

# INTERFACE DO USUÁRIO BASEADA EM VOZ COMO FERRAMENTA PARA PROMOVER O ENSINO/APRENDIZAGEM DE LÍNGUA ESTRANGEIRA

Valéria Farinazzo Martins<sup>1</sup> e André Brasileiro<sup>2</sup>

<sup>1</sup>Grupo de Processamento Gráfico – Faculdade de Computação e Informática – Universidade Presbiteriana Mackenzie

<sup>2</sup>Programa de Pós-Graduação em Engenharia Elétrica – Universidade Presbiteriana Mackenzie  
{valfarinazzo;andrebrasiliano}@hotmail.com

## Resumo

Este artigo traz um estudo do uso de interfaces do usuário baseada em voz no auxílio ao aprendizado de língua estrangeira, inserida no contexto de melhorar atividades práticas de escuta e compreensão de textos, bem como a pronúncia correta de palavras. Para tanto, através de técnicas de reconhecimento e síntese de voz, foi desenvolvida a aplicação VAL, que segue uma metodologia específica para o desenvolvimento de aplicações VUI.

**Palavras-chave:** Interface do Usuário Baseada em Voz, Ensino de língua estrangeira, Interface Homem-Máquina.

## Abstract

This paper presents a study of the use of Voice User Interface (VUI) to aid on foreign language learning – to improve the listening, understanding and pronunciation of English language. Using recognition and synthesis voice technology, an application called VAL is built; this application follows a specific methodology for developing VUI applications.

**Keywords:** Voice User Interface. Foreign language teaching, Human-Computer Interface.

## 1. Introdução

A partir do momento em que mais pessoas começam a ter computadores e dispositivos eletrônicos em geral, a usabilidade das interfaces homem-máquina torna-se requisito fundamental em todas as classes de aplicações que sejam executadas nestes equipamentos. O uso das interfaces não convencionais – tais como Realidade Virtual, Realidade Aumentada, Interfaces por Gestos e Interfaces Baseadas em Voz – surge como uma forma de tentar tornar estas aplicações mais naturais e mais fáceis de serem usadas, gerando maior satisfação aos usuários não especialistas (OLIVEIRA NETO, SALVADOR e KAWAMOTO, 2010).

A voz pode ser apontada como uma das formas mais naturais de interação entre pessoas. Desde a década de 50, estudos em Inteligência Artificial vislumbram o uso da voz como ferramenta para a interação entre máquinas e pessoas, mas as limitações de

hardware e software foram impactantes. Somente depois dos anos 90 é que a tecnologia de reconhecimento de voz contou com uma significativa melhora e sistemas puderam ser efetivamente criados e usados (MCTEAR, 2002), (COHEN, GIANGOLA e BALOGH, 2004).

As interfaces do usuário baseadas em voz – dos termos em inglês *Voice User Interface* (VUI) e *Speech User Interface* (SUI) – são interfaces em que o sistema capta as entradas por voz do usuário, realiza determinada ação correspondente ao entendimento da requisição do usuário, e gera uma saída, geralmente, por voz – pré-gravada ou sintetizada (DAMASCENO, PEREIRA e BREGA, 2005).

Tradicionalmente, o desenvolvimento de VUI tem focado em aplicações comerciais, tais como em sistemas para busca e recuperação de informações em mercados de ações, horário e reserva de vôos e reserva de hotéis, auxílio à lista telefônica, guia de restaurantes, bares e filmes. Porém, já existem algumas iniciativas voltadas para o

acesso à informação para a Educação (ESTABEL, MORO e SANTAROSA, 2006). O objetivo principal deste artigo é apresentar as possibilidades de uso de interfaces baseadas em voz nas práticas de ensino/aprendizagem de línguas estrangeiras.

Após o estudo do referencial teórico e o levantamento de requisitos, foi desenvolvida a aplicação VAL – *Voice Application Learning* - um ambiente para exercitar a escuta do idioma estrangeiro, compreensão de textos e prática de pronúncia de palavras.

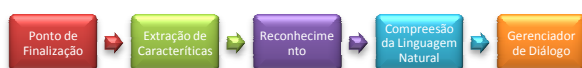
Este artigo está organizado como descrito a seguir. A seção 2 apresenta os fundamentos conceituais necessários para o entendimento da aplicação criada, a saber: reconhecimento de voz, síntese de voz e VUI; a seção 3 descreve a aplicação *Voice Application Learning*, abordando as principais fases de seu desenvolvimento. Finalmente, a seção 4 encerra o trabalho apontando as conclusões do trabalho e oportunidades de pesquisa futuras.

## 2. Fundamentos Conceituais

A fim de entender a aplicação desenvolvida *Voice Application Learning*, alguns conceitos fundamentais devem ser aqui explicados: (i) Reconhecimento de Voz, (ii) Síntese de Voz – do termo em inglês *Text-to-Speech* (TTS) – e, (iii) VUI. O item 2.4 traz os trabalhos relacionados.

### 2.1. Reconhecimento de Voz

O reconhecimento de voz, conforme apresentado na Figura 1, consiste de uma série de módulos projetados para capturar uma entrada de voz (emitida pelo usuário), entender o que foi capturado, executar as transações ou tarefas computacionais, e responder de maneira apropriada (DENG e HUANG, 2004), (COHEN, GIANGOLA e BALOGH, 2004).



**Figura 1 – Módulos de Reconhecimento de Voz**  
(Cohen, Giangola e Balogh, 2004)

O ponto de finalização – *endpointing* - detecta o início e o final da fala – através da captura do silêncio - e determina a forma de onda. A onda é empacotada e enviada para o módulo de extração das características, que transforma a demarcação do que foi ditado em fonemas e a cada um deles é atribuído um número – chamado de vetores de características. Em seguida, o módulo reconhecedor usa a sequência de vetores de características para determinar as palavras que foram ditas pelo usuário. O módulo de entendimento de linguagem natural atribui significado às palavras que foram ditas, através de um conjunto de blocos de valores. Um bloco é definido para cada item de informação que é relevante para a aplicação, ou seja, palavras-chaves. Assim, o módulo de gerenciamento do diálogo é iniciado. É o gerenciador de diálogo que determina as ações que o sistema deve fazer dentre as várias possibilidades, tais como acesso ao banco de dados ou executar uma transação.

Os principais métodos utilizados para fazer o reconhecimento automático da fala (voz) são: método acústico-fonético, uso de reconhecimento por comparação de padrões e uso de Inteligência Artificial, que são brevemente descritos a seguir (MARTINS, 1997):

**i. Método Acústico-Fonético:** considera a linguagem falada como um conjunto de unidades fonéticas finitas e diferentes. Estas unidades fonéticas são caracterizadas por um conjunto de propriedades (classificação sonoro/não sonoro, presença de ressonância nasal, localização dos formantes) que são manifestadas no sinal da fala, através do tempo e podem variar imensamente de acordo com o interlocutor e também pelos seus vizinhos fonéticos. Assim, este método consiste em identificar as unidades fonéticas que compõem a fala que será reconhecida e, a partir de sua concatenação, reconhecer a fala. Basicamente, um reconhecedor de voz que utiliza este método possui as seguintes fases: análise espectral, detecção das características fonéticas, segmentação do

signal da fala e identificação das unidades fonéticas e escolha da palavra que melhor corresponde à sequência de unidades.

**ii.** Método de Comparação de Padrões: parte da filosofia que os sistemas foram treinados para reconhecer determinados padrões. Possuem duas fases: treinamento e reconhecimento. Na primeira fase, exemplos de padrões que devem ser reconhecidos são inseridos no sistema a fim de se criar uma base de padrões. Já na segunda fase, compara-se o padrão desconhecido com a base de padrões de referência e calcula-se uma medida de similaridade. O padrão que melhor corresponder ao padrão desconhecido é eleito. Um exemplo de sistemas que trabalham desta maneira está o Modelo Oculto de Markov.

**iii.** Utilização de Inteligência Artificial: exploram os conceitos dos dois métodos anteriormente descritos. Redes Neurais podem ser enquadradas nesta classe. Utilizando as redes "Multilayer Perceptron", apresenta-se uma matriz de ponderação que representa as conexões entre nós da rede, sendo que cada saída da rede está associada a uma unidade - palavra, por exemplo - que deve ser reconhecida.

## 2.2. Síntese de Voz

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

Embora a tecnologia TTS ainda não reproduza com fidelidade a qualidade da voz humana gravada - vale à pena destacar que, até o momento, os sintetizadores de voz não conseguem representar entonação - ela tem melhorado muito nos últimos anos. Tipicamente, o uso da voz humana pré-

gravada está atrelado aos *prompts* e ao envio de mensagens para os usuários. No entanto, determinadas aplicações, como leitores de e-mail e notícias, tem dados muito dinâmicos. Nesses casos, uma vez que os textos das mensagens não podem ser previstos, pode-se usar a tecnologia TTS para criar os discursos de saída (COHEN, GIANGOLA e BALOGH, 2004).

## 2.3. Voice User Interface

Interface do usuário baseada em voz consiste na interação de uma pessoa com um sistema através de voz, utilizando uma aplicação de linguagem falada. São capazes não somente de reconhecer a voz do usuário, mas compreender o que ele diz e responder a estas entradas, geralmente, em tempo real (LAI, 2000), (SHNEIDERMAN, 2000).

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

Existem diferenças substanciais no desenvolvimento de Interfaces Gráficas do Usuário (GUI) e VUI, desde que a voz não está visível ao usuário como no caso das GUIs e também pelo seu caráter transiente, o que pode aumentar consideravelmente a carga cognitiva exigindo uma memória de curta duração do usuário mais ativa; outro ponto a considerar é que entradas por voz são muito mais rápidas que as entradas via teclado, ao passo que as saídas por voz são mais lentas do que leituras textuais feitas superficialmente. Além disto, na VUI, a comunicação tende a ser serial e com canal

único – ou seja, o usuário não é capaz de escutar e falar ao mesmo tempo, nem é capaz de ouvir mais de uma fala ao mesmo tempo. Todos estes argumentos supracitados levam a crer que projetar VUI é razoavelmente diferente de se projetar GUI e, por isto, necessitam de uma metodologia diferenciada.

## 2.4. Trabalhos Relacionados

Na literatura foram encontrados poucos trabalhos relacionados ao tema desenvolvido neste trabalho. Neste contexto, vale a pena destacar Cocciolo (2009) que descreve o desenvolvimento do projeto Meety, que visa enriquecer o ensino em salas de aulas, através da integração de internet, vídeo, som e reconhecimento de voz dentro de uma única ferramenta. Uma simulação foi realizada para testar a eficácia do Meety, onde foi exposto ao software um *podcast* com uma leitura em que o contexto era economia, e a ferramenta, através do reconhecimento de voz do *podcast*, deveria encontrar no Google e disponibilizar para alunos e professor assuntos relevantes ao contexto da leitura que estava sendo ditada. O resultado da pesquisa foi satisfatório, pois o Meety alcançou uma pontuação alta em uma escala Likert com vários recursos coletados que eram relevantes ao problema abordado no ditado.

Este artigo tem uma relação bastante próxima com este trabalho, pois versa sobre uma ferramenta que utiliza o reconhecimento de voz como um recurso adicional para enriquecer o aprendizado em sala de aula. A diferença é que o projeto Meety tem a ideia de uma interface multimodal, onde além do reconhecimento de voz, o sistema deve utilizar outros canais de comunicação para interagir com o usuário, à medida que este utiliza tecnologia de reconhecimento e síntese de voz, além da textual.

## 3. Voice Application Learning

A *Voice User Application* (VAL) consiste em uma aplicação para o auxílio no

aprendizado de língua estrangeira para pessoas não-nativas. Esta aplicação permite que qualquer pessoa interessada em treinar a escuta, a compreensão de textos e a pronúncia correta de palavras em outros idiomas possa o realizar através de seu uso.

O processo utilizado para o desenvolvimento da aplicação, baseado em COHEN, GIANGOLA e BALOGH (2004) e em LAMEL, MINKER e PAROUBEK (2000), é composto pelas seguintes fases: Definição dos requisitos, Projeto de Alto Nível, Projeto Detalhado, Desenvolvimento, Teste e Avaliação e *Tuning*.

### 3.1. Definição dos Requisitos Funcionais

Nesta etapa, o principal objetivo é adquirir um conhecimento detalhado da aplicação em questão, suas metas, características e funcionalidades desejadas, usuários finais, as motivações dos usuários para utilizarem a aplicação e o cenário de uso. Deve-se entender, também, o contexto de negócio. Sendo assim, especifica-se:

i. Definição dos usuários finais: a) Aluno: qualquer pessoa interessada em utilizar a aplicação, sem que seja necessário haver uma fase de treinamento prévio do sistema; b) Professor: responsável por inserir todos os arquivos de texto e áudio de língua inglesa em cada nível de conhecimento do aluno. Sobre estes usuários, foi possível definir:

ii. Frequência de uso: a) aluno: esporádico a diário; b) professor: esporádico;

iii. Nível de conhecimento da aplicação e de conhecimento da tecnologia: variável para ambos;

iv. Sexo e faixa etária: feminino e masculino, com faixa etária variável, para ambos.

v. Definição do Ambiente: o ambiente físico em que se dará o uso desta aplicação poderá ser uma escola de línguas ou em qualquer outro ambiente, preferencialmente em ambiente silencioso;

vi. Definição das tarefas: definem as regras do negócio. A aplicação permite que sejam introduzidos (pelo professor) áudios

sobre determinados assuntos, textos referentes ao áudio e também um questionário (texto que será transformado pelo sistema em um arquivo de áudio) que o aluno deverá responder (por voz) a fim de que se possa verificar se a compreensão sobre a historia foi suficiente. Assim, são tarefas atribuídas aos dois tipos de usuários:

**vii.** Aluno: Conforme mostrado no diagrama de Casos de Uso descrito na Figura 2, para o aluno, são suas opções:

- Só ouvir o arquivo de áudio, correspondente a uma historia;
- Ouvir o arquivo de áudio e acompanhar o texto;
- Ouvir cada pergunta do questionário disponibilizado depois do áudio e responder por voz;
- Ouvir cada pergunta e acompanhar o texto simultaneamente e depois responder por voz.

**viii.** Professor: As mesmas funcionalidades permitidas para alunos, além de (Figura 3):

- Cadastra Usuários: pode cadastrar os usuários que são alunos;
- Cadastra textos, áudio e questionários: cadastro do texto, do áudio, do grau de dificuldade deste texto e também as perguntas que estão relacionadas ao texto;

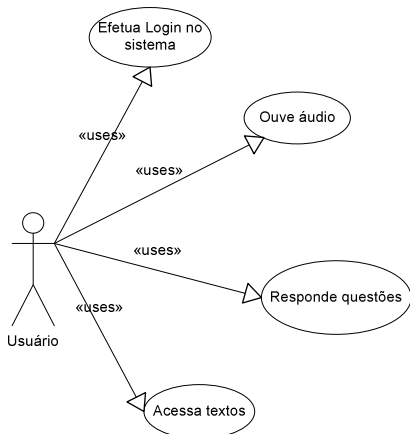


Figura 2 –Use Case para o ator Aluno

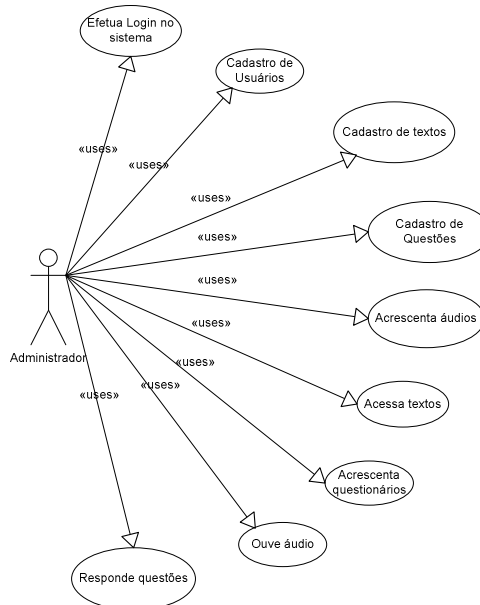


Figura 3 - Use Case para o ator Professor

### 3.2. Definição dos Requisitos não Funcionais

Ficou determinado que os requisitos não funcionais que deveriam ser atendidos nesta aplicação seriam:

**i.** Usabilidade: facilidade com que o usuário tem acesso às partes principais do programa, que fazem uso dos recursos de multimídia (quando acessa o conteúdo com os textos e os áudios) e de reconhecimento de voz, nos questionários cadastrados. O aplicativo desenvolvido faz uso de elementos simples de interface, como uma área de trabalho com diminuição da sobrecarga de informação e menus com opções objetivas, que deixam simples e auto-explicativo o uso do sistema, procurando praticidade na navegação e diminuição no número de cliques durante a navegação.

**ii.** Disponibilidade e Confiabilidade: O sistema, sendo *stand-alone*, é altamente disponível, sendo que problemas de disponibilidade poderão acontecer apenas em casos de interferência dos sistemas operacionais ou Gerenciador de Banco de Dados. Sendo assim, há pouca ocorrência de falhas no sistema, no que tange sua estrutura interna. A gramática construída para aprimorar o reconhecimento das entradas via

dispositivo de voz do usuário garante uma taxa maior de acertos do sistema quando busca em seu banco de dados as correspondências mais prováveis.

**iii. Performance:** O sistema apresentado exige velocidade moderada de transferência dentro da rede, quando o servidor de banco de dados encontra-se em estação ou servidor remoto. Não é exigida uma estação de trabalho com grande poder de processamento. Nesse caso, o tempo de resposta do sistema é baixo, uma vez que o volume de dados também é baixo. Porém existe a necessidade de uma estação que permita um processamento lógico que não demore mais que um tempo de retorno muito sensível para o usuário.

### 3.3. Projeto de Alto Nível

Esta fase é geralmente rápida, mas, tem um papel crucial, visto que cria uma experiência - do usuário - consistente, efetiva e única. O objetivo é encapsular os requisitos de forma mais concreta para guiar o projeto e tomar decisões sobre estratégias e elementos de diálogo que permeiam o projeto, alcançando consistência. Os passos desta fase para a aplicação em questão são:

**i.** Definição dos critérios-chave: a) uma interface bastante intuitiva, sem necessidade de treinamento para seu uso; b) *feedback* adequado; c) alta taxa de reconhecimento de voz.

**ii.** Estratégias do diálogo: o diálogo da aplicação consistirá em apresentar, ao usuário, perguntas sintetizadas por *text-to-speech*; a resposta a esta pergunta será dada pelo usuário por voz; o *feedback* da aplicação, a respeito da correta ou não resposta, será dada pelo usuário a cerca da compreensão do texto anteriormente apresentado ao usuário.

**iii.** Elementos pervasivos do diálogo: se a aplicação tiver uma taxa de reconhecimento baixa sobre a entrada do usuário, deve haver um mecanismo para solicitar a re-entrada do usuário.

A fim de uma melhor visualização da arquitetura do sistema, é apresentada a Figura 4, a seguir.

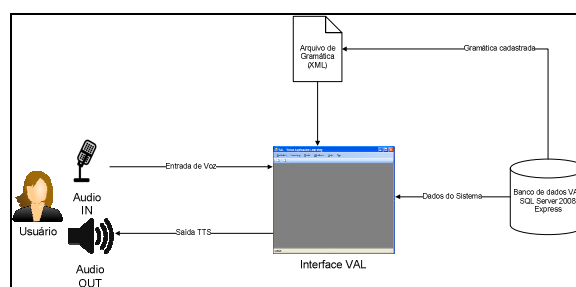


Figura 4 - Arquitetura da aplicação

A aplicação carrega o texto referente a uma pergunta e o sintetiza em voz e a dispara para o usuário. O usuário fornece a resposta correspondente a esta pergunta através de um microfone (áudio - IN). A aplicação, que já fez a carga da gramática, realiza a consulta à base de dados e devolve a resposta – confirmando se o usuário acertou ou não a pergunta – por voz.

### 3.4. Projeto Detalhado

Nesta etapa, para cada diálogo devem ser descritos alguns componentes, entre eles: os *prompts* iniciais, a gramática de reconhecimento, o gerenciamento de erros e a especificação de ação. Um exemplo deste projeto detalhado em relação à aplicação – uso de um questionário - é apresentado a seguir na Tabela 1:

SISTEMA:	Where is living John?
USUÁRIO:	He is living in Washington.
SISTEMA:	You are correct. Congratulations.
<b>Tabela 1: Exemplo de diálogo entre o sistema e o usuário</b>	

Se o sistema não entender o que o usuário diz, ou seja, se a confiabilidade do que o usuário está dizendo não alcançar um nível satisfatório – por exemplo, 70% - então o sistema deve requisitar a re-entrada da resposta do usuário (Tabela 2):

SISTEMA:	Where is living John?
USUÁRIO:	He is living in Washington.
SISTEMA:	Sorry. Could you repeat this answer, please?
<b>Tabela 2: Exemplo de erro no reconhecimento da fala do usuário</b>	

### 3.5. Desenvolvimento

O arquivo de gramática, que faz uso da tecnologia XML, é preenchido dinamicamente pelo sistema e descreve todas as palavras que estão passíveis de reconhecimento. Quanto maior o número de palavras que a gramática contiver, maior a precisão de reconhecimento de palavras, uma vez que o sistema passa a apresentar um maior número de hipóteses para as palavras ditas pelo usuário. Por outro lado, se a gramática for menor, maior é a chance de reconhecer o que o usuário fala.

O papel do banco de dados no sistema é o de armazenar a gramática entendida pelo sistema e também armazenar os dados cadastrados pelos professores – novos usuários, textos, áudios e questões. Para esta aplicação, o Sistema Gerenciador de Banco de Dados utilizado foi a Microsoft SQL Server 2008 Express.

O motor de reconhecimento de voz utilizado foi o Microsoft Speech versão 6.5, que utiliza o método de comparação de padrões para identificar a palavra que foi falada.

O cadastro dos textos realizado pelo professor, Figura 5, faz a inserção do texto, arquivo .rtf, referente ao que o aluno deve ler, seu respectivo áudio (arquivo .wav), além do grau de dificuldade (iniciante, intermediário ou avançado). Após o cadastro do texto, o professor deve fornecer as questões que comporão a parte de *listening* e interpretação do texto, conforme mostrado na Figura 6.

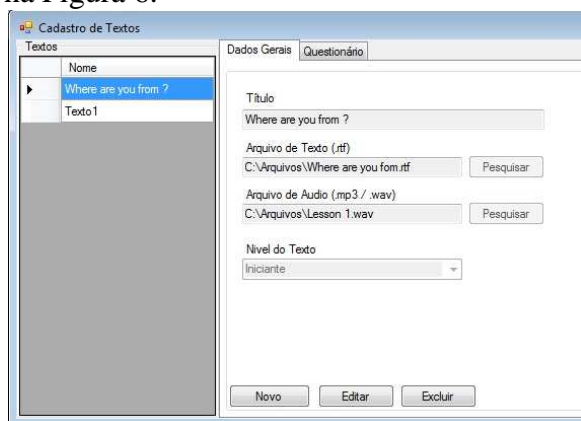


Figura 5 - Cadastro de Textos, Áudios e Grau de dificuldade do Texto

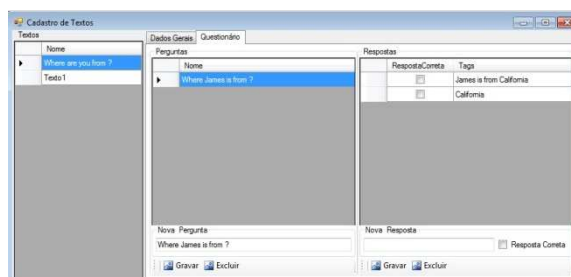


Figura 6 - Cadastro do Questionário – Perguntas e Respostas

### 3.6. Testes e Tuning

Depois de desenvolvidos, cada um dos componentes da aplicação (isto é: código de diálogo e código de interação) são testados, isoladamente. Porém, nesta fase, o teste se refere à verificação de todo o sistema. Há uma grande quantidade de testes a serem realizados antes do lançamento do sistema, incluindo teste de aplicação, teste de reconhecimento e teste de avaliação de usabilidade. Todos esses testes são executados com todo o sistema funcionando.

Depois de completar a fase de testes, já é possível instalar o sistema no ambiente do usuário. Esse procedimento geralmente é feito em duas etapas: na primeira fase é instalado o piloto, em que o sistema é lançado para um número limitado de usuários (geralmente centenas). Coletam-se os dados e o sistema é melhorado a partir desses dados. Na maioria das vezes, são realizadas três iterações de coleta de dados. Após a fase de *tuning*, o sistema é finalmente lançado para todos os usuários finais.

Esta aplicação contou com os testes de aplicação, de reconhecimento e de avaliação da usabilidade, porém, o teste com o programa-piloto foi realizado somente entre os integrantes, de maneira informal.

### 4. Conclusões

O contexto atual de desenvolvimento e pesquisa de interfaces do usuário com o computador destaca o uso da voz como uma modalidade de muita relevância na interação entre o ser humano e os dispositivos digitais

que, cada vez mais, fazem parte do cotidiano.

Embora os sistemas que utilizam VUI sejam cogitados para facilitar o uso de pessoas com impossibilidade de digitação e/ou leitura de dados na tela, pode-se verificar, através deste trabalho, a importância deste tipo de interface em usos na Educação, como proposto neste trabalho: uma aplicação baseada em voz para auxiliar o ensino/aprendizagem de língua estrangeira.

Ao longo deste trabalho foram discutidas e apresentadas as etapas do desenvolvimento, desde a análise de requisitos até os testes, e as tecnologias utilizadas na criação da aplicação VAL, isso é importante, visto que na literatura não se existem muitos trabalhos com detalhamento dos passos para desenvolver este tipo de aplicação, com exemplificação.

Podem ser identificados como trabalhos futuros:

Uma avaliação mais formal desta aplicação, junto a usuários finais – alunos e professores de língua inglesa;

Verificar como o sotaque dos usuários poderia influenciar no reconhecimento da resposta e como isso poderia ser melhorado;

Definição de critérios de usabilidade em aplicações de VUI para o ensino/aprendizagem de língua estrangeira;

Inclusão de recursos de colaboração e compartilhamento de informação; e

Combinação da modalidade voz com outras modalidades de interação, como gestos.

### Referências

BOSCH, L., OOSTDIJK, N., RUITER, J. P. Turn-taking in social talk dialogues: temporal, formal, and functional aspects, SPECOM-2004, pp 454-461, 2004.

CHAMBERLAIN, J. ELLIOTT, G., KLEHR, M., BAUDE, J. **Speech User Interface Guide**, RedPaper IBM, <http://www.redbooks.ibm.com/redpapers/pdfs/redp4106.pdf>, 2006.

COCCIOLO, A. **Using Speech Recognition Technology in the Classroom: An Experiment**. In: Computer-Supported Collaborative Learning. CSCL MAIN CONFERENCE EVENTS – POSTERS, New York, 2009.

COHEN, M. H., GIANGOLA, J. P., BALOGH, J. **Voice User Interface Design**, Addison Wesley, ISBN 0-321-18576-5, 368 pp., 2004.

DAMASCENO, E. F.; PEREIRA, T. V.; BREGA, J. R. F. **Implementação de Serviços de Voz em Ambientes Virtuais**. In INFOCOMP Journal of Computer Science, v.4, n3, p.67-73, 2005.

DENG, L., HUANG, X. **Challenges in adopting speech recognition**, Commun. ACM 47, 1, pp. 69-75, 2004.

ESTABEL, L.B.; MORO, E. L.; SANTAROSA, L. C. A inclusão social e digital de pessoas com limitação visual e o uso das tecnologias de informação e de comunicação na produção de páginas para a Internet. Ci. Inf., vol.35, n.1, 2006.

LAI, J. **Conversational Interfaces**, Communications of the ACM, Vol. 43, No 9, pp 24 – 27, 2000.

LAMEL, L.; MINKER, W.; PAROUBEK, P. **Towards Best Practice in the Development and Evaluation of Speech Recognition Components of a Spoken Language Dialog System**, Natural Language Engineering, vol 6 (3-4), United Kingdom Cambridge University Press, pp. 305 – 322, 2000.

MARTINS, J. A. **Avaliação de diferentes técnicas para o reconhecimento de fala**. Tese de doutorado. UNICAMP (Universidade de Campinas). Campinas, 1997.

MCTEAR, M. F. **Spoken Dialogue Technology: Enabling the Conversational**



User Interface, ACM Computing Surveys,  
Vol. 34, No. 1, pp. 90–169, 2002.

OLIVEIRA NETO, J. S. de; SALVADOR,  
V. F. M.; KAWAMOTO, A. L. **Aplicações  
interativas baseadas em voz na Educação:  
oportunidades e estudo de caso.** In: Anita  
Maria da Rocha Fernandes; Michelle Silva  
Wangham. (Org.). Livro de Minicursos.  
Florianópolis, 2010, v. , p. 1-26, 2010.

SHNEIDERMAN, B. The Limits of Speech  
Recognition, Communications of the ACM,  
Vol. 43, No 9, pp 24 – 27, 2000.