



UNIVERSIDADE ESTADUAL DO NORTE DO PARANÁ
CAMPUS LUIZ MENEGHEL - CENTRO DE CIÊNCIAS TECNOLÓGICAS
CIÊNCIA DA COMPUTAÇÃO

GABRIEL PETRINI MARQUES

ABORDAGEM DE APOIO AOS TESTES DE
USABILIDADE UTILIZANDO *THINK ALOUD*

Bandeirantes

2016

GABRIEL PETRINI MARQUES

**ABORDAGEM DE APOIO AOS TESTES DE
USABILIDADE UTILIZANDO *THINK ALOUD***

Trabalho de Conclusão de Curso submetido à
Universidade Estadual do Norte do Paraná,
como requisito parcial para obtenção do grau
de Bacharel em Ciência da Computação.

Orientador: Prof. Me. Thiago Adriano Coleti

Bandeirantes

2016

GABRIEL PETRINI MARQUES

**ABORDAGEM DE APOIO AOS TESTES DE
USABILIDADE UTILIZANDO *THINK ALOUD***

Trabalho de Conclusão de Curso submetido à
Universidade Estadual do Norte do Paraná,
como requisito parcial para obtenção do grau
de Bacharel em Ciência da Computação.

COMISSÃO EXAMINADORA

Prof. Me. Thiago Adriano Coleti
UENP – *Campus* Luiz Meneghel

Prof^a. Dr^a. Daniela de Freitas G. Trindade
UENP – *Campus* Luiz Meneghel

Prof. Me. Fabio De Sordi Junior
UENP – *Campus* Luiz Meneghel

Bandeirantes, ___ de _____ de 2016

RESUMO

Os testes de usabilidade demandam muito tempo para a avaliação dos observadores, pois o processo de anotar e depois verificar onde ocorreu alguma falha é trabalhoso, também, o simples fato de um usuário saber que está sendo observado pode causar constrangimento e, conseqüentemente, os testes podem obter resultados inesperados pela equipe de avaliação. O presente trabalho propõe uma abordagem de apoio aos testes de usabilidade utilizando uma técnica de extração de texto para o processamento de arquivos de áudio. A abordagem pode ser aplicada em testes de usabilidade com auxílio de imagens coletadas da tela com o objetivo de encontrar falhas após a interação com usuários tradicionais. Depois de identificadas, as falhas permitem aos projetistas corrigir ou desenvolver novamente tal interface. Após análises dos resultados obtidos, a abordagem demonstra uma redução de tempo em análises dos testes de usabilidade.

Palavras-chave: Extração de Texto, Testes de Usabilidade, Interação Humano-Computador.

ABSTRACT

Usability tests require a lot of time for the evaluation by evaluators, due to the reason that the process of noting and then verifying where a fault has occurred is also laborious, the simple fact that a user knows that it is being observed can cause embarrassment and, consequently, the tests Can get unexpected results from the evaluation team. The present work proposes an approach to support usability tests using a text extraction technique for the processing of audio files. The approach can be applied in usability tests with the help of images collected from the screen in order to find fault after interaction with traditional users. Once identified, flaws allow designers to correct or re-develop such an interface. After analyzing the obtained results, the approach demonstrates a reduction of time in analyzes of the usability tests.

Keywords: Text Extraction, Usability Testing, Human-Computer Interaction.

LISTA DE FIGURAS

Figura 1 - Etapas do processamento de áudio	26
Figura 2 - Fases do reconhecimento de voz	30
Figura 3 - Exemplo de um HMM	31
Figura 4 - Resultado 1ª Verificação.....	43
Figura 5 - Resultados 2ª Verificação	45
Figura 6 - Coletor de dados	47
Figura 7 - Percentual de acertos e erros da ferramenta.....	54

LISTA DE TABELAS

Tabela 1 - Comparativo das bibliotecas Microsoft.Speech e System.Speech.....	39
Tabela 2 - 1ª Verificação/Texto 1.....	41
Tabela 3 - 1ª Verificação/Texto 2.....	42
Tabela 4 - 1ª Verificação/Texto 3.....	42
Tabela 5 - 2ª Verificação/Texto 1.....	44
Tabela 6 - 2ª Verificação/Texto 2.....	44
Tabela 7 - 2ª Verificação/Texto 3.....	45
Tabela 8 - Resultado da Pessoa 1.....	51
Tabela 9 - Resultado Pessoa 2.....	52
Tabela 10 - Resultado Pessoa 3.....	52
Tabela 11- Resultado Pessoa 4.....	53
Tabela 12 - Resultado Pessoa 5.....	54

LISTA DE SIGLAS

ASR	<i>Automatic Speech Recognition</i>
HMM	<i>Hidden Markov Model</i>
ISO	<i>International Organization for Standardization</i>
STT	<i>Speech to Text</i>

SUMÁRIO

1. Introdução	11
1.1 FORMULAÇÃO DO PROBLEMA	12
1.2 OBJETIVOS	12
1.2.1 Objetivo Geral	12
1.2.2 Objetivos Específicos	13
1.3 JUSTIFICATIVA	13
1.4 METODOLOGIA	14
1.5 ORGANIZAÇÃO DO TRABALHO.....	14
2. Fundamentação teórica	15
2.1 Usabilidade	15
2.2 Avaliação de Usabilidade	18
2.3 Teste de Usabilidade.....	20
2.3.1 Protocolo de Testes.....	20
2.3.2 Verbalização	22
2.4 Resultados esperados	23
2.4.1 Resultados Qualitativos	23
2.4.2 Resultados Quantitativos	24
2.5 Coleta, Análise e Interpretação de Dados	25
2.6 Reconhecimento e processamento de fala	25
2.6.1 Sons da fala.....	27
2.6.2 Fonética	28
2.6.3 Reconhecimento de voz por meio do Modelo de Markov Escondido.....	29
2.7 Trabalhos relacionados	31
2.7.1 Um sistema de reconhecimento de fala contínua baseado em modelos de Markov contínuos	31
2.7.2 Reconhecimento de voz para palavras isoladas.....	33
2.7.3 Nitidez de fala dinâmica	34
2.7.4 Um ambiente de avaliação da usabilidade de software apoiado por técnicas de processamento de imagens e reconhecimento de fala.....	36
3. Materiais e Métodos	38
3.1 Técnica para a Extrator de Texto em Arquivos de Áudio	38

3.1.1	Implementação do Extrator de Texto	39
3.1.2	Primeira Verificação do Extrator de Texto.....	41
3.1.3	Segunda Verificação do Extrator de Texto.....	43
3.1.4	Considerações finais sobre o Extrator de Texto	46
3.2	Gravador de Áudio	46
3.2.1	Considerações finais sobre Gravador de Áudio	47
3.3	Ferramenta de Apoio aos Testes de Usabilidade	47
4.	Validação	50
5.	Considerações Finais.....	56
	Referências	57

1. INTRODUÇÃO

Os princípios de usabilidade são muito comentados nos dias atuais, pois muitos aplicativos ainda não seguem esses conceitos, seja por inaptidão técnica ou por ausência de interesse comercial.

Porém, algumas organizações já se atentaram para os inúmeros benefícios trazidos por meio da utilização dos fundamentos de usabilidade (NIELSEN, 1993). Um desenvolvedor de *website* que se preocupa com a usabilidade deve estar atento para manter o usuário no caminho desejado, sem confusões, que poderiam gerar grande frustração, e, conseqüentemente, a saída do usuário do site, ferramenta ou aplicativo.

A principal característica de um sistema interativo é a usabilidade. Segundo a *International Organization for Standardization* (ISO), usabilidade é a medida pela qual um produto pode ser usado por um usuário para alcançar objetivos específicos com efetividade, eficiência e satisfação em um contexto de uso específico (ISO 9241-11). A qualidade de um sistema é o grau em que o sistema satisfaz as necessidades explícitas e implícitas de seus diversos usuários. Essas necessidades das partes interessadas (funcionalidade, desempenho, segurança, facilidade de manutenção, etc.) são, precisamente, o que está representado no modelo de qualidade, que categoriza a qualidade do produto em características e sub-características (ISO 25010).

Interfaces que possuem boa usabilidade permitem que seus usuários alcancem mais produtividade, fazendo com que não tenham barreiras, obstáculos ou ruídos em suas atividades em busca de indicações para utilização da ferramenta, aplicativo ou site (PREECE; ROGERS; SHARP; 2002). Essas indicações devem ser fornecidas pelo sistema, porém, na maioria das vezes isso não é disponibilizado, assim, seus usuários desviam sua atenção em busca de orientações para a execução de suas tarefas em um ambiente externo.

Estes problemas podem ser encontrados com o apoio dos testes de usabilidade, que são métodos para verificar a facilidade de uso de uma interface para seus usuários finais. Esses usuários são incentivados a utilizar um ambiente monitorado, no local de trabalho e/ou laboratório de usabilidade, para permitir que o avaliador identifique problemas de interação do sistema com o usuário. Os problemas podem ser identificados de várias formas, uma delas é a verbalização, na

qual o usuário pode "pensar em voz alta" sobre o que está pensando do site naquele exato momento (CYBIS; BETION; FAUST, 2010).

A verbalização, em inglês, *Think Aloud*, ou pensar em voz alta, é um método no qual o sujeito é requisitado a falar em voz alta durante a execução de uma tarefa (JASPERS *et al.*, 2004), verbalizando seus pensamentos sem descrever ou explicar o que está fazendo (ERICSSON e SIMON, 1993).

Os testes que possuem o auxílio de recursos de áudio levam muito tempo para serem analisados, pois o avaliador necessita escutar todo o áudio, observar se o evento realmente aconteceu em determinado momento e avaliar esse evento de acordo com os princípios de usabilidade.

O objetivo deste trabalho é propor uma ferramenta para auxiliar na identificação de momentos insatisfatórios do usuário durante a realização do teste de usabilidade, realizando a busca de palavras específicas no arquivo de áudio, para reduzir os problemas de interface.

1.1 FORMULAÇÃO DO PROBLEMA

A análise de dados de testes de usabilidade necessita de muito mais tempo do que o próprio teste, pois o avaliador necessita rever todo o registro de áudio para poder achar um momento relevante, ou seja, depois da realização do teste o avaliador leva de 2 a 10 vezes o tempo do teste a mais para analisar cada evento que ocorreu durante o áudio (NIELSEN, 1993).

A análise de áudio dos testes de usabilidade, além da dificuldade para encontrar os eventos relevantes, também consome muito tempo para identificar os acontecimentos importantes do teste.

1.2 OBJETIVOS

Este tópico irá apresentar o objetivo geral e os objetivos específicos do trabalho.

1.2.1 Objetivo Geral

Propor uma técnica para a extração de textos de arquivos de áudio e aplicar essa técnica em testes de usabilidade, permitindo ao avaliador buscar eventos que

sejam relevantes, dentro do arquivo de áudio, por meio de palavras digitadas por ele. Posteriormente, os eventos encontrados podem ser comparados com imagens de *snapshots* registrados durante o teste de usabilidade.

1.2.2 Objetivos Específicos

Tendo em vista atingir o objetivo geral, os objetivos específicos foram atingidos, entre eles:

- Determinar uma abordagem de gravação de dados de áudio;
- Determinar uma abordagem de busca de palavras dentro de arquivos de áudio;
- Determinar uma abordagem para o registro dos *snapshots*;
- Desenvolver uma ferramenta para a prova de conceito;
- Validar a ferramenta com usuários tradicionais em testes de usabilidade;
- Analisar os resultados dos testes.

1.3 JUSTIFICATIVA

Segundo Nielsen (1993), a análise de dados de testes de usabilidade levam de 2 a 10 vezes o tempo do teste. O avaliador necessita rever todo o registro de áudio para poder achar um momento relevante, isso demanda muito tempo, pois, além de achar o momento, deve ser analisado o evento dos registros. Reduzir o tempo de análise pode ser importante para uma possível correção do site ou ferramenta.

A verbalização é uma técnica considerada importante para a realização dos testes, pois permite ao avaliador ter dados reais da opinião do usuário. Segundo Cybis (2010), o conhecimento de conteúdos lógicos e estratégicos, que partem da mente o usuário, são fundamentais para entender o que levou o participante a uma hesitação, a um bloqueio, a um desvio ou a uma situação de erro. Conhecer a opinião de um usuário para determinado produto de *software* é muito importante. Esse dado pode dar indicativos se o produto vai ser bem sucedido ou não, e com a realização dos testes pode ser atribuído o grau de interação do usuário e do software (CYBIS; BETION; FAUST, 2010).

Assim, pretende-se reduzir o tempo de identificação de problemas de interface com o apoio dos registros coletados e analisados, auxiliando o avaliador nos eventos detectados pela ferramenta, quantas vezes forem necessárias.

1.4 METODOLOGIA

A pesquisa se justifica como exploratória, pois segundo Botelho e da Cruz (2013) a pesquisa exploratória tem como principais finalidades: desenvolver, esclarecer e modificar conceitos e ideias, objetivando a elaboração de problemas mais exatos para pesquisas posteriores.

A pesquisa sugere seguir os conceitos de Usabilidade, Avaliação de Usabilidade, Testes de Usabilidade, Protocolo de teste, Coleta, Análise e Interpretação dos Dados, Processamento e Reconhecimento de Fala e Trabalhos Relacionados.

Para o desenvolvimento da ferramenta para prova de conceito, foi escolhida a linguagem *C Sharp (C#)* e o ambiente de desenvolvimento chamado *Visual Studio 2015®*.

Para verificar se a abordagem é realmente eficiente, foram convidados para a realização dos testes, alunos dos cursos de Sistemas de Informação e Ciência da Computação da Universidade Estadual do Norte do Paraná, que não possuam conhecimento específico dos conceitos de usabilidade.

1.5 ORGANIZAÇÃO DO TRABALHO

O trabalho está organizado da seguinte forma. O capítulo 2 apresenta a fundamentação teórica, mostrando os conceitos de Usabilidade, Avaliação de Usabilidade, Testes de Usabilidade, Verbalização, Processamento e Desenvolvimento de Fala e os Trabalhos Relacionados estudados para o desenvolvimento do trabalho. O capítulo 3 detalha o desenvolvimento das ferramentas necessárias para validar a abordagem. O capítulo 4 apresenta a validação da abordagem proposta. E o capítulo 5 demonstra as considerações finais referentes a abordagem.

2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção será apresentada a revisão bibliográfica dos conteúdos utilizados para a realização deste trabalho, tais como Usabilidade, Avaliação de Usabilidade, Testes de Usabilidade, Observação do Usuário, quais os resultados esperados de um teste de usabilidade, coleta dos dados, processamento e reconhecimento de fala.

2.1 Usabilidade

A usabilidade é definida como a qualidade que caracteriza a utilização dos programas e aplicações, ela não é uma qualidade intrínseca de um sistema, porém, depende do grau de utilização dos seus usuários e, também, das características da sua interface. Ela pode influenciar na satisfação de acordo com o nível de experiência de seus usuários, ou seja, a interação com um usuário experiente pode ser satisfatória, enquanto que com um usuário inexperiente pode deixar muito a desejar, assim como a arquitetura do computador, a tarefa realizada e o ambiente podem influenciar na satisfação do usuário (CYBIS; BETION; FAUST, 2010).

A usabilidade também pode ser definida como um requisito de qualidade de software necessário para alcançar a qualidade de um sistema, fazendo com que ele tenha fácil utilização (NIELSEN, 1993).

Para projetar interfaces que proporcionem usabilidade, e, conseqüentemente, a satisfação do cliente, os desenvolvedores devem conhecer seus usuários e suas necessidades, como é a estrutura dos processos cognitivos humanos, e saber que os usuários diferem entre si em termos de inteligência, estilos cognitivos e personalidades (CYBIS; BETION; FAUST, 2010).

Os desafios para projetar interfaces com usabilidade são grandes, pois a diversidade de conhecimento para usuários finais são enormes, porém, o conhecimento de cada usuário evolui com o tempo e com o uso de um sistema, ou seja, conforme os usuários percebem novas funções passam a utilizar o sistema de forma diferente e criam novas expectativas (CYBIS; BETION; FAUST, 2010).

Segundo a ISO 25010, a qualidade de um sistema é o grau em que o sistema satisfaz as necessidades explícitas e implícitas de seus diversos públicos. Essas necessidades estão representadas no modelo de qualidade, que categoriza

a qualidade do produto em características e sub-características (ISO 25010).

Para entender a ISO 25010 deve-se compreender oito características de qualidade, são elas:

- Adequação Funcional: tem como característica principal representar o grau em que um produto ou sistema fornece funções que correspondam às necessidades explícitas e implícitas quando utilizado sob condições específicas;
- Eficiência de desempenho: representa o desempenho em relação à quantidade de recursos usados com condições estabelecidas;
- Compatibilidade: grau em que um produto, sistema ou componente pode trocar informações com outros produtos, sistemas ou componentes, e/ou realizar suas funções necessárias, ao compartilhar o mesmo ambiente de hardware ou software;
- Usabilidade: definida pelo grau que um produto ou sistema pode ser usado por usuários específicos para atingir metas especificadas com eficácia, eficiência e satisfação em um contexto de uso específico;
- Confiabilidade: grau em que o sistema desempenha as funções em condições específicas por um período de tempo específico;
- Segurança: determina o quanto um produto protege suas informações e dados, determinando o grau de acesso de dados apropriado para cada nível de autorização;
- Manutenção: diz respeito ao grau de eficácia e eficiência que o sistema pode sofrer alterações para melhorar, corrigir ou adaptá-lo às mudanças no meio ambiente;
- Portabilidade: diz que o produto ou componente pode ser transferido de um hardware, software ou ambiente operacional para outro em uso.

O estudo de usabilidade pode ajudar a determinar qual o impacto que um sistema ou aplicativo representa para seu usuário. Isso é uma peça-chave para o projeto. Eles determinam o que pode ser melhorado no design de um software. Ao observar o usuário, verificando suas ações, erros e sucessos, pode-se descobrir

quais as perspectivas do usuário (LOWDERMILK, 2013).

Um bom teste de usabilidade consiste em observar o usuário, conforme eles se envolvam com o software. Seu objetivo é medir a eficiência de um recurso ou de um conjunto de recursos de um software, como o tempo para execução ou a quantidade de erros. Também pode ser realizada uma pesquisa que ajude a identificar métricas difíceis de observar como, por exemplo, a satisfação ou percepção (CYBIS; BETION; FAUST, 2010).

Algumas características são fundamentais para que um sistema tenha alto grau de usabilidade. O usuário deve fazer o esforço necessário durante a interação, o sistema deve conter informações adequadas e organizadas, deve ser fácil de aprender a utilizar e fácil de lembrar como se usa, deve ser seguro de operar na variedade de contextos em que será usado e ter utilidade (BENYON, 2011).

Gould e Lewis (1987) desenvolveram quatro princípios para alcançar a usabilidade: São eles:

- Focar desde o início nos usuários e nas tarefas. Os *designers* devem entender quem são os usuários, o trabalho que se espera realizar e em partes, fazer que os usuários participem da equipe de design.
- Medição empírica. Observar, medir as reações dos usuários por meio de cenários impressos e no próprio protótipo do projeto, e analisar.
- Design interativo. Resolver os erros encontrados em cada teste de usuário.
- Todos os fatores de usabilidade devem evoluir juntos e a responsabilidade por todos os aspectos de usabilidade devem estar sob um único controle.

Segundo Benyon (2011), a usabilidade também pode ser vista como um alcance de um equilíbrio entre quatro fatores do design de sistemas interativos que são:

- pessoas;
- atividades que as pessoas realizam;
- contexto em que estão inseridas;

- tecnologias;

Podemos observar dois lados durante uma interação, sendo que ambos podem ser otimizados, um deles é entre as pessoas e as tecnologias que se concentra na interface do usuário, e do outro lado, a relação entre pessoas e as tecnologias considerada como as atividades realizadas e os contextos das mesmas (BENYON, 2011).

2.2 Avaliação de Usabilidade

Avaliar as interfaces de um sistema é uma tarefa essencial para um *designer*, assim podem saber se o produto é agradável, atraente e se possui alguma outra qualidade. Além disso, usuários comuns gostam que um sistema seja fácil de usar e atraente, sempre visando a eficiência, eficácia e satisfação (PREECE; BORGES; SHARP, 2005).

As técnicas de avaliação de ergonomia podem ser classificadas como:

- Avaliações analíticas;
- Avaliações heurísticas;
- Inspeções por lista de verificação (*Check-list*);

Elas são embasadas em verificações e inspeções de aspectos ergonômicos das interfaces que possam causar problemas ao usuário durante as interações (CYBIS; BETION; FAUST, 2010).

A avaliação analítica, é utilizada ainda na concepção de uma Interface Humano-Computador, ela consiste em uma descrição da organização proposta das tarefas interativas, na qual já é possível verificar a consistência, carga de trabalho e o controle do usuário. Isso permite que sejam verificados os aspectos do produto antes mesmo de serem desenvolvidos. Deve ser ressaltado que essa técnica não leva em conta a hesitação do usuário, um erro ou incidente durante a interação (CYBIS; BETION; FAUST, 2010).

As avaliações heurísticas são realizadas por especialistas em ergonomia, são eles quem examina o sistema, verificando os problemas e barreiras que possam atrapalhar o usuário durante a interação. Os avaliador são apoiados por heurísticas ou padrões próprios ou já criados por outros especialistas, como Jakob

Nielsen e Christian Bastien. Podem gerar ótimos resultados quanto à rapidez de avaliação, porém, dependem muito da capacidade do especialista (CYBIS; BETION; FAUST, 2010).

Nas inspeções por meio de listas, são identificados problemas menores e repetitivos das interfaces. Consistem em uma lista de verificação que apresentam conteúdos organizados e pertinentes para a avaliação. Essa técnica possui algumas vantagens, são elas (CYBIS; BETION; FAUST, 2010):

- fornecer conhecimento ergonômico sobre os aspectos a avaliar;
- sistematizar as avaliações em se tratando de qualidade a inspecionar e de abrangência de componentes a inspecionar;
- reduzir a subjetividade normalmente associada a processos de avaliação, e;
- reduzir os custos da avaliação.

É necessário que desenvolva a lista de verificação com muita atenção, pois, listas mal elaboradas podem apresentar questões subjetivas e quantidade insuficiente, o que levam a resultados controversos ao esperado.

As avaliações heurísticas e por lista de verificação, necessitam de um pré-planejamento adequado, que envolvem algumas atividades, como definir equipes e reuniões de preparativos para a avaliação.

Podemos avaliar desde protótipos de baixo nível até de alto nível, e, também, sistemas já totalmente desenvolvidos, ou seja, a avaliação pode ser feita em qualquer etapa do desenvolvimento de um produto (PREECE; BORGES; SHARP, 2005).

Por exemplo, quando realiza-se a avaliação de alguma característica que necessite investigar se todos os requisitos necessários atingiram a meta, o teste deve ser realizado em laboratório. Mas, se realizarmos a avaliação em um ambiente natural do usuário, podem ser avaliados também os aspectos de experiência (PREECE; BORGES; SHARP, 2005).

Realizamos a avaliação de acordo com o tipo do produto, pode ser feita em um produto novo, ou seja, um conceito totalmente novo, ou um produto já inserido no mercado que está recebendo atualizações (PREECE; BORGES; SHARP, 2005).

2.3 Teste de Usabilidade

Os testes de usabilidade têm como foco de avaliação a qualidade das interações entre os usuários e o sistema. O objetivo é identificar as causas dos problemas na interface, e ver qual o impacto negativo sobre as interações. São realizados com usuários finais, ou com representativo do usuário-alvo do sistema, eles utilizam um sistema utilizando uma função específica em um contexto simulado (CYBIS; BETION; FAUST, 2010).

Para a realização dos testes, devem ser pensadas algumas características importantes que podem influenciar nos resultados obtidos, ou seja, cada etapa dos testes de usabilidade deve ser bem pensada (CYBIS; BETION; FAUST, 2010).

2.3.1 Protocolo de Testes

A realização dos testes é separada em três fases, análise contextual, montagem dos testes e execução.

Durante a análise contextual os avaliadores definem os parâmetros básicos dos testes no que se refere ao tipo de verbalização, onde serão realizados os testes e os resultados a serem obtidos, fazem um conhecimento do software e também fazem um pré-diagnostico na interface. São realizadas sessões de entrevistas com os desenvolvedores em busca de informações sobre o contexto de operação do projeto e do atual estado de desenvolvimento do projeto. Os avaliadores examinam o aplicativo, conhecem suas funcionalidades e identificam funções mais problemáticas (CYBIS; BETION; FAUST, 2010).

Na montagem dos testes devem ser definidos quais os perfis dos participantes, ou seja, qual a população-alvo do software a ser avaliado, qual o roteiro das tarefas que serão avaliadas e qual o local para a realização do teste.

Os participantes podem ser pessoas que realmente vão utilizar o sistema, podem ser considerados desde os iniciantes, que vão informar sobre a facilidade de aprendizagem e simplicidade de utilizar o sistema, até os mais experientes, que vão mostrar as informações de pertinência e a organização das funções e das informações. (CYBIS; BETION; FAUST, 2010).

Para se aproximar ao máximo de uma situação real, devem ser levados em consideração aspectos do contexto da utilização do sistema de cada usuário a formação, competência e experiência dos usuários, qual o perfil de utilização, o

ambiente operacional, ou seja, qual tipo de treinamento caso exista, e quais elementos da tarefa(CYBIS; BETION; FAUST, 2010).

Para definir o roteiro do protocolo de teste, devem ser selecionadas tarefas que envolvem os objetivos do *software*, as funções consideradas mais importantes, as funções que serão utilizadas frequentemente e as funções selecionadas no pré-diagnóstico. O tempo máximo de um roteiro deve ser de uma hora, pois o cansaço do participante pode influenciar no resultado final (CYBIS; BETION; FAUST, 2010).

O ambiente do teste deve ser montado o mais semelhante possível com o ambiente real onde o sistema vai ser utilizado pelos usuários finais, a partir de informações coletadas durante o reconhecimento do software. Por exemplo, quando o teste de usabilidade é destinado a um software de escritório, o ambiente deve reproduzir um local parecido com armários, escrivaninhas e até algumas interrupções, como telefones tocando (CYBIS; BETION; FAUST, 2010).

Antes de iniciar o teste de usabilidade, também deve ser realizado um teste-piloto para verificar se tudo o que foi previsto está funcionando, como por exemplo, a ferramenta para os registros, os questionários para o usuário e o funcionamento do sistema a ser avaliado. A execução é coordenada por ergonomistas que devem saber como proceder em caso de constrangimento por parte do participante, ou até mesmo interromper o teste por conta do mesmo. São feitas anotações do desempenho, dos erros e incidentes verificados. Durante o teste, junto ao participante, pode-se ter um facilitador, que pode desempenhar algumas tarefas e intervir conforme orientação dos ergonomistas, suas tarefas podem ser, interromper uma tarefa, ajudar o usuário, solicitar que ele passe para outra tarefa, encerrar o teste, e, também, solicitar que verbalize os pensamentos (CYBIS; BETION; FAUST, 2010).

Depois de executado, os avaliadores devem identificar cada evento relevante para verificar a usabilidade, analisar cada um e depois diagnosticar problemas de ergonomia na interface. Os resultados devem ser documentados e entregues aos projetistas, apresentando-lhes (CYBIS; BETION; FAUST, 2010):

- quais os tipos de testes realizados no que diz respeito a verbalização;
- local onde foi realizado;
- qual a forma de registro;
- informações do usuário;

- roteiros;
- como o ambiente estava montado;
- gestão do constrangimento;
- dados coletados;
- resultados obtidos relacionados a medidas de usabilidade;
- lista dos problemas de usabilidade observados;
- lista de problemas de ergonomia identificados na interface.

2.3.2 Verbalização

A verbalização surgiu na psicologia e é um método usado para apoiar os testes de usabilidade (HAAK; JONG; SCHELLENS, 2003). É considerado um dos métodos mais utilizados para o teste de usabilidade (BOREN; RAMEY, 2000; HAAK; JONG; SCHELLENS, 2003; HERTZUM; HANSEN; ANDERSEN, 2009; NIELSEN, 1993). Os pesquisadores Boren e Ramey em 2000 criaram o método de *Think Aloud* que consiste em três níveis de verbalização, que podem ser aplicados de formas diferentes e se diferenciam na carga mental do participante.

No nível 1 a verbalização acontece de maneira direta entre o que o usuário pensa e o que necessita falar. Por exemplo, o participante pensa em um número qualquer e verbaliza sem utilizar uma grande carga mental para transformar em número.

No nível 2 o participante já tem uma carga mental maior, pois o participante deve pensar em uma palavra e depois verbalizar. Por exemplo, o usuário está visualizando algum produto em um determinado site, e ele cria uma opinião sobre isso, então ele necessita realizar a transformação dessa ideia em palavra.

Já o nível 3 é considerado o mais complexo, pois a transformação se refere a lembranças ou cálculos mentais.

A verbalização é fundamental para entender o conteúdo lógico e estratégico que está na mente do usuário, pois essas informações auxiliam o avaliador a entender o que levou o participante a ter um bloqueio ou um desvio.

Existem dois tipos de verbalização (CYBIS; BETION; FAUST, 2010; PREECE; BORGES; SHARP, 2005; NIELSEN, 1993). (1) Na verbalização simultânea os usuários verbalizam o que estão pensando durante a interação, são requisitados a dizer o que estão pensando, o que estão tentando fazer, e o porquê disso. As respostas são anotadas ou armazenadas, e depois são revistas e mostradas aos

projetistas para que sejam localizadas funções que não foram compreendidas pelo participante. Essa técnica de verbalização deve ser muito bem exposta ao usuário antes do teste, pois pode causar uma sobrecarga cognitiva, porque o usuário deve dividir a atenção na atividade, pensar e verbalizar o pensamento. Porém é considerada simples e eficiente, pois, o participante sempre estará focado em sua atividade, caso o participante deixe de verbalizar o avaliador pode chamar a atenção do usuário para que ele verbalize algo (CYBIS; BETION; FAUST, 2010).

(2)Na verbalização consecutiva o participante é questionado sobre suas ações após os testes, junto aos registros de vídeo da interação, pois, ajuda a lembrar de suas razões e de suas expectativas para cada procedimento. Também são aceitos comentários sobre a interface, que podem ser boas sugestões ou críticas negativas sobre a interface ou parte dela. Porém, esse método de verbalização duplica o tempo de teste e utiliza equipamentos específicos(CYBIS; BETION; FAUST, 2010).

2.4 Resultados esperados

Os resultados dos testes de usabilidade podem ser qualitativos ou quantitativos. Os dois tipos são derivados de análises, que para este caso deve ser muito minuciosa. Demandam muito trabalho por parte dos analistas. A análise qualitativa pode ser feita com o apoio de software.

2.4.1 Resultados Qualitativos

Resultados qualitativos podem confirmar comportamentos esperados ou inesperados dos usuários, mostram o que leva o usuário a algum desvio ou como tratam situações de incidente (CYBIS; BETION; FAUST, 2010). Não são expressos de forma numérica, por exemplo, descrições, citações do entrevistado (PRECEE; BORGES; SHARP, 2005).

Durante uma análise qualitativa podem ser observados padrões ou ter impressão geral dos dados. São três tipos de análise, identificar padrões recorrentes e temas, categorizar os dados e analisar incidentes críticos.

A identificação dos padrões podem ser observadas a partir de quando o avaliador vai se familiarizando com os dados.

Os dados são categorizados a partir das entrevistas ou protocolos de verbalização, podem ter alto nível de detalhamento, ou um nível de detalhe fino, no

qual cada palavra, expressão ou gesto é analisado. Os dados são divididos em elementos, e cada elemento é categorizado. A categoria varia de acordo com a finalidade da pesquisa, mas, as categorias devem ser significativas (PRECEE; BORGES; SHARP, 2005).

A procura por incidentes críticos, na área do design de interação, tem como foco em identificar incidentes específicos significativos, e posteriormente analisá-los detalhadamente, junto a outros dados coletados que servirão de base para a interpretação. Também podem ser identificados pelos usuários em uma discussão de um evento que houve anteriormente ao teste ou pelo próprio observador durante o teste ou por meio do vídeo do teste (PRECEE; BORGES; SHARP, 2005).

2.4.2 Resultados Quantitativos

Os resultados quantitativos classificam e contabilizam a frequência e a duração dos acontecimentos em que o usuário obteve eficácia e eficiência, determinam a quantidade de usuários que obtiveram sucesso durante a tarefa, quanto tempo levou, a quantidade do tempo em que estava realizando tarefas produtivas e improdutivas (CYBIS; BETION; FAUST, 2010). São expressos em formas numéricas, ou que possam ser transformados facilmente para números (PRECEE; BORGES; SHARP, 2005).

Para obter resultados quantitativos são aplicadas algumas técnicas, que são as médias e as porcentagens. Utilizar porcentagem ajuda a padronizar os dados, principalmente se quiser comparar dois ou mais resultados. As médias são divididas em média, mediana e moda. A média diz respeito a soma dos números e dividí-los pelo número total. A mediana diz respeito ao número do meio dos dados referentes aos números classificados. E a moda é o número que mais aparece nos dados encontrados (PRECEE; BORGES; SHARP, 2005).

Antes dos dados serem analisados, é necessário que eles sejam reunidos em conjuntos de dados analisáveis. A análise inicial envolve descobrir médias, produzir gráficos dos dados para obter uma visão geral deles, identificar os padrões que os englobam. As análises também podem ajudar a identificar outras áreas que podem ser verificadas (PRECEE; BORGES; SHARP, 2005).

2.5 Coleta, Análise e Interpretação de Dados

Os dados coletados podem ter várias abordagens. As mais comuns são: anotações, gravação de áudio, fotografias e vídeos. Porém, nesse trabalho, serão utilizadas a gravação de áudio, imagens faciais e *snapshots*. As gravações de áudio permitem que o avaliador se concentre na atividade que o participante esteja realizando e não na verbalização produzida pelo usuário, fazendo com que preste mais atenção no entrevistado em vez de perder tempo em ouvir e anotar o que está se passando durante a interação entre o usuário e a interface. Copiar todo o conteúdo de um áudio nem sempre é necessário, mas, em alguns testes é necessária essa cópia. Um detalhe importante é que quando a abordagem for a gravação de áudio como a principal técnica utilizada, é importante que a qualidade seja boa (PRECEE; BORGES; SHARP, 2005).

O tipo de análise depende de qual abordagem dos dados será utilizada, que pode influenciar no objetivo da abordagem de apoio aos testes de usabilidade. A partir dos resultados destes, obtemos resultados quantitativos ou qualitativos, ou também a combinação dos dois, que pode gerar resultados abrangentes (PRECEE; BORGES; SHARP, 2005).

Na fase de interpretação os resultados devem ser apoiados pelas conclusões feitas pelo observador/avaliador, e não podem ser influenciados por pré-julgamento do avaliador. Além disso, é comum cometer alguns erros durante a interpretação, como fazer afirmações que vão além dos dados, também, tomar cuidado nas conclusões ao utilizar palavras como "muito" ou "frequentemente". O observador de maneira alguma deve influenciar no resultado, ele deve ser imparcial e ser objetivo para obter resultados verdadeiros (PRECEE; BORGES; SHARP, 2005).

2.6 Reconhecimento e processamento de fala

Um teste de usabilidade que utiliza a verbalização *Think Aloud* faz com que o participante de um teste de usabilidade diga o que está pensando ou o que está fazendo durante aquele determinado momento do teste. Nessa hipótese a utilização do reconhecimento de voz permite que o avaliador ganhe tempo para observar mais o usuário e deixe de anotar o que o participante verbalize.

Os estudos em reconhecimento de voz, ou reconhecimento automático da fala (*Automatic Speech Recognition - ASR*), avançaram muito nos últimos anos. Já existe uma ampla diversidade de sistemas que utilizam o reconhecimento de fala que envolvem o uso de vários níveis de vocabulários (REBINER; JUANG, 1993).

Antes de citar os tipos de sistemas reconhecedores, é importante saber quais etapas, ou técnicas, acontecem durante o reconhecimento da fala, conforme ilustrado na Figura 1. Primeiramente é a entrada, ou seja, a coleta do áudio, por meio de microfone ou por um arquivo de áudio já existente, em seguida é realizado o processamento, em outras palavras, a transformação do som de sinal analógico para sinal digital, e por fim, a fala é retornada por meio de texto, junto com seu instante e sua confiança que é sempre disponibilizada pela ferramenta ou biblioteca que realiza o processamento de áudio.

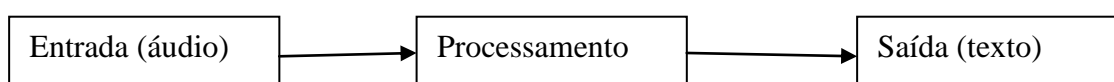


Figura 1 - Etapas do processamento de áudio

Existem três tipos de reconhecimento de voz, são elas (REBINER; JUANG, 1993):

- Reconhecimento de palavras isoladas: é simples, porém, seu vocabulário pode ser pequeno (até 10 palavras) ou médio (de 10 até 1000 palavras). Para os vocabulários pequenos a chance de acerto é de 99%, já para os vocabulários médios é de 95% das palavras. Porém, as palavras devem ser ditas pausadamente.
- Reconhecimento de palavras conectadas: é mais complexo do que o anterior, utilizam fonética padrão e fazem o reconhecimento de forma natural, onde não existe pausa entre as palavras, porém a frase deve ser bem pronunciada.
- Reconhecimento de voz contínua: que é o mais complexo entre eles, exige mais capacidade do programador ao implementar, pois, deve tratar todas as características e vícios de linguagem.

Para desenvolver sistemas de reconhecimento de voz é necessário observar 5 fatores que influenciam na implementação do sistema, são eles (LARA; GRANDE; SANCHES, 1991):

- **Locutor:** que neste trabalho é representado pelo participante. Ele nunca vai pronunciar uma locução sempre da mesma maneira devido à diversas situações. Existem diferenças entre os locutores, pois diferem em gênero, idade, situação social ou região de origem. Porém, para os testes deste trabalho, serão selecionados usuários de uma faixa etária específica e de uma região específica. A abordagem será validada por participantes de um mesmo grupo social.
- **Forma de falar:** o ser humano pronuncia de forma contínua e produz o som com efeitos coarticulatório. Junto às variações introduzidas pela pronúncia, as palavras possuem uma diferença entre uma mesma palavra dita no início, no meio e no fim de uma frase.
- **Gramática:** a existência de uma gramática dentro de um sistema reconhecedor de voz, ajuda a aumentar a taxa de reconhecimento, ao eliminar ambiguidades e diminuir a carga computacional ao limitar o número de palavras em uma fase de reconhecimento.
- **Ambiente:** este pode ser fundamental para o reconhecimento, pois ele é o responsável pela adição do ruído. Os ruídos podem ser ocasionados por tosses, respiração forte ou até mesmo por pessoas que estão no mesmo local. Por isso, é necessário implementar um sistema robusto que trate o ruído.

Para obter bons resultados no reconhecimento de voz, é necessário ter um conhecimento das características das unidades fonéticas, que são descritos nas seções seguintes.

2.6.1 Sons da fala

Um fonema é a menor unidade sonora do sistema fonológico de uma língua (TUFANO, 1998). A mudança de um fonema em uma palavra, a transforma em outra palavra.

Abaixo, alguns conceitos segundo Tufano (1998):

- **Fonética:** é a parte da gramática que se preocupa em estudar a parte sonora da Língua.

- Fonema: é a menor fração sonora , representado dentro destas barras / ê/, /é/, /ã/.
- Sons vocálicos: são fonemas que ao serem pronunciados, não sofrem pressão nenhuma no aparelho reprodutor fonético. Experimente pronunciar /a/, /é/, /i/, /ó/,/u/.
- Vogal e Semivogal: toda sílaba em Língua Portuguesa tem sempre uma vogal. Se tiver mais de um som vocálico na mesma sílaba, um deste som vocálico será Semivogal, aquela pronunciada com menor intensidade.
- Consoante: é o fonema produzido com a pressão, ou interferência do aparelho fonador, em qualquer uma de suas partes. De acordo com o local que é pressionado é que é classificado.
- Oclusivas: são consoantes pronunciadas com a interrupção do ar, ao pronunciá-las existe a explosão de ar em algum local do aparelho, são elas as oclusivas: /p/, /t/, /q/, /b/, /d/, /g/.
- Fricativas: lembra a fricção. O ar passa exprimido por duas articulações: ex. lábios / dentes. Na prática, lembra um sopro de vento.

2.6.2 Fonética

A unidade fonética empregada por um sistema reconhecedor de voz pode ter uma entre várias características (BISPO, 1997), e para a escolha das unidades, devem ser analisados estes dois critérios:

- Consistência: diz que as características devem ser similares.
- Treinabilidade: as amostras para o treinamento e para a criação de um modelo com bons resultados nos testes devem ser suficientes.

São utilizados vários tipos de unidades, serão descritas as mais utilizadas seguindo esses dois critérios (BISPO, 1997).

Palavras. São caracterizadas como a unidade mais natural da voz, pois são elas que queremos reconhecer. O treinamento do sistema é fortemente influenciado pelas unidades básicas, sendo assim demonstram ter mais desempenho. Porém, o tempo de busca dessas unidades em vocabulários grandes pode ser lenta. A medida em que o vocabulário aumenta, conseqüentemente, o tempo de busca aumenta (BISPO, 1997).

Fones são unidades fonéticas menores que palavras.

Fones independentes do contexto. É assumido que o fone de um contexto é igual a outro fone de qualquer contexto. O que não é verdade, pois uma palavra não é apenas uma concatenação de fones (BISPO, 1997).

Fones dependentes da palavra. O contexto das palavras depende das palavras onde os fones ocorrem. Portanto, se uma palavra não aparece frequentemente no treinamento, os parâmetros podem ser alternados com os parâmetros dos sistemas construídos a partir dos fones independentes do contexto (BISPO, 1997). Segundo Bahl *et al.* (1988), estes sistemas possuem maior desempenho perante aos sistemas baseados em palavras.

Difones e Trifones (Fones dependentes do contexto). São semelhantes aos fones dependentes da palavra, porém não se considera a palavra como um contexto, e sim como vizinhos. Existem 3 modelos para considerar contextos diferentes: (BISPO, 1997)

- modelos de difones à esquerda: que consideram somente o fone à esquerda;
- modelos de difones à direita que: somente consideram o fone à direita;
- modelos de trifones: que consideram à esquerda e à direita do fone.

Segundo Bahl (1988) a utilização dos sistemas embasados em trifones possuem uma taxa de erro reduzida em 50% em relação aos sistemas baseados em palavras.

2.6.3 Reconhecimento de voz por meio do Modelo de Markov Escondido

A busca por padrões na área de reconhecimento de voz pode ser caracterizada como um problema, pois, os padrões são achados a partir de vários parâmetros que definem um sinal de voz. Para realizar o reconhecimento de padrões, conforme ilustra a figura 2, é necessário verificar duas etapas, que são o treinamento e o reconhecimento a partir de comparações. Depois das duas etapas surge o conhecimento, que é definido a partir de quando a voz for conhecida, através dos padrões. Na fase de comparação, são comparadas a fala com os padrões já conhecidos do sistema que foram absorvidos na fase de treinamento do

sistema e depois são classificados de acordo com a qualidade e o acerto dos padrões (RABINER, L; JUANG, B.; 1993).

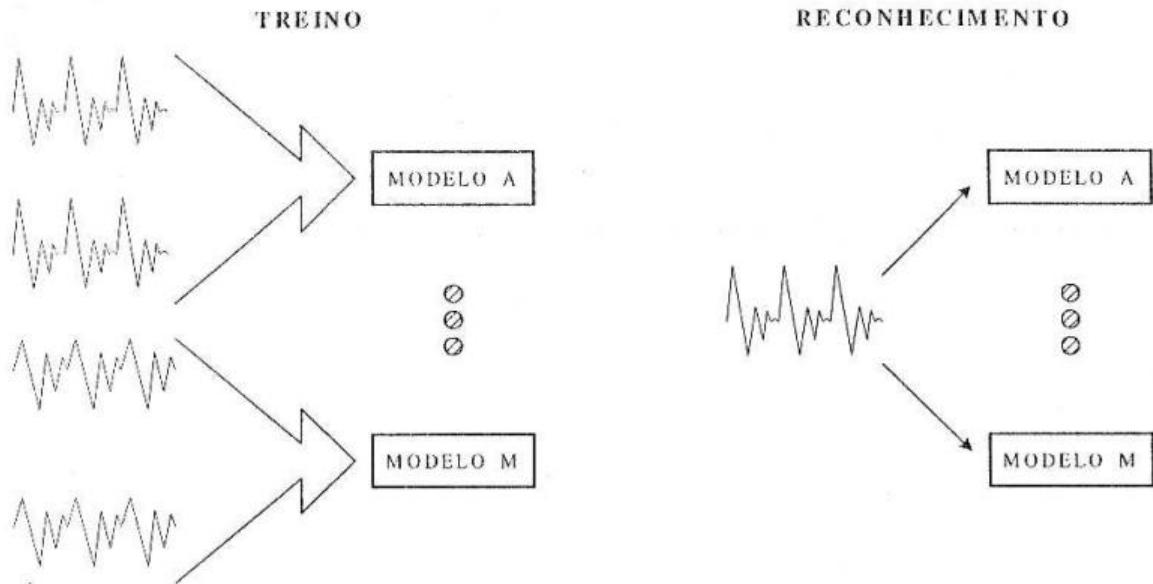


Figura 2 - Fases do reconhecimento de voz

O Modelo de Markov escondido, ou em inglês, *Hidden Markov Model*(HMM) é um método para reconhecimento de voz. O HMM é utilizado para representar as unidades da fala. Ele é utilizado para reconhecer palavras em vocabulários pequenos e sub-palavras em grandes vocabulários.

O modelo é excelente para a representação da fala, devido a carga computacional, eficiência e flexibilidade.

O HMM, tem demonstrado eficiência na caracterização das propriedades temporais e espectrais do sinal de voz, seu uso é baseado nas seguintes afirmações (GALES, 1991):

- A fala pode ser segmentada, dividida em estados.
- A probabilidade de um novo treinamento ser gerado depende do estado atual.

Os HMMs podem ser utilizados para representar qualquer unidade da fala. Para sistemas de reconhecimento de voz com vocabulários pequenos, são utilizados para modelar diretamente as palavras. Já para os sistemas com vocabulários grandes são utilizados para modelar sub-palavras.

Este modelo é representado, pelas máquinas de estados finitos, que são: conjunto de nós (estados) e arcos (transições entre os estados). Utiliza uma

topologia chamada esquerda-direita, na qual só ocorre transição para o mesmo estado ou para o estado seguinte, não é permitido que pule algum estado (RABINER, 1989).

A Figura 3 demonstra um exemplo de um HMM, que poderia ser utilizado para uma palavra pequena, composta por apenas três partes estacionárias. Neste caso, busca explicar alguns parâmetros que são responsáveis pela caracterização: o conjunto de estados q_n , as densidades de probabilidade de emissão, $p(x_i/q_n)$, associadas a cada estado e as probabilidades de transição, $p(q_j/q_n)$, para cada transição permitida, do estado q_n para o estado q_j .

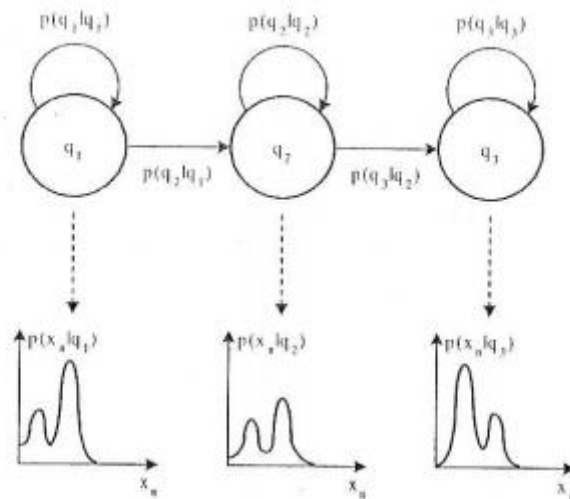


Figura 3 - Exemplo de um HMM

2.7 Trabalhos relacionados

2.7.1 Um sistema de reconhecimento de fala contínua baseado em modelos de Markov contínuos

Este trabalho foi desenvolvido no Laboratório de Processamento Digital de Fala da Faculdade de Engenharia Elétrica e de Computação da UNICAMP. Seu principal objetivo foi desenvolver um sistema para reconhecimento de fala que fosse independente do locutor, reconhecesse fala contínua e que pudesse ter vocabulários grandes. Os desenvolvedores do sistema também disponibilizaram o trabalho para que mais pesquisadores pudessem usá-lo a fim de reduzir o tempo de outras pesquisas paralelas (YNOGUTI, VIOLARO; 2000).

O sistema é baseado nos modelos de Markov contínuos, e foi modelado por sub-unidades fonéticas. A avaliação foi realizada por um total de 40 locutores, dividindo igualmente os gêneros, ou seja, 20 homens e 20 mulheres. Cada locutor teve que pronunciar 200 frases e delas foram identificadas 694 palavras diferentes, criando um vocabulário de porte médio.

O sistema possui quatro módulos, são eles: o primeiro módulo é de extração de parâmetros, na qual é captado o sinal de voz no formato WAV ou em binário. O segundo módulo é de treinamento, na qual é utilizado o algoritmo de treinamento chamado Baum-Welch. Para a realização do treinamento é necessário viabilizar ao sistema algumas informações, como por exemplo: as sub-unidades a serem utilizadas, a transcrição das locuções, utilizando estas sub-unidades e as locuções de treinamento em formato WAV. Depois é gerado o modelo de linguagem a ser utilizado. E por último é o módulo de reconhecimento, que consiste em fazer o mapeamento dos parâmetros acústicos correspondentes às entradas em suas ortografias. Nesta etapa devem ser fornecidos ao sistema, além da locução, o modelo de linguagem, ou seja, a gramática empregada, também os modelos HMM das sub-unidades acústicas, o vocabulário e o modelo de duração de cada palavra.

A base de dados utilizada para este trabalho foi criada a partir de 40 locutores, já citados acima. Todas as gravações, ou seja, as pronúncias dos participantes foram realizadas em um ambiente silencioso, com microfone de boa qualidade. A transcrição fonética foi realizada manualmente para cada locução. Todas as sub-unidades utilizadas no trabalho somam um total de 36.

Os testes foram realizados com e sem modelos de duração de palavras. Foram realizados quatro tipos de testes, com dependência e sem dependência do locutor, e com locutores masculinos e femininos. Que visaram estabelecer qual a melhor configuração para o sistema e o seu desempenho em termos de taxa de acerto e tempo de processamento. O modelo de duração das palavras proporcionou melhor desempenho do sistema para todos os testes, menos para os testes dependentes do locutor.

O sistema foi testado exaustivamente e os pesquisadores já visavam novas ideias e algoritmos para empregar no sistema. Os autores visam ao estudo de novas maneiras para reduzir o tempo de processamento e reconhecimento de voz.

2.7.2 Reconhecimento de voz para palavras isoladas

Este trabalho, realizado na Universidade Federal de Pernambuco, teve como objetivo desenvolver um sistema de reconhecimento de voz com alta taxa de acerto. Para aplicações com esta finalidade, é ideal que se inicie com vocabulários pequenos, isso permite desenvolver uma base de conhecimento necessária para trabalhar em aplicações complexas. Assim como o trabalho relacionado anterior, é utilizado o modelo de Markov (SILVA, 2009).

Para poder realizar o processamento, cada unidade é dividida em quadros, assim como a duração de tempo, extraindo de cada quadro os seus parâmetros, para depois criar os HMM's para as palavras do banco de dados. O trabalho foi implementado no MATLAB.

O processamento é dividido em algumas etapas, elas serão descritas abaixo.

Na fase inicial é realizado um pré-processamento, uma análise espectral onde o som é analisado e determina a sua frequência e intensidade da pessoa que está falando e depois a extração dos parâmetros da voz. Na base de dados, cada exemplar é representado por um conjunto de vetores, que inclui os parâmetros extraídos da fala.

Posteriormente, é feito o treinamento do sistema, é considerada pelos autores como a principal parte do processamento, pois poderá decidir se os resultados obtidos serão bons ou não. Ainda nesta etapa são previstos o HMM para cada palavra do vocabulário.

Na fase de reconhecimento, quando é realizada uma ação, ou seja, quando é feita uma locução por algum participante, para descobrir qual o modelo que tem mais chance de gerar. Ainda não são realizados cálculos e nem a geração do modelo, é feita somente a extração das características que serão utilizadas no algoritmo de reconhecimento.

Determinaram que o algoritmo utilizado é o *forward* que indica a probabilidade de cada modelo de palavra ser uma certa palavra dita pelo participante. Então o modelo em que determinar a maior probabilidade é escolhido como correspondente à palavra falada, levando em consideração que todas as palavras tem a mesma probabilidade de ser reconhecida. Também pode ser utilizado

o algoritmo de *Viterbi*, porém, este só faz uma aproximação para a probabilidade das observações.

Para obter resultados aceitáveis, foram realizados vários testes, para cada teste foi trocado o número de estados e o número de misturas gaussianas dos estados, com o objetivo de conseguir um número ideal destes parâmetros.

A construção da base de dados é determinante para o desempenho do sistema.

Os experimentos foram realizados de duas maneiras diferentes, e sua base foi dividida em dois grupos: uma para treinamento e outra para testes. No treinamento, foi feito com 1080 locuções e no grupo de testes, foram realizados 480 locuções. Os experimentos, tiveram por objetivo, encontrar um padrão para o número de estados, e assim verificar qual o tamanho do modelo ideal para utilizar no reconhecimento das palavras.

Este trabalho buscou desenvolver um sistema de reconhecimento de voz para palavras isoladas, além de facilitar a pesquisa para trabalhos futuros. Também foram observados alguns aspectos importantes para o reconhecimento, como, a influência do ambiente para a precisão do reconhecedor, e que o algoritmo que verifica os extremos da fala, ou seja, o final da pronúnciação.

Os resultados foram satisfatórios, mas os autores sugerem algumas melhorias, são elas: gerar bases de dados maiores e em locais em que o sistema vai ser realmente utilizado, utilizar vocabulários diferentes e pesquisar outras técnicas de pré-processamento para testar no sistema proposto por eles.

2.7.3 Nitidez de fala dinâmica

Este trabalho¹, consiste em um sistema para o reconhecimento de voz, que para cada usuário pode gerar uma ou várias interpretações da verbalização, na qual irá reconhecer os fonemas do som captado. É utilizada uma gramática para fazer a ligação dos fonemas com as sílabas e palavras, e também para incluir novos elementos de ligação e reduzir o tempo de busca na gramática. As interpretações podem passar por diversas técnicas de processamento para que seja escolhida a que tenha mais confiabilidade.

¹ <https://www.google.com/patents/US7634409>

A área de reconhecimento de voz tem sua demanda cada vez maior, e nos celulares isso é cada vez mais utilizado. Em vez dos usuários utilizarem os dispositivos móveis, com entradas manuais, ou seja, através de botões, podem passar a usar através da fala, de forma prática e eficiente. Porém, muitos métodos, ou algoritmos não são eficientes.

O objetivo deste trabalho é reduzir o tempo de processamento de áudio e também interpretar com mais precisão a fala humana. O grande problema é o reconhecimento para vocabulários em sistemas embarcados, que pode ser melhorado com o reconhecimento a partir de uma sequência de fonema. O sistema utiliza uma ou mais técnicas de processamento, para chegar a melhor confiabilidade possível.

Este mecanismo pode gerar várias interpretações, que tenham maiores resultados em confiabilidade da verbalização. As informações obtidas são armazenadas em matrizes. Durante a execução do sistema, o participante vai emitir o som, ou seja, uma sequência fonética, e em cima disso é comparado com um modelo de semelhanças para identificar um modelo que esteja mais próximo do que foi pronunciado pelo participante. Cada exemplar que o participante emitir, pode ser novamente interpretado para que possa aumentar o nível de confiabilidade e aumentar a precisão do sistema.

Existe um módulo de política que consiste em colocar um ou mais agentes para representar o conhecimento. Esses agentes competem entre si, utilizando um modelo para revisar uma interpretação preliminar para um determinado contexto. O agente também pode adicionar parâmetros de domínio para determinar uma possível interpretação.

O sistema também contém um módulo de análise de histórico de interpretações que ajuda o sistema a adicionar agentes de domínio com base em uma análise dos reconhecimentos das verbalizações anteriores. Também possui um módulo de frequência, que calcula a frequência que cada palavra já foi dita, quais as possíveis combinações entre palavras, fonéticas pronunciadas, e erros na interpretação da verbalização.

Toda informação gerada é armazenada afim de aperfeiçoar as futuras interpretações.

2.7.4 Um ambiente de avaliação da usabilidade de software apoiado por técnicas de processamento de imagens e reconhecimento de fala

Este trabalho, realizado na Universidade de São Paulo, teve como objetivo desenvolver uma abordagem de apoio a testes de usabilidade utilizando verbalização e as reações faciais dos participantes. Com o apoio destes, após um processamento, a abordagem disponibiliza as imagens das interfaces do momento, antes e depois de um evento, com o intuito de permitir uma análise do evento registrado. O objetivo principal da abordagem é reduzir o tempo de identificação de eventos negativos durante uma interação Humano-Computador (COLETI, 2013).

A abordagem consiste em quatro etapas: definir a estrutura do sistema, coleta, validação e análise dos dados.

Definir a estrutura do sistema consistiu em identificar ferramentas que pudessem ser úteis ao sistema. Também foram analisados recursos que pudessem se enquadrar na ferramenta, para evitar o desenvolvimento de novos algoritmos complexos.

A ferramenta foi desenvolvida na linguagem de programação C Sharp (C#) no ambiente de desenvolvimento *Microsoft Visual Studio Express*. Este ambiente permite fácil integração com componentes selecionados para apoiar os reconhecedores faciais e de fala. Para o reconhecimento de fala foi utilizada a ferramenta Coruja, devido a sua gratuidade, capacidade de integração com ambientes de desenvolvimento e o material disponibilizado pelos criadores. Para apoiar o reconhecimento facial, utilizou a biblioteca *OpenCV* que realiza o processamento em tempo real ou *offline*. Foi desenvolvida na linguagem C/C++ e assim como a ferramenta Coruja permite integração com o ambiente de desenvolvimento escolhido.

A coleta de dados foi realizada com a utilização da ferramenta *ErgoSV* desenvolvida para prova de conceito da abordagem proposta, que possui os recursos necessários para aquisição das imagens faciais, *snapshots* e do áudio. Depois serão processados e registrados em uma base de dados para posteriormente estarem a disposição dos avaliadores que irão fazer as observações juntos aos dados processados.

A validação foi realizada para verificar a eficácia da abordagem, portanto, foram realizados testes de usabilidade, com a ferramenta batizada como *ErgoSV*, os

testes foram aplicados em um site de comércio eletrônico, um site de um curso de pós-graduação e um software de edição de imagens por 12 estudantes de computação e usuários tradicionais.

Depois de coletados os dados, o objetivo era verificar se a abordagem de fato iria reduzir o tempo dos testes de usabilidade para identificar possíveis problemas de interface.

O trabalho permitiu o desenvolvimento de uma ferramenta para a avaliação de usabilidade, que permite a identificação de interfaces com problemas de usabilidade ou consideradas agradáveis. Também foi notada a eficiência no auxílio dos eventos onde o participante deveria avaliar a interface para identificar possíveis problemas. As imagens também obtiveram resultados satisfatórios pois auxiliaram a identificar qual o local em que poderia estar um determinado problema de interface. O objetivo principal foi atingido, pois foi possível reduzir o tempo de análise dos dados. Porém, foi identificado que seriam necessários mais dados para a análise de imagem, pois as imagens não puderam mostrar a posição para classificar se o participantes estavam ou não satisfeitos.

Portanto, o ambiente demonstrou limitações, são elas: as palavras chaves dependiam de uma configuração do dicionário fonético, e a forma de configurar podia influenciar no reconhecimento. Os participantes realizaram os testes de forma tensa, pelo fato de estarem realizando experimentos e por serem observados.

3. MATERIAIS E MÉTODOS

Para o desenvolvimento do presente trabalho, foi realizada uma pesquisa sobre técnicas de processamento de áudio para apoiar na geração de dados de um teste de usabilidade com o intuito de reduzir o tempo de análise dos dados dos testes de usabilidade e métodos para gravação de dados de áudio para realização da coleta de dados destes testes.

Para poder alcançar os objetivos determinados neste trabalho, foram desenvolvidos um extrator de texto e um gravador de áudio para coletar os dados necessários para a realização deste.

Nas seções seguintes serão detalhados o desenvolvimento das ferramentas e apresentados os resultados de uma verificação do extrator de áudio.

3.1 Técnica para a Extrator de Texto em Arquivos de Áudio

Inicialmente, ficou estabelecido que seria desenvolvido uma abordagem de extração de texto, porém, devido as dificuldades para esse desenvolvimento, ficou determinado o uso de recursos computacionais que facilitassem o reconhecimento de áudio.

Antes de iniciar o desenvolvimento do extrator de texto, foram realizadas pesquisas sobre o processamento de fala. Como linguagem de programação foi definida a linguagem C# (C Sharp), porém, não foi seguido esta definição devido a dificuldade em encontrar uma ferramenta ou biblioteca que realizasse o reconhecimento de voz. Então, se iniciou uma busca por bibliotecas ou ferramentas que realizassem o reconhecimento de voz e convertesse em texto, em inglês, *Speech to Text* (STT), porém, houve muita dificuldade em encontrá-las.

Após algumas buscas, foram encontradas as bibliotecas *Microsoft.Speech* e *System.Speech*, houve uma grande dificuldade em entender o funcionamento e a diferença entre as duas bibliotecas, então se iniciou um estudo para entende-las e definir qual delas seria utilizada no extrator de texto.

Para melhor entendimento, a Tabela 1 possui as diferenças entre as duas bibliotecas.

Tabela 1 - Comparativo das bibliotecas Microsoft.Speech e System.Speech

Microsoft.Speech	System.Speech
Instalação separada	Parte do Sistema Operacional
Utilização de gramáticas	Usa gramáticas ou ditado livre
Sem treinamento	Treinamento específico
Possui Português	Não possui Português

Para verificar qual das bibliotecas se enquadrava mais no perfil do extrator, foi desenvolvida uma ferramenta utilizando a biblioteca *System.Speech*, e outra ferramenta utilizando a biblioteca *Microsoft.Speech*. Após desenvolvidas as ferramentas, as duas foram comparadas a fim de verificar com qual delas o extrator de texto seria utilizada. Então, ficou decidido que a biblioteca *Microsoft.Speech* seria utilizada, por não necessitar de treinamento e por possuir a língua portuguesa, assim deixou-se de lado a biblioteca *System.Speech*, justamente por necessitar de treinamento específico para cada usuário, e só realizar o reconhecimento em inglês e outras línguas, sem reconhecer o idioma português.

3.1.1 Implementação do Extrator de Texto

Antes do início do desenvolvimento necessitou da instalação de um motor de reconhecimento, que se encontra na página². Foram instalados motores que realizam o reconhecimento em Inglês e Português. Depois de instalados, se iniciou o desenvolvimento do extrator. Para realizar o reconhecimento são realizados os seguintes passos:

- Realizar a instância de um `SpeechRecognitionEngine`;
- Definir a entrada;
- Criar e carregar uma gramática;
- Criar um manipulador de eventos, e;
- Iniciar o reconhecimento.

Para um melhor entendimento dos passos descritos, cada um deles serão apresentados:

² <https://www.microsoft.com/en-us/download/details.aspx?id=27224>

- O primeiro passo é realizar uma instância `SpeechRecognitionEngine`, "`SpeechRecognitionEngine sre = new SpeechRecognitionEngine();`", todos os métodos da parte de reconhecimento estão contidas nela.
- Após, deve-se definir a entrada, pode ser definido o microfone, e o reconhecimento é realizado no instante, porém, neste caso será por meio de áudio no formato `.wav`. A entrada é feita da seguinte forma: "`sre.SetInputToWaveFile(caminho desejado);`".
- Então, o próximo passo passa a ser criar uma gramática e carregá-la. Na gramática serão adicionadas todas as palavras que o avaliador deseja encontrar dentro do arquivo de áudio e então ela é carregada para que o reconhecedor possa fazer seu trabalho. O avaliador determina quais palavras serão reconhecidas pela ferramenta antes de iniciar a avaliação do teste. O reconhecimento só acontece com as palavras previamente estabelecidas na gramática, ou seja, o extrator de texto só procura no áudio as palavras contidas na gramática. Seus comandos são: "`GrammarBuilder gb_resultado = new GrammarBuilder();`", "`gb_resultado.Append(ch_palavraDigitada);`", "`Grammar resultado = new Grammar(gb_resultado);`" e "`sre.LoadGrammarAsync(resultado);`", que cria uma gramática, adiciona as palavras desejadas na gramática, e carrega a gramática para poder iniciar o reconhecimento.
- Posteriormente, é criado um manipulador de eventos, que no extrator, ele vai capturar de cada evento encontrado, a palavra, desde que ela esteja na gramática, a confiança em que a palavra foi dita, e o tempo em que a palavra foi pronunciada. O início do processamento é realizado pelo comando: "`sre.SpeechRecognized += new EventHandler <SpeechRecognizedEventArgs>(leitor);`". O parâmetro "leitor" é o manipulador de eventos.

Para o manuseio do extrator, o usuário deve escolher a língua desejada, português ou inglês, selecionar o áudio em que irá extrair os textos, adicionar a gramática todas as palavras em que deseja buscar, e iniciar o processamento. Após ter o áudio processado, a ferramenta dispõe os números de ocorrências encontradas de cada palavra previamente adicionada a gramática. Então, o usuário poderá selecionar uma das palavras e verificar em quais momentos elas ocorreram, e, também verificar qual o grau de confiança encontrada pela ferramenta.

Após desenvolvida a ferramenta, foi iniciada uma verificação para observar qual o grau de acerto da ferramenta. Que será descrito na seguinte seção.

3.1.2 Primeira Verificação do Extrator de Texto

Para verificar a eficiência do extrator de texto, foram selecionado três textos aleatoriamente, para que quatro pessoas realizassem a leitura e, posteriormente, a ferramenta realizar a busca de 10 palavras selecionadas aleatoriamente. Abaixo seguem os resultados dos 3 textos.

Vale lembrar que as células denotadas por "X" são os resultados em que a ferramenta não encontrou nenhuma palavra no áudio de determinada pessoa, já as denotadas por "Errado" são os resultados em que o extrator encontrou alguma palavra, porém encontrou a palavra errada, e por fim denotado por um tempo/confiança, que indica qual o tempo de duração em que a palavra foi encontrada e qual o grau de confiança em que a ferramenta encontrou a palavra. Deve ser ressaltado que os métodos da biblioteca que indicam o tempo e confiança de cada palavra encontrada.

Na tabela 2, pronunciaram rapidamente, fazendo com que uma palavra dita ficasse muito próxima à palavra seguinte, possivelmente, causando um baixo índice de acerto , ou seja, apenas 27,5% de acertos.

Tabela 2 - 1ª Verificação/Texto 1

Texto 1				
Palavra	Pessoa 1	Pessoa 2	Pessoa 3	Pessoa 4
Canto	X	00:02.55/50%	00:02.41/50%	X
Completa	Errado	X	Errado	Errado
Alegre	X	00:08.06/50%	00:08.46/50%	00:12.70/50%
Poeta	Errado	Errado	X	Errado
Irmão	X	X	Errado	Errado
Vento	X	00:20.44/50%	Errado	Errado
Desmorono	X	Errado	X	00:33.13/50%
Permaneço	X	X	00:22.82/50%	Errado
Passo	00:20.25/69,56%	Errado	Errado	00:47.34/50%
Eterno	X	00:35.37/79,37%	Errado	Errado

Já na tabela 3, onde são apresentados os resultados do segundo texto, os participantes também verbalizaram de forma rápida e obteve um número abaixo do que o texto 1, apenas 17,5% das palavras pronunciadas.

Tabela 3 - 1ª Verificação/Texto 2

Texto 2				
Palavra	Pessoa 1	Pessoa 2	Pessoa 3	Pessoa 4
Distante	X	X	00:04.57/73,91%	X
Malvada	X	Errado	Errado	X
Prova	X	X	X	X
Floresta	01:01.40/50%	01:46.19/88,54%	X	01:38.86/69,18%
Anoitecer	X	Errado	X	X
Anões	X	02:53.49/65,43%	Errado	X
Acordou	X	03:06.41/89,67%	X	X
Apresentaram	Errado	Errado	X	X
Feliz	Errado	Errado	X	Errado
Envergonhado	Errado	03:30,65/58,91%	Errado	Errado

E o ultimo texto, conforme a tabela 4, obteve uma média de 20% a 40% das palavras procuradas.

Tabela 4 - 1ª Verificação/Texto 3

Texto 3				
Palavra	Pessoa 1	Pessoa 2	Pessoa 3	Pessoa 4
Alemanha	Errado	Errou	X	X
Biblioteca	Errado	X	00:24.17/64,30%	X
Regularidade	Errado	X	X	01:00.36/81,92%
Instituição	Errado	00:54.04/30,18%	X	X
Individualista	01:16.53/79,18%	01:35.63/83,75%	Errou	01:49.55/93,53%
Dimensões	Errado	X	02:50.20/46,57%	Errado
Questionamentos	02:41.52/80,49%	X	03:16.95/74,56%	01:46.36/71,19%
Desenvolvimento	X	X	Errou	X
Intelectual	03:41.63/67,34%	Errou	04:28.28/80,43%	X
Científicas	Errado	X	X	X

Abaixo segue o resultado em porcentagem para melhor entendimento.

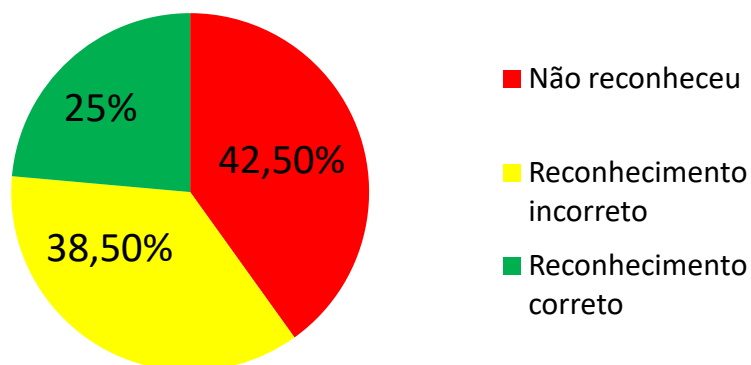


Figura 4 - Resultado 1ª Verificação

Esta primeira verificação, não demonstrou resultados esperados, pois a porcentagem de acerto da ferramenta foi de apenas 25%. Após algumas observações foi possível notar que a forma de falar e a velocidade influenciam fortemente no resultado da verificação, assim como possíveis ruídos próximos do local onde os áudios foram gravados, que possivelmente atrapalharam no momento do reconhecimento.

Portanto, ficou definido que seria realizada uma nova verificação, onde cada participante foi orientado a verbalizar o texto de forma pausada e o ambiente possuiu um controle maior em relação à movimentação de pessoas e aparelhos ligados. Na próxima seção será descrita a segunda verificação.

3.1.3 Segunda Verificação do Extrator de Texto

Antes de realizar esta segunda verificação, foram realizados testes com pronúncia pausada e clara, e o extrator apresentou alto índice de acerto. Então tendo em vista o baixo índice de acertos do extrator de texto, foi realizada uma segunda verificação, na qual cada participante realizou a leitura dos mesmos textos, que foram utilizados na primeira verificação. As palavras escolhidas também foram as mesmas, pois o objetivo desta segunda verificação era ver o quanto aumentaria a taxa de acerto de acordo com a velocidade e a clareza em que o participante pronunciava o texto.

Abaixo estão disponíveis as tabelas com os resultados da verificação. Como já descrito na primeira verificação, os resultados com "X" são os que não encontraram a palavra dentro do arquivo de áudio, com "Errado" são os resultados

que encontraram, porém estavam errado. E por último denotado por um tempo/confiança, onde encontrou corretamente a palavra pronunciada.

Já no Texto 1, da tabela 5, é possível notar uma taxa de acerto maior que em qualquer outra da primeira verificação. Os participantes foram orientados em se preocupar também com a pronúncia correta para cada palavra, lembrando que eles não sabiam quais palavras foram escolhidas para a verificação, com isso, o índice de palavras procuradas aumentou para 60%.

Tabela 5 - 2ª Verificação/Texto 1

Palavra	Pessoa 1	Pessoa 2	Pessoa 3	Pessoa 4
Canto	00:01/56%	00:03/87%	00:02/77%	X
Completa	Errado	00:10/78%	00:08/90%	00:10/94%
Alegre	00:08/44%	00:12/66%	X	00:13/79%
Poeta	X	00:15/50%	X	00:17/68%
Irmão	Errado	00:17/73%	00:14/65%	00:19/66%
Vento	X	00:27/50%	Errado	X
Desmorono	X	00:29/87%	00:26/78%	00:34/77%
Permaneço	00:25/37%	00:33/86%	00:32/84%	Errado
Passo	Errado	Errado	00:39/80%	00:48/50%
Eterno	00:36/75%	Errado	X	Errado

Como no primeiro texto, o segundo, da tabela 6, obteve uma boa taxa, porém a pessoa 3 e 4 tiveram números extremamente baixos, que possivelmente, podem ter sido atrapalhados por falas próximo a gravação ou devido a pronúncia do participante.

Tabela 6 - 2ª Verificação/Texto 2

Texto 2				
Palavra	Pessoa 1	Pessoa 2	Pessoa 3	Pessoa 4
Distante	Errado	Errado	Errado	Errado
Malvada	00:33/75%	Errado	00:40/60%	00:36/67%
Prova	Errado	Errado	X	Errado
Floresta	01:27/78%	01:23/80%	X	Errado
Anoitecer	02:20/84%	02:49/90%	X	Errado
Anões	X	X	Errado	Errado
Acordou	02:36/77%	03:05/83%	X	Errado
Apresentaram	02:46/90%	03:18/80%	X	Errado
Feliz	02:49/93%	Errado	X	Errado
Envergonhado	02:57/61%	03:32/79%	Errado	Errado

E no texto 3, da segunda verificação, como apresenta a tabela 7, somente a pessoa 3 obteve baixo número de acertos. A pessoa 4 teve todas as palavras identificadas, pois foi claro e alto tom da voz, contribuindo para o índice de 60% das palavras encontradas.

Tabela 7 - 2ª Verificação/Texto 3

Texto 3				
Palavra	Pessoa 1	Pessoa 2	Pessoa 3	Pessoa 4
Alemanha	00:06/57%	00:04/92%	00:02/73%	00:04/93%
Biblioteca	Errado	X	00:19/77%	00:24/69%
Regularidade	01:37/91%	Errado	X	01:04/37%
Instituição	01:49/92%	01:38/79%	X	01:13/51%
Individualista	02:52/91%	02:47/87%	X	02:01/91%
Dimensões	Errado	04:00/69%	X	02:45/58%
Questionamentos	05:30/83%	Errado	X	03:46/86%
Desenvolvimento	06:45/60%	06:18/31%	Errado	04:34/77%
Intelectual	07:28/83%	Errado	Errado	04:56/84%
Científicas	Errado	Errado	Errado	05:08/38%

Para esta verificação, os participantes foram orientados a verbalizar mais lentamente do que anteriormente, com isso houve um aumento de 26,67%, ou seja, os acertos foram de 25% para 51,67%. Lembrando que para alcançar este número, necessitou de um ambiente mais controlado, para evitar ruídos, e a questão da verbalização.

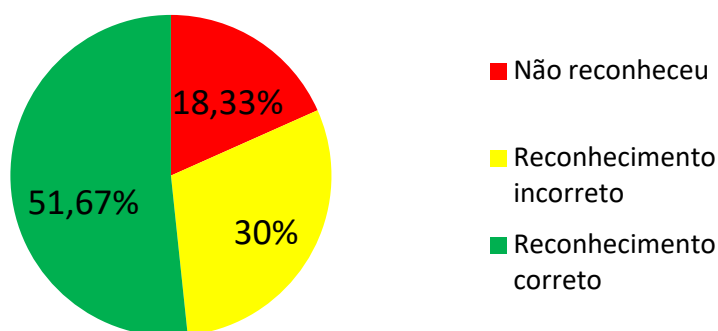


Figura 5 - Resultados 2ª Verificação

Com um resultado mais satisfatório do que a primeira verificação, esta obteve bons números. Porém, para a realização dos testes de usabilidade, deverá ser levado em conta, o ambiente em que será realizado. Deve ter muito cuidado para

que não possua muitos ruídos e que os participantes realizem a verbalização de forma pausada, para que isso não interfira nos resultados dos testes de usabilidade.

3.1.4 Considerações finais sobre o Extrator de Texto

O desenvolvimento desta parte da abordagem, consistiu em utilizar recursos já existentes de bibliotecas que tratassem os arquivos de áudio. Com o apoio desta, evitou o desenvolvimento de códigos complexos que realizassem o reconhecimento de voz, e como consequência se obteve foco maior na realização da abordagem.

As verificações realizadas tiveram foco em obter o maior número de acertos, na primeira verificação se obteve um número baixo, porém foram notados ruídos, o que levou esta verificação a atingir um número baixo. Já para segunda verificação, foi levado em conta todo e qualquer objeto, aparelho eletrônico, a circulação de pessoas, ou o que pudesse fazer com que os áudios possuíssem ruídos, com isso o número de acertos dobrou. E foi possível observar qual a forma em que se deve orientar os participantes a verbalizar e o que devia controlar no ambiente de coleta de dados.

3.2 Gravador de Áudio

Para o desenvolvimento do gravador de áudio, foram pesquisadas bibliotecas que auxiliassem nas gravações e técnicas para a captura de *snapshots*. Logo, foi encontrada e utilizada a biblioteca *winmm.dll* para a gravação dos dados de áudio. Após um breve estudo sobre seus comandos, ficou determinado que seriam utilizados para a gravação de áudio seguintes comandos "*open new type waveaudio alias Som*" e "*record Som*" que iniciam um novo áudio e iniciam a gravação, respectivamente. E para o encerramento do áudio são executados os comandos "*pause Som*", "*save Som* diretório" e "*close Som*", que pausa o som, salva no diretório escolhido e fecha a gravação após salvar, respectivamente.

Depois de realizados testes para verificar se o gravador de áudio funcionava adequadamente, foi acrescentado um método para captura de *snapshots*, onde é capturado uma imagem da tela a cada segundo de execução da ferramenta com o auxílio de um *timer*. A ferramenta salva cada imagem com o formato "*hhmmss.png*", indicando a hora, minuto e segundo do teste, a fim de facilitar a busca feita pela Ferramenta de Apoio aos Testes de Usabilidade, que será descrita na próxima seção.

Para a utilização desta ferramenta, o participante deve escolher um diretório onde deseje que fique salvo os dados, então a aplicação cria um arquivo neste diretório e divide em duas partes, uma para o arquivo de áudio e outro para as imagens coletadas. Após selecionar o local desejado, o usuário deve inserir o nome em que deseja salvar o teste de usabilidade, e dar início ao teste. Após realizado, o teste deve ser finalizado.

Abaixo segue a imagem do gravador de áudio, ou melhor, o coletor de dados da abordagem.



Figura 6 - Coletor de dados

Suas indicações, são as seguintes:

1. Campo do diretório, o participante indica qual local será salvo os dados
2. Indica qual o nome da pasta que deseja criar, dentro do diretório escolhido o item 1.
3. Após selecionado o diretório e indicado o nome, o botão inicia a gravação do áudio e a captura dos *snapshots*.
4. Depois de realizado o teste, o campo Salvar: pausa e salva o áudio, e finaliza a captura das imagens.

3.2.1 Considerações finais sobre Gravador de Áudio

O gravador de áudio demonstrou eficiência ao ser utilizado, pois, ele organiza o diretório separando o áudio das imagens, facilitando caso o avaliador deseje visualizar as imagens no diretório ou por meio do extrator de texto que será adaptado para receber as imagens.

3.3 Ferramenta de Apoio aos Testes de Usabilidade

Com o desenvolvimento de um gravador de áudio e um extrator de texto, foram feitas modificações no extrator de texto de acordo com a abordagem idealizada. O gravador de áudio, foi idealizado para armazenar os dados coletados de uma forma que simplificasse para a Ferramenta de Apoio aos Testes de Usabilidade, ou seja, o extrator de texto foi modificado para que receber os dados

coletados através do gravador de áudio e dispor aos avaliadores de forma eficiente e clara, então surgiu esta ferramenta.

A ferramenta após receber a gramática, primeiro ela analisa o áudio e retorna os eventos encontrados, quando o avaliador seleciona um evento, a ferramenta busca a imagem no diretório utilizando o tempo como parâmetro de busca com o formato "hmmss.png" que indicam a hora, minuto e segundo do teste.

A Figura 7 demonstra como a ferramenta dispõe as informações para os avaliadores.

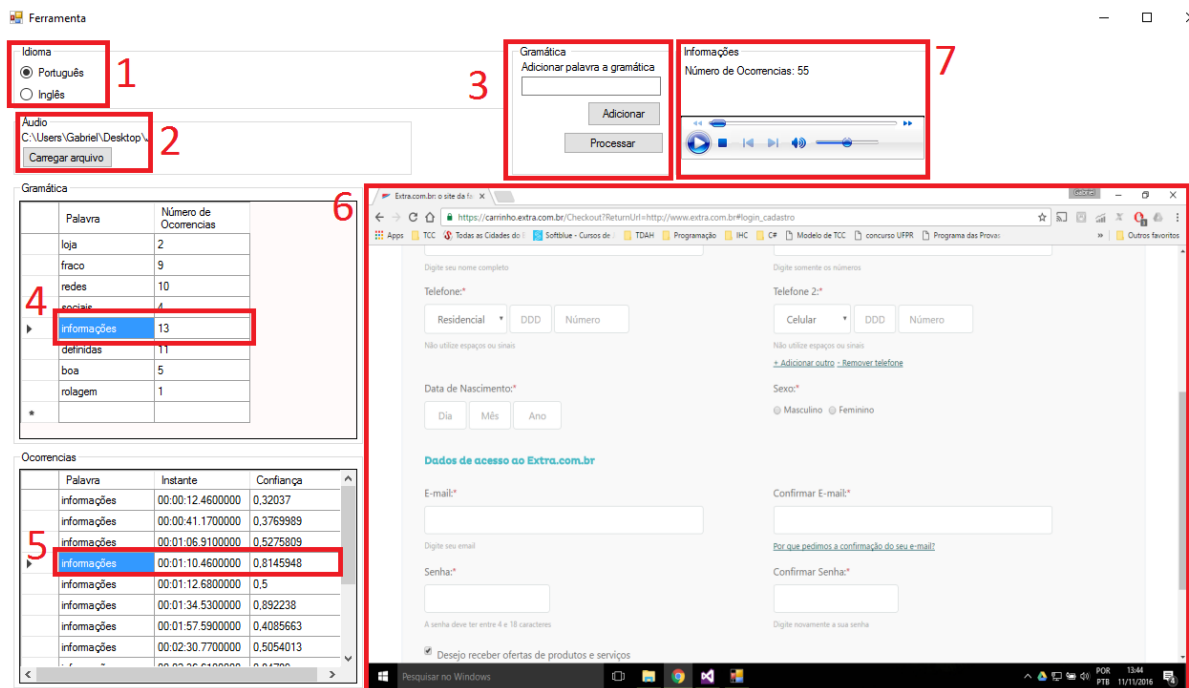


Figura 7 - Indicações da Interface

Abaixo serão descritos os itens marcados na imagem acima:

1. Seleção do idioma, para reconhecer adequadamente o avaliador deve selecionar o idioma em que foi realizado o teste de usabilidade, estão disponíveis em Português e Inglês.
2. Indicar qual o diretório em que está salvo o teste, neste diretório deverão estar salvos as imagens e o áudio do teste realizado.
3. Este item indica o campo da gramática, nele o avaliador deve inserir toda palavra que deseja encontrar no áudio do teste realizado, após a inserção de todas as palavras ele deve indicar o início do processamento.
4. Lista de palavras da gramática, nela está disposta todas as palavras inseridas na gramática e quantas ocorrências de cada palavra foram encontradas.

5. Lista de ocorrências, após escolher uma palavra na lista da gramática a ferramenta atualiza a lista de ocorrências e mostra quais instantes e qual a confiança de todas as ocorrências encontradas de determinada palavra.
6. Imagem do teste, após selecionar um instante, a ferramenta busca qual a imagem mais próxima da ocorrência.
7. Este item mostra o total de ocorrências encontradas e permite que o avaliador ouça o áudio do teste para que ele possa selecionar palavras ditas pelo participante afim de encontrar possíveis eventos que sejam relevantes para o teste.

4. VALIDAÇÃO

Antes de iniciar a validação em si, foram realizados testes com a ferramenta pronta, afim de verificar se as imagens condizem com o tempo em que são salvas de acordo com o momento do teste, então a partir disso iniciou-se a validação da abordagem.

Nielsen (1993) disse que, para realizar avaliações de usabilidade com resultados satisfatórios, os testes devem ser feitos com no mínimo 4 ou 5 participantes e suas tarefas feitas separadamente para não ter influencias das outras pessoas. Tendo em vista esta afirmação, foram selecionados 5 estudantes de cursos de tecnologia, com conhecimento técnico intermediário de computação, porém nenhum deles possui conhecimento da disciplina de IHC.

Foram realizados testes de usabilidade de *websites* que ficaram a critério dos participantes escolher quais eles queriam avaliar, pois o foco do trabalho era verificar se a abordagem proposta seria eficiente e não avaliar *website* escolhido.

Todos os participantes escolheram *websites* de *e-commerce* para realizar os testes, eles foram orientados a realizar operações básicas durante suas interações com a interface, como inserir alguns produtos no carrinho de compra e fazer inserção de alguns dados nos portais, e verbalizar o que estavam pensando, o que estavam tentando realizar ou até o porque de estar realizando aquilo. Também foram alertados de que toda informação pessoal inserida seria inútil para o trabalho e também ninguém teria acesso a ela.

Os testes duraram em média de 6 a 10 minutos, pois alguns participantes realizaram as orientações rapidamente e paravam o teste. Porém todos tiveram grande quantidade de dados de áudio, ou seja, todos verbalizaram bastante, desde opiniões pessoais até possíveis problemas de interface. Porém, em alguns momentos foram lembrados que eles deviam verbalizar, pois ficavam por mais de 30 segundos sem verbalizar nada, ou devido não verbalizar entre uma página e outra.

A validação da abordagem consistiu em analisar os dados de áudio, em busca de possíveis eventos relevantes para o teste, através das verbalizações, quando encontrado um possível evento, era realizado uma busca nas imagens através do tempo do áudio, e analisado a imagem a fim de encontrar problemas de usabilidade. Posteriormente era carregado cada teste de usabilidade na ferramenta, e novamente o arquivo de áudio era analisado em busca de eventos, porém eram

utilizadas as mesmas palavras da análise manual, e então realizava o processamento na ferramenta que indicava onde estavam cada evento, e por fim suas imagens eram analisadas em busca de problemas de usabilidade.

As tabelas 8, 9, 10, 11 e 12 indicam as palavras que poderiam ser um evento relevante para o teste de usabilidade de cada um dos participantes, indicando o tempo e a confiança que a ferramenta indicou. Lembrando que todas elas foram analisadas manualmente, e com a ferramenta foram analisados somente as palavras encontradas.

Na validação realizada com a pessoa 1, foi notado que, ela buscava por informações e se preocupava com a maneira em que os dados eram dispostos na página e algumas questões de *feedback*. A tabela 1 apresenta o eventos analisados manualmente, e posteriormente na ferramenta, tendo um alto índice de acerto, pois o participante verbalizava de forma clara, e o teste não possuiu muitos ruídos.

Tabela 8 - Resultado da Pessoa 1

Pessoa 1		
Palavra	Tempo	Confiança
Gostei	00:08	89%
Boa	00:38	92%
Amostras	-	-
Frete	01:53	87%
Carrinho	01:54	90%
Anúncios	03:37	59%
Definidas	04:41	84%
Redes	05:43	80%
Sociais	05:43	80%
Formatados	-	-
Preenchimentos	07:02	75%

Já na Tabela 9, a pessoa 2 indicou meios que facilitassem a forma de utilização, como ter o CEP do cliente salvo, para que não precisasse informar ao *website* toda vez que utilizasse o carrinho de compras. Ele verbalizou sempre com confiança, possivelmente, causando grande número de acertos e alta confiança indicada pela aplicação.

Tabela 9 - Resultado Pessoa 2

Pessoa 2		
Palavra	Tempo	Confiança
Design	00:27	70%
Login	00:32	75%
Fraco	00:34	86%
Opção	00:42	83%
Formulário	01:06	93%
Ofertas	01:38	89%
Boa	01:43	80%
Cansativo	02:20	94%
Informações	02:30	44%
Importantes	02:44	50%
Fotos	03:30	91%
Faltou	04:14	89%
Recursos	-	-

A pessoa 3, conforme mostra um número baixo de acerto, possivelmente, devido ao participante verbalizar com voz baixa. Neste caso, o número de eventos avaliados manualmente e com a utilização da ferramenta, obteve muita diferença, pois os eventos encontrados avaliados algumas vezes.

Tabela 10 - Resultado Pessoa 3

Pessoa 3		
Palavra	Tempo	Confiança
Facebook	-	-
Interativo	01:11	39%
Notificações	-	-
Busca	-	-
Fonte	-	-
Pequena	-	-
Relacionados	-	-
Fonte	-	-
Cidade	-	-
Estado	04:43	59%
Telefone	-	-
Formatados	-	-
Corretas	06:15	42%
Erro	-	-
Editar	07:51	66%

A pessoa 4, testou funcionalidades que pudessem facilitar a utilização da página como preencher campo da cidade após preencher o CEP, também realizou buscas para verificar o *feedback*. Verbalizou de forma clara e teve alto índice de acerto, podendo, avaliar com a utilização da ferramenta quase todos os eventos que foram encontrados manualmente.

Tabela 11- Resultado Pessoa 4

Pessoa 4		
Palavra	Tempo	Confiança
Opções	-	-
Confirma	00:27	63%
Procurar	-	-
Produto	00:32	90%
Imagens	00:57	89%
Certa	-	-
Inserir	02:05	78%
Busca	03:18	31%
Aparelho	03:55	88%
Interativa	04:24	40%
Estado	04:58	70%

Observa-se que o baixo número de acertos do teste com a pessoa 5, provavelmente, foi devido a velocidade e a altura em que o participante pronunciou. Este verbalizou muitas informações importantes, informando erros e testando as funcionalidades presentes em cada página, assim como o carrinho de compra, caracteres especiais e *login*.

Tabela 12 - Resultado Pessoa 5

Pessoa 5		
Palavra	Tempo	Confiança
Histórico	-	-
Frete	01:42	94%
Produto	-	-
Ruim	-	-
Letras	08:48	79%
Dígitos	-	-
Números	-	-

Conforme na Figura 7, é notável um número maior que os números encontrados nas verificações do extrator de texto. Todos os eventos foram selecionados manualmente, levando em conta a frase dita pelo usuário e considerados possíveis eventos que pudessem representar algum problema de usabilidade, e analisados, posteriormente, todos eles foram buscados pela ferramenta. Foram escolhidos no total 55 possíveis eventos relevantes, 35 delas foram encontradas na ferramenta, totalizando 64% e as outras 20 palavras somaram 36% dos eventos que não foram encontrados e conseqüentemente não foram avaliados.

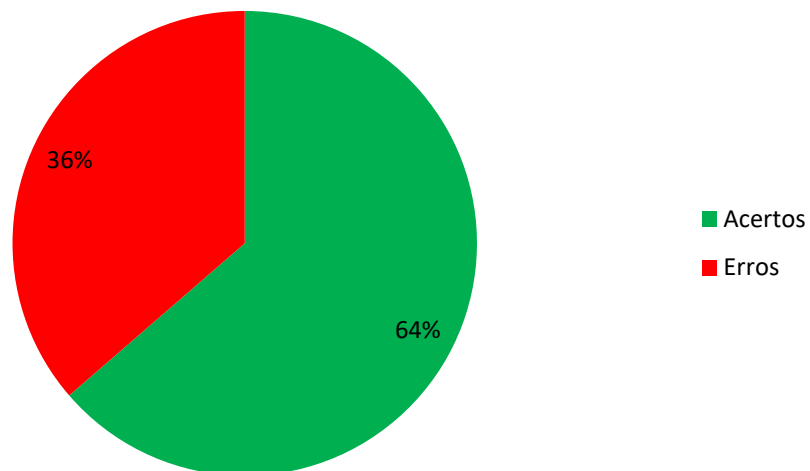


Figura 7 - Percentual de acertos e erros da ferramenta

O número de erros da ferramenta nos Testes de Usabilidade podem ser referentes a ruídos próximos do local do teste, a velocidade em que o participante verbalizou ou a forma em que verbalizou. As orientações foram muito importantes

para obter um número maior de acertos, pois nenhum deles tinha conhecimento de como realiza o reconhecimento de voz.

A tabela abaixo relaciona os tempos de análise com o tempo de teste e qual participante.

Tabela 13 - Resultado Analise Manual x Analise com a Ferramenta

	Tempo de teste	Tempo de analise manual	Tempo de analise com ferramenta
Pessoa 1	07 minutos e 43 segundos	20 minutos	13 minutos
Pessoa 2	05 minutos e 59 segundos	16 minutos	11 minutos
Pessoa 3	08 minutos e 03 segundos	19 minutos	13 minutos
Pessoa 4	05 minutos e 53 segundos	18 minutos	14 minutos
Pessoa 5	09 minutos e 31 segundos	25 minutos	9 minutos

Algumas considerações são válidas:

- Pessoa 1, 2 e 4: Verbalizaram de forma clara e objetiva, gerando um número baixo de erros pela ferramenta.
- Pessoa 3 e 5: Verbalizaram pouco mais rápido em relação aos demais, o que pode ter causado um maior número de erros.

Os testes mostraram que a abordagem auxilia o avaliador na busca por eventos levando em conta o tempo, ela faz com que o tempo da analise pela ferramenta diminua em relação ao tempo de analise manual. Porém é necessário observar que os eventos não encontrados não foram avaliados na ferramenta, diminuindo o tempo de teste diminuindo o tempo da analise.

O tempo de analise pela ferramenta foi inferior ao tempo de analise manual devido a algumas palavras não encontradas, porém os testes que tiveram mais palavras acertadas também tiveram um tempo reduzido, devidamente, ao processo de indexação das imagens terem sido automatizadas.

5. CONSIDERAÇÕES FINAIS

A abordagem foi proposta com o intuito de reduzir o tempo de análise dos Testes de Usabilidade, onde cada participante poderia verbalizar o que estava em sua mente, sem ter que decorar palavras chaves ou avaliativas. Para realizar o trabalho necessitou de separar em etapas, como determinar e desenvolver métodos para capturar *snapshots* e gravar o áudio do usuário, também como buscar as palavras dentro de um arquivo de áudio e verificar se a abordagem contribuiria para os Testes de Usabilidade.

Foi decidido que a ferramenta seria desenvolvida com a linguagem C# devido a facilidade de aprendizado e por possuir métodos de processamento de áudio, assim poderia reduzir o tempo de desenvolvimento e evitar desenvolver um método para a gravação de áudio e para reconhecimento de voz, o que é mais complexo de implementar. Também, para o registro de *snapshots* foram utilizados métodos prontos para reduzir o tempo de desenvolvimento e focar mais na validação da abordagem.

A ferramenta proposta utilizou recursos computacionais já existentes para reduzir o tempo de desenvolvimento de algoritmos complexos. E validada a partir de testes de usabilidade, com participantes inexperientes em relação aos conceitos de usabilidade e IHC.

O percentual de acerto da Ferramenta foi satisfatório, pois foram maiores que os números das verificações do extrator de texto. Porém deve ser levado em conta que as palavras que não foram avaliadas com o apoio da ferramenta fizeram com que diminuísse o tempo, justamente por não ter como avaliar os eventos que não foram encontrados, mas nos testes em que a maioria dos eventos foram encontrados o tempo também diminuiu.

Como trabalho futuro e conforme os estudos em reconhecimento de voz forem se atualizando, revalidar a abordagem proposta utilizando novas abordagens de extração de textos ou até mesmo as bibliotecas existentes, com o intuito de aumentar a porcentagem de acerto e poder aumentar o número de dados a serem analisados.

REFERÊNCIAS

ALENCAR, V. F. S.; **Atributos e domínios de interpolação eficientes em reconhecimento de voz distribuído**. 2005. Pós-graduação em Engenharia Elétrica. PUC-Rio. Rio de Janeiro.

BENYON, D. **Interação humano-computador**. [S.l.]: Pearson Brasil, 2011.

BISPO; S.; **Reconhecimento de voz contínua para o português utilizando modelos de Markov escondidos**. 1997. Dissertação (Doutorado em Ciências). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

BOREN, M. T.; RAMEY, J. **Thinking aloud: Reconciling theory and practice**. *IEEE Transactions on Professional Communication*, 2000.

BOTELHO, J. M; CRUZ, Vilma Aparecida Gimenes da. **Metodologia: Estudo e ensino**. São Paulo: Pearson Education do Brasil, 2013. 135 p.

COLETI, T. A. **Um ambiente de avaliação da usabilidade de software apoiado por técnicas de processamento de imagens e reconhecimento de fala**. 2014. 155 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2013.

CYBIS, W; BETION, A. H.; FAUST, R. **Ergonomia e Usabilidade - Conhecimentos, Métodos e Técnicas**. [S.l.]: Novatec Editora, 2010.

HAAK, M. J.; JONG, M. D. T. de; SCHELLENS, P. J. **Retrospective vs. concurrent think aloud protocols: testing the usability of an online library catalogue**. *Behavior & Information Tecnology*, 2003.

HERTZUM, M.; HANSEN, K. D.; ANDERSEN, H. H. **Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload?** *Behavior & Information Tecnology*, 2009.

ISO 25010: **Software engineering. Software product Quality Requirements and Evaluation (SQuaRE)**. *Software and quality in use models*.

ISO 9241-11: **Requisitos ergonômicos para o trabalho com dispositivos de interação visual**.

LOWDERMILK, T; **Design Centrado no Usuário: um Guia Para Desenvolvimento de Aplicativos Amigáveis**. Novatec, 2013.

NIELSEN, J. **Usability Engeneering**. [S.l.]: Morgan Kaufmann, 1993.

PREECE, J.; BORGES, Y.; SHARP, H. **Design de Interação, Além da interação homem computador**. [S.l.]: Bookman, 2005.

PREECE, J.; ROGERS, Y.; SHARP, E.; **Interaction Design: Beyond Human-computer**. Interaction. New York, NY: John Wiley & Sons. 2002.

RABINER, Lawrence R. **A tutorial on hidden Markov models and selected applications in speech recognition**. *Proceedings of the IEEE 77.2* (1989): 257-286.

SILVA, A. G.; **Reconhecimento de Voz para Palavras Isoladas**. 2009. Graduação em Engenharia da Computação. Universidade Federal do Pernambuco. Recife.

TUFANO, D; **Estudos de Língua e Literatura - Volume 1**. Moderna. 1998.

VOICEBOX TECHNOLOGIES, INC. (Estados Unidos). **Dynamic speech sharpening**. USA nº US7634409, 31 dez. 2005, 31 dez. 2009. RABINER, Lawrence; JUANG, Biing-Hwang. **Fundamentals of speech recognition**. 1993.

YNOGUTI, C.A.; VIOLARO, F. **Um sistema de reconhecimento de fala contínua baseado em Modelo de Markov Contínuos**. In: Simpósio Brasileiro de Telecomunicações, 2000.