comunidade de IHC que os sistemas interativos se adaptem às necessidades do usuário e, não o contrário, como é possível observar em tantos sistemas presentes no cotidiano das pessoas. Sendo a voz uma das formas mais naturais de comunicação entre pessoas, a possibilidade de ser, também, a voz, uma forma mais natural de comunicação entre o homem e a máquina, é necessário que esta forma de interação seja bastante intuitiva e “amigável” aos usuários finais. Sendo assim, a preocupação de se avaliar se os Sistemas de Transcrição Automática de Laudos em Radiologia vem de encontro a esta preocupação da comunidade de IHC.

8.3 Trabalhos Futuros

A primeira contribuição como trabalho futuro é sugerida para que se tenha mais automatização dos testes, tanto com especialistas em Usabilidade, quanto com usuários voluntários e finais, que não era o foco deste trabalho. O uso de agentes inteligentes que possam captar os laudos, ditar os laudos através da sintetização da voz e verificar os pontos de não reconhecimento da voz permitiria que a base de testes fosse relativamente maior. Isso diminuiria imensamente o tempo de avaliação por inspeção de software, considerada um dos pontos problemáticos para esta metodologia.

Alguns problemas encontrados em relação à diferença de grafia entre os laudos utilizados e o laudo gerado pelo sistema, poderiam ser padronizados, a fim de que a verificação das diferenças entre os dois textos não contivesse “pseudo-erros” que tiveram que ser desconsiderados pelos especialistas em avaliação de maneira manual.

A utilização de um *corpus* mais variado, não formatado e amplo da área Radiológica poderia contribuir significativamente para a confirmação ou reajustes de alguns valores coletados, tais como a precisão do reconhecimento da voz.

Além disso, a aplicação desta metodologia a uma quantidade maior de hospitais e clínicas radiológicas que tenha o Sistema de Transcrição Automática em Radiologia poderia dar uma visão mais real de como estes sistemas estão atendendo, atualmente, a critérios ligados à usabilidade. Seria, então, possível propor melhorias ligadas à Usabilidade e geração de políticas de incentivo ao uso destes sistemas nos hospitais.

Esta visão mais real do quadro nacional no uso destes sistemas poderia proporcionar uma melhoria destes sistemas por parte do fabricante, entendendo melhor as necessidades dos usuários finais.

Por outro lado, a aplicação desta metodologia a outras classes de sistemas de reconhecimento de voz poderia ser realizada fazendo alterações nos requisitos específicos da aplicação. Assim, com poucas modificações – inserção e remoção de alguns requisitos específicos -, esta metodologia poderia ser utilizada para se testar outros tipos de sistemas

VUI, tais como sistemas de consultas: sistemas de reservas de passagens aéreas, reserva de hotéis, consulta de horários de vôos e similares.

Pensando numa abordagem mais humana no uso de VUI para a área de Radiologia, seria desejável verificar a existência de - ou construir - um reconhecedor de voz brasileiro que tenha um custo baixo ou inexistente para o seu uso. Assim, seria possível construir um Sistema de Transcrição Automático de Laudos em Radiologia voltado a hospitais com poucos recursos financeiros e que atendam à comunidade de baixa renda. Isso é desejável a fim de agilizar as consultas e os resultados dos exames de Radiologia, proporcionando um maior bem-estar a estas pessoas.

O uso desta metodologia de avaliação, aqui utilizada para avaliar sistemas já disponíveis no mercado, poderia ser adaptada para ser incorporada ao processo de desenvolvimento do software, a fim de melhorar os produtos que são vendidos no mercado, principalmente atendendo os requisitos ligados à satisfação do usuário. Isso seria possível, inserindo os usuários finais no processo de desenvolvimento do software. Este aspecto é bastante comentado pela comunidade de Engenharia de Software, porém, como é de conhecimento da área, sempre são gastos recursos financeiros e de tempo insatisfatórios para melhorar os aspectos ligados à Usabilidade de sistemas.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABNT – NBR 10152/1987. Disponível em <http://www.filecrop.com/NBR-10152.html>. Acesso em 26/12/2010.
- AGFA. Disponível em <http://www.agfa.com/brazil/bu/healthcare.jsp>. Acesso em 23/12/2010.
- AMARAL, M.B.; SALVADOR, V.F.M. Gestão de Sistemas de Informação Hospitalar. In: *Administração Hospitalar – Instrumentos para a gestão profissional*, São Paulo: Loyola, p. 455-475, 2005.
- ALAPETITE, A.; BOJE, A.H.; MORTEN, H. Acceptance of speech recognition by physicians : a survey of expectations, experiences, and social influence, In: *International journal of human-computer studies*, ISSN 1071-5819, v. 67, n. 1, p. 36-49, 2009.
- ALLEN, J., PERRAULT, C. Analysing intention in utterances. In: *Artificial Intelligence* 15, p. 143 - 178, 1980.
- AVOURIS, N.M. An introduction to software usability. Workshop on Software Usability. In: *Proceedings of the 8th Panhellenic Conference on Informatics*. v. 2, p. 514-522, 2001.
- BERNSEN, N.O.; DYBKJAER, L. Usability evaluation in spoken language dialogue systems and their components. In: *Proceedings of the workshop on Evaluation for Language and Dialogue Systems*, v. 9, Toulouse, France, 2001.
- BHAN, S.N.; COBLENTZ, C.; NORMAN, G.R.; ALI, A.H. Effect of voice recognition on radiologist reporting time. In: *CARJ*, v. 59, n. 4, October 2008.
- BOSCH, L.T., OOSTDIJK, N., RUITER, J.P.D. Durational Aspects of Turn-Taking in Spontaneous Face-to-Face and Telephone Dialogues. In TSD(2004) 563-570
- CHAMBERLAIN, J. ELLIOTT, G., KLEHR, M., BAUDE, J. Speech user interface guide, redpaper, s. l.: IBM, 2006.
- CHIN, J. P., DIEHL, V. A., NORMAN, K. L. Development of an instrument measuring user satisfaction of the human-computer interface. In: *ACM CHI'88 Proceedings*, p. 213-218, 1988.

- COHEN, M. H., GIANGOLA, J. P., BALOGH, J. *Voice User Interface Design*, Boston: Addison Wesley, ISBN 0-321-18576-5, 368 páginas, 2004.
- CURTIS, B. How to choose good metaphors. In: *IEEE Software*, Austin, v. 11, n. 3, p. 86-88, 1994.
- DELAMARO, M. E., MALDONADO, J.C., JINO, M. *Introdução ao teste de software*, Rio de Janeiro: Campus, 2007.
- DENG, L., HUANG, X. 2004. Challenges in adopting speech recognition. In: *Commun. ACM* v. 47, n. 1, p. 69-75, 2004.
- DESURVIRE, H. W. Faster, Cheaper!! Are usability Inspection Methods as Effective as Empirical Testing?. In: J. Nielsen (ed.) *Usability Inspection Methods*, New York: John Wiley, 1994.
- DIAH, N.M.; ISMAILI, M.; AHMAD, S.; DAHARI, M.K.M. Usability testing for educational computer game using observation method. In: *CAMP'10 is the first international conference on Information Retrieval and Knowledge Management*. Shah Alam, Malaysia, 2010.
- DYBKJAER, L.; BERNSEN, N.O. Usability issues in spoken language dialogue systems. In: *Natural Language Engineering, Special Issue on Best Practice in Spoken Language Dialogue System Engineering*, v. 6 Parts 3 & 4, p. 243-272, 2000.
- DYBKJAER, L.; BERNSEN, N.O. Usability evaluation in spoken language dialogue systems. In: *Proceedings of the ACL 2001 Workshop on Evaluation Methodologies for Language and Dialogue Systems*, 2001.
- DYBKJAER, L.; BERNSEN, N.O.; MINKER, W. New challenges in usability evaluation - beyond task-oriented spoken dialogue systems. In: *INTERSPEECH-2004*, p. 2261-2264, 2004a.
- DYBKJAER, L.; BERNSEN, N.O.; MINKER, W. Usability evaluation of multimodal and domain-oriented spoken language. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, v. 5, Lisbon, Portugal, p. 1695-1698, 2004b.
- DOSVOX. Disponível em <http://intervox.nce.ufrj.br/m/>. Acesso em: 27/12/2010.

- DURLING, S.; LUMSDEN, J. Speech Recognition Use in Healthcare Applications. In: *Proceedings of MoMM2008*, MoCoHe, ACM 978-1-60558-269-6/08/0011, 2008.
- FENELON, S. Aspectos ético-legais em imaginologia. *Radiol Bras* v.36 n.1 São Paulo Jan, Feb, 2003.
- FISCHER, G. Communication requirements for cooperative problem solving systems. In: *Information Systems*, New York, v. 15, n. 1, p. 21-36. Jan, 1990.
- FOLEY, J. D., VAN DAN, A. *Computer graphics: principle and practice*. Reading: Addison Wesley, 1990.
- GIBBON, D., MOORE, R., WINSKI, R. (eds.): *Handbook of standards and resources for spoken language systems*. Mouton de Gruyter, Berlin, New York, 1997.
- GOPAKUMAR, B. et al. Reengineering radiology transcription process through voice recognition. In: *IEEE 2008 International Conference on Industrial Engineering and Engineering Management*, Singapore, December 8-11, 2008.
- GOLD, J. D., LEWIS, C. Designing for usability: key principles and what designers think. In: *Communications of the ACM*, v.28, n. 3, p.300-311, 1985.
- GUTIERREZ, A. J., MULLINS, M. E., NOVELLINE, R. A. Impact of PACS and voice-recognition reporting on the education of radiology residents. In: *Journal of Digital Imaging*, v. 18, n. 2, June, 2005.
- HARTIKAINEN, M.; SALONEN, E.; TURUNEN, M. Subjective evaluation of spoken dialogue systems using SER VQUAL method. In: *INTERSPEECH-2004*, p.2273-2276, 2004.
- HUANG, X.; ACERO, A.; HON, H.W. *Spoken language processing*. New Jersey: Prentice Hall. ISBN 0-13-022616-5, 2001.
- HUMAN FACTORS AND ERGONOMICS SOCIETY CODE OF ETHICS. Disponível em <http://www.hfes.org/web/AboutHFES/ethics.html>. Acesso em 28/12/2010.
- HUNDT, W. et al. Speech processing in radiology. In: *Eur Radiol* n.9, p. 1451-1456, 1999.
- HUNT, A., WALKER, W. *A fine grained component architecture for speech application development*, Sun Microsystems, Inc., Technical Report: TR-2000-86, Mountain View, CA, USA, 2000.

- IBM Research. Disponível em: <http://www.ibm.com/us/en/>. Acesso em 02/03/2009.
- ICHIKAWA,T. et al. Radiological Reporting That Combine Continuous Speech Recognition with Error Correction by Transcriptionists. In: *Tokai J Exp Clin Med.*, v. 32, n. 4, p. 144-147, 2007.
- IEEE 829-2008. Disponível em http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4578383. Acesso em: 19/12/2010.
- ISO 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs) Part 11: Guidance on usability. ISO 1997.
- KAMM, C., WALKER, M., RABINER, L. The role of speech processing in human computer intelligent communication. In: *Speech Communication* n.23, p. 263 - 278, 1997.
- KANAL, K.M. et al Initial evaluation of a continuous speech recognition program for radiology. In: *Journal of Digital Imaging*, v. 14, n. 1, p 30-37, 2001.
- KAWAMOTO, A.L.; SALVADOR, V.F.M.; OLIVEIRA NETO, J.S. Requirements and guidelines for the evaluation of voice user interfaces. In: Aleksandar Lazinica. (Org.). *User Interfaces*. : InTechweb, 2010, v.1, p. 32-55.
- KIMBERLY, D. V., A methodology of error detection: Improving speech recognition in radiology. PhD thesis, Simon Fraser University, Burnaby, Canada, 2006.
- KOMATANI, K.; UENO S.; KAWAHARA, T; OKUNO, H. G. Flexible guidance generation using user model in spoken dialogue systems. In: *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, p. 256-263, July, 2003.
- LAI, J. Conversational Interfaces. In: *Communications of the ACM*, v. 43, n. 9, p. 24 - 27, September, 2000.
- LAMEL, L.; MINKER, W.; PAROUBEK, P. Towards best practice in the development and evaluation of speech recognition components of a spoken language dialog system. In: *Natural Language Engineering*, v. 6, n. 3-4, United Kingdom Cambridge University Press, p. 305 - 322, 2000.

- LARSEN, L.B. Issues in the evaluation of spoken dialogue systems using objective and subjective measures. In: *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU'03)*, St. Thomas, U.S. Virgin Islands, USA, p. 209-214, 2003.
- LAUESEN, S. *User Interface Design: A software engineering perspective*, Pearson Education, Great Britain, ISBN 0 321 18143 3, 2005.
- LEE, C. *A practitioner's guide to software test design*. Artech House, 2003.
- LEVIN, E., LEVIN, A. Evaluation of spoken dialogue technology for real-time health data collection. In: *Journal of Medical Internet Research*, v.8, n. 4, 2006.
- LIANG, T. User interface design for decision support systems: a self-adaptive approach. In: *Information & Management*, New York, v. 12, n. 4, p. 181-193, April, 1987.
- MACSYM disponível em www.macsym.com.br/medical. Acesso: 10/12/2010
- MARCONI, M.A.; LAKATOS, E.V. *Metodologia científica*. São Paulo: Atlas, 2007.
- McGURK, S.; BRAUER, K.; MACFARLANE, T.V.; DUNCAN, K. A.: The effect of voice recognition software on comparative error rates in radiology reports. In *The British Journal of Radiology*, n. 81, p. 767–770, 2008.
- MCTEAR, M.F. Spoken dialogue technology: enabling the conversational user interface. In: *ACM Computing Surveys*, v. 34, n. 1, p. 90–169, March, 2002.
- MICROSOFT – Reconhecimento de Voz. Disponível em: <http://www.microsoft.com/portugal/windowsvista/features/foreveryone/speech.msp>. Acesso em 10/04/2010.
- MITCHELL, P.P., *A step-by-step guide to usability testing*, Lincoln, NE: iUniverse , 2007.
- MOHR, D.N. et al. Speech recognition as a transcription AID: A randomized comparison with standard transcription. In: *Journal of Medical Informatics*, v. 10, n. 1, p. 85-93, 2003.
- MÖLLER, S. A new taxonomy for the quality of telephone services based on spoken dialogue systems. In: *Proc. 3rd SIGdial Worksh. on Discourse and Dialogue*, US-Philadelphia, p. 142-153, 2002.

- MÖLLER, S. *Quality of telephone-based spoken dialogue systems*, New York, NY: Springer, 2005.
- MYERS, B.A. User interface software tools. In: *ACM Transactions on Computer-Human Interaction*, v. 2, n. 1, p. 64-103, March, 1995.
- NIELSEN, J. Usability engineering at a discount. In: G. Salvendy et al (eds.). *Designing and using human-computer interfaces and knowledge based systems*. Amsterdam: Elsevier, 1989.
- NIELSEN, J. *Usability Engineering*. Academic Press, Cambridge, MA, 1993.
- NIELSEN, J. Heuristic Evaluation. In: J. Nielsen (ed.) *Usability Inspection Methods*. John Wiley, New York, 1994.
- NIELSEN, J. *Designing web usability*. Indianapolis: News Riders, 2000.
- NIELSEN, J. (2000, March 19). Why You Only Need to Test with 5 Users. Disponível em: <http://www.useit.com/alertbox/20000319.html>. Acesso em 12/12/2010.
- NUANCE, Disponível em: <http://nuance.com/naturallyspeaking/>. Acesso em 20/12/2010.
- OLIVEIRA NETO, J. S. de ; SALVADOR, V. F. M. ; KAWAMOTO, A. L. . Aplicações interativas baseadas em voz na Educação: oportunidades e estudo de caso. In: Anita Maria da Rocha Fernandes; Michelle Silva Wangham. (Org.). Livro de Minicursos. Florianópolis, v.1, p. 1-26, 2010.
- OLIVEIRA NETO, J. S.; SALVADOR, V. F. M. Projeto de interfaces baseadas em comandos de voz: conceitos, ferramentas de desenvolvimento e aplicações. ERI-RJ, 2007.
- ORTH, A. I. *Interface Homem-Máquina*. Porto Alegre: AIO, 2005.
- PARASURAMAN, A., ZEITHAML, V.A.; BERRY, L.L., SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. In: *Journal of Retailing*, v.64, n. 1, 1988.
- PAULETT, J.M., LANGLLOTZ, C.P. Improving language models for radiology speech recognition. In: *Journal of Biomedical Informatics*, 42, p. 53-58, 2009.

- PERNICE, K., NIELSEN, J. Eyetracking methodology: how to conduct and evaluate usability studies using eyetracking, NIELSEN NORMAN GROUP, 2009. Disponível em: <http://www.useit.com/eyetracking/methodology>. Acesso em 20/12/2010.
- PHILIPS HEALTHCARE. Disponível em <http://www.healthcare.philips.com/br/>. Acesso em 23/12/2010.
- PRESSMAN, R.S. *Software Engineering*. 7th ed., New York: Addison-Wesley, 2004.
- PRICE, P. Evaluation of spoken language systems: the ATIS domain. In: *Proceedings of the Third DARPA Speech and Natural Language Workshop*, Morgan Kaufmann, 1990.
- RECANTO DAS LETRAS. Disponível em <http://recantodasletras.uol.com.br/gramatica/152254>. Acesso em: 20/12/2010.
- ROBBINS, A.H. et al. Speech-controlled generation of radiology reports. In: *Radiology* n. 164, p. 569-573, 1987.
- ROCHA, H. V., BARANAUSKAS, M. C. C. *Design e avaliação de interfaces humano-computador*, NIED, Instituto de Computação, Unicamp, Campinas, 2003.
- RODRIGUES, R. Manual de pautas para el establecimiento de sistemas locales de informacion. In: *OPAS – Organização Panamericana de Saúde*, 1996.
- RODGER, J. A., PENDHARKAR, P. C., A field study of database communication issues peculiar to users of a voice activated medical tracking application. In: *Decision Support Systems*, v. 43, n. 1, p. 168-180, 2007.
- RUBIN, J.; CHISNELL, D. *Handbook of usability testing: how to plan, design and conduct effective tests*. 2nd ed., New York: John Wiley & Sons, 2008.
- SALVADOR, V.F.M.; OLIVEIRA NETO, J.S.; PAIVA, M. G. Evaluating voice user interfaces in ubiquitous applications. In: Milton Mendes; Pedro Fernandes. (org.). *Designing Solutions-Based Ubiquitous and Pervasive Computing: New Issues and Trends*. Natal, RN: IGI Global, 2010.
- SALVADOR, V.F.M.; MOURA JR, L.A. Heuristic evaluation for automatic radiology reporting transcription systems. In: 10th International Conference on Information Science, Signal Processing and their applications - ISSPA 2010, Kuala Lumpur. Proceedings of 10th

International Conference on Information Science, Signal Processing and their applications, 2010.

SALVADOR, V.F.M.; MOURA JR, L.A. Evaluation methodology for automatic radiology reporting transcription systems. In: 13th World Congress on Medical and Health Informatics - MedInfo 2010, Cape Town. Proceedings of 13th World Congress on Medical and Health Informatics - MedInfo 2010.

SALVADOR, V.F.M.; BRITTO, M.; MOURA JR, L.A.; ALMEIDA JR, J.R. Qualidade de dados para gestão de conhecimento na área de saúde. In: *XI Congresso Brasileiro de Informática em Saúde*, Florianópolis, 2006.

SALVADOR, V.F.M., ALMEIDA FILHO, F.G.V., A área de Informática em Saúde sob aspectos éticos. In: *Revista Cadernos – Centro Universitário São Camilo*, v. 12, n. 2, abril; junho 2006, ISSN 0104-5865.

SALVADOR, V. F. M.; OLIVEIRA NETO, J. S. ; KAWAMOTO, A. L. . Requirement engineering contributions to voice user interface. In: *First International Conference on Advances in Computer-Human Interaction*, Sainte Luce, p. 309-314, 2008.

SALVADOR, V. F. M.; OLIVEIRA NETO, J. S., GUIMARAES, M. P. Levantamento de heurísticas para avaliação de interfaces do usuário baseadas em voz. In: *Interaction South America*, São Paulo, 2009.

SAN-SEGUNDO, R. et al. Knowledge-combining methodology for dialogue design in spoken language systems. In: *International Journal of Speech Technology* n. 8, p. 45-66, Springer Science + Business Media, 2005.

SEBASTIAN, D., Development of a field-deployable voice-controlled ultrasound scanner system, MS Thesis, Worcester Polytechnic Institute, Worcester, Mass., USA, 2004.

SELIGMAN, M., DILINGER, M., Usability issues in an interactive speech-to-speech translation system for healthcare. In: *Proceedings of the First International Workshop on Medical Speech Translation*, New York, NY, USA, June 4-9, 2006.

SHNEIDERMAN, B. The limits of speech recognition. In: *Communications of the ACM*, v. 43, n. 9, p. 24 – 27, September, 2000.

- SHORTLIFFE, E. *Medical informatics – Computer applications in medical care*. Reading, MA Addison-Wesley, 1990.
- SKOV, M. B.; STAGE, J. Supporting problem identification in usability evaluations. In: *Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*, Canberra, Australia, 2005.
- SNOMED. Disponível em: <http://www.ihtsdo.org/snomed-ct/>. Acesso em 07/03/2009.
- SOMMERVILLE, I. *Software Engineering*, 6th ed., Addison Wesley, ISBN 0-201-39815-X, 2001.
- SONNENBERG FA, BECK JR. Markov models in medical decision making: a practical guide. *Med Decis Making*, v. 13, n. 4, pp. 322-38, 1993.
- SPEECH MAGIC™. Disponível em <http://www.myspeech.com/>. Acesso em 20/12/2010.
- SUHM, B., MYERS, B., WAIBEL, A. Multimodal error correction for speech user interfaces. In: *ACM Transactions on Computer-Human Interaction*, v. 8, n. 1, p. 60-98, 2001.
- SUTCLIFFE, A., GAULT, B. Heuristic evaluation of virtual reality applications. In: *Interacting with Computers* 16, New York: Elsevier, 2004.
- VAVASSORI, F.B., LOH, S. Método heurístico para avaliação e projeto de interfaces homem-software. Trabalho de Conclusão, Pelotas: Universidade Católica de Pelotas, 1995.
- VOLL, K., ATKINS, S., FORSTER, B. Improving the utility of speech recognition through error detection. In: *Journal of Digital Imaging*, v. 21, n 4, p. 371-377, 2008.
- VORBECK, F., BA-SSALAMAH, A., KETTENBACH, J., HUEBSCH, P. Report Generation using digital speech recognition in radiology. In: *Eur. Radiol.* 10, Springer-Verlag, p. 1976-1982, 2000.
- WALKER, M. PARADISE: A framework for evaluating spoken dialogue agents. In: *Proc. of the Association of Computational Linguistics (ACL)*, p. 271 - 280, 1997.
- WALKER, M. A., PASSNNEAU, R., BOLAND J. E. Quantitative and qualitative evaluation of Darpa communicator spoken dialogue systems. In: *Proceedings of the 39rd Annual Meeting on Association for Computational Linguistics*, Toulouse, France, 2001.

WHITE, K. S., *Speech recognition implementation in radiology*, New York: Springer-Verlag, 10.1007/s00247-005-1511-x, 2005.

ZUKERMAN, I.; LITMAN, D. Natural language processing and user modeling: synergies and limitations. In: *User Modeling and User-Adapted Interaction* n. 11, p.129-158, Kluwer Academic Publishers, Netherlands, 2001.

APÊNDICES E ANEXOS

**APÊNDICE A - QUESTIONÁRIO DE SATISFAÇÃO DE USUÁRIOS DE
SISTEMAS DE TRANSCRIÇÃO DE LAUDOS EM RADIOLOGIA**

QUESTIONÁRIO DE SATISFAÇÃO DE USUÁRIOS DE SISTEMAS DE TRANSCRIÇÃO DE LAUDOS EM RADIOLOGIA

Este serviço tem como objetivo principal melhorar permanentemente a qualidade de Sistemas de transcrição automática de laudos.

Seção 1: Perfil do Usuário

1: Informações pessoais

1.1 Sexo: () M () F

1.2 Idade: _____

1.3 Nível de escolaridade: () Graduação () Pós-Graduação () Mestrado () Doutorado

1.4 Cidade Natal: _____

2: Experiência do usuário com computadores

2.1 Quantas horas por semana, em média, você utiliza o computador?

() menos de 2 horas

() entre 2 e 5 horas

() entre 5 e 10 horas

() mais de 10 horas

3: Sua experiência com Sistemas de Transcrição de Laudos em Radiologia

3.1 Como você classifica o uso do sistema de transcrição automática de laudos, quando o utiliza

() funciona de maneira incrivelmente boa

() funciona bem

() funciona regularmente

() funciona de maneira deplorável

() nunca utilizei

Seção 2: Avaliação da usabilidade através de Questionário

Avalie o uso do sistema com os critérios descritos abaixo marcando com um X.

Experiência de uso do sistema					
	Concordo Fortemente	Concordo	Indeciso	Discordo	Discordo Fortemente
Esta aplicação tem uma apresentação gráfica agradável e legível					
É fácil entender o que eu preciso fazer na aplicação					
A interface gráfica é bastante intuitiva					
A aplicação informa o que está acontecendo ao usuário (por exemplo, se uma palavra não foi reconhecida durante o ditado)					
Foi fácil aprender a utilizar a aplicação					
Quanto à velocidade de resposta (transcrição do ditado em texto) é suficiente					
Mesmo sem experiência no uso deste sistema, é fácil utilizá-lo					
Acredito que a aplicação funciona de maneira adequada quanto ao reconhecimento da minha pronúncia					
Há mecanismos no sistema, que acesso através de mouse ou teclado, que proporcionam um melhor desempenho na transcrição (tais como acessar frases já prontas através de um menu, enquanto dito)					
Quando erro no ditado, é fácil fazer o sistema reconhecer o que falo e se adaptar.					
O sistema possui atalhos para que eu possa realizar minha tarefa de maneira mais eficiente					
As formatações, tais como “ponto final”, “vírgula”, “negrito”, são de fácil assimilação, não me fazendo pensar muito para utilizá-las.					
Quando faço hesitações, tais como "mmh", "Aaah", o sistema ignora, captando somente palavras pertencentes a meu vocabulário para o laudo.					
É fácil fazer a formatação do texto enquanto dito um laudo, tal como mudar de linha e pontuar.					
O sistema de ajuda conseguiu esclarecer minhas dúvidas sobre o programa.					
Consigo entender perfeitamente o som gravado do ditado que fiz.					
Considerações sobre a aplicação					
Pontos positivos da aplicação:					
Pontos negativos da aplicação:					
Sugestões ou reclamações sobre a aplicação:					

**APÊNDICE B - CARTA DE APRESENTAÇÃO PARA OS
PARTICIPANTES DO TESTE**

CARTA DE APRESENTAÇÃO PARA OS PARTICIPANTES DO TESTE

Prezado(a) Senhor(a),

Esta é uma carta de apresentação sobre os testes de usabilidade a respeito de um sistema. Sou aluna de doutorado da Escola Politécnica da Universidade de São Paulo (USP) sob orientação do prof. Dr. Lincoln de Assis Moura Jr. Estou em fase de testes de um sistema e necessito da colaboração de voluntários para auxiliar nesta etapa de testes.

Venho, através desta, pedir, gentilmente, sua participação como voluntário para o teste do sistema.

Trata-se de um sistema de ditado – ou seja, tudo o que disser utilizando um aparelho com microfone, aparecerá na tela do computador. Este aparelho é semelhante a um microfone para a captação do áudio, e também possui funções de mouse.

Serão disponibilizados alguns laudos radiológicos para os testes sem a identificação do paciente.

Todos os testes serão explicados previamente a cada sessão, depois de explicados, o observador somente coletará os dados com base em suas ações, sem responder a qualquer pergunta.

Serão 4 sessões de testes. Os testes durarão cerca de 45 minutos e conterão: questionário pré-teste, explicação, sessão de testes e questionário pós-teste.

Todos os dados coletados nestes testes serão mantidos em sigilo absoluto. Nenhum experimento colocará o usuário em risco. As imagens serão divulgadas apenas em veículos impressos em artigos científicos.

Atenciosamente,

Valéria Farinazzo Martins

APÊNDICE C – AUTORIZAÇÃO PARA O USO DE IMAGEM E SOM

AUTORIZAÇÃO DE USO DE IMAGEM E VOZ

Pelo presente instrumento, _____,

(nome completo)

_____, _____, _____, portador(a) da Cédula de
(nacionalidade) (estado civil) (profissão)

Identidade RG nº _____, inscrito(a) no CPF/MF sob nº _____,

com endereço à Rua _____, nº _____, complemento _____,

Bairro _____, _____/____ autoriza a autora deste trabalho, Valéria Farinazzo Martins,

RG 22.543019-8,

a **usar, reproduzir e exibir, no Brasil ou no exterior :**

sua imagem e sua voz,

captadas em fotografia ou outros meios aptos à reprodução, autorizando-a também a editá-las, alterá-las ou adaptá-las, a seu exclusivo critério, bem como, sincronizá-las com músicas ou outros sons. A pesquisadora também autorizada a utilizar, reproduzir e exibir a imagem mencionadas, isoladamente ou em conjunto com outras imagens, nos seguintes materiais: folhetos em geral (encartes) , anúncios para revistas e jornais, reportagens, filmes, cartazes, mídia eletrônica, programas televisivos, materiais didáticos (para aulas, palestras, conferências, congressos etc.), e outros suportes materiais que podem ser reproduzidos por terceiros, tais como, DVD, CD-Rom, DVD-Audio, CD etc., todos destinados à divulgação das campanhas e das atividades desenvolvidas pela pesquisadora bem como, dos serviços por ela prestados, que poderá ser feita ao público em geral. A presente autorização é concedida por prazo indeterminado e a título gratuito, definitivo, irrevogável e irretratável, abrangendo o uso, a reprodução e exibição da imagem e da voz acima mencionadas em todos os veículos de comunicação existentes, nada podendo ser reclamado a título de pagamento ou de indenização, obrigando a todos por si e por seus herdeiros.

São Paulo, de de 2010 .

(Assinatura)

Testemunhas:

1- _____ 2 - _____

RG nº

RG nº

APÊNDICE D – PLANO DO OBSERVADOR

PLANO DO OBSERVADOR

Nome da Métrica	Naturalidade da Fala do Usuário
Participante	
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™
Roteiro	<p>a) Através da leitura de laudos:</p> <ul style="list-style-type: none"> i) Verificar a diferença de precisão entre o usuário falar mais pausadamente ou mais rapidamente; ii) Verificar a diferença de precisão entre o usuário falar com tom baixo ou alto; iii) Usuário lê um laudo e usuário fala um laudo sem ler; <p>b) Através da leitura de laudos:</p> <ul style="list-style-type: none"> i) Verificar a diferença de precisão entre o usuário falar mais pausadamente ou mais rapidamente; ii) Verificar a diferença de precisão entre o usuário falar com tom baixo ou alto;
Resultados	
Análise	

Nome da Métrica	Recuperação de Erros
Participante	
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™
Roteiro	<p>Verificar, através de observação, como o sistema age perante os seguintes erros:</p> <ul style="list-style-type: none"> a) O usuário usa palavras que não estão no dicionário da aplicação – um laudo será modificado para atingir este objetivo. b) O sistema não consegue reconhecer o que o usuário dita c) O usuário erra o laudo e deseja refazê-lo.
Resultados	
Análise	

Nome da Métrica	Carga Cognitiva
Participante	
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™

Orientações aos Participantes	Se você não fez o treinamento inicial da fala, faça-o agora; a) Leia um laudo sem a pontuação. b) Leia um laudo e faça a pontuação correta.
Roteiro	
Resultados	
Análise	

Nome da Métrica	Adequação do <i>Feedback</i> e Visibilidade do Sistema
Participante	
Material Necessário	1 laudo de radiologia, 1 equipamento SpeechMike™
Roteiro	<p>a) o sistema não deve fornecer <i>feedbacks</i> que atrapalhem a capacidade de raciocínio do usuário:</p> <ul style="list-style-type: none"> • O sistema interrompe o ditado; • O sistema inicia processos de alertas; <p>b) deve estar presente, de maneira que o usuário, em tempo oportuno, saiba de erros que tenham ocorrido no sistema enquanto ele ditava o laudo:</p> <ul style="list-style-type: none"> • O sistema permite, de alguma forma, que ocorreu uma falha no reconhecimento de alguma palavra;
Resultados	
Análise	

**APÊNDICE E - RESULTADOS DA APLICAÇÃO DO QUESTIONÁRIO
PARA 6 VOLUNTÁRIOS**

RESULTADOS DA APLICAÇÃO DO QUESTIONÁRIO PARA 6 VOLUNTÁRIOS

Seção 2: Avaliação da usabilidade através de Questionário

Avalie o uso do sistema com os critérios descritos abaixo marcando com um X.

Experiência de uso do sistema					
	Concordo Fortemente	Concordo	Indeciso	Discordo	Discordo Fortemente
Esta aplicação tem uma apresentação gráfica agradável e legível	2	3			1
É fácil entender o que eu preciso fazer na aplicação	4	1	1		
A interface gráfica é bastante intuitiva	1	4		1	
A aplicação informa o que está acontecendo ao usuário (por exemplo, se uma palavra não foi reconhecida durante o ditado)	1				5
Foi fácil aprender a utilizar a aplicação	4	2			
Quanto à velocidade de resposta (transcrição do ditado em texto) é suficiente		3		1	2
Mesmo sem experiência no uso deste sistema, é fácil utilizá-lo	2	3			
Acredito que a aplicação funciona de maneira adequada quanto ao reconhecimento da minha pronúncia	4			2	
Há mecanismos no sistema, que acesso através de mouse ou teclado, que proporcionam um melhor desempenho na transcrição (tais como acessar frases já prontas através de um menu, enquanto dito)		1			5
Quando erro no ditado, é fácil fazer o sistema reconhecer o que falo e se adaptar.		2		4	
O sistema possui atalhos para que eu possa realizar minha tarefa de maneira mais eficiente			1	2	3
As formatações, tais como “ponto final”, “vírgula”, “negrito”, são de fácil assimilação, não me fazendo pensar muito para utilizá-las.	3	1		2	
Quando faço hesitações, tais como "mmh", "Aaah", o sistema ignora, captando somente palavras pertencentes a meu vocabulário para o laudo.	1	2			2
É fácil fazer a formatação do texto enquanto dito um laudo, tal como mudar de linha e pontuar.	3	1		2	
O sistema de ajuda conseguiu esclarecer minhas dúvidas sobre o programa.				4	2
Consigo entender perfeitamente o som gravado do ditado que fiz.	6				

Considerações sobre a aplicação

Pontos positivos da aplicação:

É fácil de ser utilizada

É fácil de ser utilizada e amigável

Reconhece a maioria das palavras ditadas na minha velocidade normal de fala.

Interface simples e intuitiva – não necessita de muitos passos para iniciar

Fácil assimilação e memorização

Reconhece bem a voz, mesmo com ruído

Fácil reconhecimento de palavras do vocabulário técnico e do cotidiano

Interface simples e intuitiva

Reconhece bem a voz mesmo pronunciada em tom mais alto e mais baixo, mais rápido e mais lento.

Pontos negativos da aplicação:

A aplicação não reconhece minha voz quando falo baixo;

A aplicação “se confunde” quando falo rápido;

A aplicação não apresenta uma maneira de corrigir um erro;

Não considerei nenhum ponto negativo

Velocidade de transcrição

Não reconhece nada se o volume da voz estiver baixo

Apenas melhorar os pontos negativos

Não reconhece bem palavras fora do vocabulário médico (Radiologia)

Atraso grande

Homônimas são reconhecidas errado, por ex, “ponto” e “.”

O sistema não possui mecanismo de reparo de erros durante o ditado

Sugestões ou reclamações sobre a aplicação:

Tratamento de erros – prevenção: Marcar as palavras que foram escritas com certo grau de indecisão aumentaria o grau de confiabilidade na aplicação.

Pode ser utilizado em muitas aplicações

Falta de atalhos vocais

Tratamento de erro – algum mecanismo para reparar algum erro, solicitado pelo usuário

**APÊNDICE F - PLANO DE OBSERVAÇÃO PARA O USO DO SISTEMA DE
TRANSCRIÇÃO AUTOMÁTICA DE LAUDOS EM RADIOLOGIA**

PLANO DE OBSERVAÇÃO PARA O USO DO SISTEMA DE TRANSCRIÇÃO AUTOMÁTICA DE LAUDOS EM RADIOLOGIA

DELAY DO SISTEMA						
	Medição 1	Medição 2	Medição 3	Medição 4	Medição 5	Média
(Segundos)						

NATURALIDADE DA FALA	
O usuário fala de maneira natural, ou seja, nem muito pausado nem muito rápido	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	
O usuário insere pontuação explicitamente	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	
O sistema coloca pontuação automaticamente	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	
O usuário fez interjeições durante a gravação do laudo	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	
O sistema ignorou as interjeições do usuário	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	
O usuário falou em tom normal	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	

INTERFERÊNCIA SONORA						
	Medição 1	Medição 2	Medição 3	Medição 4	Medição 5	Média
(db)						

TEMPO PARA O LAUDO FICAR PRONTO	
Tempo de ditado:	
Observações:	
Tempo para correções:	
Observações:	
Tempo para impressão:	
Observações:	
Tempo de verificação e assinatura:	
Observações:	
Tempo Total	

RECUPERAÇÃO DE ERROS	
O usuário refez parte do texto depois do ditado	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	
Quantas palavras foram editadas	
Observações:	
O usuário inseriu palavras novas no sistema	<input type="checkbox"/> Sim <input type="checkbox"/> Não

Observações:	
O usuário formatou o texto	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	

SOBRECARGA COGNITIVA	
O usuário olhou para a tela enquanto ditava o laudo	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	
O usuário fez hesitações quando o texto apresentava <i>delay</i> para aparecer na tela	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	
O usuário dita o laudo e já faz a formatação explícita (ditada) do laudo	<input type="checkbox"/> Sim <input type="checkbox"/> Não
Observações:	

**ANEXO A - TABELA DE REFERÊNCIA DE INTERFERÊNCIA SONORA
DESEJÁVEL PARA AMBIENTES**

TABELA DE REFERÊNCIA DE INTERFERÊNCIA SONORA DESEJÁVEL PARA AMBIENTES

São seguidas as disposições da ABNT - NBR 10152 e as normas brasileiras correspondentes (ABNT - NBR 10152, 2010)..

Quadro 73: Medição de Ruído

NB-95 NBR 10152/1987 - Tabela 1	dB(A)	NC
HOSPITAIS	35 - 45	30 - 40
Apartamentos, Enfermarias, Berçários, C.Cirúrgicos	40 - 50	35 - 45
Laboratórios, Áreas para uso do público	40 - 50	35 - 45
Serviços	45 - 55	40 - 50
ESCOLAS		
Bibliotecas, Salas de música, Salas de desenho	35 - 45	30 - 40
Salas de aula, Laboratórios	40 - 50	35 - 45
Circulação	45 - 55	40 - 50
HOTÉIS		
Apartamentos	35 - 45	30 - 40
Restaurantes, Salas de Estar	40 - 50	35 - 45
Portaria, recepção, Circulação	45 - 55	40 - 50
RESIDÊNCIAS		
Dormitórios	35 - 45	30 - 40
Salas de Estar	40 - 50	35 - 45
AUDITÓRIOS		
Salas de Concerto, Teatros	30 - 40	25 - 30
Salas de Conferências, Cinemas, Salas de Múltiplo Uso	35 - 45	30 - 35
RESTAURANTES		
Restaurantes	40 - 50	35 - 45
ESCRITÓRIOS		
Salas de Reuniões	30 - 40	25 - 35
Salas de Gerência, Projetos e Administração	35 - 45	30 - 40
Salas de Computadores	45 - 65	40 - 60
Salas de Mecanografia	50 - 60	45 - 55
IGREJAS E TEMPLOS		
Cultos Meditativos	40 - 50	35 - 45
LOCAIS PARA ESPORTE		
Pavilhões fechados para espetáculos e Atividades Esportivas	45 - 60	40 - 55

**ANEXO B - LAUDOS UTILIZADOS NOS TESTES E INSPEÇÕES
DE USABILIDADE**

LAUDOS UTILIZADOS NOS TESTES E INSPEÇÕES DE USABILIDADE

Quadro 74: Laudos utilizados nos testes

Laudos	Texto
1	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste).</p> <p>Laudos: Osteofitose dorsal. Seios e cupulas diafragmaticas livres. Vascularizacao pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca normal. Aorta alongada. Arco médio normal. Parênquima pulmonar de transparência normal.</p>
2	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste).</p> <p>Laudos: Osteofitose dorsal. Esternorragia. Obliteração dos seios costo-frênicos. Vascularização pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca aumentada. Aorta alongada. Arco médio normal. Estria densa no ápice do hemitorax direito.</p>
3	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste).</p> <p>Laudos: Costela cervical bilateral. Osteofitose dorsal. Obliteração do seio costo frênico posterior esquerdo. Vascularização pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca normal. Aorta alongada com calcificação na crossa. Arco médio normal. Parênquima pulmonar de transparência normal.</p>
4	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste).</p> <p>Laudos: Esternorragia. Prótese de valva mitral.</p>

	<p>Seios e cúpulas diafragmáticas livres. Aumento do hilo direito e da vascularização pulmonar. Imagem cardíaca aumentada. Aorta normal. Arco médio retificado. Parênquima pulmonar de transparência normal.</p>
5	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste). Laudo: Osteofitose dorsal. Obliteração dos seios costo-frênicos posteriores. Vascularização pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca aumentada. Aorta alongada. Arco médio normal. Parênquima pulmonar de transparência normal.</p>
6	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste). Laudo: Estruturas ósseas e tecidos moles da parede torácica, sem alterações. Seios e cúpulas diafragmáticas livres. Vascularização pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca normal. Aorta normal. Arco médio normal. Parênquima pulmonar de transparência normal. mediastino sem alterações. conclusão: tórax normal.</p>
7	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste). Laudo: Osteofitose dorsal. Seios e cúpulas diafragmáticas livres. Vascularização pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca normal. Aorta normal. Arco médio normal. Parênquima pulmonar de transparência normal.</p>
8	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste</p>

	<p>longa (baixo contraste). Laudo: Costela cervical à direita. Seios e cúpulas diafragmáticas livres. Vascularização pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca normal. Aorta alongada. Arco médio normal. Parênquima pulmonar de transparência normal.</p>
9	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste). Laudo: Estruturas ósseas e tecidos moles da parede torácica, sem alterações. Seios e cúpulas diafragmáticas livres. Vascularização pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca normal. Aorta normal. Arco médio retificado. Parênquima pulmonar de transparência normal.</p>
10	<p>Técnica: Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste). Laudo: Estruturas ósseas e tecidos moles da parede torácica, sem alterações. Seios e cúpulas diafragmáticas livres. Vascularização pulmonar normal. Hilos de configuração, topografia e dimensões normais. Imagem cardíaca aumentada. Aorta alongada. Arco médio normal. Parênquima pulmonar de transparência normal.</p>